COMPUTING SCIENCE

ALARP Explored

Felix Redmill

TECHNICAL REPORT SERIES

No. CS-TR-1197 **March 2010**

Bibliographical details

REDMILL, F.

ALARP Explored
[By] F. Redmill
Newcastle upon Tyne: University of Newcastle upon Tyne: Computing Science, 2010.

(University of Newcastle upon Tyne, Computing Science, Technical Report Series, No. CS-TR-1197)

Added entries

UNIVERSITY OF NEWCASTLE UPON TYNE
Computing Science. Technical Report Series. CS-TR-1197

Abstract

This paper explores the ALARP Principle. It explains its tenets, presents its history, examines its practical application, and discusses the concepts which underpin it and to which it is related. Whereas the paper's narrative is continuous, each chapter covers its topic fully, is largely self-contained, and may be read in isolation – except that Chapter 4 is a necessary preliminary to Chapter 5. Chapter 1 introduces the subject of risk, the variability of its perception between individuals, and the nature and need of risk-tolerability decisions. Chapter 2 introduces and explains the ALARP Principle, and Chapter 3 recounts its development from a legal perspective. Chapter 4 derives the preparatory processes essential to the Principle's application, and Chapter 5 describes the processes necessary for its central purpose – the making of risk-tolerability decisions. Section 6 offers a discussion of the main risk-based concepts on which the Principle depends and with which it interacts. Finally, Chapter 7 draws conclusions and discusses the relationship between the Principle and the law of the land.

About the author

Felix Redmill has been with the Centre for Software Reliability (CSR) since 1991, when he became the inaugural Co-ordinator of the Safety-Critical Systems Club. He holds degrees in Electrical Engineering and Computation and is a Fellow of both the Institution of Electrical Engineers and the British Computer Society. He has studied, written on, and lectured on the subjects of Software Engineering, Project Management, and Safety Risk Engineering and Management, and his current primary research interest is safety risk.

Suggested keywords

RISK
TOLERABLE RISK
ALARP PRINCIPLE
SFAIRP
SAFETY DECISIONS
LEGAL SAFETY REQUIREMENTS
HEALTH AND SAFETY EXECUTIVE

TECHNICAL REPORT SERIES

No. CS-TR-1197 March, 2010

ALARP Explored

F. Redmill

Abstract

This paper explores the ALARP Principle. It explains its tenets, presents its history, examines its practical application, and discusses the concepts which underpin it and to which it is related. Whereas the paper's narrative is continuous, each chapter covers its topic fully, is largely self-contained, and may be read in isolation – except that Chapter 4 is a necessary preliminary to Chapter 5. Chapter 1 introduces the subject of risk, the variability of its perception between individuals, and the nature and need of risk-tolerability decisions. Chapter 2 introduces and explains the ALARP Principle, and Chapter 3 recounts its development from a legal perspective. Chapter 4 derives the preparatory processes essential to the Principle's application, and Chapter 5 describes the processes necessary for its central purpose – the making of risk-tolerability decisions. Section 6 offers a discussion of the main risk-based concepts on which the Principle depends and with which it interacts. Finally, Chapter 7 draws conclusions and discusses the relationship between the Principle and the law of the land.

Printed and published by the University of Newcastle upon Tyne,
Computing Science, Claremont Tower, Claremont Road,
Newcastle upon Tyne, NE1 7RU, England.

ALARP Explored

by Felix Redmill

Felix.Redmill@ncl.ac.uk

Abstract

This paper explores the ALARP Principle. It explains its tenets, presents its history, examines it practical application, and discusses the concepts which underpin it and to which it is related.

Whereas the paper's narrative is continuous, each chapter covers its topic fully, is largely self-contained, and may be read in isolation – except that Chapter 4 is a necessary preliminary to Chapter 5.

Chapter 1 introduces the subject of risk, the variability of its perception between individuals, and the nature and need of risk-tolerability decisions. Chapter 2 introduces and explains the ALARP Principle, and Chapter 3 recounts its development from a legal perspective. Chapter 4 derives the preparatory processes essential to the Principle's application, and Chapter 5 describes the processes necessary for its central purpose – the making of risk-tolerability decisions. Section 6 offers a discussion of the main risk-based concepts on which the Principle depends and with which it interacts. Finally, Chapter 7 draws conclusions and discusses the relationship between the Principle and the law of the land.

Contents **Page**

Acknowledgment

During the preparation of this paper, parts of it were published as a series of seven articles in the UK Safety-Critical Systems Club's newsletter, *Safety Systems*, between September 2008 and September 2010 – Vol. 18, Nos. 1, 2 and 3, Vol. 19, Nos. 1, 2 and 3, and Vol. 20, No. 1.

CHAPTER 1

Introduction: Risk, Its Perception and Its Tolerability

1.1 Risks of a New and Different Character

Attached to every decision and every action that we take, there are risks. In the pursuit of every goal there is risk to avoid, overcome or accept. The technologies we develop and the systems we create bring not only benefits but also risks. Historically, these have mostly been easily identifiable and limited both spatially and temporally, but in the last century they have broken through these limitations.

Beck [1992] lists features of modern technological hazards that were not traditionally common: they are often invisible and irreversible, they are not restricted to their places of origin but threaten all forms of life in all parts of the planet, and in some cases could affect those not alive at the time or place of an accident. Moreover, many modern hazards are surrounded in considerable uncertainty. Beck goes on to say that to start with they may exist only in the (scientific or anti-scientific) knowledge about them, and that they can be 'changed, magnified, dramatized or minimized within knowledge, and to that extent they are particularly open to social definition and construction'. He maintains that such a large proportion of political and economic resource is being devoted to these new risks that we are heading towards becoming a 'risk society'.

As long ago as 1982, Coates [1982] said that Americans now live in a 'totally human-made world', and that people who build, design, plan, execute, sell and maintain complex systems do not know how they *may* work. This observation points to the complexity of much of modern technology, and to the fact that the causes of safety breaches are not limited to known failure modes. It is even more telling in the light of Galbraith's [1979] warning that we make survival second to production, even though the latter is based on arms which are thoughtfully designed to be able to wipe us all out.

According to Beck [1992], 'we are concerned no longer exclusively with making nature useful, or with releasing mankind from traditional constraints, but also and essentially with problems resulting from techno-economic development itself'. And Gould et al [1988] remind us that 'while modern science and technology are responsible for prolonging and enriching the lives of most people in the world, they are also responsible for shortening the lives of some and for threatening the lives or well-being of all.'

The systems that give rise to modern technological risks are often complex, and two of the attributes that frequently typify them are uncertainty and high potential consequences. When uncertainty is low, it may be possible to estimate the risks with reasonable confidence. Then the question is, what is tolerable, given the benefits? But when uncertainty is high, there is an added gamble attached to a decision to proceed with the technology or system in question. Then, those in favour of progress, or of deriving the perceived benefits, tend to be willing to accept greater risks than others. They argue that the technology should be assumed safe if it cannot be proved otherwise. On the other hand, those more cautious about the risks, or who are more circumspect about the value of the benefits, are likely to argue the opposite – that the technology or system should be assumed unsafe if it cannot be proved to be safe. Perhaps the majority stand between these two, waiting apprehensively and praying for happy outcomes but believing that some day the worst will occur. What is deemed to be tolerable depends on who is doing the deeming.

1.2 The Tolerability of Risk

Kaplan and Garrick [1981] argue that in isolation there is no such thing as acceptable risk; that because of its very nature risk should always be rejected, and it is only the chance of benefits that leads us to tolerate certain risks. At first sight, this might suggest that full societal acceptance of a risk would depend on a universally equitable distribution of both benefits and risks. If everyone stood to gain uniformly from the benefits, homogeneity of opinion might be expected. But (as discussed below) the perception of risk is not consistent among individuals. Moreover, exposure is not uniform across society, and life involves involuntary exposure to risks that others have created and that exist because of benefits to which the individual may not have access or is unable to afford.

Further, the creation of risks, and the making of risk-tolerability decisions, may be stimulated by, among other motives, need, greed, and the desire for power or pleasure. As the influence of each of these forces, and the opportunities to gratify them, differ between individuals, so does the willingness to tolerate specific risks.

Who then should decide which benefits are worth the risks, what risk levels are tolerable, and what criteria [Redmill 2000a] should be used in making judgements?

One answer is that the experts in the technology in question should decide. Typically, engineering and scientific experts consider risk to be defined by the probability of an event occurring and the potential consequences if it did occur, and risk analysis is intended to derive values – quantitative or qualitative – for these two variables.

But the opinion that only technical experts should make risk-tolerability decisions is not universally held. Lowrance [1976] identifies two components of risk management – risk measurement and the judgment of safety – and, though he says that the former is the preserve of the technical experts, he defines the latter as a matter of public values to be resolved in the political process.

If this is accepted, the results of expert risk analyses would not necessarily be deciding factors but, rather, just one source of input to the decision-making process. The other input of consequence would be public values. But how can values be determined, and then represented in the political process, if they are not ubiquitous?

Studies by psychologists, pioneered by Slovic, Fischhoff and Lichtenstein [e.g. 1980, 1985], showed that lay people do not perceive risk in the same way as experts do. Their judgements are influenced by concerns other than purely technical risk characteristics. This finding was not taken to mean that lay people are irrational, but that they implicitly use a broader definition of risk than experts when making judgements about what is of concern to them.

It was found that experts' judgements of risk were based on the expected numbers of fatalities whereas lay people's risk perceptions were founded, in addition, on qualitative components of the risks. One influential factor in perception was found to be 'disaster potential' – i.e. the magnitude of a disaster in a catastrophic year rather than an average year – which particularly affected the perception of nuclear energy. Other characteristics associated with high risk included: involuntary, delayed, unknown, uncontrollable, unfamiliar, potentially catastrophic, dreaded, severe. When correlated together, these were found to be subordinate to three higher-level factors: dread (the fear of a risk as opposed to an ability to live with it), familiarity, and the number of people exposed to the risk.

In addition, since Wynne [1980] showed a significant factor to be trust – whether or not the public has trust in those managing the risks – this has been recognised as crucial (see, for example, Bier's [2001] review).

The work summarised in Slovic's [1992] review of the research into risk perception, and Kahneman, Slovic and Tversky's [1982] compendium on work on human heuristics

and biases, focused on lay people's rejection of risks that the experts proclaim to be small. Yet, the reverse can occur: lay people may accept risks that experts deem, or would deem, to be significant. Whereas there are ample observations of this, there is scope for experimentation to provide empirical data. Mary Ann Evans (*aka* George Eliot) wrote in *Silas Marner*, 'The sense of security more frequently springs from habit than from conviction, and for this reason it often subsists after such a change in the conditions as might have been expected to suggest alarm. The lapse of time during which a given event has not happened is, in this logic of habit, constantly alleged as a reason why the event should never happen, even when the lapse of time is precisely the added condition which makes the event imminent. A man will tell you that he has worked in a mine for forty years unhurt by an accident as a reason why he should apprehend no danger, though the roof is beginning to sink; and it is often observable, that the older a man gets, the more difficult it is to him to retain a believing conception of his own death.' [Eliot 1861]

1.3 Protection of the Public

In the United Kingdom (UK), individuals must make their own decisions about the tolerability of personal risks. But risks that affect others are subject to legislation and regulation.

The UK's Health and Safety at Work, Etc. Act 1974 [HSW Act 1974] requires that risks imposed on others (employees and the public at large) should be reduced 'so far as is reasonably practicable' (SFAIRP), but it offers no guidance on how to determine what is SFAIRP in any given circumstance. What risks are tolerable under the law? What limits should be placed on the risks that may be imposed on others? In some cases, the government legislates fixed limits, for example on exposure to nuclear radiation. But there are so many different plants, products and processes, that it would be impossible to define limits for each under all possible circumstances. Prescriptive rules are therefore impossible, and the Health and Safety at Work, Etc. Act calls on duty holders to take responsibility for their risks – to analyse and assess them properly, to reduce them appropriately, and not to impose on others risks that could reasonably be reduced further. The Act implies a 'goal-setting' regime.

But can all duty holders be trusted? How can agreement be achieved? What equation should there be between risk reduction and its cost? There can be no fixed formula for answering such questions, for answers differ according to circumstances. What is required is a defined approach to answering them – an approach that is both consistent and understandable to all and, importantly, communicated to and understood by those duty holders who impose risks on others. Beyond that, duty holders need not only to be able to apply the defined approach but also to show (for example, in safety cases) that they have asked, and competently and adequately answered, the appropriate questions.

For this, the Health and Safety Executive (HSE – the Regulator in most fields of functional safety) exists. It carries the remit and authority to ensure consistency and provide guidance on safety matters, enforce correction when risks exceed reasonable practicability, investigate accidents, and initiate prosecutions when breaches of the Act appear to have been committed.

1.4 The Health and Safety Executive's Guidance

The HSE designed the ALARP (as low as reasonably practicable) Principle as a guide to making risk-tolerability decisions [HSE 1988, 1992, 2001]. It employs the concept of 'tolerable risk', defining it as indicating 'a willingness to live with a risk so as to secure

certain benefits' [HSE 2001]. As such, a tolerable risk is not necessarily one that would, in the absence of the possible benefit, be judged acceptable. Nor is it likely to be accepted unthinkingly, for it is the price paid for gaining the benefit offered by the system or enterprise in question and, as such, must be tested for appropriateness and value. Thus, the HSE distinguishes 'tolerable' from 'acceptable'.

The ALARP Principle is not difficult to understand (see Chapter 2), but its application is non-trivial (see Chapters 4 and 5), often due to ignorance or uncertainty of the likelihood or consequences of supposed risks, and always because risk-tolerability decisions depend on circumstances, which may change. Further, risks are not necessarily considered tolerable as presented or perceived, and decisions must be made on whether they require reduction before acceptance, and how and by how much they should be reduced. The triangular balancing of risks, the costs of reducing them, and the potential benefits of doing so, is a source of debate, dispute, and even despair for many safety engineers and managers. We can never be certain that our judgement at the time of acting will find favour with a court, sitting later, with retrospective knowledge.

