Grasping at the Shadow of Safety and Missing the Substance[1]

Felix Redmill
Redmill Consulting, London UK

Abstract

Safety must not only be achieved but also demonstrated, and the processes involved require safety engineers and managers to possess both judgement and an understanding of fundamental principles. Yet it is common for practitioners to grasp at safety's shadow and miss its substance by suspending their judgement in favour of rules, and by replacing basic principles with standards and other tools. This paper examines some relevant examples.

1 Engineering, Management and Professionalism

There are two prongs to system safety: achieving it and demonstrating its achievement. Whereas they can be separately identified, their activities are interwoven. Safety is achieved by doing the right things well; it is demonstrated by presenting a logical argument, with evidence to support it, that the right things have been done well. Thus, the activities of achievement and demonstration must be integrated, and both must be planned in advance.

The modern approach to determining which activities are the right ones and, thus, what evidence is appropriate, is to take a risk-based approach. Risk analysis, followed by assessment of risks against tolerability criteria, provides the basis for judgement of how safe a system needs to be and, thus, what activities are 'the right ones'.

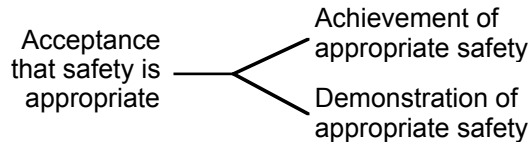

Figure 1: Safety's two prongs

It is usual to think of safety-achievement activities as safety engineering. But activities that are technical do not always represent good engineering. Engineering implies control, and control demands management. Thus, an integral part of engineering is management, and the engineer who does not manage what he or she does is a technician. But if management extended only as far as the immediate control of activities, engineering would be ineffective. It also requires an infrastructure within which to function, and this must be put in place by management at various levels. At the project level, the infrastructure is the responsibility of the project manager, and it defines such things as roles and responsibilities, communication mechanisms, and documentation structures. It is intended to facilitate the control of activities and to ensure both their effectiveness and their efficiency. But a project infrastructure also requires a context, and this is provided by an organisation's policies, strategic plans and culture, which are the responsibility of senior management. An organisation needs a

[1] This was the Invited Paper at the Sixth International Symposium on Programmable Electronic Systems in Safety Related Applications, Cologne, Germany, 4-5 May 2004

collective attitude to safety that is an integral part of its culture. Engineering and management go hand in hand, and operational effectiveness requires both.

Figure 2: Engineering and management are complementary

If safety engineers are to achieve and demonstrate safety, then safety management must make it possible for them to do so – by installing an appropriate infrastructure and developing and nurturing a safety culture. But too many engineers do not recognise the importance to their work of management. And a significant cause of this is that management does not recognise its own importance in the achievement of safety. In many organisations, even those engaged in safety-critical activities, senior management does not include safety on the agenda of directors' meetings.

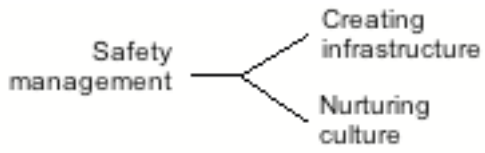

Figure 3: Key safety management roles

Engineering is perceived to be about 'doing' and 'making' things. In the present context it is about making safe systems. Whereas technicians perform activities according to specifications and rules, engineers use their judgement to apply basic principles to particular requirements in the context of a particular environment. Their decisions are based on understanding, and it is they who create the designs from which technicians must work, and the rules within which they must work. But as well as understanding fundamental principles and their applications, engineers must also employ techniques that have been derived from the principles. In this way they make their efforts efficient as well as effective.

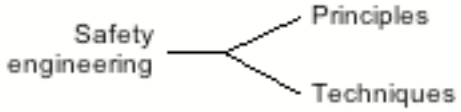

Figure 4: Key engineering fields of understanding

Safety engineering and safety management may be defined separately, as in the above paragraphs. Indeed, the engineers who carry out risk analyses are not likely to put in place their project's infrastructure, and the senior managers who define safety policy are unlikely to be involved in system development. So the two parties may see themselves as belonging to different disciplines. Yet, education and training that emphasise only safety engineering are likely to produce engineers who do not recognise or understand their dependence on management and who do not fully recognise their own responsibilities to manage, control, and take responsibility for their activities. Further, there is a dearth of education and training

on safety management, so there is seldom a formal reminder to managers of their own responsibilities for safety. Thus, we need education and training that address both safety engineering and safety management, show the interdependence between them, and emphasis their importance equally. It is recommended that educators and training providers give consideration to satisfying this need.

