The University of Manchester
Manchester
Business School

Working Paper Series

Human Reliability Analysis:
A Review and Critique

Sondipon Adhikari, Clare Bayley. Tim Bedford, Jerry Busby, Andrew Cliffe, Geeta, Devgun, Moetaz Eid, Simon French, Ritesh Keshvala, Simon Pollard, Emma Soane, David Tracy, Shaomin Wu

Manchester Business School Working Paper No 589

Manchester Business School

Manchester Business School
The University of Manchester
Booth Street West
Manchester M15 6PB

+44(0)161 306 1320
http://www.mbs.ac.uk/research/workingpapers/

ISSN 0954-7401

Author and Affiliation

Simon French
Manchester Business School
Booth Street West
Manchester
M15 6PB
Phone: +44 161 275 6333
Email: simon.french@mbs.ac.uk

Abstract

Few systems operate completely independent of humans. Thus any study of system risk or reliability requires analysis of the potential for failure arising from human activities in operating and managing this. Human reliability analysis (HRA) grew up in the 1960s with the intention of modelling the likelihood and consequences of human error. Initially, it treated the humans as any other component in the system. They could fail and the consequences of their failure were examined by tracing the effects through a fault tree. Thus to conduct a HRA one had to assess the probability of various operator errors, be they errors of omission or commission. First generation HRA may have used some sophistication in accomplishing this, but in essence that is all they did. Over the years, methods have been developed that recognise human potential to recover from a failure, on the one hand, and the effects of stress and organisational culture on the likelihood of possible errors, on the other. But no method has yet been developed which incorporates all our understanding of individual, team and organisational behaviour into overall assessments of system risk or reliability.

Keywords

Cynefin categorisation of tasks; normative and summative risk and reliability analyses; high reliability organisations; human reliability analysis (HRA); Swiss cheese model.

JEL Classification

How to quote or cite this document

French et al. (2009). Human Reliability Analysis: A Review and Critique. *Manchester Business School Working Paper, Number 589* available: http://www.mbs.ac.uk/research/workingpapers/

This report was produced by the *Rethinking Human Reliability Analysis Methodologies* project funded by the UK Engineering and Physical Sciences Research Council under contract: EP/E017800/1. The contract was led by Manchester Business School, the University of Manchester and involved Cranfield University, the Universities of Kingston, Lancaster, Nottingham, Strathclyde and the University of Wales, Swansea.

The project aimed to:

- survey a range of relevant literatures to identify knowledge, theories and models that should be incorporated into an updated human reliability analysis (HRA) methodology;
- explore several case studies to identify some of the interactions between people, organisations, management and cultures that led to 'human error' in the broadest of senses;
- explore recent suggestions for enhancing HRA by the inclusion of some of the qualitative understandings of the contexts and mechanisms that affect human error;
- identify required elements for an enhanced quantitative methodology for HRA.

In addition to this report the project also produced the following papers:

- T. Bedford and C Bayley (2008) 'Sensitivity analysis and the CREAM method for Human Reliability' Department of Management Science, University of Strathclyde, Glasgow, G1 1QE.
- S. Wu and S. J. T. Pollard (2008) 'Human reliability analysis has a role in preventing drinking water incidents' School of Applied Science, Cranfield University, Cranfield, MK43 0AL.

Human Reliability Analysis:

A Review and Critique

Final report of the EPSRC funded project

Rethinking Human Reliability Analysis Methodologies

Sondipon Adhikari[1], **Clare Bayley**[2]**, Tim Bedford**[3], **Jerry Busby**[4], **Andrew Cliffe**[5], **Geeta Devgun**, **Moetaz Eid**, **Simon French**[6], **Ritesh Keshvala**, **Simon Pollard**[7], **Emma Soane**[8], **David Tracy**, and **Shaomin Wu**

Abstract

Few systems operate completely independent of humans. Thus any study of system risk or reliability requires analysis of the potential for failure arising from human activities in operating and managing this. Human reliability analysis (HRA) grew up in the 1960s with the intention of modelling the likelihood and consequences of human error. Initially, it treated the humans as any other component in the system. They could fail and the consequences of their failure were examined by tracing the effects through a fault tree. Thus to conduct a HRA one had to assess the probability of various operator errors, be they errors of omission or commission. First generation HRA may have used some sophistication in accomplishing this, but in essence that is all they did. Over the years, methods have been developed that recognise human potential to recover from a failure, on the one hand, and the effects of stress and organisational culture on the likelihood of possible errors, on the other. But no method has yet been developed which incorporates all our understanding of individual, team and organisational behaviour into overall assessments of system risk or reliability.

[1] University of Wales, Swansea

[2] Manchester Business School

[3] University of Strathclyde

[4] University of Lancaster

[5] University of Nottingham

[6] Address for Correspondence: Simon French, Manchester Business School, the University of Manchester, Booth Street West, Manchester, M15 6PB, UK. Email: simon.french@mbs.ac.uk

[7] Cranfield University

[8] University of Kingston

In this report we explore these issues, surveying briefly both the HRA literature and related literatures on behavioural and organisational theories. We conclude that:

(i) no single HRA method is likely to suitable for all purposes and contexts;

(ii) the range of theoretical bases for modelling human reliability needs to be extended to take account of, e.g., more recent understandings of human cognition and behaviour as well as organisational effects on performance;

(iii) much more comparative research is needed on the coherence, strengths and weaknesses of different HRA methods proposed to date.