As the regulator's preferred model of how the taking of risk-tolerability decisions may (or even, should) be approached, the ALARP Principle carries considerable influence in the UK – and in some other countries whose laws and regulatory guidance are based on it. Yet it is not always well understood. The following chapters explore the ALARP Principle (Chapter 2), its origins (Chapter 3), its application (Chapters 4 and 5), and the factors that are relevant to its application and on which it depends (Chapter 6).

CHAPTER 2

Explanation of the ALARP Principle

2.1 The ALARP Model

A simple representation of the ALARP model is given in Figure 1, where risk increases up the vertical axis. The two horizontal lines define thresholds and create three 'regions' (or categories) of risk. The lower horizontal line is the 'broadly acceptable' threshold, up to which risk may be considered tolerable without reduction. The upper horizontal line is the 'limit of tolerability' threshold, above which risk is considered unacceptable, except under exceptional circumstances. The absence of definitive numeric risk values at the thresholds suggests the need for calibration of the model, and this is discussed later.

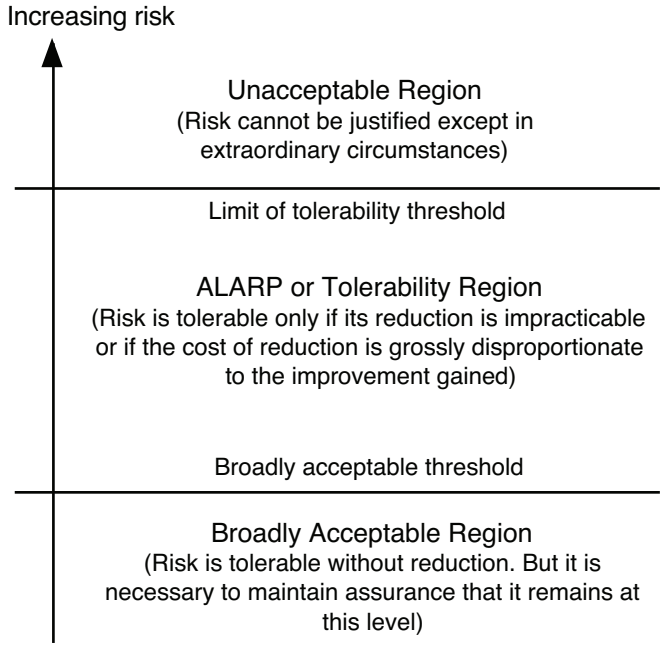

Figure 1: A Simple Representation of the ALARP Principle

According to the Principle, the way in which we think of, and act on, a risk depends on which of the three regions it lies in. (Thus, application of the Principle depends on risk values already having been estimated, and this is discussed further in Chapter 4.) In the 'unacceptable', region, a risk is too high to be tolerable and *must* be reduced – at least into the next region (the ALARP region) – if the system posing the risk is to be deployed. In the 'broadly acceptable', region, a risk is considered to be negligible and may be accepted without reduction – though it should be monitored in case it increases with time and circumstances. The HSE suggests that risks up to this lower threshold equate to those that we accept without concern in everyday life – which is why the word 'acceptable' is used when referring to risks in this region. In the 'tolerability' region, a risk should be reduced until it is 'as low as reasonably practicable' (ALARP), which is the point at which the cost of reducing it further would be grossly disproportionate to the gain achieved by the

further reduction. A risk in this region may be tolerated as it is, reduced to the broadly acceptable threshold or further, or reduced to any intermediate level, depending on the 'practicability' of reducing it.

Decisions in the two extreme regions are, in theory, straightforward (given an estimate of the value of the risk), because their tolerability is defined in absolute terms by the two thresholds. But decisions on risks in the ALARP region are potentially complex, for their tolerability depends on the practicability of reducing them; they require, in each case, comparisons of the gains to be achieved by reducing a risk against the sum of the various costs of effecting the reduction. A number of factors may contribute to the decisions on which mode of risk reduction to employ and how much reduction is reasonably practicable. The steps in the decision process are discussed primarily in Chapter 5 but also in Chapter 4.

The ALARP model provides a framework for risk assessment (or evaluation) and its application depends on risk values having already been estimated. It reminds its users of the need for risk identification and analysis, but it is not a technique for carrying them out.

2.2 Rationale for the Model

The HSE claims that the ALARP model mirrors the way in which we view risk in everyday life, consciously or unconsciously. Some risks we reject out of hand; for example, we would not set out to cross a road if there was a fast-moving lorry bearing down on us. Other risks we accept without thinking; for example, if there is no traffic in sight we cross the road at our leisure. Then there are risks that may or may not be accepted, depending on the benefits to be gained; for example, we may dodge traffic to cross the road if we are in a hurry, but wait for a lull or a favourable traffic light if we are not. If the prize is great enough, we accept a higher and higher risk in an attempt to win it – until a point is reached when the risk is just too great. This point is the 'limit of tolerability' threshold and it differs depending on our personality, the prize, and the circumstances.

This 'everyday' decision-making process is based on three main tests, which the HSE [1992] applies in regulating industrial risks:

a) Is a given risk so great or the outcome so unacceptable that it must be refused altogether?

b) Is the risk, or has it been made, so small that no further precaution is necessary?

c) Has a risk, which falls between these two states, been reduced to the lowest level practicable, bearing in mind the benefits flowing from its acceptance and taking into account the costs of any further reduction?

The ALARP model reflects these three tests and may therefore be applied in any circumstances: it is a general model. Its conscious application can assist us in understanding, assessing, and making decisions about risks, whether they are man-made or natural and whether they are externally imposed or voluntarily taken.

2.3 Application to the Law and Safety Regulation

The ALARP Principle is the regulator's model of how compliance with the Health and Safety at Work, Etc. Act 1974 [HSW Act 1974] should be addressed and assessed.

Safety regulation affects those who, in running businesses and factories, operating plant and equipment, and creating and selling products, pose risks to others – both employees and the general public. The tolerability region of the model recognises that the potential benefits to be gained from businesses, factories, plant, and products may be desirable, useful, and wealth-producing, at least to some people, and the model is

intended to facilitate the reduction to tolerable levels of the risks created in pursuit of those benefits.

For the HSE, the ALARP Principle serves two roles: it models the process that risk creators are recommended to observe in determining the tolerability of their intended risks, and it makes transparent the processes used by regulators in making assessments. It is referred to as a tolerability of risk (TOR) framework [HSE 2001]. 'In the end,' says the HSE [1992, Para 26], 'it is always a question of applying one of the three main tests.'

But who should apply the tests and make tolerability decisions? The HSE says that all stakeholders (e.g. the risk creator, those at risk, and the regulator) should be involved. As acceptance of risk depends on the potential benefits, there are likely to be differences between stakeholders in perception and opinion. For example, if those at risk will not share in the potential benefits of accepting the risk, they are unlikely to be as enthusiastic for its acceptance as those who will reap the rewards. Thus, it is important for all to be involved in determining what is tolerable in given circumstances. (It may be noted here that the regulator is a representative of the general public and acts for its protection.)

In the Health and Safety at Work, Etc. Act, there is no equivalent of the ALARP Principle's limit-of-tolerability threshold. Thus, in theory, there is no upper limit to the risk that may be imposed on others. However, the legal concept of reasonableness applies not only to the practicability of reducing risk but also to its imposition, so it is sensible, in assessing risk, to define a boundary beyond which it would be unreasonable to step – and to retain the threshold when using the model.

2.4 The 'Carrot' Version

Although Figure 1 offers an outline of the ALARP Principle, the HSE uses the 'carrot' model of Figure 2 in its literature. The two models are, essentially, identical, but the HSE's usage has made the superimposed triangle iconic. The original intent of the breadth of the triangle, which increases with risk value, was to indicate the obvious criterion that the higher the risk the higher would need to be the cost of risk reduction for the unreduced risk to be justified as tolerable. This is not to say that the reduction of a significant risk necessarily requires significant expenditure. Indeed, in many cases, the cost of a risk-reduction action can be small, or even trivial. For example, a ship's journey could be navigated riskily via a narrow gap in a reef, or at much reduced risk by sailing round the reef at a cost of only a few extra kilometres.

The ALARP principle implicitly recognises that zero risk is not an option in any enterprise. In the context of safety risk, the 'broadly acceptable' threshold is sometimes taken to indicate the 'safe' level. However, it should be noted that the threshold is not set at zero risk and, even at the point of the carrot in Figure 2, the risk is considered negligible rather than non-existent. Similarly, the 'limit-of-tolerability' threshold is sometimes considered to define 'unsafe' – but it is not an indicator of certain catastrophe or even of maximum risk. Indeed, in the HSE's recent representations of the 'carrot diagram' (e.g. [HSE 2001]), the threshold is not defined, and the change of region is represented by a graduation of colour, showing that the point of change is not definitive but subject to judgement. Further, between the two thresholds, in the tolerability region, risk may be tolerable or not, depending on the circumstances. Thus, safety is not absolute but relative. To refer to something as 'safe' is to say that we consider it 'safe enough' or 'appropriately safe' *in the circumstances*. Other people, or the same people under different circumstances, may judge the same risk differently – and this shows the advantage of using a model that facilitates consistency in risk assessment.

In the absence of a model, with only language for describing the assessment of risk, not only would tolerability decisions be arbitrary, but also communication would more easily

be misunderstood. Thus, the ALARP model, with its defined thresholds, facilitates not only judgement but also communication.

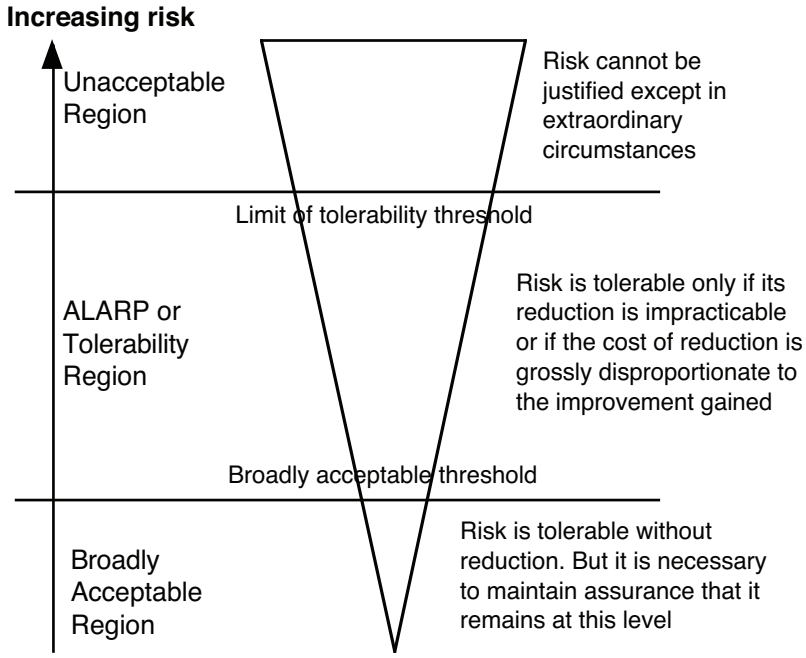

Figure 2: The Health and Safety Executive's ALARP Model

2.5 Calibration of the Model

The HSE's recommendation to the nuclear industry [HSE 1988, 1992] was for the risk value at the 'broadly acceptable' threshold to be one in a million (10^{-6}) per year and that at the 'limit of tolerability' threshold to be one in a thousand (10^{-3}) per year for voluntary employees and one in ten thousand (10^{-4}) per year for the public who have the risk imposed on them 'in the wider interest of society'. Both of these risk values refer to the death of an individual *from any particular cause* – which is a reminder that it would not be sufficient for a risk creator to claim that the average of a number of risks was ALARP; it is necessary to demonstrate that every individual risk (i.e. the risk from every particular source in a system) is, or has been reduced, ALARP.

Since the HSE's initial publication for the nuclear industry, these values have been taken to be more general recommendations, used in many industry sectors and safety-related applications – although the HSE itself points out that 'these criteria are merely guidelines to be interpreted with commonsense and are not intended to be rigid benchmarks to be complied with in all circumstances. They may, for example, need to be adapted to take account of societal concerns or preferences.' [HSE 2001]

If indeed the model reflects the way in which we typically make risk decisions, it is a tool for general use and must be calibrated for each separate application. The thresholds may be moved up or down, and the unit of loss may need to change. In non-safety-related applications, units of currency would usually replace the numbers of deaths. Moreover, fatality is not the only criterion of safety; the potential consequences of many safety risks are more likely to be injury than death. Thus, a failure of the ALARP model, as presented by the HSE, is that it is only calibrated in terms of fatalities and that it does not address, or even acknowledge, other harmful outcomes of safety risks.

2.6 Postscript

The ALARP model is not the only philosophy of risk tolerability, and others are used in other parts of the world. Though mostly applied to safety risk, it defines a general way of thinking and a framework that facilitates risk communication. Its principles are not difficult to understand, but its application requires the collection of a great deal of information and substantial judgement in analysing it.

CHAPTER 3

Legal and Historical Background

3.1 Reasonableness in the English Common Law

Some legal offences are defined by prescription, for example a stipulated threshold (such as a speed limit), which it is illegal to exceed. But in the majority of cases there is no definitive prescription, the boundary between criminal and acceptable behaviour depends on the circumstances, and there may be plausible arguments on both sides – for both guilt and innocence. Traditionally in such cases, under the English Common Law, judges attempted to determine what is 'reasonable' in the circumstances. And their way of doing this was to appeal to the concept of 'the reasonable man', asking, in effect: How would a reasonable person determine this case? In order to suggest that this notional appeal was not to someone above the status of the appellants, but, in fact to an 'ordinary person' and, by implication, their peer, judges often referred to 'the man on the Clapham omnibus' – i.e. to the common man.

The tradition of judges determining what is 'reasonable' – in a manner understandable and acceptable to those engaged in the case – led to both legal precedents and statutes that hinge on the interpretation of that word. For example, it is permitted in law to use reasonable force to repel an assailant, or a burglar, and in such cases the criterion of reasonableness is proportionality to the threat imposed by the aggressor.

In the Corporate Manslaughter and Corporate Homicide Act 2007, a breach of a duty of care is defined as a 'gross' breach 'if the conduct ... falls far below what can *reasonably* be expected of the organisation in the circumstances.' Responsibilities of the court must therefore be, first to understand the circumstances of the case, second, to determine what is or would be reasonable, and, on the results of these findings, to pass judgement.