In establishing and maintaining control, a necessary requirement on both engineers and managers is decision-making. Safety life-cycle processes include critical decision points, at which both engineering and managerial practitioners are required to exercise considerable judgement, particularly as, in many cases, both risk and uncertainty are high.

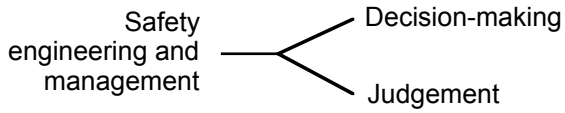

Figure 5: Key requirements of engineers and managers

In summary, safety engineers and managers must possess understanding as well as knowledge. They are responsible not merely for carrying out activities but, importantly, for determining which activities are appropriate and for making them possible by putting an appropriate control infrastructure in place. It is not sufficient for them merely to work within rules, for it is they who must decide what rules should apply and, for this they must use their judgement. They carry the responsibility for safety. Yet, engineers and managers often put aside both fundamental principles and their own judgement, by working to rules that they have not validated in the prevailing circumstances, and asking their staff to work to these rules. Such substitutions of rules for judgement are examples of a search for short cuts instead of safety – of grasping at the shadow and missing the substance. This paper examines five examples in which professionals:

- Set their goal as compliance with a standard rather than the achievement of safety;
- Focus on deriving safety integrity levels rather than understanding the risks;
- Attempt to derive numeric risk values, even when the evidence to support a quantitative approach is absent;
- Concentrate on equipment risks while ignoring the greater risks posed by human operators and managers;
- Create automated safety functions but fail to change human behaviour.

2 Comply with a Standard or Seek to Achieve Safety?

Standards perform two functions. They remind us of what we should do and they make what we do repeatable. Since many people want to avoid judgement and be told what to do, standards can be very useful. But someone has to decide what should be done. In using a standard we transfer responsibility for the decision to its authors. If we are the authors, we can have confidence that the standard defines processes that we have proved in practice. But if we use a standard that others have written, we make the assumption that their decisions are better than ours and appropriate to our circumstances. In broad terms this may be true, but each application of the standard requires interpretation, so the assumption can have considerable influence on our work.

By the time a standard is produced, it is likely to be out-of-date. It needs to be reviewed regularly, with respect both to current good practice and to its appropriateness to the team's work. The more detailed the standard, the more frequent the reviews should be and the

greater the amount of work involved in maintaining it. If it only provides a framework, by defining the outline of a working process, it may remain valid for a relatively long time. The more detail it contains on how things should be done, the more rapid is its obsolescence.

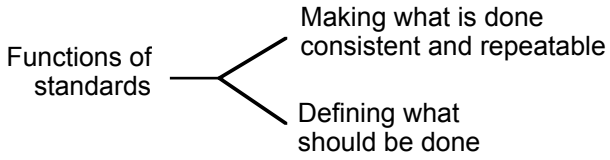

Figure 6: The roles of standards

Introducing a standard from outside throws up a number of issues.

- It is unlikely to represent the organisation's way of working and will need to be tailored for use. In addition, the organisation's processes may also have to be changed.
- It is certain to present difficulties of understanding and interpretation. The organisation needs someone who knows the standard well to provide support to its users.
- A 'generic' standard (e.g. IEC 61508) has a wide field of coverage, so many of its clauses will be unnecessary in any given application. The user organisation must tailor the standard by removing unnecessary clauses and editing others.
- It is likely not to represent up-to-date good or best practice. The delays between drafting parts of the standard and obtaining agreement on them, between combining the parts and obtaining agreement on the whole, between agreement and publication, and between publication and an organisation's adoption and use, can be considerable, particularly in the case of international standards.

