Making Robust ALARP Decisions for In-Service Systems

A.J. Rae, M. Nicholson*

* Department of Computer Science, University of York, U.K. (E: andrew.rae, mark.nicholson@cs.york.ac.uk)

Keywords: ALARP, Operational Safety, Risk Acceptability

Abstract

In the United Kingdom, and in similar jurisdictions, it is a legal requirement that safety risk associated with a system be demonstrably As Low as is Reasonably Practicable (ALARP). Whilst a complete ALARP justification can be prepared in parallel with design activities, the operation of a system brings change and new information which can invalidate this justification. This paper presents a framework for addressing issues associated with operational ALARP, and examines some of the ALARP questions arising from in-service incidents.

1 Introduction

This paper confronts the issues faced by operators of systems in maintaining acceptable levels of safety risk presented by those systems. Simplistically, it may appear that a system which has been declared safe at the time of introduction to service needs only to be maintained in accordance with the requirements and assumptions of the safety case used for initial certification. However, acceptance into service can only ever be justified in terms of forecast risk based on imperfect information. Operation brings new data about the configuration, role and environment of operations, the reliability of components, the safety of designs, and the social acceptability of the safety risks presented. Over time, new mitigations to risks may become available, challenging the justification used for initial design decisions.

Section 2 of the paper provides some basic concepts of risk, uncertainty, and risk acceptability. Section 3 uses the As Low as Reasonably Practicable (ALARP) principle as an example to show how the acceptability of safety risk presented by a system can deteriorate over time. In Section 4 an ethical decision-making process is used as a framework to establish requirements for maintaining risk acceptability. Sections 5 and 6 examine existing regulations and recent incidents to illustrate the complexity of the issues. Finally, the paper sets out objectives for further research on this topic.

2 Risk, Uncertainty, and Outcomes

The decision to enter a new system into service is informed by a forecast of the level of safety risk presented by that system. This forecast may be based on historical risk data, analysis and test of the system, or some combination of these. The actual risk presented by a system during its operational life will necessarily vary from the forecast due to uncertainties in the information sources used for the estimate.

To fully understand this, it is necessary to distinguish between three concepts: probability, uncertainty, and outcomes [14]. If offered a wager on the outcome of the roll of a die, you may forecast that the probability of rolling a “six” is one-sixth. There is uncertainty in this estimate. To make such a forecast you must assume that the die has six differently numbered sides, is fair, and that the method of rolling will preserve this fairness. The actual probability of rolling a “six” may be one-tenth. Now, if the die has been rolled, the outcome may be a “five”. The fact that you now know that a “six” was not rolled, does not change the fact that the initial probability was one-tenth.

When making decisions about the acceptability of system deployment, risk is forecast, not known. Even after a system has served for many years and been decommissioned, it is still not possible to know with certainty the level of risk that those who interacted with the system faced. With enough instances of the system in use, and records of the outcomes, upper and lower bounds on the risk can be determined with a specified level of confidence.

Decisions are further complicated by the fact that risk acceptability is seldom based on the absolute magnitude of the forecast risk. Frameworks such as “As Low as Reasonably Practicable (ALARP)” [11], “Globelment au moins aussi bon (GAMAB)” [3] and to some extent “Minimum Endogenous Mortality (MEM)” [3] compare the forecast of risk presented by a system with either forecast benefits or with alternate risks.

Thus, risks are typically judged to be acceptable not through an objective timeless measurement, but by comparing a forecast to a changeable standard. An otherwise sound decision to introduce a system into service can be called into question through new information or through shifts in the viability and risks of alternatives.

3 Fragility of ALARP Determination

The fact that a system has been approved as safe by a certification body at the time of acceptance into service is not evidence that the system is currently safe. Under some legal frameworks, including the United Kingdom’s “As Low as