Thus, reasonableness has traditionally been, and remains, a criterion of judgement in courts of law. At the same time, it is a variable whose meaning must be determined against criteria extracted from the circumstances of the case.

In the context of risk, what is reasonable depends, among other variables, on what is to be gained by taking the risk. But because the potential benefits are usually not equally distributed between those creating a risk and those having it imposed on them, disagreements are likely to arise and to come to court for resolution.

3.2 Edwards vs. The National Coal Board

An explanation of how to conform to the obligation to do what is 'reasonably practicable' to reduce risk was enunciated in a case in 1949 which gave definition to the way of thinking that is now referred to as 'risk-based' and which eventually resulted in the development of the ALARP Principle. The plaintiff, Edwards, alleged negligence by the National Coal Board, claiming that organisations have a duty of care to their employees and that the Coal Board failed to discharge this duty by not adequately shoring-up its mine shafts so as to protect its employees from harm. In an attempt to explain how 'reasonable practicability' should be determined, the judge, Lord Asquith, said: 'A computation must be made in which the quantum of risk is placed on one scale and the sacrifice, whether in money, time or trouble, involved in the measures necessary to avert the risk is placed in the other, and that, if it be shown that there is a gross disproportion

between them, the risk being insignificant in relation to the sacrifice, the person upon whom the duty is laid discharges the burden of proving that compliance was not reasonably practicable.' [Edwards vs. The National Coal Board 1949]

3.3 The Robens Report

The next significant development in the evolution of ALARP occurred in the early 1970s, when the government initiated a fundamental review of the regulation and management of occupational risks. The resulting Robens Report [Robens 1972], named after the committee's chairman, recognised that the extent of occupational risk was such that health, safety and welfare at work could not be ensured by an ever-expanding body of legal regulations enforced by an ever-increasing army of inspectors. It therefore recommended that the primary responsibility for ensuring health and safety should lie with those who create risks and those who work with them. The report further recommended that to create such an environment the law should provide a statement of principles and definitions of duties of general application, with regulations setting more specific goals and standards.

The Robens Report's recommendations for both legislation and regulation were implemented shortly afterwards. The Health and Safety at Work, Etc. Act 1974 [HSW Act 1974] was introduced. It contained the basic legislation and it allocated the responsibility for the regulation of health and safety risks to the Health and Safety Commission (HSC) and the Health and Safety Executive (HSE). What is to be done about any particular risk is defined in the Act, not absolutely but in terms of proportionality between the costs of reducing the risk and the benefits to be gained by doing so. According to the Act, risks should be reduced 'so far as is reasonably practicable' (SFAIRP), which demands no more than is reasonable and possible to deliver, with the onus being on the imposer of a risk to justify its residual value by demonstrating that further reduction would not be reasonably practicable. The principle of reasonable practicability, as well as that of gross disproportion, formerly enunciated in case law [Edwards vs. The National Coal Board 1949], thus became enshrined in statute.

It should be noted that the Act does not offer a justification for the acceptance of huge risks on the grounds that their reduction would be prohibitively expensive. Yet, unlike the ALARP model [HSE 1992], which includes a 'limit of tolerability' threshold (see Figure 2), the Act does not explicitly state that there is a threshold beyond which risk should not be accepted. However, accepting a very high risk demands exceedingly strong justification, particularly with respect to reasonableness, and so, in attempting to comply with ALARP, the assumption of a 'limit of tolerability' threshold is a wise precaution. If the value on the risk scale of this threshold has not been determined on an industry-sector basis, it should be determined for the particular case in question, and demonstration of good judgement in this could subsequently be evidentially significant in a court of law.

3.4 Publication of ALARP

In 1987, a Public Inquiry into the proposed Sizewell B nuclear power station [Layfield 1987] recommended that the HSE should formulate and publish 'guidelines on the tolerable levels of individual and social risk to workers and the public from nuclear power stations'. Not only was the concept of such guidelines on tolerable risk new, so was Layfield's recognition that risk analysis involves both scientific assessment and social values. One of his recommendations was that 'the opinion of the public should underlie the evaluation of risk; there is at present insufficient public information to allow

understanding of the basis for the regulation of nuclear safety'. The result was the publication, by the HSE, of the ALARP Principle as guidance to the nuclear industry on the approach to reducing risks SFAIRP [HSE 1988, updated 1992].

Since the 1980s the HSE has not only given a great deal of advice to risk creators about their duties of care, but also taken many initiatives to inform the public both of what risk management should entail and how the HSE itself goes about its duty of assessment and regulation (in particular [HSE 2001]).

Since the ALARP Principle came into being in 1988, its application has been extended by the HSE from its original context in the nuclear industry. It is now the model on which regulation across all industry sectors is based, and it sets the terms of reference for making risk tolerability decisions by all who would impose risks on others. Its precepts are not hard to understand, but though its model (Figure 2) appears simple, it is not trivial to apply in practice, for it depends heavily on interpretations, in the circumstances, of what is reasonably practicable.

3.5 Self-regulation

The system defined by ALARP, of risk determination, assessment, reduction and management by those who create risks, amounts to self-regulation – as had been recommended by the Robens Report. Thus, the Act is not prescriptive but goal-based. The creator of a risk must set and justify the risk-reduction goals and demonstrate that they have been achieved; an independent assessor must assess and approve (or reject) the goals and verify their attainment. The regulator, if involved, has the right, on behalf of the public and the law, to assess the discharge of both the risk-creator's and the assessor's responsibilities.

3.6 Practicability not Possibility

The HSE [2008b] has pointed to the words of Lord Asquith in his judgement in the case of Edwards vs. the National Coal Board: '"Reasonably practicable" is a narrower term than "physically possible" …'. To be practicable, risk-reduction action must, of course, be possible, but what is intended is that the cost of reducing the risk should be taken into account, and that the result should be reasonable in the circumstances. Borrowing heavily from the words of Lord Asquith, the HSE (1992, Para 32) defined 'reasonably practicable' in the following terms: 'Legally speaking, this means that unless the expense undertaken is in gross disproportion to the risk, the employer must undertake that expense'. However, recognising that there will always be some risk and that the question is, how much, the HSE goes on to say that 'reasonably practicable' also means that there is a point beyond which the regulator should not press the operator to go – 'but that they must err on the side of safety'.

3.7 Application in Law

The Health and Safety at Work Act is unusual in English law in that it places the onus on the defendant to demonstrate innocence – by proving that everything was done, so far as was reasonably practicable, to ensure safety by the reduction of risk. When a case is brought to court, the things that have been done, or not done, by the accused are matters of fact. Whether it was reasonable to have done them, or not done them, are matters of judgement, to be made by the court.

When risk is involved, reasonableness depends on what it was practicable to do to reduce the risk, or risks. And practicability comprises three factors: what it was possible to do and, for each possibility, the costs of doing it, and the benefits gained. As the costs become greater, there may come a point at which they become 'grossly disproportionate' to the benefits gained, at which point a defendant demonstrates that further reduction would not be reasonably practicable and that the risk in question was reduced SFAIRP.

The point of gross disproportionality varies for the different possible courses of action to reduce a risk, so a court would need to be convinced that all avenues had been explored before a decision was made on what should be done. Further, as the point of gross disproportion is a matter of judgement, we cannot be certain that our determination of it now will find favour with a court later; therefore, it is wise to err on the side of caution, particularly when the risk is high. (The HSE suggests that 'a disproportion factor (DF) of more than 10 is unlikely' and that 'the duty holder would have to justify use of a smaller DF' [HSE 2008e].)

In addition, although the ALARP model represents a sensible way of attempting to comply with the law, it is not a part of the law. The values of risk at its two thresholds are not defined in law. Indeed, the thresholds themselves do not appear in the Act; they exist only in the HSE's model in order to emphasise the fact that different types of decision-making occur at different levels of risk (see Chapter 1 for the three tests defined by the HSE). Even then, they are matters of calibration. The HSE's recommendations for calibration in the nuclear industry may offer sensible values for the thresholds, but there is no guarantee that a court will take these to be definitive in any given case.

Moreover, where the principle of reasonableness is invoked, no case provides an absolute precedent for others; each must be judged on its own circumstances. Suppose, for example, that a drug has a side effect that causes five deaths per million of those who take it. If the drug is efficacious and saves many other lives, and the cost of reducing the risk is significant, a court may hold that it is reasonable for the manufacturer not to reduce it further. But suppose a vegetable supplier sprays cabbages with an additive that makes them look fresher but which causes the deaths of five per million cabbage eaters. Would a court accept that there is no need for the supplier to reduce such a risk on the grounds that it is too expensive to do so? It is at least conceivable that the court would find that there is no justification for the risk, small as it is, and that it should be avoided altogether. It is also possible to imagine a situation in which it is deemed unreasonable to reduce a risk that lies in the model's tolerability region – if the method of reduction has an unintended consequence that throws up a new and greater risk. The ALARP model is a tool that should be used with understanding and discretion.

3.8 The ALARP Principle's Role as a Bridge

The ALARP Principle arose out of legal concepts, precedents and requirements. It is not a technical concept, and its purpose is not to meet engineering requirements. Rather, it is an attempt to bridge the gap – sometimes perceived as an abyss – between those responsible for the safety of technological systems and the requirements placed on them by UK law. It is a framework to facilitate good judgement on safety-related risk-reducing actions (technical or otherwise) in order to comply with the law. In the matter of interpreting the law, it is a bridge from the legal to the engineering; in fulfilling the requirement for risk reduction to be SFAIRP, it is a bridge from the engineering to the legal.

CHAPTER 4

Essential Preliminaries to ALARP Decision-making

4.1 Background

Many who apply the ALARP Principle do so with regard to relatively well understood hazards, in the context of well understood systems, with uncertainty of potential consequences low, and concern unlikely to be expressed by the public and media. Under such circumstances, risk control is often considered to be a simple process – of hazard identification and analysis, followed by appropriate risk reduction, with the ALARP model merely a yardstick against which to confirm compliance with the law. But all users are not in this camp. Many are in fields where there is considerable uncertainty of risk probabilities, and even consequences, relative ignorance of the relevant science or technology and, perhaps, active public and media concern. Then, because there is likely to be a keenness to employ the technology or proceed with the intended project, there is also an urgent need for risk-tolerability decisions. Examples may be found in the food and drugs industries, the planning of action on climate change, biotechnology, and many other fields. In such cases, risk-tolerability decision-making is not a mere nicety in an already well understood process of risk management. And, because in such circumstances there is less information available than is necessary for confident decision-making, ALARP testing must be seen by the public to be conducted thoroughly and honestly.

Decision-making in simple situations also encourages the notion that everything for the ALARP Principle's effective application is already in place. But its use must include the creation of infrastructure and of processes appropriate to the most demanding cases. Information about the risks must be derived, with bounds on any uncertainties in their estimation and statements of the levels of confidence in their accuracy. Further, societal concerns should be explored.

This chapter outlines the processes that are essential to carrying out these tasks, prior to the effective application of the ALARP Principle. The extent to which these general processes may be tailored in circumstances where both ignorance and uncertainty are low is, then, a matter of judgement.

The processes are summarised in the flow chart of Figure 3 (which is extended in the next chapter to include ALARP testing itself). They are portioned into three subdivisions in the figure and described in the corresponding sections of text.

4.2 Creating the Infrastructure

A decision-making infrastructure is a strategic tool for use throughout a project. With tailoring of processes, it can be made applicable in other projects, and in the case of workplace risks it is likely to be useful at any time in the future. An infrastructure alone offers no guarantee of good decisions, for an inadequate hazard identification or analysis would provide an unsound input of information. However, without an appropriate infrastructure, confidence in risk-tolerability decisions must be low. The creation of an infrastructure for making such decisions is as important as the decision-making itself. The bracketed numbers in the following sub-headings relate to the corresponding processes in Figure 3.

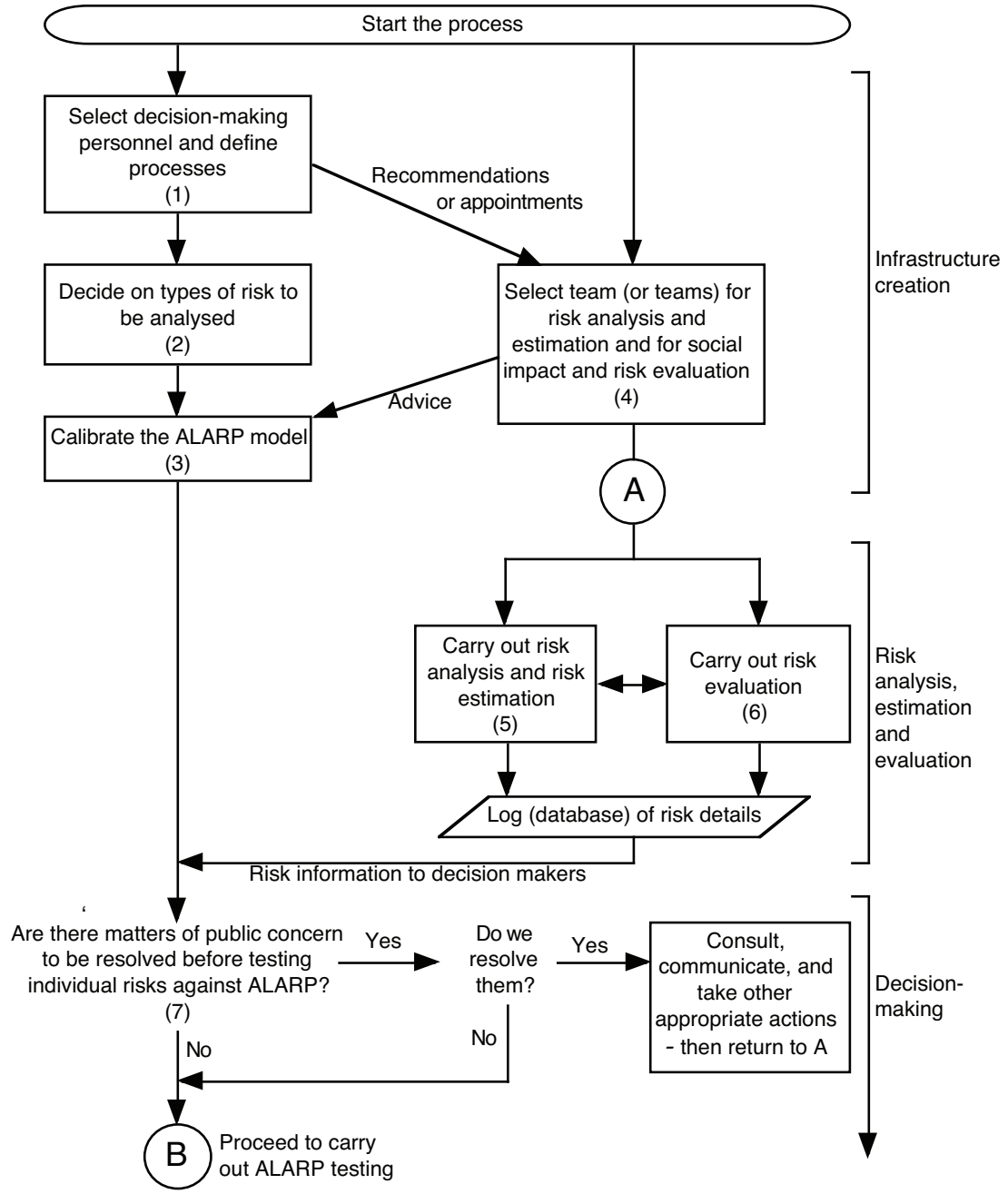

Figure 3: A Flowchart of the Processes Preliminary to Making ALARP Decisions

4.2.1 Decision-making personnel and processes (1)

The IEC [1998] calls for risk tolerability to be determined by discussion among stakeholders, including regulators, those posing the risks, and those exposed to them. The HSE [1992] says that all affected parties should be involved. In some cases, e.g. of workplace risks, participants may be limited to the employer and employees (or their representatives), but in many projects – and in policy decisions in which risk features –

there are several stakeholders among whom responsibilities and activities may be distributed.