Only then will we be able to build comprehensive system risk and reliability analyses in which a reasonable degree of trust may be placed.

Keywords: Cynefin categorisation of tasks; normative and summative risk and reliability analyses; high reliability organisations; human reliability analysis (HRA); Swiss cheese model.

Contents

1 Introduction

Reliability analysis [1, 2] and risk analysis [3-5] are two subjects with a great deal of overlap: the former tending to deal with systems subject to repeated failures and the need for maintenance policies to address these; the latter tending to deal with more catastrophic one-off failures that may write-off a system with concomitant impacts elsewhere. But both essentially are concerned with anticipating possible failures and assessing their likelihood. Human reliability analysis (HRA)[9] relates to methodologies for anticipating and assessing the effect of those failures which relate to human action or inaction, and not the failure of some physical component. It should be noted that human error is a major contributor to the risks and reliability of many systems: over 90% in the nuclear industry [6], over 80% in the chemical and petro-chemical industries [7], over 75% of marine casualties [8] and over 70% of aviation accidents [9]; see also Hollnagel [10]. A survey of failures in drinking water distribution undertaken as part of this project examined 61 cases, and found that of 172 distinct error conditions, 39 (23%) arose from physical or environmental causes, 65 (38%) from human error, 62 (36%) from latent design errors and 6 (3%)from consumer or third party failings [11]. Thus HRA is an essential part of any wider risk or reliability analysis.

In this paper we survey the current state of HRA, arguing that it is ill-suited for the needs of contemporary society and business. The growth of service industries with new business models implies a greater dependence of economies on reliable human interactions. For instance, recently human checks and balances failed to detect some dubious investment behaviour of a trader at *Société Générale* and led to a loss of some €4.9bn, large enough to have economic and financial effects beyond the bank. The current 'credit crunch' owes not a little to misjudgement and error in the banking and finance industries, indicating the growing interdependence of many disparate parts of the modern global economy and implying a yet wider requirement for HRA [12, 13].

We make our case that current practices in and uses of HRA are not fit for the complexities of modern society on a number of grounds.

- Historically HRA methodologies have treated the failings of humans in systems in a manner similar to those of physical components. So-called second and third generation methodologies have tried to recognise the difference between human failure and physical breakdown. However, their success is at best partial.

- The widely quoted Swiss Cheese Model (see below), which offers a qualitative understanding of system failure, fails to recognise many interactions and correlations between human errors and the operational environment. It emphasises a reductionist approach to HRA and may thus 'wrong-foot' the users of reliability analysis methodologies leading them to miss some of the key factors and mechanisms that should be built into their models; and, perhaps, put too much trust in the combined effect of several safety barriers.

- Human behaviour is complex and often counter intuitive. For instance, it seems sensible to use modern technological advances to make the physical components of a

[9] A list of abbreviations is given in Appendix A.

system safer. But there is some evidence that making subsystems safer could make the overall system less safe because of the propensity of humans to take less care personally when a system takes more care [10, 14]. High reliability organisational design in part uses teams to check and double check each others' activities, but again there are cases in which each falsely assumed that the others had checked something.

- The very concepts of human *error* and *reliability* are misleading. Human errors and faults are socially defined events: a perfectly reasonable action to one person may be an unreasonable failure to another [15]. Furthermore, however well judged a decision may be *a priori*, it may through 'ill fortune' lead to unwanted outcomes. Hence what may seem an error in hindsight may not be the outcome of irrational or erroneous choices. We should focus more on human *behaviour* as individuals, groups and organisations and recognise its potential involvement in system failure – without the pejorative judgement of whether that behaviour is aberrant in any sense.

- The roles of risk and reliability analysis in general and of HRA in particular, are often misunderstood by system designers, managers and regulators. In a sense they believe in the models too much and fail to recognise the potential for unmodelled behaviours – physical or human – to lead to overall system breakdown: *cf* [16]. There are two ways in which HRA may be used.

  - When HRA is incorporated into a *summative* analysis, its role is to help estimate the overall failure probabilities in order to support decisions on, e.g., adoption, licensing or maintenance. Overconfidence in the models leads to overconfidence in the estimated probabilities and possible poor appreciation of the overall risks.

  - There are also *formative* uses of HRA in which recognising and roughly ranking the potential for human error can help improve the design of the system itself and also the organisational structures and processes by which it is managed. Effective HRA not only complements sound technical risk analysis of the physical systems, but also helps organisations develop their safety culture and manage their overall risk. Indeed, arguably it is through this that HRA achieves its greatest effect.

- Existing HRA models have been developed (primarily in the nuclear industry) for use in assessing operator deviations from nominal operating modes. However, a more holistic view of the impact of humans requires a wider modelling framework which accounts for differing human roles in the system from the strategic to the operational. Design flaws introduce further risks to the system that arise from limitations inherent in human foresight. Thus it may be the case that no single HRA methodology is appropriate to all contexts and all parts of a full system risk analysis; and thus there may be a need to draw on several methodologies in analysing the system.

Few of our comments are new: several have their origins more than a quarter century ago. But to a large extent they point to issues that remain to be circumvented and ones that are often ill-appreciated outside professional risk and reliability communities. NUREG-1792: *Good Practices for Implementing Human Reliability Analysis* [17] has made several of the points made here, although we would claim that our critique is more fundamental in nature. We do not believe that adopting current best practice will address the most serious of issues that concern us.