Thus, an organisation needs to treat the introduction of a standard as a project, training senior managers who must make decisions about its introduction and use, determining what tailoring of the standard's requirements is necessary and carrying out the tailoring, training potential users, providing a help desk to deal with users' queries, and planning a continuous review of the standard's appropriateness and a change mechanism (Redmill 1988). The introduction of the standard demands preparation, its use requires discernment.

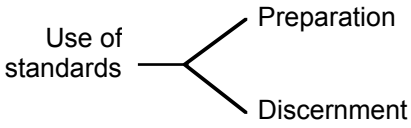

Figure 7: Necessities for the introduction and use of a standard

Unfortunately, introduction is seldom treated as a project, so the following are likely to present problems.

- Use of a standard to define what the organisation should do removes the need for judgement and replaces it with rules.
- Users of the standard focus on achieving compliance with it instead of on achieving safety.
- Many standards emphasise technical activities at the expense of managerial ones, so even managers think of them only as technical standards.
- There is often no one to address queries about the standard or to make necessary changes to it.

- A standard may define a process that has never been used or validated. Its various parts are often derived from different sources, and processes composed of several parts may not have been proved to be sensible or even practical.

The introduction of a standard requires planning and preparation, and its use requires discernment and support. These requirements represent the judgement that is the responsibility of engineers and managers. This judgement is necessary in considerable measure when the standard is first introduced, and then, perhaps to a lesser extent, each time it is used. It represents a comparison of the standard's rules against fundamental principles, in order to tailor the standard so that its users do not have to go back to the principles at each step of their work. In making the comparisons, the engineers and managers must determine to what extent compliance with the standard equates to the achievement and demonstration of safety, and, where there is a shortfall, they themselves must supplement the standard by defining what must be done. Otherwise compliance with the rules will not amount to adequate safety engineering and management. For example, the international standard IEC 61508 (IEC 2000) does not give advice on how to build up a convincing argument that safety has been achieved. An organisation that only complies with the standard may find it difficult to convince independent safety assessors of their case for safety. Compliance may get credit from a safety assessor, but it may not be sufficient for the organisation to obtain approval to operate their equipment.

3 Derive Safety Integrity Levels or Understand the Risks?

Three questions that are fundamental to safety-critical-system development are:
1. How safe must the system be?
2. How can we achieve this level of safety (or, what must we do to derive confidence that we have achieved it)?
3. How can we demonstrate that we have achieved the required level of safety?

In many modern standards, particularly IEC 61508 (IEC 2000), answering the first question requires making a quantitative calculation of the risks attached to the 'equipment under control' (EUC), carrying out an assessment of the risks against tolerability criteria, and translating some function of the result into a safety integrity level (SIL). The SIL then defines the required level of safety. In IEC 61508, the SIL is defined by the amount by which a risk needs to be reduced. In other standards, the SIL may be derived differently.

When systematic faults are dominant, the standards' answer to the second question is to equate development processes to the derived SIL, on the premise that confidence in safety is derived from employing appropriate processes. The SIL concept is summarised in the 'bowtie' diagram of Figure 8 (Redmill 1998).

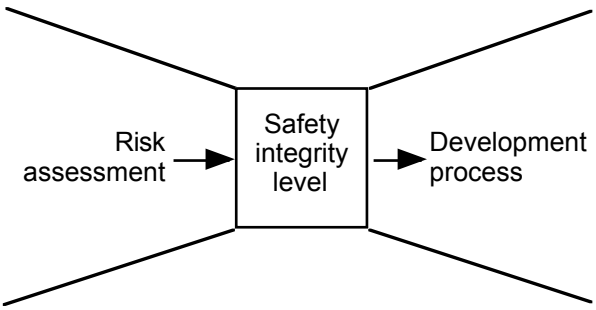

Figure 8: The 'bowtie' summary of the SIL concept

A reason for translating the initial numeric risk value into a qualitative SIL category is that numerical risk analysis is mostly impossible for software. (It is worth mentioning that the standards do not offer advice on how to carry out quantitative analysis of software-based EUCs or their control systems. If, because of the difficulty of deriving numeric risk values, confidence in them is low, the SILs that are translated from them must themselves be suspect.) Other misuses and dangers of SILs are outlined by Redmill (2000).