If the local public is to be involved (e.g., via focus groups), a long period of communication and trust-building may be required.

Participants in ALARP decision-making should therefore be identified early, and appropriate processes defined, both for decision-making itself and for communication between those involved in it. The participants should be advised of the processes and, if necessary, trained in their operation. What is done should be documented in the safety case, for the definition of appropriate processes and the selection of appropriate decision-makers could later be evidentially significant.

4.2.2 What types of risk? (2)

The Health and Safety at Work, Etc. Act [HSW Act 1974] and the Health and Safety Executive [HSE 2001] address safety risks to people. However, other legislation (such as the Control of Major Accident Hazards Regulations [COMAH 1999]), as well as society, is also concerned with risks to the environment. Indeed, harm to the environment, such as the inappropriate disposal of toxic substances (in the sea, in rivers, or on land) can pose risks to people, directly as well as indirectly, and in both the short and long terms. Further, some risks are indirectly (and not, at first, obviously) safety-related – for example, security risks are both critical in themselves and may create safety vulnerabilities.

But a modern risk assessment process should not necessarily be restricted to risks to the safety of humans and the environment. It may also be appropriate to address others, such as financial risks and those of loss of reputation (for example, following an environmental disaster). It is therefore the responsibility of stakeholders to define what types of risk are to be included in the analysis and assessment processes. Non-safety risks are not subject to health and safety regulation or HSE guidance, but in all cases the safety implications should be identified and analysed.

This discussion does not include project risks. The project-risk register should be separate from the logs of system-based risks, for it is addressed by different people, over different time scales.

4.2.3 Calibrating the ALARP model (3)

The two thresholds on the ALARP model (see Figure 2) must be defined (for an industry sector, a company, a particular type of work, or a project). The HSE's [1992, 2001] recommendations to the nuclear industry for risk to individuals (one death per million persons per annum for the broadly acceptable threshold and, for the limit of tolerability threshold, one death in a thousand per annum for relevant workers and one in ten thousand for the public at large) may be considered reasonable guides in other circumstances, if the risk involved is of death, but the decision should explicitly be made.

When the risk is not of death, calibration needs to be in terms of the particular risks. For factory workers whose equipment risks may be of damage to a hand, or of impairment to hearing, it may, for example, be in terms of the acceptable and intolerable numbers of accidents per worker per year. (Such non-death risks are typically addressed by relevant good practice.)

In the case of risks to the environment, losses are, typically, expressed in financial terms – though there are difficulties in credibly determining these (as discussed later) because, in many cases, market values do not apply.

In all cases of safety risks, decisions should be justified and the justifications, with supporting evidence, documented, preferably in a safety case. (It should be noted that modern practice considers risks to the environment under the heading of 'safety risks'.)

As the ALARP model is general-purpose, it may be employed to reflect the commercial risks to the company if a project should fail – for which purpose it would be calibrated in monetary units. But, of course, this type of risk is not subject to the law or to HSE regulation.

Whatever the types of risk, there may be different opinions on calibration among the decision-making stakeholders. But, clearly, for each type, each threshold can only have one value, so compromises must be reached.

And there's a further point to be made. Because risk tolerability is dependent on context, it is necessarily dependent on the value of the potential gains for which risk is being taken. And these may be reduced, or changed, whenever the objectives of an enterprise are changed, or if it becomes clear that they cannot be met in full. Under such circumstances, the ALARP model's calibration – and the organisation's policy on risk tolerability for the enterprise – should be reviewed.

4.2.4 Selecting teams for risk estimation and evaluation (4)

Even in small companies it is expected that risks must be estimated by persons having an understanding of both the system under consideration and risk analysis. The more significant the risks are, or may be, the greater the requirement for suitably qualified and experienced personnel. The selection of a team to identify and analyse the hazards and estimate the risks is therefore crucial to effective application of the ALARP Principle – and justification of appropriateness and adequacy of the team should be documented in the safety case. In certain instances, such as when uncertainty is high and there is, or may be, significant public concern, it is wise to include independent experts.

It should be remembered that the risk analysts are seldom the makers of decisions on risk tolerability (though they may provide advice – see Figure 3). They should therefore possess the ability to frame and communicate the risk information [Redmill 2007] and, crucially, its limitations (by including confidence and uncertainty bounds) so that the decision makers can understand it.

In addition, there needs to be consideration of the social impact of the risks in question – and of their perception in the eyes of a public who may be ignorant (or misinformed) of the technology on which they are based. There have been many instances in which the lack of such consideration, in the policies and actions of both government (e.g. bovine spongiform encephalopathy (BSE)) and engineering organisations (e.g. Shell with the disposal of its Brent Spa oil platform), has resulted in mistakes, conflicts and bad publicity that could have been avoided. For well-understood systems and their hazards, the analysis team may include the expertise to carry out an adequate study of possible social impacts. However, when uncertainty is high and public and media scrutiny are likely, a separate team, which includes specialist experts, may be necessary.

4.3 Understanding the Risks

The Advisory Committee on Novel Foods and Processes found that 'scientific and consumer issues are best settled side by side, not consecutively, as used to be the case. The previous approach: "First sort out the science, and then look at the consumer issues" simply does not work.' [Burke 1997] The Department of Environment (DoE) [1995] delivered a similar message. Its diagrammatic illustration (see Figure 4) shows four components of two studies being conducted in parallel. Whereas the DoE's risk estimation was done by technical risk analysts, its risk evaluation (to use their term) required economists and sociologists. Reflecting these lessons, Figure 3 shows risk estimation and evaluation being carried out separately but concurrently – with collaboration between

them – and it is recommended that in all cases the two be completed and their results combined and considered together before a project is proceeded with and presented to the public.

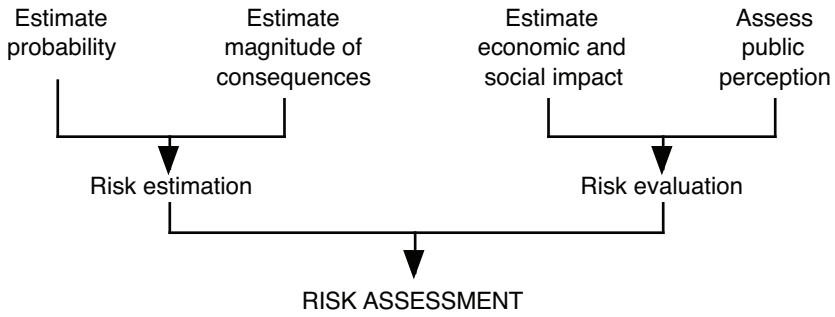

Figure 4: Risk Assessment Based on Both Risk Estimation and Evaluation (Department of Environment 1995)

4.3.1 Risk analysis and estimation (5)

ALARP decisions implicitly depend on the assumptions that risks, and the degree of their reduction, can be determined accurately. If these assumptions were valid, ALARP decisions would be a simple matter of comparing accurately derived numbers. But risk is of the future and, therefore, may be estimated, more or less accurately, but not directly measured. All risk values are estimates. In cases of well-understood hazards, the estimates may be relatively accurate, but in all cases there is uncertainty. Estimates of the future may be based on either a model or statistical inference and, when there is insufficient data for the latter to be reliable and insufficient knowledge of underlying mechanisms for confident construction of the former, predictions are subject to considerable uncertainty – as has been the case, for example, in the field of climate change. Such uncertainty should be recognised and acknowledged.

The details of risk analysis are outside the scope of this paper, but the following points are worth noting.

- The risk-analysis process must include hazard identification, hazard analysis, risk estimation, and the estimation of levels of uncertainty and confidence.
- Hazard identification is crucial and must be thorough, for hazards not identified are not analysed and the risks posed by them are not reduced.
- Identification should address hazards that cause losses not only during normal operation but also at all other times, 'including construction, commissioning, operation, maintenance, foreseeable modifications and eventual decommissioning or disposal' [HSE 2008c].
- The identification process should not only uncover the obvious hazards, and those previously observed, but also those that are 'reasonably foreseeable' [IEC 1998].
- Foresight is not reasonable if it is merely that of an intelligent layman; it must be appropriate to a professional in the field and so requires an appropriately selected team.
- All hazards, those identified in a formal process and those discovered in other ways, should be included in a hazard log and analysed.
- Hazard analysis demands the acquisition of information that contributes to developing an understanding of the likelihood and potential consequences of each hazard maturing into an accident, the synthesis of the information, and the analysis of the results.
- Whether quantitative or qualitative, hazard analysis is an estimation process and can

never be considered wholly accurate. Indeed, in some cases, numeric risk values can be without foundation and misleading.

- Risk estimates should be based on no added risk-reduction functions being in place. This avoids false, over-optimistic assumptions, and it facilitates more reliable determination of the safety functions that are required to reduce risks ALARP.
- Every risk-mitigation proposal should be justified by evidence that demonstrates, with high confidence, that it would in fact reduce its target risk ALARP.
- All risk estimates should be accompanied by bounds on confidence and uncertainty.
- The effort invested in hazard identification and analysis and risk assessment should be proportional to the significance of the hazards [HSE 2004].

4.3.2 Risk evaluation (6)

When risks are novel, technologies new, or uncertainty high, public concerns are often raised – and amplified by the media. Then, to proceed without public consent could be suicidal to projects, businesses, government policies, and even governments. Examples in recent years include the BSE incident, disposal of the Brent Spa oil platform, the sale of foods derived from genetically modified plants, stem-cell experimentation, actions (and non-action) on climate change, control of diseases in embryos, and many others. Studies on potential economic and societal impacts need to be carried out and are shown in Figure 3 as being concurrent and collaborative with technical risk analysis.

4.4 Decision-making

In the present context, decision-making may be considered as consisting of two types. First there are the decisions on what to do about safety risks, based on their positions on the ALARP model. Then there are the more strategic decisions that may need to be made if the risks in question are, or may become, of significant public concern. If the public – or the media – perceives the overall risks of an enterprise to be unacceptable, and the public has not been reassured (even if the experts on the enterprise believe the risks to be tolerable), it is wise to defer deployment of the system – or postpone the enterprise – until communication, consultation, and other necessary actions have achieved resolution and acceptance in the public mind. Compliance with the law, as laid out in the Health and Safety at Work Act, only requires the first type of decision; forcing businesses to take social concerns into account in making safety decisions would increase the likelihood of their becoming risk averse, over-engineering their systems, incurring excessive costs, and becoming uncompetitive. Yet, the wellbeing of a company, and even its survival, may depend on attention to the second type of decision; businesses could suffer even greater costs as a result of public opprobrium.

In the BSE, Brent Spa, and genetically modified foods sagas, the UK government, Shell, and the seed producers, respectively, all protested that risks were tolerable (or, indeed, insignificant), but public scepticism (referred to by the HSE as 'societal concern') nevertheless did great damage to those bodies and their projects. Further research showed that Shell's claims were correct – and the leading protester, Greenpeace, quickly acknowledged this – but the damage had already been done. Shell later made a point of announcing that the event had taught them of the importance of taking public concern into consideration in their decision-making (e.g. [Wilkinson 1997]), and the decision point (7) on Figure 3 reminds those engaged in risk-tolerability decision-making of this. When the answer to the question at (7) is 'No', or when a conscious decision has been taken not to act on matters perceived to be of public concern, it is then appropriate to commence ALARP testing of individual risks – which is the subject of the next chapter – and why the

bracket around this portion of the figure is not closed.

4.5 Summary

Effective use of the ALARP Principle requires it to be applied, within an appropriate decision-making infrastructure, to risks whose values have been derived expertly and thoroughly. Only then can its results be credible in a safety case – or in a subsequent court of law. Further, an organisation's objectives and projects, or even its future, may depend on taking public perceptions of their risks into consideration, so these should be addressed in the decision-making process. The flow chart of Figure 3 shows these processes; it is general and covers all circumstances, but it may be tailored for those in which ignorance and uncertainty are low. How this is done is a matter of judgement by an organisation's or project's decision-makers.