We begin in Section 2 with a brief review of the development of HRA methodologies from the 1960s to the present day. In doing so we offer a discussion of the so-called *first, second*, and emerging *third generation* HRA methodologies, categorising these not so much chronologically as by their characteristics in terms of their underlying concepts, models and applicability. While we recognise that these methodologies have sought to incorporate more sophisticated understandings of human behaviour, we believe that they are still lacking, given the many developments that have occurred in, *inter alia*, behavioural, cognitive and organisational sciences. To support this belief we offer a review of such developments in Section 3. In Section 4 we discuss several case studies, reflecting on them in the light of the developments discussed in Section 3 and seeking to draw out lessons to shape the development of HRA. In Section 5 we discuss a number of contextual issues relating both to the tasks being under taken in the system under study and the purpose of the risk or reliability analysis itself. Our conclusion is that there are a plethora of issues to be taken account of in modelling and analysing human behaviour. Given this, we argue in Section 6 that we may need to draw on several HRA methodologies within a single overall system risk or reliability analysis, at each point choosing an appropriate HRA method for the specifics of the particular context. However, we also recognise that we may not have sufficient comparative data on current HRA methodologies to be able to delineate clearly the circumstances for which each is appropriate. Thus we outline a route-map for future R&D to build this portfolio methodology. We draw together the threads of our argument in Section 7.

Before beginning we should remark further on the language of error and failure. Perhaps for reasons of history and culture, it is common to talk of the failings and errors of others, to seek to attribute blame to some and exonerate others. But to do so may miss much. As noted above, we shall argue that it is more helpful to talk about behaviours without any negative attribution of failing or error. Moreover, we should talk of the behaviours of groups, teams and organisations as much as of human operators. System failure may arise from perfectly reasonable behaviours and interactions of any of these. Thus later in the paper we shall move away from the language of human error and failure. However, in the early sections in which we review previous and current perspectives, we use the somewhat pejorative language of error still found in much of the literature.

2 The development of quantitative HRA

2.1 Introduction

HRA has its roots in the early probabilistic risk assessments performed as part of the US nuclear energy development programme in the 1960s [4, 18]. Early first generation HRA methods were very similar to those in other areas of reliability analysis: namely, the probability of a human error is assessed via a simple fault (or event) tree analysis. The fault tree simply listed an initiating event, which might be an indication of a system error reaching the human operator or an intention on the part of the operator to perform an action, and then considered a series of subsequent events which could lead to the ultimate success or failure of the operator in achieving his/her goal. Essentially, the human operator is treated as a component in the system much as any other component such as a microchip or load-bearing structural element. Hollnagel [10] refers to this general approach as decomposition. Operators failing to respond to events were termed errors of *omission*, while unintended human actions were labelled errors of *commission*.

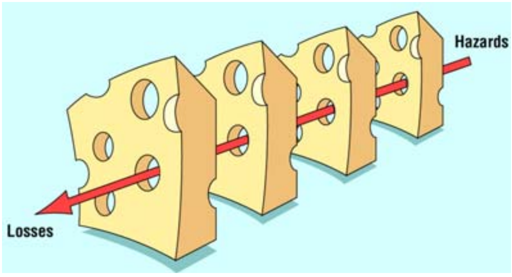


Figure 1 Reason's Swiss Cheese model [19]

However, such a simplistic dichotomy flies in the face of current qualitative understandings of human cognition, motivation and decision making, including the effects of stress, emotion, training, group interactions, organisational structures, cultures and so forth. Research in these fields has shown that there are systematic influences on decision making and behaviour that cannot be categorised as simply as omissions or commissions. Human failure is far more complex than the failure of, say, a steel support beam or a hard disk. We return to this issue in Section 3.

Reason [19] likened system failure involving human error to slices of Swiss cheese: see Figure 1. Essentially this suggests that systems do not fail because of a single failure, but because several elements fail near simultaneously, as if the holes in slices of Swiss cheese have aligned. In a way the Swiss Cheese model itself is too mechanistic: one is led to imagine a fixed number of slices, sliding backwards and forwards relative to each other until a series of holes align.

In safety studies one talks of the number of safety barriers or layers between normal operation and system failure; and, in a sense, the Swiss Cheese model picks up on this. Systems are designed with a set number of safety barriers and these barriers are intended to be independent. But human behaviour can correlate the risks of failure of two or more barriers. Their behaviour and propensity to failure varies in complex ways with, e.g., their tiredness, stress and general emotional state, which may well be influenced by external events leading to a common cause which may disrupt several safety barriers simultaneously. For instance, the Chernobyl Accident (Section 4.1) was in large measure caused by the imperative to conduct an engineering experiment within a fixed time, leading to stress in the operators and behaviour that effectively compromised some of the safety barriers together. Another potential unsafe behaviour is to discover an indication of a ‘hole’ in one layer and to defer further investigation, relying on the ‘cover’ offered by other layers: such behaviour lay at the heart of a system failure in the Sellafield case study (Section 4.1). On the positive side, humans have the ability to recover, to respond to the unexpected, to think ‘out of the box’, and so on, effectively repairing a compromised layer or even introducing a new one. In terms of the Swiss Cheese model, many of these failings correspond to varying the size of the holes, perhaps in a correlated fashion and maybe varying the number of layers over time. Wu and Pollard [11] suggest an extra layer representing consumers and third parties is needed to understand water industry failure. Reason himself discusses similar criticisms [20]; but the simpler mechanistic thinking implicit in Figure 1 still pervades thinking in much of reliability engineering.