Defining development processes as being appropriate to the SILs is based on the assumptions that the level of reliability, or safety, achieved depends on the development processes, and that the most effective processes are also the most expensive to employ. But the lack of empirical evidence to substantiate these assumptions is pointed out by Thomas (2003), who argues for the abandonment of SILs, at least in their present form.

Emphasis on the SIL concept has caused many practitioners, supported by their organisations, to focus on deriving SILs rather than seeking means of achieving the appropriate levels of safety. Further, it is not uncommon for practitioners' assumptions to produce errors in favour of a pre-intended SIL. This is not good engineering. The SIL concept can in some instances be useful, but it is a means to an end, a tool, and therefore only the shadow. The substance is the achievement of safety. Safety engineering and management need to apply basic principles first and then employ the tool when appropriate.

Considering whether the SIL concept should be abolished, let us examine its necessity by expanding Figure 8's bowtie diagram, as in Figure 9. This shows that risk assessment produces the 'necessary risk reduction', R (using IEC 61508 as an example), from which a SIL is deduced. It is apparent that if the translation from R to SIL is omitted, then R would define the development processes, thus bypassing the need for the SIL. Thus, at first sight, the SIL concept is unnecessary.

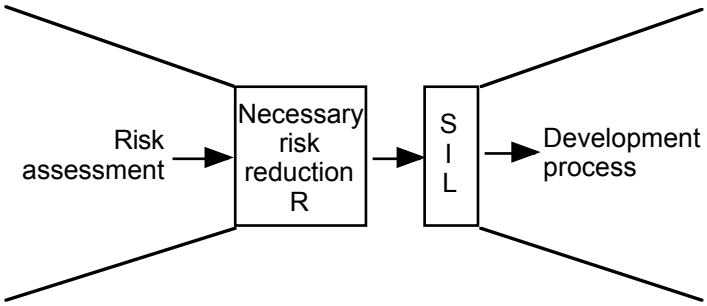

Figure 9: Is the SIL concept necessary?

However, if the risk analysis and assessment are qualitative, then R will be a categorisation, as in the MISRA (1994) Guidelines – i.e. a SIL by another name. If R is numeric, there is an infinite range of values that it could possess. This in itself does not preclude its use for defining the development processes, but there is not a process for every possible value of R, and if the appropriateness of processes is related to R at all, it must be related to broad categories. Thus, for defining development processes, R would have to be converted into a categorised value – such as a SIL. The SIL concept is therefore implicit in the assumption that the achieved level of safety depends on the development processes. So, we must ask the more useful question of whether safety is correlated with development process.

Starting from an engineering perspective, it is apparent that no process can deliver software integrity on its own. It requires conscientious application. If management is negligent and fosters a poor culture, or if the competence of practitioners is low, poor results can be expected, whatever the process. Further, although some processes force their users into structured work, which should induce a lower fault rate, there is no evidence that any particular method or process delivers a defined integrity level, even if well managed and competently applied (Thomas 2003). However, failed projects, in which 'good' and 'appropriate' processes have been used, provide evidence that the reverse is true. It is more likely that good and professional practitioners, with good culture and the intent to do a good job, will be successful, in spite of the processes that they use. Thus, confidence of consistently achieving a SIL by the development processes defined in the standards is low.

The true need is to demonstrate safety according to the three prongs of Figure 10. When the SIL concept is useful in meeting this need, it is right that it should be used – but always as a tool; it should never be allowed to become the principal focus.