CHAPTER 5

Making ALARP Decisions

5.1 Preamble

This chapter addresses the central purpose of the ALARP Principle: making risk-tolerability decisions. Drawing on HSE literature, it presents a route through the process by means of the flowcharts of Figures 5 (which continues from Figure 3) and 6. The preliminary processes previously explained include the estimation of risk values, so at the point of entry (B) to Figure 5, it is already known where on the ALARP model (Figure 2) each risk is placed.

Hazard analysis and risk assessment should be carried out at a number of points in the life cycle of a system, and some of these will include ALARP decisions. The goal is for all risks to be ALARP when a system is deployed and subsequently during deployment, and it is the responsibility of the system's stakeholders to ensure this. Which stakeholder carries primary responsibility depends on circumstances; further, the distribution of responsibilities differs between stakeholders' organisations. Thus, this chapter addresses what needs to be done in making ALARP decisions, but not exactly who should make them or at what point in a system's life cycle they should be made; the general term 'duty holder' is used.

The regions of Figure 5 designated by the large square brackets on the right are addressed in corresponding sections of the text below, and the text should be read with reference being made to the figure.

The numbers in round brackets in the figure (which continue sequentially from those in the flowchart of Figure 3) are indexed by their equivalents in the headings and text.

5.2 Some General Requirement

The ALARP Principle requires duty holders not only to reduce risks appropriately but also to judge what reduction is appropriate. Making such judgments demands knowledge and competence, and duty holders are expected to possess them. A further requirement is for organisational, and not merely individual, safety consciousness and behaviour. Senior management must set appropriate safety policy, define and nurture a proper safety culture, promulgate appropriate standards and practices, and manage the selection and training of competent, safety-conscious workers. In all cases, these are fundamental requirements, and their absence necessarily raises questions about the credibility of an organization's ALARP claims.

5.3 Risks Outside the Tolerability Region

Although the ALARP Principle covers the entire ALARP model (see Figure 2), the term 'ALARP decision' is often applied to the consideration of risks whose estimated values lie within the model's Tolerability Region – though it should be remembered that such comparisons depend on the calibration of the ALARP model, as discussed above. As well as addressing risks in the Tolerability Region, decisions must be made about those lying outside of the two thresholds.

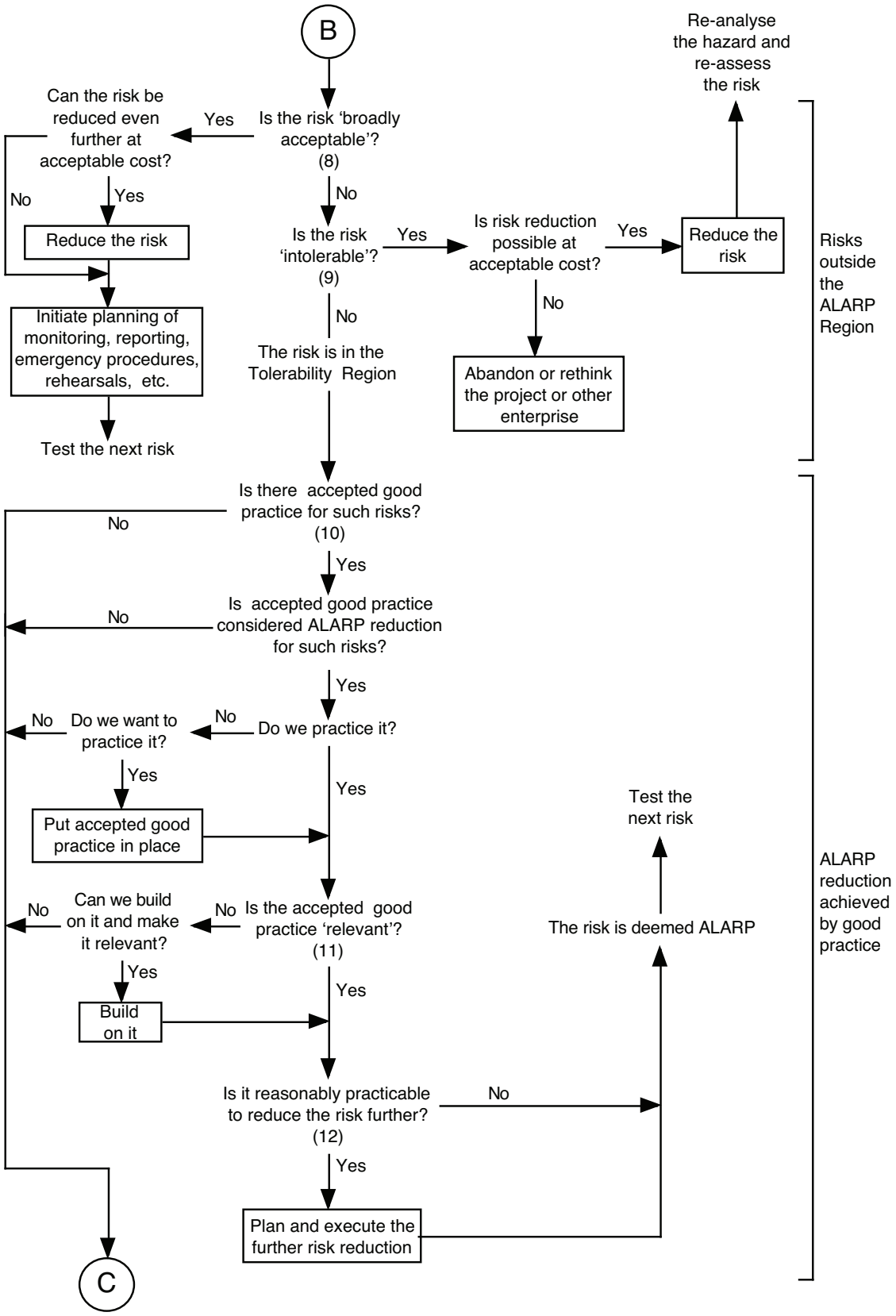

Figure 5: Decisions for Risks Outside the Tolerability Region and for which Good Practice Provides Reduction ALARP

5.3.1 Broadly Acceptable risks (8)

Even risks already in the 'broadly acceptable' region may be reduced further, and they should be if reduction can be achieved at reasonable cost. For example, the risk of loss of life due to fire in a building may be deemed broadly acceptable if the walls are of concrete to prevent a fire starting and doors are fire-proofed to stop it spreading. But further reductions could be achieved by introducing a smoke detection system, placing a fire extinguisher in every room, putting evacuation notices on the walls, including the rescue service's telephone number on the notices, conducting evacuation rehearsals, and more.

Should these additional risk-reduction measures be taken? Almost certainly some or all of them should if their additional costs are small – and particularly if the building is open to the public. They should not be discarded as unnecessary just because the risk is already deemed to be low, but should be assessed for their value and cost.

Once it is decided not to reduce a broadly acceptable risk further, plans should be put in place to monitor it to detect any increase over time, due to change in either the system or its environment.

5.3.2 Intolerable risks (9)

A system must not be deployed if it is deemed to carry an intolerable risk. If the risk cannot be reduced, or if it cannot be reduced at a cost acceptable to the duty holder, either the objectives of the project must be altered so as to avoid the hazard that throws up the intolerable risk, or the enterprise must be abandoned. Continuation is therefore not merely a safety matter but also a commercial one; the question arises: is it worth reducing the risk sufficiently to save the project?

If the risk is reduced into the Tolerability region, the new residual risk must be resubmitted to full analysis, starting with hazard identification, because it is not unusual for risk mitigation activity to throw up new hazards.

5.4 Good Practice

The HSE reports that in most situations ALARP decisions 'involve a comparison between the control measures a duty-holder has in place or is proposing and the measures we would normally expect to see in such circumstances, i.e. relevant good practice' [HSE 2008a]. Thus, great reliance is placed on good practice, which the HSE [2008d] defines as 'the generic term for those standards for controlling risk which have been judged and recognised by HSE as satisfying the law when applied to a particular relevant case in an appropriate manner'. In short, it is inefficient and costly to re-invent risk-reduction measures for the same equipment or activities, and it makes sense to employ measures that have already been proven.

Good practice may be documented, e.g. as HSE guidance, an approved code of practice, or a standard; or it may be unwritten and still recognised, e.g. if it is 'well-defined and established standard practice adopted by an industrial/occupational sector' [HSE 2008d].

The HSE refrains from requiring the use of 'best' practice for various reasons. For example, because there is seldom agreement on what practice is best; because what may be best in one circumstance may not be best in a slightly changed setting; and because a code of practice that is new but better than all others is unlikely to be widely used, standardised, or even widely known. However, a conscientious duty holder will strive to attain best practice, whatever it is perceived to be.

According to the HSE, the use of accepted good practice is, in most cases, a sufficient

basis for a claim that a risk is ALARP, so it makes good sense for a duty holder to explore the appropriateness of good practice. The question is therefore asked, at point (10) in Figure 5, whether accepted good practice exists for the management of the sort of risk that is under consideration – and then, if it does exist, whether, in general, it is considered to reduce such risks ALARP. Then, as it is crucial that the good practice is actually practiced by the duty holder, the further question of whether the stakeholder's organisation uses it (or will now use it) is asked. The answer 'no' to either question precludes an ALARP claim on the basis of good practice.

But even when the answer to these questions is 'yes', the further question of relevance arises – at point (11) in Figure 5. Practice only becomes *accepted* as *good* after proven success in use with particular equipment in defined circumstances. The amalgam of rules, tasks and competences that compose accepted good practice in one circumstance would almost certainly require change if either the equipment in use or its operational environment were different. New hazards could be presented by a new equipment model that works similarly to its predecessor, or by existing equipment operated in a changed environment or put to a new use. So any such changes require study of whether the practice remains valid or whether it too requires amendment.

For the same reason, the proper use of standards, procedures, and other documented recommendations requires knowledgeable governance [Redmill 2000b]. Every application – and every changed application – demands specific tailoring and guidance.

Thus, accepted good practice must be demonstrated to be relevant to the present circumstances, where 'relevant' is defined by the HSE [2008c] as 'appropriate to the activity and the associated risks, and up to date.' If it is not relevant, it may, in some cases, be tailored to become so; otherwise other risk reduction measures must be employed. Further, even if accepted good practice is employed and relevant, the decision that it reduces the risk under consideration ALARP should not be automatic; it requires consideration, acceptable evidence, and justifiable argument. It may require consultation with experts or approval from a higher authority.

Finally, no risk should be deemed ALARP unless the possibility of further reduction has been explored, as indicated by the question at point (12) in Figure 5. Once good practice has been followed, consideration should be given to whether more could be done to reduce the risk. If there is more, the presumption is that duty-holders will implement the further measures, but the reasonable practicability of doing so requires the application of first principles to compare the further risk reduction to be achieved with the sacrifice involved in achieving it.

5.5 Beyond Good Practice

There are numerous risks, across all industry sectors, for which good practice is inapplicable – for example, in the case of a new technology, or when elements of existing technologies or systems are integrated. Then ALARP decisions must be made by assessing and comparing options for risk reduction. The main steps in this process are shown in Figure 6 (which is a continuation of Figure 5), with the following supplementary notes expanding on them.

5.5.1 Identifying and specifying options (13, 14)

The first step is to identify all feasible options. This requires good knowledge of the system under consideration and understanding of the mechanism of the risk in question. Without such knowledge and understanding, it is likely not only that decisions will be

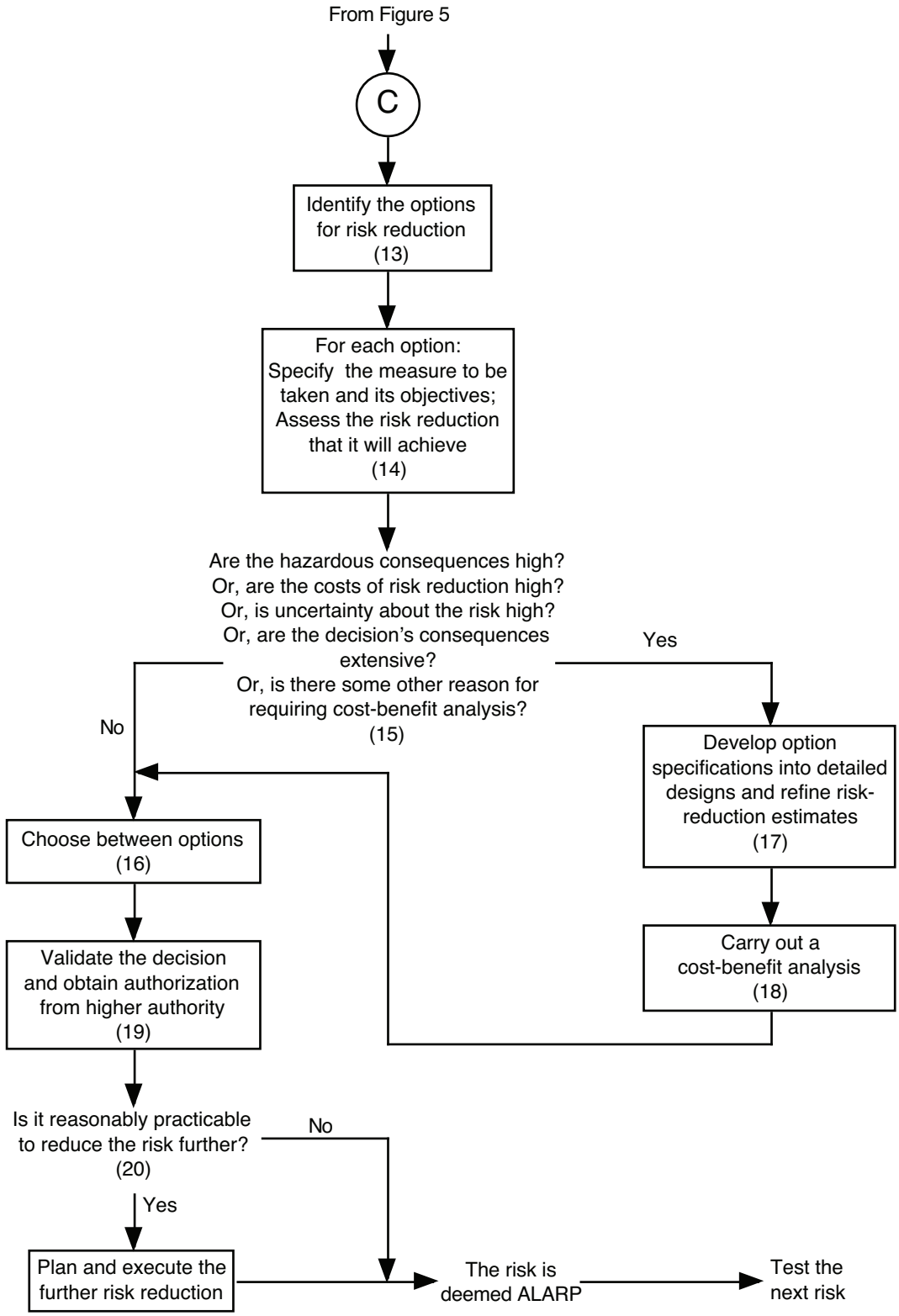

Figure 6: Decisions for Risks in the Tolerability Region for which Good Practice Does Not Provide Reduction ALARP

questionable but also that they will later be questioned, particularly if an accident has occurred. It is therefore important that all whose expert judgment is relied on should possess adequate and relevant competence. Indeed, this should be documented (in the safety case, if there is one) so that decisions will later be seen to be credible.