During the 1990s many of these issues raised by Reason [19, 21] and others (e.g. Janis [22], Perrow [13], Roberts [23], Weick [24]) led to calls to revise HRA methodologies and adopt more sophisticated models and understandings of human error. Thus recent, *second generation* methodologies have attempted to develop a more sophisticated approach to human reliability, particularly the ability of humans to recover and prevent some or all the consequences of the threat of impending system failure. Note that the term second generation was coined by Doughty [25]. Emerging *third generation* methodologies go further modelling a range of behaviours in to recover and avert failure. However, we should note that one can easily be led into categorising HRA methods according to the chronology of their first development. It is perhaps better to characterise the different methodologies according to their characteristics, as we do below.

Summaries of several different HRA methodologies may be found Appendix B; see also [10, 26-28]. We recognise that our selection is somewhat UK/European centric: it is where our experience lies. However, we do not believe that this rather partial selection invalidates any of the general points that we make in this report. We close this section with a brief generic summary of the main stages of HRA, recognising that particular methods may omit some of these and include others.

Stages in an HRA.

1. Examine the system in detail, perhaps through a plant visit if it exists already or by a careful study of the plans and design.
2. Review information from any risk analyses of the physical system and, in particular, any fault trees.
3. Talk through and brainstorm where any human activity or inactivity may lead to a potential system failure.
4. Detail the tasks and activities in which the humans are involved and which may lead through omission, commission or deliberate intervention to system failure, building an HRA fault tree to explore the interactions that lead to failure. When focusing on operator error, stages 3 and 4 are referred to as a *task analysis*.
5. Assign *human error probabilities* (HEPs) to the fault tree. These are essentially the probabilities that the human activity cause specific branches in the HRA fault tree to be taken. They may be derived from tables collated from observations or from expert judgement.
6. Modify the HEPs by applying *performance shaping factors* (PSFs) which seek to allow for the moderating effects of stress, tiredness or similar.
7. Assess and allow for any probabilistic dependence and correlations between the human activities. This is almost inevitably a matter for expert judgement.
8. Use the HRA fault trees, HEPs, PSFs and assessed correlations to calculate the overall system failure probabilities.
9. Consider potential recovery actions and repeat stages 3 – 8 above on these, using the result to modify the overall system failure probabilities.
10. Conduct sensitivity analyses on all the calculations and use these to inform the systems designers and risk managers of the potential for failure arising from human activities.

2.2 Characteristics of HRA methodologies

Some ‘first’ generation methods had a quite sophisticated approach to human reliability, perhaps more so than some ‘second’ generation ones. Here we review a range of methodologies according to their characteristics. As Boring [29] has noted, chronology is not a good guide to whether a method is considered first, second or third generation. Context and cognition are the two features that second generation methods are supposed to contain – yet context is certainly modelled to some extent in first generation methods. Third generation methods contain more dynamic simulation, and have to be implemented on a computer.

The classification into different generations of methods is a sign of a need to classify and categorize the whole family of approaches that have been developed. Unfortunately there is little clarity about why different models have been developed, and why the specific modelling choices have been made, Furthermore, we really need to categorize not only the type of situation being modelled, but the purpose of the model, the end user of the model, the resources required etc. We are some way from having such a framework, although we return to this topic later in the paper.

In order to make a start towards such a holistic classification, we begin by describing some of the features of important methods. The table below considers a number of features:

- *Task analysis*: is the method based on a preliminary task analysis?
- *Dependencies*: are we able to model statistical or other kinds of dependencies between different events?
- *Performance shaping factors* (PSF): these are typically used to describe aspects of the environment in which the human is acting. We describe if and how such PSFs are included in the model.
- *Decomposition*: this describes how the method breaks down a collection of actions comprising a task.
- *Time effects*: these describe how time can be incorporated into the model, for example when there is a time given to complete a task.
- *Error classification*: this indicates whether there is a classification scheme used that would direct the user to apply the method differently depending on the class of error identified.
- *Expert opinion required*: this indicates in what ways the analyst is required to be a source of data or make judgements, or (where only limited judgement is required), how much the method guides the analyst.
- *Calibration*: this covers the extent to which the model outputs have been or are calibrated to ensure numerical accuracy.
- *Uncertainty in outputs*: does the output give point values only or does it also include an assessment of a range of credible values?

- *Sensitivity analysis*: is there a clearly developed sensitivity analysis procedure for the model?
- *Context modelling*: in what ways is the context of the human activity modelled within the method?
- *Operator control modes*: does the model allow for a range of different types of operator behaviour in different situations?