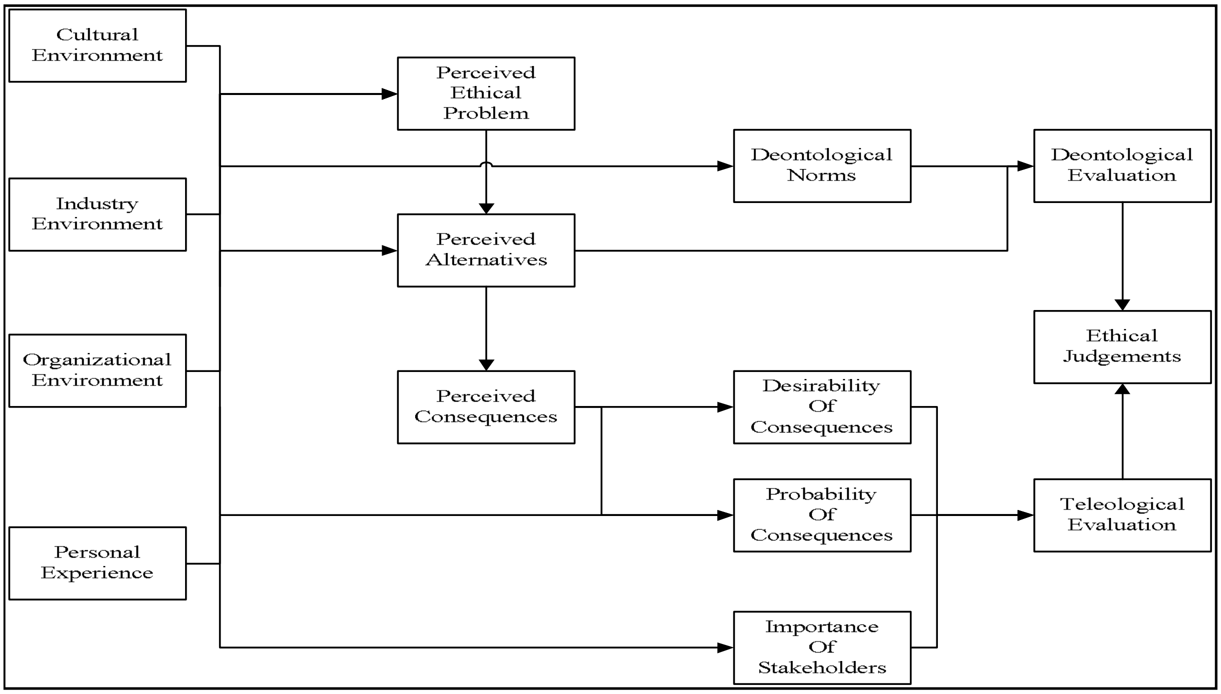

Figure 1- Ethical Decision Making Process

Reasonably Practicable (ALARP)" regime, a system may become unsafe even in the absence of adverse evidence.

To establish the truth of this claim, consider that a complete argument demonstrating ALARP must necessarily contain the following elements:

1. Specification of a particular system design
2. Specification of a particular configuration, role and environment (CRE) in which the design will operate
3. An identified list of hazards associated with operation of the system
4. A claim that the list of hazards is sufficient
5. An assessment of the safety risk presented by each of the hazards, including estimates and assumptions used to quantify the safety risk
6. A claim that for all mitigations which were not included in the system design or CRE, the mitigations were not reasonably practicable to implement.

This list of elements provides a framework for showing how risk which was demonstrated to be ALARP can become no longer ALARP.

(a) The system design may change
(b) The CRE may change beyond the constraints considered or assumed by the ALARP argument
(c) New hazards may be postulated or realised
(d) Failure to adequately implement a search process for new hazards may invalidate the claim of hazard list sufficiency
(e) Operational or maintenance data may contradict or otherwise invalidate estimates and assumptions used to assess the safety risk
(f) New mitigations may become available, or changes may occur to the practicality of known mitigations.

4 Maintaining Robust ALARP Justification

Preserving safety risk acceptability is a decision-making process. Since it may cause harm or benefit to others, it meets the definition of "moral" or "ethical" decision-making. Whilst "moral" and "ethical" can have varied and distinct meanings, here we use them interchangeably as descriptors of a type of decision to be made, rather than as a value judgement on the quality of the final decision. Hunt and Vitell [12] present a model of how ethical decisions are made and carried out. The parts of this model dealing with decision making are shown in **Figure 1**. Their model is descriptive rather than normative: that is, it shows how decisions are actually made rather than an ideal of how they should be made.

From this model, it is possible to derive the required elements of a system for maintaining safety risk acceptability. Firstly, it is necessary to have an information gathering system listening to the cultural, organisational and industry environment. Specifically, this system must be keyed to identify any of the events (a) to (f) above.