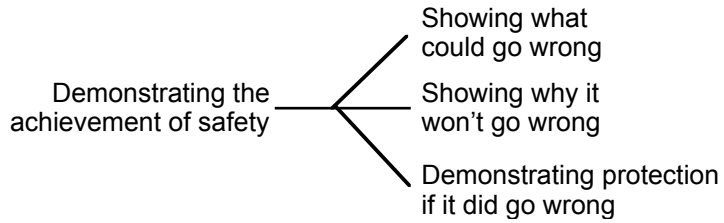

Figure 10: The three prongs of demonstrating safety

4 Calculate Numbers or Analyse the Risks?

The origins of risk analysis lie in reliability theory, which was developed when almost all equipment was electromechanical or mechanical and failed randomly. Historic failure data provided the basis for the derivation of statistical distributions that could reasonably be expected to predict future failure rates. Thus, knowledge of the operational histories of a system's components permitted calculations, via fault trees, of the probabilities of occurrence of the system's various failure modes. Of course there was subjectivity involved in such risk analyses (Redmill 2002a, 2002b). Their accuracy depended on the thoroughness of the identification of failure modes, the choice of design of the fault trees, and other variables. But, in the main, the results were considered to be accurate. Almost invariably risk analysis employed quantitative methods and produced numeric results.

But, in general, today's systems are likely to include electronic components, many of which are software-based and more complex. Such systems cannot be tested exhaustively in cost-effective time, and, although random failures may occur, their faults are predominantly systematic. When these conditions maintain, attempting to derive numeric probabilities of failure can be speculative, particularly when the software is bespoke and recently developed and there is no history of operation. Yet many risk analysts continue to derive numeric results, even though they may only be translations of qualitative risk estimates. The trouble is that decision-makers are likely to believe quantitative risk values that are communicated to them, if the assumptions that were made in their derivation are absent.

But the apparent precision of numbers is not the same as accuracy. The accuracy of results depends on the correctness of information, and confidence in this depends on the pedigree of its source. A reckless search for numbers is not a good substitute for an attempt to understand the risks. In the absence of adequate evidence, or when our confidence in

evidence sources is low, it is often better engineering to replace quantitative calculation with qualitative estimation, and to communicate:

- Our assumptions;
- An assessment of the pedigree of the data on which our analyses are founded;
- An assessment of the accuracy of the estimate of each risk value.

Such information would remind decision-makers that the risk values are estimates and that judgement is required in using them. Safety managers should expect to find such information associated with risk values, and if it is absent they should ask for it.

5 The Risks Posed by Managers

Safety standards, even modern ones, give almost no advice on how to include in analyses the risks posed by humans. IEC 61508 (IEC 61508) mentions 'human factors', but, although it provides detailed instruction on dealing with hardware and software risks, there is no equivalent guidance on what human factors are or how to identify, analyse, and assess their risks. Similarly, the standard defines both hardware and software safety integrity but does not propose an equivalent method of stating the requirements on human elements of systems. Yet it is accepted that the human is an integral part of most systems. Should there be a 'human safety integrity' to define the necessary credentials of system operators?

It is often said that there is a human cause, or partial cause, of almost every accident. If this is true, then risk analyses that do not include the risks posed by a system's human elements cannot be complete. Indeed, they would exclude what is, potentially, the major risk factor, and their results would be optimistic.

During the drafting of IEC 61508, engineers were, typically, not familiar with the subject of human factors or with the human reliability assessment (HRA) techniques that had been emerging since the 1970s. Even now, when an increasing number of organisations are attempting to include operator error in risk analyses, the use of these techniques is not widespread, for they are still mainly the preserve of the psychologists and ergonomists who developed them. But if intuition is our only tool in addressing the human component, we miss the opportunity to employ the techniques that were designed for the job (Redmill 2002c). Some HRA techniques require the construction of databases of information on human error, both physical and cognitive, in order to provide as sound a foundation as possible for estimating the likelihood of error in the future. Their analysis is essentially qualitative and, although many techniques include translations of non-numeric results into quantitative risk values, their merit lies in taking a methodical approach rather than in arriving in accurate numeric results.

However, the greatest deficiency in our analyses may be the total omission of the risks posed by management, and, particularly, senior management. We now include human factors in some risk analyses, and are learning about HRA, but we have not yet begun to address management risks. Yet, evidence that senior management's leadership and decision-making failures are primary causes of accidents is provided abundantly in the reports of accident inquiries. In her examination of the 1986 Space Shuttle Challenger disaster, Diane Vaughan (1996) shows how deviant behaviour within NASA became the norm and caused systematically flawed decision-making.

The safety policies that management create, senior management's leadership in safety matters, and the attitudes that management promote in nurturing corporate culture, are all recognised as having considerable influence on functional safety. These topics are being addressed in academe under the heading of 'safety culture' (for a review see Gadd and Collins 2002). In some organisations they are being addressed as a way of improving safety.