Method	Task Analysis	Dependencies	Performance Shaping Functions (PSF)	Decomposition	Time Effects	Error classification	Expert opinion required for application
THERP	Assumed	Special dependency method used to couple probabilities in event sequence, but this is numerical adjustment rather than true dependency model	Yes, used to modify nominal HEPS	ET used to present results of task analysis	Yes, for abnormal events	EOC/EOO	Limited judgement for class selection
SLIM	Assumed	Not explicitly modelled. Implicit modelling through PSFs	Yes, used to score impact on situation	Not in method	No	No	Highly driven by EJ
HEART	Assumed	Not explicitly modelled. Implicit modelling through PSFs	Yes, but called EPCs which are used to modify generic HEPS	ET suggested as method by which calculations for different HEART tasks are linked together	No, but shortage of time is an EPC	Generic tasks detail level of complexity of task, time available, supervision etc	Limited judgement required
HCR	Assumed	No explicitly modelled. Implicit modelling through PSFs	Yes, used to influence median time required for task	Models time to task completion	Yes	SRK - used to determine median response time	Limited judgement required
ATHEANA	Assumed		Yes, but called EFCS				
CREAM	Assumed	Scoring includes dependency through adjustment of CPC scores for dependent CPCs in the basic method. Probabilistic dependency implicit through CPCs	Yes, but called CPCS	Not explicitly in method, but assumes that appropriate logic model will be used	No. but control modes are related to time available, and *available time* is a CPC	Generic failure types classified through cognitive function (observation, interpretation, planning and execution)	Assessment of CPCs and generic failure types.

Table 1 Characteristics of Different HRA methodologies
(Abbreviations are listed in Appendix A)

Method	calibration	Uncertainty in outputs	Sensitivity analysis	Context modelling	Operator control modes
THERP					
SLIM	Calibrated by two "known" HEPs	No	No	Context rating of task on PSF scale and PSF weight	No
HEART	Error bands calibrated to published studies	Uncertainty bands included, with method for propagating	No	Context through task analysis and EPCs	No
HCR	Calibrated to simulator tests ?				
ATHEANA					
CREAM	Control modes in basic model, and basic failure probabilities in extended model judged to be similar to other HRA methods	Uncertainty ranges for outputs of basic and extended models	No	Context through task analysis and CPCs	Defines different control modes, but these are model outputs rather than inputs

Table 2 Characteristics of Different HRA methodologies (continued)
(Abbreviations are listed in Appendix A)

3 A review of relevant developments in behavioural, cognitive, management and organisational sciences

3.1 Relevant developments in behavioural and cognitive science

3.1.1 Introduction

Approaches to modelling errors and biases in human and organisational decision making tend to categorized human error into three groups.

1. Systematic errors related to individuals, e.g. biases in information processing, information overload, fatigue.
2. Systematic errors related to organisational systems or structures, e.g. ineffective communication structures within and across organisational hierarchies.
3. Random errors of people or organisational systems which are difficult to predict or quantify.

As indicated above, we are concerned that such categorisations use somewhat pejorative terminology. More important, perhaps, is a concern that the categorisation misses some important issues that influence human reliability. In the next section we discuss an alternative approach to the conceptualization and modelling of human behaviour in organisations.

3.1.2 Modelling error

We consider decision making and behaviour in a broad sense rather than focusing in on error. Understanding errors requires a more holistic approach to understanding decision making behaviour for three reasons.

First, the error focus on HRA models may be too narrow [26, 30]. Error behaviours are a subset of individual behaviours. To consider error alone might be to commit a type II error on our part. Errors are just one of a range of behavioural products of a number of individual and organisational precursors. Errors are not a class of behaviours that are entirely distinct from other behaviours and should not be considered in isolation. In the organisational context, it is often an external system or judgement that categorises a behaviour as an error rather than the behaviour itself being inherently and indisputably wrong.

Second, models of HRA that include people factors typically focus on cognitive aspects of decision making, such as the ATHEANA model. Recent developments in the modelling of decision making emphasise the dual influences of cognition and emotion on decision outcomes [31]; Loewenstein, Weber, Hsee and Welch's model of decision making is one such example [32]. The addition of emotions to cognitive models of decision making is highly relevant to safety critical industries since the consequences of accidents are frightening as well as costly. Furthermore, work roles and the interpersonal nature of work have emotional antecedents and consequences.

Third, the use of high reliability systems designed and engineering to minimize errors and hazards has both benefits and disadvantages. It is of course important that systems are designed to be as safe as possible. However, the reliance on such systems can cause biases and flaws in decision making. A high profile example is the leak in the THORP plant at Sellafield (Thermal Oxide Reprocessing Plant) that was discovered in 2005: see Section 4.1 below. This relatively modern plant had been designed to a high standard of safety. Information that indicated a system problem was available for some months and yet went unnoticed. Despite previous incidents in 1998 and earlier in 2005, the information that should have suggested a leak, or at least a problem requiring investigation, was misinterpreted. The prevailing attitude was one of an error-free system and information that could suggest the contrary was ignored or dismissed. This type of decision making behaviour has well researched in other contexts for some years: see, e.g. Janis & Mann, [33]. An effective system, therefore, should be both safe and yet not perceived as error proof.

3.1.3 A holistic approach to decision making

In view of the above reasons why it is relevant to take a more holistic view of decision making, we propose an alternative to the current HRA approach to error modelling that draws upon the concept of self-regulation. A self-regulatory approach to decision making behaviour has its roots in three related areas of psychology: risk, individual self-regulation, and models that incorporate the concept of optimal levels of functioning.

One approach to modelling risk is the *risk thermostat* model [10, 14]. Adams [34] proposed that there is a dynamic interaction between actors' perceptions and behaviours, and their environment. The Adams model proposes that the interaction between risk propensity, perceptions of societal risk, perceived danger, and positive and negative outcome expectancies influence risk behaviour such that people will adjust their behaviour to be more or less risky, as appropriate for their preferences and their situation. For example, Adams claimed that improvements in road safety and car design have led to greater risk taking in driving.