Then, in order for comparisons to be made, a clear specification should be produced for each option, defining what measures would be taken (e.g. a guard will be installed on the machine), what the measures are expected to achieve (e.g. the guard will prevent damage to operators' hands), and an assessment carried out to determine what risk reduction is anticipated (e.g. the frequency of damage to hands will be reduced from N per year to M per year). It is also necessary for each option to be supported by sound evidence of why the stated risk reduction can reasonably be expected.

5.5.2 Mode of decision between options (15)

Neither UK law, which requires risks to be reduced 'so far as is reasonably practical' (SFAIRP), nor the ALARP Principle, provides definitive guidance on when the cost of reducing a risk is disproportionate – or grossly disproportionate – to the benefits derived from reducing it. Ultimately it is a decision for a court of law; meanwhile, it must depend on expert judgment. In most cases, the circumstances of the risk are understood, the costs and consequences are not excessively high, and the effects of the decision are not wide-ranging. Then a first-principles approach, supported by knowledge and experience, is likely to lead to an ALARP decision that is justified by the evidence and logical argument.

But in some cases, for example in complex situations, in high-hazard industries and new technologies, and when the effects of the decision would be extensive, decision-making is less straightforward and comparisons between options require more detailed evidence. Then cost-benefit analysis (CBA) is used to provide additional information – not on the technicalities of risk reduction, but to support decisions on economic proportionality. Point (15) on Figure 6 is where it is decided whether or not CBA is required.

5.5.3 ALARP decisions by professional judgment (16)

In seeking to reduce industrial risks, there is a recognized order of preferences in the choice of measures to be taken:

- Avoid the hazard altogether, for example by replacing a toxic chemical with a non-toxic one;
- Address the hazard at its source by creating an inherently safe design, for example by introducing an interlock to prevent the hazard from giving rise to an accident;
- Introduce engineering controls, such as by collecting dangerous waste products and separately managing their safe disposal;
- Apply collective protection measures, for example regulatory controls such as exposure limits – and in any case, any applicable regulatory rules should be applied and adhered to;
- Provide specific protection for individuals exposed to the risk, for example special clothing and equipment;
- Impose protective procedures, such as signing-in on entry.

For a given risk, a measure of the first type should be chosen if one exists, unless its cost is grossly disproportionate to its benefits. If none exists, or it is grossly disproportionate, a measure of the second type should be selected unless one does not exist or it is grossly disproportionate. And so on – though the application of a higher-level measure does not preclude the additional application of a lower-level one. Thus, the onus is on the duty holder not simply to balance the costs of each possible measure against its

benefits, but to implement the safest measure unless it involves a grossly disproportionate sacrifice – in which case it may be deemed not reasonably practicable.

When the circumstances, as at point (15) on Figure 6, are appropriate, an expert, using this hierarchy of preferences as guidance, could in most cases arrive at a sensible and justifiable ALARP decision – at point (16) – without the use of cost-benefit analysis. It is important, then, to document the argument and evidence for the decision, to seek validation of it by another expert, to obtain authorization from an appropriate level of seniority – as at point (19) – and to document all this information for inclusion in the safety case.

Finally, as observed earlier, no risk should be deemed ALARP unless the possibility of further reduction has been explored – see point (20) in Figure 6.

5.5.4 Cost-benefit analysis (17, 18)

When the answer to any of the questions at point (15) in Figure 6 is 'yes', the HSE [2001] recommends employing a cost-benefit analysis (CBA). Then, in order to facilitate derivation of the best possible estimates of costs and benefits, and so increase confidence in comparisons, the specifications prepared at point (14) must be developed into detailed designs and the estimates of expected risk reductions must be further refined.

The costs to be aggregated in a comparison are all those that are 'necessary and sufficient to implement the measure' [HSE 2001]. Included are the costs of implementation of the measure, operation and maintenance, and any losses in time or productivity that are directly due to the measure. From them should be subtracted any gains resulting from the measure, for example those arising from improved productivity and a decrease in accident costs. The costs of safety assessment and justification of the measure should not be included. The Rail Safety and Standards Board has provided a table of advice on the applicability of costs and benefits [RSSB 2008].

Like costs, some benefits may be readily translatable into monetary terms – when they are based on marketable goods. But, in the field of safety, benefits may take the form of reductions in numbers of deaths or injuries, whose values are not determined from what they were sold for in the recent past. Yet, comparisons between costs and benefits depend on both being expressed in the same units. Thus, health and safety benefits must be translated into monetary terms. There is no universally agreed way of effecting this translation, and all methods are based on 'judgments', are open to misinterpretation or manipulation, and are often contentious. Ideally, whatever equation is used should be acceptable to all stakeholders, including, at least, the duty holder, those at risk, and the appropriate regulator.

A method often employed is 'contingent valuation', in which representative samples of the population are asked what they would be prepared to pay for an incremental increase in a defined form of safety, with values being derived from analysis of their answers. This is the approach taken by the UK's railway industry, which annually reviews its 'value of preventing a fatality' (VPF). The VPF is then applied to other safety benefits via the equivalence formulae: 1 fatality is equivalent to 10 major injuries, 200 reportable minor injuries, and 1000 non-reportable minor injuries [RSSB 2008]. The 2009 VPF, stated by Anson Jack, Director of Policy, Research & Risk, to railway colleagues in June 2009 in a letter posted on the internet (http://www.rssb.co.uk/pdf/reports/vpf_2009.pdf) was £1,661,000.

Since costs and benefits may occur at any point in a system's life, accuracy of comparison demands that all be translated not only into monetary terms but also into present values. This requires discounting factors, whose choices are, necessarily, based on assumptions that may not reflect reality – or they may appear reasonable at the time but be undermined later by changes in economic circumstances. It should also be observed

that results, and therefore comparisons, are sensitive to the selected values. Care should therefore be taken in their choice.

The results of a CBA are intended to be justifiable estimates of the costs and benefits of the identified options for risk reduction. However, given the several variables whose values are subjectively chosen and to which the results are sensitive, they cannot be considered definitive.

5.5.5 ALARP decisions informed by CBA

A decision informed by the results of a CBA cannot avoid expert judgment. Indeed, a CBA is not the sole decider, but only one source of information input to the decision-making process. The hierarchy of preferred options, introduced above, is still the primary basis of choice. CBA results should provide evidence to inform and support judgments on the proportionality between the costs and the benefits of an option that has been chosen from the perspective of safety.

When uncertainty about a risk is high, a CBA should not be used to reject all options for risk reduction on the grounds of cost. Lack of proof, or evidence, of risk should not be taken to imply the absence of risk, and any concessions should be made on the side of safety. In other words, the precautionary principle should be applied.

Finally, as in choosing between options without CBA, the final choice of risk-reduction measure should be validated, authorization from a higher authority should be sought, and the possibility of further reasonably practicable risk reduction should be explored (see points (19) and (20) on Figure 6).

5.5.6 New hazards created

In making ALARP decisions, attention should be paid to the potential side effects of risk-reduction measures. It is a bonus if an effect is the reduction of another risk. But it may also happen that another risk is increased, in which case the costs of the further measures needed to reduce the second risk must be included in the costs of mitigating the first. Analysts must remain aware of the possibility of such adverse side effects by taking a holistic approach [HSE 2008c] and carrying out impact analysis to determine the potential undesired consequences of the intended risk-reduction measures.

If a new hazard is created by a risk-reduction measure, the HSE [2001] advises that it should be treated separately – except that, if it is found not to be reasonably practical to reduce it ALARP, the first measure cannot be taken.

CHAPTER 6

Discussion of ALARP-related Topics

6.1 Introduction

In previous chapters, reference has been made to a number of concepts. They are legal, economic, sociological, and risk-based. Some appear easy to understand, some are vague, some controversial, and all are tricky to apply. In the application of the ALARP Principle, all will at some time come into play, and their conscious or unconscious interpretations, and the assumptions included in their interpretations, will affect the outcomes of risk-tolerability decisions. It is therefore important to recognise and understand them and their effects of ALARP decisions. This chapter attempts to identify them, briefly to enquire into the difficulties that they throw up, and to raise questions and make suggestions.

6.2 Reasonableness and Practicability

The ALARP Principle is considered to be a single concept. But 'ALARP' includes two concepts, reasonableness and practicability, whose influences on stakeholders are different and often separate: risk mitigation that is practicable to one party may not be reasonable to another.

Having determined what to them is reasonably practicable, the management of a duty-holding company will often conclude (perhaps implicitly) that it is reasonable for the resultant 'tolerable' risk to be imposed on other parties. In other words, those imposing risks are likely to address ALARP compliance from the perspective of practicability.

On the other hand, laypeople, who typically possess no knowledge of risk analysis, the ALARP Principle, or the technical means of risk reduction, judge compliance from the perspective of reasonableness. They ask such questions as, 'Is it reasonable for this risk to be imposed on me, or my community?' and 'What do I get in return for accepting this risk?' Reasonableness, therefore, is judged within the qualitative feelings and beliefs of individuals and the public at large.

The bond between a company and the public is trust. The HSE recognises this in Stage 4 (Adopting decisions) of its 'Approach to reaching decisions on risk' [HSE 2001]. It points out that success in ALARP decisions depends on ensuring that interested parties (stakeholders) are satisfied with (among other things) 'the way uncertainty has been addressed', 'the plausibility of the assumptions made', and 'how other relevant factors such as economic, technological and political considerations have been integrated into the decision-making process'. It would be useful if every duty holder involved all other stakeholders, as a matter of course, in such validation of processes and decisions.

Trust is created and sustained by right communication and right product. It is proposed that the attributes of the latter in the present context are appropriate quality, reasonable pricing, and adequate safety. For the public, who are not party to the details of what goes on within a company, the integration of all three into a product or service is the basis of reasonableness. If the quality of a product or service falls below what the public deems appropriate, or pricing is unreasonable, or safety inadequate, and the company bases the reason on a trade-off with another attribute, the public loses trust and is likely to deem the trade-off unreasonable. In the event of an accident, these feelings are exacerbated.

It is therefore crucial to a company that no stakeholder should perceive its ALARP decisions as a triumph of cost cutting over reasonableness. When risk uncertainty is low, it is usually not difficult for decisions to appear both practicable and reasonable. But when uncertainty is high, it may not be so easy, particularly if trades-off are proposed. In such circumstances, it is recommended that ALARP decision-makers should give active consideration to whether each type of stakeholder would be likely to approve of both the practicability and the reasonableness (and particularly the latter) of each decision. It may be useful to employ a facilitating device, such as the matrix of Figure 7, to ensure that they have all been duly considered.

Stake-holder	*Practicability agreed by stakeholder?*	*Solution reasonable to stakeholder?*	*Why reasonable to stakeholder?*
Duty-holder			
Employees			
Disinterested neighbour			
User			
General public			
Etc.			

Figure 7: A Facilitator of Checks for Both Practicability and Reasonableness

6.3 Gross Disproportion or just Disproportion?

In his Edwards vs. the National Coal Board ruling, Lord Asquith expressed the need for gross disproportion between a higher cost of risk reduction and a lower saving achieved by the reduction, if a claim that the reduction is too expensive is to be supported. In drafting its advice on complying with the ALARP Principle, the HSE could not unilaterally define a lesser differential. Yet, no other authority has called for 'gross' difference, and, with no guiding criteria for judging when disproportion is gross, expert judgement is arbitrary.

Meanwhile, in modern times, the UK government has frequently called for proportionality with respect to risk – in applying restrictions, imposing punitive measures, and so on. For example, of the many references in the government's report on risk [Cabinet Office 2002] to the need for proportionality, not one mentions 'gross', not even in the list of 'principles of managing risks to the public', which includes: 'Proportionality and precaution'.

In another legal instance – that of Judicial Review – the test of proportionality, rather than gross disproportionally, is made in determining if, in acting against a wrong, the state has exerted too great a power.

There is no doubt that compliance with the ALARP Principle demands gross disproportionality; the HSE says so, and ALARP and its requirements are defined by the HSE. And there is no doubt that Lord Asquith's legal precedent exists. Hitherto, practitioners have accepted the HSE's requirement. But now the UK's Rail Safety & Standards Board [RSSB 2007] has challenged the meaning of and requirement for gross disproportion. It asks if the requirement for gross disproportion is itself a disproportionate response to risk, and points out that Her Majesty's Railway Inspectorate's guidance on ALARP says that 'factors of disproportion' (rather than gross disproportion) should be

applied. Thus, given the modern emphasis on mere proportionality, it may be that a jury, or an appeal judge, would find that disproportionality rather than gross disproportionality is required to justify accepting a risk without further reduction.

If such a possibility exists, it is time for a debate on the 'gross' criterion, for an incorrect belief in the law's requirement for gross disproportionality would lead (and perhaps does lead) to an excessively risk-averse attitude, with tolerable risks being refused, opportunities missed, and over-engineering being employed and excessive costs being suffered in many cases. But an incorrect belief in the need only for disproportionality would lead to breaches of the law and convictions. It's time for a review.