But even so, Kletz (2000) says that managers do not realise that they could do more to prevent accidents. He points out how little they feel the need to get involved in the details of safety and, by contrast, how detailed their involvement is in production and cost. He shows (see Figure 11) the disproportionately small effort that they accord to their own failure as opposed to equipment failure.

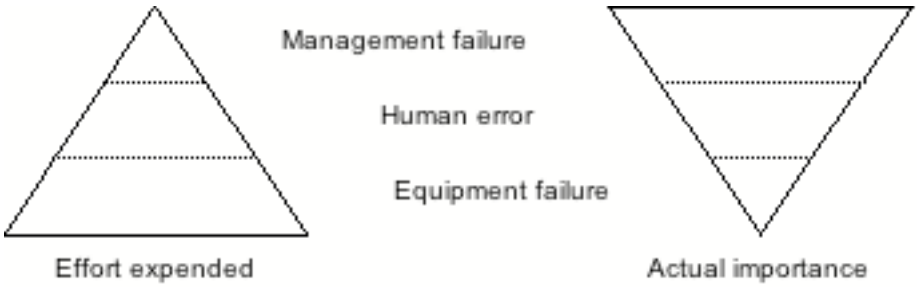

Figure 11: Types of failure: comparing importance against management effort

Risk analysts examine the smallest details of how equipment could fail. Yet, although management failures can predispose whole organisations to accident, their potential is not analysed. Risk analyses may provide a more or less accurate estimate of the likelihood of equipment failure, but, if they do not include what is perhaps the most significant element of safety risk, they almost certainly provide an underestimate of the likelihood of accident.

6 The Need to Change Behaviour

A strategy for dealing with risk is to purchase insurance. In respect of purely financial loss, insurance cover may provide a like-for-like replacement. But when it comes to safety, insurance can only provide compensation. It does not improve safety. Yet, in many organisations, the attitude towards the insurance of safety risk is the same as that towards the insurance of financial risk: it is thought that the risk has been covered and that it is acceptable to proceed. Insurance is a risk-management activity, but when it is treated as a risk-reduction activity it gives a false sense of security.

Similarly, other risk-management activities, including risk-reduction activities, are often assumed, without validation, to achieve the results that they are intended to achieve. But many do not. Unintended consequences may introduce new risks, in some cases greater than those being managed. The types of unintended consequences of interest here are those concerning the behaviour of humans.

Wilde (1994) points out that when people are made safer, they are likely to trade their increased safety for improved performance, and he offers numerous examples of this, mainly in the field of road traffic. He refers to studies of the use of seat belts, anti-lock breaking systems, crash bars, and other equipment-based safety functions, which the authorities assumed would achieve accident reduction. He shows that they are only successful if ‘all other things remain equal’ and that, in fact, other things do not remain equal, for the safety functions have the effect of inducing drivers to drive faster and closer to other vehicles, take corners more sharply, and leave breaking later. The result is that the accident rate typically is not reduced and is often increased. Wilde also points to accident migration. For example, seat belts may lead to the deaths of fewer vehicle occupants, but the accompanying change in driver behaviour may lead to more deaths of cyclists and

pedestrians. In one case, the imposition of a speed limit on a German expressway resulted in a 21% reduction in accidents but was accompanied by a 29% increase on a parallel road on which there was no limit.

From these studies, Wilde deduces that, when humans are involved, improving their safety leads to more reckless behaviour that returns the experienced level of risk approximately to its former level. Thus, a genuine safety improvement requires not only safer equipment but also countermeasures that change human behaviour. We must attempt to increase people's motivation towards safety.

We need to learn from this. Because our risk analyses are mostly concerned with equipment risks, our countermeasures take the form of changes in equipment design and the addition of protective functions. When we make changes to working procedures, we tend to assume that humans will adhere to them. Yet many accidents are caused by intentional violations of procedures, made because the procedures are perceived to be inappropriate or restrictive. We need to pay more attention to human behaviour. We should consider the behavioural changes that might take place when safety improvements are introduced, and also the changes that may need to be made in order to achieve safety improvements. For both of these, we should attend to motivation. We must ensure that:

- Procedures are appropriate, safe, and understood and approved by their users;
- Man-machine interfaces encourage safe working;
- Measures to improve safety include the motivation of people to want safety;
- 'Good' safety culture is rewarded.