Individual self-regulation is defined as the internal and behavioural adjustments that function to maintain factors such as cognitions, emotions and performance within acceptable limits [35]. This approach to modelling behaviour proposed that behaviour is goal orientated and there are internal, hierarchical processes that enable people to put thoughts into actions [36] through activation and inhibition of decision making processes. Some of the decision processes take place at a subconscious level and never reach conscious deliberation, a process called automaticity [37]. Thus, like safety critical organisations, there is a dynamic interaction between people and their environment that is designed for effective behaviour.

Models of decision making and behaviour that incorporate optimal levels of functioning have a long history and a range of organisational applications. For example, Yerkes and Dodson published their inverted U model of the association between performance and arousal in the early 20th century. More recent models of work performance show similar patterns: some effort and pressure can be effective, too much of both leads to burnout [38].

The literature on decision making heuristics and biases is also relevant [31, 39, 40]. Numerous studies have demonstrated the existence of systematic and robust cognitive biases, and are well summarized by Bazerman [41]. For example, emotionally-laden or otherwise individual salient information is recalled easily and likely to be considered as significant to a decision when more objective evidences shows that other types of information are more important to a decision. The case of the Sellafield operators and managers dismissing information about a leak in the plant is a similar kind of bias. However, the processes that drive biases have arisen for a reason – we cannot take into account all the information that surrounds us and so we need to select information to attend to in order for any action to be taken. The work of Gigerenzer and colleagues has shown that some heuristics can improve decision making. For example, making fast decisions based on almost no information can yield better results that having some prior information which can bias decision making [42].

Finally, the organisational context must be considered both as an influence on individual level decision making and as an integral outcome of individual and group decision making processes. Choices are made at all levels of organisational design that are subject to the same processes of automaticity, flawed biases and self-regulation as individual decision making.

3.1.4 Conclusion

To conclude, a model of human reliability at the individual level and high reliability at the organisational level needs to integrate decision making and the dynamic interaction of actors and their environment. Using the models discussed above points to consider for future developments of HRA models. The models also lead to implications for improving individual and organisational safety and reliability, including the following.

1. Understand the emotional and cognitive influences on decision making without having a blinkered approach to error management.
2. Take into account the interaction between individuals and the environment.
3. Debias decision making: be familiar with safety procedures not complacent about hazards.
4. Create an environment where people understand their role in the overall process and there is open discussion of a wide range of organisational issues.

3.2 Studies of high reliability organisations

3.2.1 Introduction

The past 20 years has seen several studies of *high reliability organisations* (HROs), which Roberts [23] defined as organisations failing with catastrophic consequences less than one time in 10,000. These studies recognise that certain kinds of social organisation are capable of making even inherently vulnerable technologies reliable enough for a highly demanding society. An HRO encourages a culture and operating style which emphasises the need for reliability rather than efficiency [24].

As organisations, HROs emphasise a culture of learning, although they clearly do not rely in any sense of learning from mistakes! Instead, HROs resort to learning from imagination, vicarious experience, stories, simulations and other symbolic representations [24]. They also emphasise a culture of sharing their learning and knowledge, their mental models: 'heedful inter-relating' [43], 'collective mindfulness' [44], 'extraordinarily dense' patterns of cooperative behaviour [45] and 'shared situation awareness' [46].

Weick and Roberts [43] argue that in HROs develop aggregate mental processes exhibit qualities as noticing, taking care, attending and concentrating. Moreover, instead of a rigid division of labour, there are true teams in which all members share information and above all accept a joint responsibility for safety and reliability.

Usually HROs apply a strategy of redundancy [47] with teams of operators 'watching each others backs'. As noted, teams share common mental models of both their internal organisational processes and the external world. Redundancy may increase complexity of operations as it makes the operations system less easily understood or opaque [13, 48]. However, redundancy also increases the probability or chance of getting adequate information to solve probable dangers, consequently reducing the risks arising from complexity rather then increasing them.

When necessary, HROs try to decentralize the authority of senior teams or management responsible for decision making. La Porte and Consolini [49] describe patterns of authority in air traffic control changing between a normal, bureaucratic mode, a different mode in high tempo operations in which hierarchical rank defers to technical expertise, and a pre-programmed emergency response mode. Rijpma [50] suggests that HROs use decentralisation to enable those working closest to any problems to address and solve them as they emerge or become apparent. Using this method rapid problem solving is achieved, resulting in an increase in reliability and reduction of the risk of accidents occurring in highly critical situations. This decentralisation may increase the complexity of the organisation as knowledge and lines of authority need to be distributed, but La Porte [45] suggests the balance of these opposing effects can lie in the direction of higher reliability. Along with team-based organisational structures with devolved authority, HROs tend to have the capacity to change rapidly when circumstances demand [51].

3.2.2 How do HROs relate to HRA

There are some fundamental differences between the fields of HROs and HRA.

- HRO theory is essentially descriptive, whereas HRA is normative, suggesting techniques to deal with variability and error in the human performance.
- HRO theory looks at the organisations, teams, collective qualities like collective mindfulness [44], whereas HRA tend to focus on tasks undertaken by individuals.
- In many respects, HRO theory emerged as a critique of normal accident theory [13]; HRA has largely emerged as a way of filling the human gap in technical risk assessment.
- In perspective HRO is fundamentally optimistic in believing that social organisations can produce high reliability in the most demanding of circumstances, while HRA is fundamentally pessimistic because it concentrates on the human capacity to make errors.