Secondly, there must be a trigger system to recognise when the safety justification has been challenged by incoming information. This is equivalent to the box in the model of **Figure 1** which shows a "perceived ethical problem". At a minimum, a decision-making process must be invoked:

- When an assumption or estimate made as part of the safety argument or evidence is invalidated by actual operational data. For example, if an accident co-effector is judged to be "incredible" as part of a safety assessment, and there is a reported occurrence of the co-effector, that part of the analysis is invalidated.
- When a design change (including a change in maintenance or usage procedures) alters any part of the system that was subject to detailed safety analysis, or any interface with a part of the system subject to detailed safety analysis.
- When an serious accident, incident or near-miss occurs (this is a special case of assumption invalidation, as systems are seldom deployed when serious accidents are forecast to be anything but incredible).

Thirdly, there must be a mechanism for complete and honest listing of options (shown as "perceived alternatives" in **Figure 1**). This mechanism must include:

- Recognition that removal from service is an option
- Recognition that imposing operational limitations is an option
- Recognition that selecting an option that cannot be instantaneously implemented involves accepting exposure to risk beyond that forecast at the time of certification
- Recognition that seeking further information or analysis is an option and not exclusive of other options.
- Recognition that communication of risk information is an option and not exclusive of other options.

The fourth and fifth pillars of the system provide for deontological and teleological evaluation of the options. As defined by Broad [2], deontological evaluation is concerned with the inherent rightness or wrongness of options. Examples of deontological thinking include judging acts according to the motives or intent of the actors, or judging acts according to rights or duties. Teleological evaluation, by contrast, is concerned with the outcomes of options. This need not be utilitarian cost-benefit analysis, but may include aspects of fairness, consistency and equity.

Engineering ethics as well as safety regulation require both deontological and teleological evaluation. Elaboration of a full method for performing and reconciling these evaluations is beyond the scope of this paper. Recent promising work on trade-off decision-making for dependability includes Despotou's "Factor ANalysis and Design Alternatives (FANDA)" and associated "Trade-Off Methodology (TOM)" [4].

5 Existing Regulation

5.1 Civil Aerospace – European

The European Union mandates that type-certificate holders maintain a "system for collecting, investigating and analysing reports of and information related to failures, malfunctions, defects or other occurrences" [6]. Airworthiness Directives are issued when an unsafe condition is detected that is likely to exist or develop in aircraft other than the one where the initial condition was detected.

This reactive system is, at face value, insufficient for ALARP maintenance. It does not implement a search process for new hazards or available safety measures, and has no explicit triggering mechanism when assumptions made in the safety case are violated (unless those violations also constitute directly unsafe conditions). The type-certificate system does involve explicit triggers for changes in design or CRE.

Guidance material associated with the regulations serve:
"(a) To postulate basic principles which should be used to guide the course of actions to be followed so as to maintain an adequate level of airworthiness risk after a defect has occurred which, if uncorrected, would involve a potential significant increase of the level of risk for an aircraft type.

(b) For those cases where it is not possible fully and immediately to restore an adequate level of airworthiness risk by any possible alleviating action such as an inspection or limitation, to state the criteria which should be used in order to assess the residual increase in risk and to limit it to an appropriate small fraction of the mean airworthiness through life risk." [5]

These principles are based on explicit recognition that removal from service and imposing operational limitations are options, and provide quantitative acceptability limits for additional exposure to risk. These limits are in turn based on the understanding that multiple periods of increased risk are inevitable for commercial aircraft, and can be included in consideration of through-life acceptable risk.

Table 1 taken from the guidance material shows the flying or calendar time within which a defect should be corrected. The assumptions are that the aircraft life is 60,000 hours and that there will be 10 'catastrophic event' campaigns through the life of the type.

Estimated catastrophe rate to aircraft due to the defect under consideration (per aircraft flight hour)	**Average reaction time for aircraft at risk (hours)**	**On a calendar basis**
4 x 10-8	3750	15 months
5 x 10-8	3000	12 months
1 x 10-7	1500	6 months
2 x 10-7	750	3 months
5 x 10-7	300	6 weeks
1 x 10-6	150	3 weeks
1 x 10-5	15	Return to base

Table 1: Maximum allowable correction time [5]

5.2 Civil Aerospace – American

The FAA [7] has similar requirements for type-certificate holders to the European Aviation Safety Agency with respect to reporting failures and defects, but without the requirement to maintain a system for collecting and investigating reports and information.