7 Discussion and Some Recommendations

The achievement and demonstration of appropriate system safety require not only safety engineering but also safety management. The two are not separate but integrated. Management must create the safety infrastructure within which technical activities are carried out; and feedback from technical activities, such as risk analysis, must provide information to inform decision-making. Both engineers and managers require knowledge of safety principles and both need to exert control over project and operational activities. Both are responsible for safety decisions that require sound judgement.

However, the need to make good practice teachable and repeatable means that activities must often be defined in terms of procedures and rules. But the rules must be securely founded on basic principles, and it is the engineers and managers who must make them. When engineers and managers discard principles in favour of standards and other tools, they also discard their professionalism. They transfer decision-making to the makers of the standards and tools, and they cease to be guarantors of safety. This paper has shown a few examples of how this can happen.

Yet, safety engineers and managers cannot escape the responsibilities that their positions carry. As a reminder of the professional standards required of them, and of the integrated way in which safety engineering and management should function, here are a few recommendations.

- Senior managers need to develop a greater understanding of safety principles and a greater awareness of their safety responsibilities. Management training needs to emphasise safety issues.
- Engineers must recognise their dependence on safety management systems, and their responsibilities in creating and maintaining them. Engineering training should address the management and control aspects of engineers' responsibilities.
- Educators and trainers should teach safety engineering and safety management as complementary, interdependent and integrated.

- We need clear distinctions between the engineers and managers who define procedures and set rules and the technicians and other staff who adhere to them. The decision-makers must recognise their responsibilities, and they should be expected to display sound judgement and to possess an understanding of fundamental safety principles.
- The introduction of a standard or other tool (including conceptual tools such as the SIL concept) should be treated as a project. Subsequently, each application should be planned with reference to basic principles, carried out with discretion, and monitored.
- Engineers need to develop a better understanding of human cognition, and to learn to apply the human reliability assessment techniques that exist for addressing the risks attached to human factors. At the same time, engineers should also work with psychologists and ergonomists who are expert in the application of such methods.
- We need to consider not only how management risks may be reduced but also how they may be included in risk analyses.
- In risk-reduction and other risk-management activities, we should pay more attention to the unintended consequences of human behaviour. We should promote safe behaviour by the motivation of safe attitudes.

This list is not exhaustive. It merely offers a few suggestions for the improvement of safety engineering and management practice.

8 References

Gadd S and Collins A M (2002). Safety Culture: A review of the literature. Health & Safety Laboratory, Sheffield

IEC (2000). Functional safety of electrical/electronic/programmable electronic safety-related systems (standard in seven parts). International Electrotechnical Commission, Geneva

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

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

Redmill F (1988). The Introduction, Use and Improvement of Guidelines. Proceedings of SAFECOMP '88, Fulda, Germany. Pergamon Press, Oxford

Redmill F (1998). IEC 61508 – Principles and Use in the Management of Safety. Computing & Control Engineering Journal, 9 (5), October

Redmill F (2000). Safety Integrity Levels – Theory and Problems. Lessons in System Safety, Proceedings of the Eighth Safety-critical Systems Symposium, Southampton, UK (Ed. Felix Redmill and Tom Anderson). Springer-Verlag, London

Redmill F (2002a). Risk Analysis – a subjective process. Engineering Management Journal, 12 (2), April

Redmill F (2002b). Exploring Subjectivity in Hazard Analysis. Engineering Management Journal, 12 (3), June

Redmill F (2002c). Human Factors in Risk Analysis. Engineering Management Journal, 12 (4), August

Thomas M (2003). Issues in Safety Assurance. Computer Safety, Reliability, and Security – 22nd International Conference, SAFECOMP 2003, Edinburgh, UK (Ed. Stuart Anderson, Massimo Felici, Bev Littlewood). Springer

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

Wilde G (1994). Target Risk. PDE Publications, Toronto