The two fields share a focus that is both human and social, suggest that any thinking about how HRA should develop might benefit from asking what is relevant about HROs. The particular 'reliability' to which their titles allude is remarkably similar. In both cases it has a strong connection not merely to things working as they should but also to notions of safety and hazard.

A common tone in writings about high reliability organisation is that high reliability originates in adversity. Thus, for example, Weick [24] points out that reliability tends to increase when performance pressures are high, not low. Early ideas about HROs rather contradicted this notion of strength in adversity [49]; subsequent work, however, emphasises that HROs accept and deal with fluctuations, striving for resilience rather than invariance [52]. This quality of achieving particularly high reliability because of, not despite, the high stakes and demanding circumstances now tends to permeate HRO studies.

3.2.3 The critique of HROs and reliability generally

There are several challenges that have been mounted to the HRO line of work. First, some suggest that HRO perspectives are heavily functionalist and neglect politics and group interests [48, 53, 54]. A second criticism relates to the empirical studies underpinning HRO theory [48, 54, 55]. Critics argue that the context of some of the most important HRO studies, e.g. on the flight decks of aircraft carriers, is misleading, only evidence of safety in simulated rather than actual operations. Others argue that the mechanisms and qualities that are said to underlie the achievement of high reliability are neither particularly characteristic of HROs nor unequivocally good for reliability. Finally some challenge that what reliability means is somehow obvious, unitary and absolute. The argument always seems to be about such questions as whether organisations 'really' are reliable, whether this reliability can be called 'high', and what 'actually' produced the reliability, rather than what might be interesting and problematic about reliability as a concept.

4 Case studies

4.1 Nuclear

The nuclear industry offers some of the best recorded incidents in which human error either led to a system failure or had the potential to do so. For instance, we note briefly some events that indicate significant human potential for errors of commission or omission.

- *Three Mile Island*. The accident happened on 28 March 1979, at the Three Mile Island nuclear power plant near Harrisburg, Pennsylvania [56]. There was no significant release of radiation, but a full core melt was only just averted. The causes of the accident continue to be debated to this day, but one thing is clear. The initiating event, the formation of a hydrogen bubble which forced down cooling water exposing the core, had not been anticipated in the reactor's design or safety studies. The operators not only did recognise what was happening, but also had never anticipated that it might. It was an incident beyond their experience and imagination: in a very real sense outside of scientific and engineering knowledge as it stood then. A key learning point in relation to HRA is that the operators behaved entirely sensibly and in accordance with their mental models of what they believed was happening. There was no error in their behaviour in this respect, not at least in the sense of human error of HRA theory.

- *Chernobyl*. The Chernobyl accident occurred on 26 April 1986 and did involve an enormous release of radioactivity [57]. The accident itself, contrary to the TMI accident, was caused by a deliberate act, an experiment that caused an explosion and fire. The experiment involved running the plant outside its design parameters at very low power. The personnel of the Chernobyl station were familiar with the experiment because they carried it out previously at the Chernobyl-3 reactor and the Kursk station in Russia [58]. However, the personnel responsible for the experiment had been working for some 15-20 hours when the experiment started because of delays in handing the reactor over for the experiment. They were tired and under pressure to complete the experiment quickly and return the reactor to normal energy production. During the experiment, they deliberately turned off three separate safety systems and switched to manual control. This was against safety instructions, but probably they had done it frequently before. In this case, human failure on on-line operations was caused by the effect that seemingly "freak infringement of rules" which did not cause an accident in the past lead to more violation of rules in the future [59]. The accident occurred without any component failure. However, the design of the reactor depended on the operators following certain safety instructions. As seen, the operating and regulatory regime in place was inadequate. Attempts by operators to recover the situation triggered flash boiling of water, which in turn led to a breach of fuel can or containment, and the exposure of the hot fuel element to water. Within seconds, a major chemical explosion occurred which destroyed the reactor and caused the worst nuclear catastrophe in history. While there is no doubt that deliberate acts were key in causing the accident, there is equally no doubt that the tiredness, stress and poor safety culture within the operating team were contributing factors.

- *Doonreay Shaft*. A 65m deep shaft at the Doonreay nuclear power plant, originally dug to remove rock from a pipe discharging treated water was used from 1958 to 1977 as a low level waste pit. In 1977 there was a major explosion in the shaft caused by the reaction of sodium potassium (NaK) alloy with water. Some radioactive waste was spread over a large area. One can wonder about many things to do with the management of this pit, but one thing is clearly incredible: the deposition of NaK, albeit encased in cast iron, into the pit. The scientists and engineers involved would surely have known that: (i) the shaft was wet – it linked to the discharge pipe and went below sea-level; (ii) NaK reacts explosively with water; and (iii) cast iron corrodes. Put these facts together and the 'accident' was completely predictable.
- *Sellafield pigeons*. In 1998 it was discovered that pigeons were transferring radioactivity from the Sellafield nuclear site to the surrounding region[10]. Although there had been many risk analyses at the plant over the years, this potential route for contamination to be taken off-site had not been anticipated.