6.4 Cost-benefit Analysis

It is generally accepted that whether it is worth obtaining something may be determined by comparing the cost of its acquisition with its value. This is the principle of the market, whose effectiveness depends on the ability to obtain, with confidence, reliable valuations of all components of the equation. But when some or all of these components are not marketable goods, their valuations cannot be obtained from the markets and must therefore be arbitrarily – or imaginatively – derived. And, when dealing with safety and environmental risks, many components, particularly on the cost side of the equation are not marketable goods.

For determining values, economists use contingent valuation (CV) techniques, which consist of asking people how much they are willing to pay (WTP) to save, or willing to accept (WTA) for the loss of, the thing in question (e.g. a rare insect's habitat, a human life) or some attribute assumed to be related to it. (For a review of CBA and CV, see [Redmill 2001a].) Typically, in the real world, a good (e.g., a book) or a right (e.g., the right to use property) is obtained by paying the value demanded by the good- or right-holder – which implies the use of WTA for obtaining the holder's valuation. But this could defeat a cost-benefit analysis, for a single very high valuation could set a price beyond the budget of many intended projects. So economists justify the use of WTP methods – in which a population (perhaps including some right-holders) are asked what they would pay to retain (i.e. not to lose) their right (rather than what they would expect as compensation for its loss) – which result in lower valuations more acceptable to the decision makers.

Then, in creating CBA equations, assumptions are necessary, and many of these can have significant effects on the ultimate results. For example, some costs are treated as 'externalities' (i.e. are excluded) or undervalued. In one study – on whether a road should be built through green-belt land – permission was achieved by attributing to the land the low value of 'land without planning permission'. Similarly, in dealing with the possibility of death, the costs of injuries and suffering may be excluded altogether.

In the fourth of his 2009 series of Reith Lectures, Professor Michael Sandel [2009] referred to 'the spurious science of cost-benefit analysis' and provided several examples of results that are obviously absurd. Of course, what may appear to ordinary intelligent individuals as absurd are often the very results that study commissioners want. CBA can easily be – and often is – manipulated.

If assumptions are held constant across the studies, CBA offers means of comparison between options for reducing a risk. But its use for a single study requires great caution. Indeed, many, including Sandel [2009] and Adams [1992] contend that it is in many cases unsuitable for the purpose to which it is put. When both uncertainty and the stakes are high, the decision is, properly, a political one and requires value judgements. A reason for decision makers (particularly politicians) using CBA is often to avoid such judgements and to give the impression that decisions are scientific, or objective, or made by economic experts rather than those with responsibility for them. Yet, the suggestion that CBA is

objective is false – as is the notion that all negative effects can, or should, be regarded simply as financial losses. When both uncertainty and the stakes are high, politicians and senior management should accept the responsibility for their decisions – with the results of a CBA being one of several inputs rather than, necessarily, the definitive one. Then, too, if senior decision makers find the prevailing uncertainty to be too great for them to be confident that their decisions would lead to adequate public safety, they should admit that an option is not to proceed with the proposed enterprise and to employ the precautionary principle.

6.5 Uncertainty and the Precautionary Principle

Application of the ALARP Principle depends on being able to determine values, with reasonable confidence, of risk, the amount of reduction to be achieved, and the cost of achieving the reduction. In most cases – for example, in the use of factory equipment – the hazards and their potential consequences are well understood. But sometimes, new risks arise, either in the world (e.g. climate change) or from new technologies (e.g. the effect of frequent use of mobile telephones on the brain) about which there may be considerable uncertainty. At first, their potential consequences may only be guessed at, and it may not even be certain that adverse consequences could occur. What is to be done? Should progress be denied because of fear or excessive risk aversion? Or, should devastating consequences be chanced in the hope of progress? On the one hand, there is the argument that the lack of proof of safety risk is sufficient justification for the assumption that there is none; on the other, it is argued that safety risk should be assumed until there is proof of its absence.

Figure 8 offers a guide to taking the first step to determining the level of uncertainty. It brings together the state of available knowledge and the degree to which experts have formed a consensus on it, and invites decision, within the cells of the matrix, on the extent to which it is justified to proceed with the plan to use the technology or create the intended system. Clearly, the bottom right cell of the matrix is the only point at which there is strong agreement on an adequate volume of knowledge. However, it should be noted that the matrix, as constructed, is for the assessment of risks associated with proceeding. Inaction may also give rise to risks, and these may also need to be addressed.

In the top-right and bottom-left cells, it is suggested that discretion might permit a planned attempt to manage the risks using the 'Precautionary Principle'. This has traditionally been represented by the slogan, 'Err on the side of caution', but it was codified at the Rio de Janeiro Earth Summit [1992] for the protection of the environment: 'Where there are threats of serious or irreversible environmental damage, lack of full scientific certainty shall not be used as a reason for postponing cost effective measures to prevent degradation'.

This form of the principle can be (and is) employed not only with respect to the environment but generally, and it declares against reckless gambling in the face of uncertainty when the stakes are high.

Yet, the precautionary principle is not impregnable. It offers protection but not infallible protection. It is vulnerable to misuse, misinterpretation or varied interpretation, underestimation of or failure to recognise risks, failure to apply or police it, application that is too late, illegal breaches, delaying tactics, and so on. Of itself, it is not a panacea. It must be used wisely, within regulation. It must be accompanied by persuasion of everyone to adopt the culture of protecting the future by, if necessary, foregoing or deferring apparent progress if that progress embodies too much uncertainty and might result in significant or irreversible damage. So, in the end, what is it really? It is the decision to err on the side of caution.

Degree of Consensus \ Information and Knowledge	*Little or none*	*High*
Disputed	Both research and discussion required. Do not proceed.	Information is not yet agreed as knowledge. Discussion required on interpretation. Perhaps further experiment and analysis needed. Proceed with caution, using the Precautionary Principle, but only with a strong supporting argument.
Agreed	Research required. If believed consequences are not high, perhaps proceed - with a great deal of caution - using the Precautionary Principle	Proceed according to the ALARP Principle.

Figure 8: Initial Steps in Addressing Uncertainty

6.6 Unintended Consequences

Every action gives rise to consequences other than those for which it was taken, and risk-reduction actions are not exceptions. For example, medicines and surgery have side effects, weed killers and insecticides kill other wild life and pollute the environment, and fences erected to keep spectators from invading a football pitch caused their deaths by preventing them escaping when crushed from behind.

Thus, the management of risk is not confined to a choice between risk and no risk, or even to the balancing of risks against benefits, but includes the balancing of risks against risks.

New risks thrown up by risk-reduction action are sometimes referred to as 'countervailing', and the HSE refers to them as 'transferred'. The HSE [2008b] recommends that, when a transferred risk arises from a different hazard (i.e. not the hazard for which the action was taken), it should be treated as separate, to be reduced ALARP in its own right. This seems a reasonable approach, as long as the transferred risks are small in number and they do not themselves give rise to transferred risks. Otherwise, there is the added risk that unknown and undocumented failure modes and causal chains will exist and that, later, the new hazards will be activated by small changes in a system, its environment, or its operational procedures or parameters.

It is recommended that impact analysis be carried out on risk-reduction measures in an attempt to detect unintended consequences, and that whenever a transferred risk is detected, sensitivity analysis be carried out on its hazard, with the results documented for future use. A structured technique would facilitate this, one that has been proposed being Graham and Wiener's [1997] RTA (risk trade-off analysis). This guides the user first to recognise the unintended consequences of risk-reduction proposals, then to model them, and finally to apply risk weightings so as to facilitate an optimised outcome.

6.7 Societal Concerns

The HSE [2001] takes societal concerns into consideration in making ALARP decisions. It says that the triangle in the ALARP model represents increasing level of risk 'measured by

the individual risk and societal concerns it engenders' [Para 122]. It later says that special consideration should be given to accidents carrying the potential for multiple fatalities [Para 136]. But the view that individual companies should respond in their safety decisions to society's anxieties, over and above their responses to risks, is not universally held. The UK's Rail Safety and Standards Board [RSSB 2006] questions that view and reports on sponsoring research which concluded that society does not respond to multiple-fatality accidents with greater anxiety than to the equivalent number of single-fatality events.

It was pointed out in Chapter 4 that a company would be well advised to consider allaying public anxiety, caused by its policy or product, before pressing on with its policy or product. Not to do so would pose a risk to the company's well-being. But such a step may be costly and should be at the company's choice; it should not be enforced by a safety regulator. Perhaps it is time for the matter of societal concerns to be reconsidered and debated in the context of the ALARP Principle, for it is doubtful that the law requires it.

6.8 Decision-making

Decisions are almost never based on a single source of evidence. Typically, among others, there are technical, economic, political, and personal inputs. The ALARP Principle recognises this by including both technical and economic components in practicability.

The level of confidence that can be justified in an ALARP decision necessarily depends on the confidence that is held in the various components of knowledge and information on which it is based, as well as in the decision-making process itself. Confidence in technical risk estimates is discussed below. It has already been emphasised that no decision should be based on cost-benefit analysis alone and that, when uncertainty is high, political input is essential. And, of course, confidence in political aspects depends on the knowledge and competence of the decision makers.

Importantly, the personal component, though often unconsciously contributed, is never absent from decision-making. Psychologists have conducted a considerable amount of research into the subject, and their results show that we are hugely influenced by our biases (created in us by our upbringing and other experiences) and by the mental shortcuts (heuristics) taken unconsciously by our brains in carrying out complex tasks. Although in recent times the results of psychological research into human error have been brought into system safety engineering, under the heading of 'human factors', the vast body of knowledge on decision-making under risk has been confined mostly to the fields of economic and financial activity [e.g. Thaler 1992]. It is time for this deficiency in the domains of engineering and health and safety to be made good, and it is recommended that decision-making as a subject, and the results of this research, should be brought into the education and training of all engineers. It is beyond the scope of this article to elaborate on the subject, but a review of it and its relevance is offered in [Redmill 2001b].

6.9 Calibration of the ALARP Model

For the calibration of the ALARP model, many – perhaps most – practitioners think only of the HSE's guidance (see Chapter 2). But, being in terms only of fatalities, this carries two principal disadvantages. First it leads many practitioners to place excessive emphasis on fatality as an indicator of risk, and even to believe that other hazard consequences are irrelevant. Second, it offers no advice to those dealing with hazards with non-fatal, or a mixture of fatal and non-fatal, consequences.

In industries, such as road transport, where accidents tend to be high-impact and to

result in more brain-damaged victims than deaths, addressing risk tolerability only in terms of fatalities is an inadequate and potentially misleading approach.

The UK railway industry recognises this in its 'value of preventing a fatality' (VPF, see Chapter 5). However, its equation of one fatality equals 10 major injuries or 200 minor injuries is crude, for it assumes equality between all types of major injury and all types of minor injury. It does not, for example, distinguish between injuries from which recovery is almost certain (e.g. a broken bone) and those that are irreversible (e.g. serious brain damage). Nor does it address environmental risk.

The ALARP Principle was originally drafted by the HSE for the nuclear industry, in which death might be the most likely result of an accident, but its use in other industries requires a more refined approach to hazard consequences. It is now time for research to be carried out into the development of a system of calibration of the ALARP model – and, more generally, of tolerability assessment – that integrates all categories of hazard consequence: fatalities, major and minor injuries, environmental destruction of all sorts, and other losses that may be quantified financially.

6.10 Confidence in Risk Estimates

Being of the future, and expressing potential rather than inevitability, risk estimates must necessarily carry uncertainty.

In many cases, the system under study is well understood, with both its failure modes and their potential consequences known, so there may be high confidence in the risk estimate. In other cases, such as in the context of a new technology, uncertainty concerning risk may be high (as discussed above). Then, risk estimates must be speculative and therefore must carry low confidence.

But there are other factors that lead (or should lead) to low confidence. A list of examples (though not an exhaustive list) is offered:

- Unsafe conditions can arise not only from component failures but also from unforeseen interactions between components that are functioning as designed.
- Hazard identification is incomplete, due to lack of information or to unreliable or incorrect information.
- The branch of risk assessment that covers human operators – human reliability assessment (HRA) – is not only incomplete in its development but also under-used in the field of system safety engineering. So far, it is confined mostly to the domain of psychologists and ergonomists, but it needs to be taught and employed more in health and safety engineering generally. A review of the topic is offered in [Redmill 2002].
- Very many accident investigations show management – via policy, documented procedures or the lack of them, strategy, leadership, safety culture or the lack of it, and many other manifestations – to be a major cause. Yet, risk analyses almost never address the potential for such effects of management [Redmill 2006].

These examples are instances of, if not systemic then at least frequent, underestimates of risk. Thus, when such conditions obtain, risk analysts in all fields are advised to take care to double-check their assessments and look for ways in which they might have erred.

It should be noted that risk can also be over-estimated, leading to excessive risk-management activity and costs. This can occur, for example, due to over-cautions assumptions or the over-estimation of the likelihood of unsafe failures, both of which can be represented in the subjective construction of fault trees (see Funtowicz and Ravetz [1990], Fischhoff, Slovic and Lichtenstein [1978] and a discussion in Redmill [2001b]).

Thus, it is always unwise to allow certainty, or even high confidence, in risk estimates to develop. The more cautious belief that something must have been overlooked is far more in keeping with the profession of a risk analyst or manager.

CHAPTER 7

Conclusions

7.1 Review

The principles of modern safety thinking that underpin the ALARP Principle – such as a risk-based approach, self-regulation, and the need to achieve and demonstrate appropriate safety in advance – began their evolution within the English Common Law, were made explicit by Lord Robens in his review of the regulation and management of occupational risks [Robens 1972], and were then enshrined in statute in the Health and Safety at Work, Etc. Act [HSW Act 1974], which requires risks imposed on others to be reduced 'so far as is reasonably practicable' (SFAIRP). A law merely states what must be done; it is left to those who would abide by the law to determine how to do it and to other authorities (assessors, regulators and the courts) to determine whether it has been done in conformity with the law. In the absence of other guidance on how to go about reducing risks SFAIRP, the UK's principal safety regulator, the Health and Safety Executive, proposed the ALARP Principle, which is a strategic approach based on a model.