FAA Order 8040.4 provides that safety risk management be treated on a cost-benefit basis. Associated guidance [8] indicates that this should be performed for safety using Value of Statistical Life (VSL) calculations. This is a strictly utilitarian evaluation process, ignoring any deontological considerations. Policy does not appear to mandate consideration of removal from service and operational limitation as options.

5.3 Nuclear

The Health and Safety Executive (HSE) in its guidance for Nuclear Directorate Inspectors [10] recognises that older nuclear plants may meet ALARP requirements at higher risk levels than newer plants. HSE requires that where certain standards are not met, the requirement for ALARP translates into a plan to reduce the risk to acceptable levels "within as short a period as reasonably practicable".

6 ALARP in the Real World

6.1 Case Study One – Boeing 777-200 Icing Induced Engine Failure

Rolls-Royce is currently developing a modification for the fuel-oil heat exchanger on their Trent 800 engines after investigators managed to replicate the icing-induced constriction suspected of causing British Airways Boeing 777-200 crash at London Heathrow in 2008 [1]. Through their tests investigators have shown, with "reasonable repeatability" that a layer of ice can accumulate inside the fuel-delivery pipes, greater than the quantities needed to block the heat exchanger if flushed downstream. The investigators at the AAIB recommended that Boeing and Rolls-Royce jointly reviewed the 777's engine fuel system and developed changes which prevent ice restricting fuel flow at the heat exchanger.

This accident shows a number of the difficulties associated with determining whether a product remains ALARP after an accident. First, there was a challenge to the claim of independence of failure between the two engines. In this case the coupling was via an element outside the control of the engine developer. This was a "systems of systems" failure, so the question arises as to which systems no longer are justifiably ALARP. Civil aircraft and engines are separately certified but have to work together to provide overall airworthiness so it would appear reasonable that both are deemed to have their ALARP status challenged.

An initial response to the Heathrow accident was to change the flight operations of the affected aircraft and to undertake extra actions during flight. This alters the operation performance and risk profile of aircraft operations thus potentially challenging ALARP further. The Rolls-Royce redesign is a more permanent solution to the issue, although delayed in its application; it should enter service approximately two years after the Heathrow accident. In both circumstances the ability for a product to claim to be ALARP is addressed by a mixture of responses by a variety of parties. So, how much can a claim by third parties be relied on for an ALARP claim?

6.2 Case Study Two – Snatch Land Rover

The mother of a soldier killed by a roadside bomb in Iraq has gained permission for a judicial review challenge into the Defence Minister's decision not to hold a public inquiry into the use of the Snatch Land Rover [9]. Thirty-eight soldiers have been killed by roadside bombs whilst travelling in the Snatch vehicle.

This is a case where the operational circumstances have changed. First, combat operations rather than peacetime operations are being undertaken. Second, the Iraqi insurgents and Taleban fighters have altered their tactics to take note of the lightly armoured Snatch. Third, safety risk acceptability for the armed forces appears to be changing in UK society.

How quickly are the duty-of-care holders required to react to these changes? How does the duty of care express itself in the form of ALARP when different risk environments need to be addressed through time? How can responses to this change be addressed within the ALARP framework? For instance avoiding use of the Snatch before a suitable alternative is available may cause overall more safety risk than continuing to use the vehicles. How much of this responsibility lies with the MoD (who procured and operate the vehicle) and how much lies with the developers of the Snatch?

6.3 Case Study Three – Cairns Tilt Train

On 10 April 2009 Queensland Rail announced that:

"... preliminary information has raised issues concerning the specification, design and subsequent testing and construction of the Cairns Tilt Train.

We have commenced work immediately with our technical and engineering experts to provide assurances to the integrity of design and construction.

We have acted on the side of caution and cancelled services until the nature of the issues raised can be addressed." [16]

On 30 April, less than three weeks later, Queensland Rail issued a further press release which began:

"The Cairns Tilt Train will return to full service from Monday (4 May) following formal notification today by the Rail Safety Regulator that QR has addressed all concerns raised in the advice to QR on 9 April." [17]

Whilst the exact nature of the issues is not in the public domain, this is notably a case where an organisation was uncertain that an ALARP justification held, and took the precaution of removing the system from service until the organisation was confident that the concerns were unfounded.