This paper commenced with an introduction to risk tolerability, which was followed by an overview of the ALARP Principle and then its historical and legal origins. The Principle's application in all types of risk-tolerability situations were then analysed. Finally, a number of topics that necessarily interact with the ALARP Principle, and on which the Principle depends, were discussed, and it was shown that each of these topics and their interactions with the ALARP Principle offer opportunities for further research.

Following this exploration of the ALARP Principle, its purpose, its background and its application, there are two important questions to be asked: how efficient is it, and how effective is it?

7.2 Efficiency of the Principle

This paper has so far been concerned with the Principle's efficiency. Though easy to understand, the ALARP Principle is not so easy to apply. It offers a straightforward model to facilitate risk-tolerability decisions, but its use requires a great deal of professional knowledge and judgment and depends on the handling of a number of risk-related factors.

The decision-making processes were presented in three flowcharts. As these are intended to cover all situations, they contain a large number of decisions points. However, in engineering safety, uncertainty is mostly low and the determination of appropriate ALARP decisions follows readily from the proper identification and sensible analysis of available options. Indeed, according to the HSE [2008a], most ALARP decisions are based on the use of accepted good practice.

But when uncertainty is not low, decisions are not so easily arrived at, and confidence in them may not be high. To derive them, not only is a sound knowledge of risk and its analysis required, but so also is knowledge of the factors to be considered in applying the ALARP Principle, many of which are not within engineering or risk curriculums and not within the compass of most engineers. Some (e.g. reasonableness, practicability and gross disproportion) are legal; some (e.g. cost-benefit analysis) are economic; some (e.g. societal concerns) are sociological; and some (e.g. the effects of human biases on risk analysis,

decision making) are psychological. Thus, in order confidently to arrive at just, socially acceptable, legal ALARP decisions, there is a requirement not only for knowledge in the various subjects but also for an open, observant and analytical approach to bring them together in a manner appropriate to the situation in hand.

Then there are questions of confidence – in the results both of risk analysis and decisions taken – and how to determine it and how to express it.

Applying the ALARP Principle when uncertainty is high is a non-trivial exercise. Not only are risk and its analysis tricky subjects, but, as already pointed out, engineers and other risk analysts and managers may possess no more than a superficial knowledge in the domains of the related factors.

7.3 Effectiveness: ALARP and SFAIRP

But what of the ALARP Principle's effectiveness? What confidence can there be that a risk deemed ALARP would also be judged to have been reduced SFAIRP? Can the two concepts be said to be identical? They cannot. As already pointed out, they were defined by different parties (the law-makers and the safety regulator) for different purposes (stating a legal requirement and offering guidance on a strategic approach to meeting it). But does one imply the other? No. Meeting either is never an incontrovertible fact, proven by some irrefutable logic; rather, it is a matter of judgment. It may reasonably be assumed that the party (an individual or a team) that judges a risk to be ALARP would, a moment later, judge the same risk to have been reduced SFAIRP. But there are several factors why a risk judged ALARP now may not, later, be judged to have been reduced SFAIRP.

First, there is the matter of time. The law is likely to be invoked after an accident has occurred, and this would be some time after a system has been deployed (with its risks deemed ALARP). ALARP is now; SFAIRP judgements are made later, in a court of law, to settle a civil dispute or a criminal prosecution.

Second, the two judgments are made by different people. The ALARP judgment is likely to be made by practitioners, assessed by third-party independent safety assessors and, perhaps, approved by a regulator, whereas a legal SFAIRP judgment is made by a jury or judge.

Third, there is the matter of information. The legal judgment is made in retrospect, based not only on what informed the ALARP decision but also on what has come to light since and on opinions offered by expert witnesses. Different questions asked by barristers and different assertions made by expert witnesses could make all the difference to a judgement by a jury or judge. The scrutiny exercised in a court of law may also lead risk creators to recognise that there was information that they should have collected or considered but didn't and actions that they should have taken but neglected to take.

Fourth, there is the matter of perspective. Practitioners making ALARP decisions are likely to start with the desire to demonstrate that their system is adequately safe – i.e. that its risks are all tolerably low – whereas a jury is likely to look for evidence that it was not. The difference in perspective can be significant.

Fifth, there is the matter of subjectivity. Whatever the strength of other factors, ALARP and SFAIRP decisions depend on judgment and therefore involve subjectivity. A practitioner may place greater or less emphasis on this or that piece of information, or on the opinion of this or that advisor, while a jury or judge may be swayed by the persuasiveness of the advocate for the prosecution or the defence, by the logic of this or that expert witness, or by the framing of this or that piece of evidence. The greater the uncertainty surrounding a risk, the greater is the influence of subjectivity in a tolerability decision. Subjectivity is an essential ingredient in professional judgment, but in court, later, a jury may subjectively conclude that the decision-maker should have inclined in a

different direction.

Sixth, there is the matter of context. Engineers and other duty holders must make risk decisions that meet legal requirements; juries and judges must deliver determinations of whether or not those decisions do, or did, in fact satisfy the law. Practitioners make their decisions in a technical context and juries and judges make theirs in a legal context. As mentioned earlier, the ALARP Principle is intended as a bridge between the legal and the engineering to assist the practitioner to interpret the legal requirements, and then as a bridge back to the legal in making the claim that the requirements have been fulfilled. In the two crossings, there is plenty of scope for misunderstanding. A record of earlier ALARP judgment may be evidentially useful, but it cannot guarantee a favourable court finding later.

The ALARP Principle is not a formula for the provision of consistent or repeatable results. Nor is the law. Both are designed to cover all possible situations and, therefore, both require judgment whenever they are employed. They both depend on determinations of practicability and reasonableness, which, as explained above, differ according to the case in hand and, in all cases, are subject to judgment. There can be no guarantee that the same ALARP decision would be arrived at by two different practitioners, and certainly none that an ALARP decision arrived at now in an industrial context would, later be judged by non-engineers in a legal context to have met the SFAIRP test. Indeed, whereas the foregoing discussion contrasts ALARP and SFAIRP decisions, it could, as easily and without change, be used to contrast ALARP decisions made by different people at different times.

7.4 The Importance of Confidence

It would seem that, in a trial on the legal, SFAIRP, side of the ALARP bridge, a good lawyer is worth more than a good engineer. But what must the engineer provide in order to give a good defence lawyer an advantage over the prosecution's good lawyer? The answer is: evidence that all was done that could reasonably have been done in the circumstances. Thus, there needs to be a record, preferably in a safety case, of what lies behind every ALARP decision, i.e. sound evidence of due diligence: not only the details of risk analysis and risk-management actions, but also the assumptions made and reasons for their validity, and an assessment of the evidence underpinning each decision and the justified level of confidence in it.

In particular, it is important to provide reasons not only for action but also for doing nothing. When an accident has led to an inquiry or a prosecution, inaction may, in retrospect, give the impression of negligence, and negligence may be hard to refute. There needs to be evidence, first that it was not negligence and, second, that inaction was justified. More than that, the engineer should be able to claim to have been 'as confident as reasonably practicable' (ACARP) in each ALARP decision. To increase the likelihood of favourable SFAIRP judgments, ALARP decisions should be supported by ACARP arguments.

7.5 Further Work and Final Words

The ALARP Principle is a tool for the assessment of risk tolerability, and its operation depends on the manner in which its users define and address other aspects of risk and its analysis and management (the more important of such topics were discussed in Chapter 6. Such definitions and manners of address create assumptions that become implicit in each application of the ALARP Principle. Further, they vary between practitioners for a number

of reasons, including a lack of knowledge of risk principles and of other relevant factors. Thus, improved use of the ALARP Principle – and, indeed, improvement in the assessment of risk tolerability in general, whether or not the ALARP Principle is employed as a tool – calls for better education in the subject of risk, better knowledge of the related topics, and research into some of those topics.

Education would be improved by teaching risk to all science and engineering undergraduate and postgraduate students (as well as others, such as those in medicine, law and psychology). Such teaching should provide adequate instruction in the theory; it should also emphasise that techniques are not the principal components of the subject but tools to be chosen appropriately and only used with understanding. Regarding research, suggestions for this into topics related to the ALARP Principle were made in Chapter 6.

Finally, it should be mentioned that the ALARP Principle is not essential or mandatory. It is quite possible to carry out risk tolerability assessment and to satisfy the law without employing it. Indeed, many do so, for many who must reduce their risks have never heard of it. It is the principle of taking a risk-based approach and being able to demonstrate that all risks have been reduced SFAIRP that are fundamental, not the use of a particular technique or model.

References

Adams J [1992] Horse and Rabbit Stew. In Coker A and Richards C (Eds): *Valuing the Environment*. Belhaven Press, London

Beck U [1992] *Risk Society*. Sage Publications, London

Bier V M [2001] On the State of the Art: Risk Communication to the Public. *Reliability Engineering and System Safety*, Vol. 71, pp 139-150

Burke, Derek [1997] The Regulatory Process and Risk: a Practitioner's View. *Science, Policy and Risk*. The Royal Society, London

Cabinet Office [2002] *Risk: Improving government's capability to handle risk and uncertainty*. Strategy Unit Report, Cabinet Office

Coates J F [1982] Why Government Must Make a Mess of Technological Risk Management. In Hohenemser C and Kasperson J X (eds): *Risk in the Technological Society*. American Association for the Advancement of Science, Washington D.C.

COMAH [1999] *The Control of Major Accident Hazards Regulations*. HMSO

Department of Environment [1995] *A Guide to Risk Assessment and Risk Management for Environmental Protection*. HMSO, London

Earth Summit [1992] *United Nations Conference on Environment and Development (UNCED).* Rio de Janeiro, Brazil

Edwards vs. The National Coal Board [1949] 1 All ER 743

Eliot G [1861] *Silas Marner*. Penguin Classics (1985)

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

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

Galbraith K [1979] *The Affluent Society*. Third edition, Penguin Books

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*. Russell Sage Foundation, New York

Graham, John D and Wiener, Jonathan B (Eds.) [1997] *Risk Vs. Risk*. Harvard University Press

HSE [1988] Health and Safety Executive: *The Tolerability of Risk from Nuclear Power Stations*. Discussion Document, HMSO, London

HSE [1992] Health and Safety Executive: *The Tolerability of Risk from Nuclear Power Stations*. HMSO, London

HSE [2001] Health and Safety Executive: *Reducing Risks, Protecting People*. HSE Books.

HSE [2004] Health and Safety Executive: *A Guide to Risk Assessment Requirements (INDG218).* HSE Leaflet (available in pdf from http://www.hse.gov.uk/)

HSE [2008a] Health and Safety Executive: *ALARP At A Glance.* HSE Leaflet (available in pdf from http://www.hse.gov.uk/) (updated 08.04.08)

HSE [2008b] *Principles and Guidelines to Assist HSE in its Judgements that Duty-holders have Reduced Risk As Low As Reasonably Practicable*. Health and Safety Executive leaflet (available in pdf from http://www.hse.gov.uk/) (updated 08.04.08)

HSE [2008c] *Policy and Guidance on Reducing Risks as Low as Reasonably Practicable in Design*. Health and Safety Executive leaflet (available in pdf from http://www.hse.gov.uk/) (updated 08.04.08)

HSE [2008d] *Assessing Compliance with the Law in Individual Cases and the Use of Good Practice*. Health and Safety Executive leaflet (available in pdf from http://www.hse.gov.uk/) (updated 08.04.08)

HSE [2008e] *Cost Benefit Analysis (CBA) Checklist*. http://www.hse.gov.uk/risk/theory/alarp1.htm (updated 08.04.08)

HSW Act [1974] *The Health and Safety at Work, Etc. Act*. HMSO, London

IEC [1998] IEC 61508: Functional safety of electrical/electronic/programmable electronic safety-related systems (standard in seven parts). International Electrotechnical Commission, Geneva

Kahneman, Slovic and Tversky [1982] *Judgement Under Uncertainty: Heuristics and biases*. Cambridge University Press

Kaplan S and Garrick B J [1981] On the Quantitative Definition of Risk. *Risk Analysis*, Vol. 1, page 1

Layfield, Sir Frank [1987] *Sizewell B Public Inquiry Report*. HMSO, London

Lowrance W [1976] *Of Acceptable Risk*. Los Altos, Kaufman

Redmill, Felix [2000a] How Much Risk Reduction is Enough? *Journal of System Safety*, Vol. 36, No. 1, pp 7-8

Redmill, Felix [2000b] *Installing IEC 61508 and Supporting its Users - Nine Necessities*. Fifth Australian Workshop on Safety Critical Systems and Software, Melbourne, Australia

Redmill, Felix [2001a] *Cost-benefit Analysis*. Unpublished Report

Redmill, Felix [2001b] *Subjectivity in Risk Analysis*. Unpublished Report available at: http://www.csr.ncl.ac.uk/FELIX_Web/new_index.html

Redmill, Felix [2002] Human Factors in Risk Analysis. *Engineering Management Journal*, Vol. 12, No. 4, August, pp171-176

Redmill, Felix [2006] Understanding the Risks Posed by Management. In Redmill F and Anderson T (Eds.): *Developments in Risk-based Approaches to Safety: Proceedings of the Fourteenth Safety-critical Systems Symposium, Bristol, UK,* pp 155-169

Redmill, F [2007] Risk Communication. *eJournal of System Safety*, Vol. 42, No. 3

Robens, Lord [1972] *Safety and Health at Work: Report of the Committee, Cmnd 5034.* HMSO, London

[RSSB 2006] *Valuing Safety*. Rail Safety and Standards Board, UK

[RSSB 2007] *The Route to 'Taking Safe Decisions'*. Rail Safety and Standards Board, UK

RSSB [2008] *Taking Safe Decisions*. Rail Safety and Standards Board, UK

Sandel, Michael [2009] *Reith Lectures 2009: A New Citizenship*. BBC, UK

Slovic P [1992] Perceptions of Risk: Reflections on the Psychometric Paradigm. In Krimsky S and Golding D (eds): *Social Theories of Risk*. New York, Praeger

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?* pp 117-152, 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

Thaler, Richard H [1992] *The Winner's Curse*. Princeton University Press, New Jersey

Wilkinson, Angela [1997] Perception and Authority. *Science, Policy and Risk*. The Royal Society, London

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