7 Open Questions

7.1 Accidents and Incidents as Change Drivers

Many operational or design changes which are reported in the press occur as a result of major accidents or incidents. There are several possible explanations for this:

1. Accidents and incidents are reliable indicators that forecast risk is inaccurate, and techniques for responding to accidents and incidents are better developed than techniques for gathering and responding to more subtle indicators.
2. Changes in public risk perception and risk acceptability are driven by media stories, as a result of the availability heuristic [13]. This increases public demand for operational or design changes.
3. Operational and design changes happen frequently in response to subtle indicators, but are not publicly reported.

Empirical research is required to determine the nature and extent of this phenomenon. Are changes in response to accidents driven primarily by new information, or by different perceptions? How can risk forecasts best be modified without resorting to accidents as crude indicators, and how can responses to this information be reliably triggered?

7.2 Availability Cascades and Single-Voice Alarms

An availability cascade is a feedback loop where the perceived importance of an issue rises rapidly due to its increased availability in the public discourse [15]. Whilst this phenomenon is readily recognised, an appropriate course of action in response is not well defined.

A parallel situation exists where responsible decision makers are informed of potential risks that are not readily quantifiable. With hindsight after an accident, ignoring such "warnings" appears grossly negligent, but determining and justifying the best course of action in response to alarms is another area where guidance is lacking.

7.3 Risk Decision Making Under Uncertainty

The framework presented in this paper does not directly address the process of ethical judgement of options in response to operational safety risks. Whilst we have hinted that FANDA [4] may be a starting point, it needs to be combined with theories of decision making under high uncertainty to be widely useful for safety risk decision making.

7.4 Structuring Organisational Communication

Another part of the framework which requires development is structuring organisations and communication channels to effectively identify challenges to safety risk forecasts, thereby triggering explicit decision making. If staff in the appropriate roles do not comprehend the assumptions on which the risk forecasts depend, they will not recognise challenges to those assumptions, even if they are presented with suitable data.

References

[1] Air Accidents Investigation Branch, Interim Report: Accident to Boeing 777-236ER, G-YMMM at London Heathrow Airport on 17 January 2008, 2009.

[2] C.D. Broad, *Five Types of Ethical Theory*. New York: Harcourt, Brace and Co., 1930.

[3] CENELEC, EN50126 Railway applications: The specification and demonstration of reliability, availability, maintainability and safety (RAMS), 1999.

[4] Georgios Despotou and Tim Kelly, "An Argument Based Approach for Assessing Design Alternatives and Facilitating Trade-offs in Critical Systems," *Journal of System Safety*, vol. 43, no. 2, March - April 2007.

[5] European Aviation Safety Agency, Executive Director Decision 2003/1/RM, 2003.

[6] European Union, "COMMISSION REGULATION (EC) No 1702/2003," *Official Journal of the European Union*, September 2003.

[7] Federal Aviation Administration, Code of Federal Regulations, Title 14 Section 21, 2009.

[8] Federal Aviation Administration, Economic Analysis of Investment and Regulatory Decisions - Revised Guide, 1998.

[9] Frances Gibbs, "Soldier's mother wins court fight over Snatch Land Rovers," *The Times*, July 11 2009.

[10] Health and Safety Executive, Nuclear Directorate

Guidance on the Demonstration of ALARP, 2009.

[11] Health and Safety Executive, *Reducing Risks, Protecting People: HSE's Decision-Making Process.*, 2001.

[12] Shelby D. Hunt and Scott Vitell, "A General Theory of Marketing Ethics," *Journal of Macromarketing*, vol. 6, no. 5, 1986.

[13] Daniel Kahneman, Paul Slovic, and Amos Tversky, *Judgement under uncertainty: Heuristics and Biases.*: Cambridge University Press, 1982.

[14] Frank H Knight, *Risk, Uncertainty and Profit.*: Houghton Mifflin Company, 1921.

[15] Timur Kuran and Cass R Sunstein, "Availability Cascades and Risk Regulation," *Stanford Law Review*, vol. 51, no. 4, 1999.

[16] Queensland Rail, Press Release: Cairns Tilt Train cancelled until further notice, 10 April 2009.

[17] Queensland Rail, Press Release: Cairns Tilt Train safe to return to full service, 30 April 2009.