A recent leak at Sellafield is particularly relevant because many human and organisational behaviours interacted and led to the incident. We therefore describe it in greater detail. On April 20th 2005 a leak was detected in the Sellafield Thermal Oxide Reprocessing Plant (THORP) after a video camera revealed that approximately 83,000 litres of radioactive waste, or dissolver liquid, had leaked into the base of the cell. Closer inspection revealed that a feed pipe to accountancy vessel V2217B had fractured. This is believed to have been due to fatigue stresses induced by excessive movement of the

[10] http://news.bbc.co.uk/1/hi/uk/55612.stm

vessel to which the pipe is connected. It is estimated that the pipe suffered major failure around the 15th of January, 2005, but may have started to leak as early as July 2004.

Subsequent investigations indicated the following errors and behaviours all contributed to the leak itself and a failure to detect it sooner.

- The vessel and pipework design was changed during the construction of the plant. The original design may well not have fractured in the same way. However, during the preparation of the safety case for the plant, assumptions and precautions in relation to seismic activity were changed and the designed changed to allow for this. So a design fault that had originally been engineered out was re-introduced in another phase of the design process.
- There were earlier indications that there might be a leak from accounting of the input and output to and from the cell. However, these were largely ignored because of a 'new plant' culture: see earlier remarks in Section 3.1. There was a belief that such a modern plant could not suffer from leaks or other failures. Some of the written operating instructions were ambiguous, leaving too much to the interpretation of the operating staff. In the context of the 'new plant' culture and other management imperatives, it was too easy to ignore inconclusive but pertinent readings and observations. It is also noteworthy that this 'new plant' culture was implicated in two previous smaller incidents elsewhere in THORP in 1998 and earlier in 2005.
- Even when a decision was taken to investigate the leak, misunderstandings among senior management led to operations being continued to meet production targets longer than they should have been.

The Board of Inquiry report [60] on the incident repeatedly makes the point that the 'new plant' culture was at the heart of many of the failings of the operators, management processes and organisation.

4.2 Railway

4.2.1 Lambrigg Derailment

On 23rd February 2007, a Virgin train travelling from London to Glasgow derailed between Preston and Carlisle, at Lambrigg Ground Frame crossover located near Cumbria [61]. Of the 108 passengers and 4 crew members travelling aboard, one fatality resulted with a further 22 individuals requiring hospital treatment. The immediate cause of the derailment was identified as faulty points on the track; this was in turn the result of a fault in the stretcher bar of the points which consequently led to the left and right switch rails to become disconnected. Two securing bolts were also detached from the stretcher, one of which was lying next to the points while the other was missing entirely. The nuts which secure the stretchers together, having being tightened with the incorrect equipment, became free due to dynamic loadings; this fault failed to be identified in a subsequent inspection due to unauthorised splitting of patrol groups. As such faults could have potentially been identified and corrected through remedial action, prior to derailment, the incident could have thus been prevented as it was highlighted that the deterioration of the points had occurred some time before.

The causes highlighted above, which were the most apparent causes of the derailment, were further identified in the post-incident investigation, as having been caused by a number of underlying contributory factors based on human reliability and error.

1. There existed a number of deficiencies in the inspection and maintenance regime ultimately causing the points to fall into disrepair and the fault thus being unidentified. Such deficiencies included:

   - A breakdown in the local management structure responsible for inspection and maintenance. Inspections carried out were found to be non compliant with set standards and procedures and supervisors tended to reinforce this behaviour as acceptable by employing unsafe inspection arrangements.

   - Track patrolling regime’s systematically failed to inspect the area adequately. Routine inspections routinely cover a required mileage of track; due to management incompetence the area of the track containing the fault was overlooked one week prior to the derailment; the subsequent patrol report was nevertheless authorised, with a gap in the inspection going unnoticed.

2. Mandated standards were not communicated or executed in the required manner, with a lack of sample checking of the track to test inspection quality and arrangements. Between maintenance and track supervisory management there was evidence of a split ‘them and us’ culture which had consequential effects on the way in which operations were conducted.

3. Patrolling of the track was poorly managed; patrollers were allocated to random patrol lengths thus compromising understanding of certain areas of the track and many of the patrollers’ certification of competence had lapsed with lack of evidence to suggest any assessment of monitoring; despite this lapse being highlighted to local management, it was ignored and this behaviour thus became acceptable. There was no review of patrols and there was no definite method by which defects on site were marked, with checklists to identify these being used inconsistently.

4. The quality assurance regime did not recognize failures in the reliability of inspection regimes or in the application of best practice. Personnel were not briefed about any new standards requiring compliance and staffs’ competency in following practice was not managed. Failure to follow rules and standards went unreported and not acted upon; such unacceptable behaviour was encouraged as it was enacted too by higher level supervisors.

While the accident had a clear immediate physical cause in the faulty points, it is clear the real cause of the accident was human, managerial and organisational. No single individual failed. Rather many human factors contributed to the accident ranging from the managerial to the cultural.

4.2.2 Saxmundham Collision (*User worked crossing collision*)

On 22nd May 2006, a freight train collided with a car trying to cross at a User Worked Crossing (UWC) near Saxmundham, England. No one was injured as a result of the collision and the train was not derailed, however both vehicles did suffer minor damages at the area of impact. The concerned UWC is situated on a private road which from the north to south side, leads to private dwellings. Permitted users of the UWC include residents of the dwellings, farmers of the surrounding farmlands and users with authorised access from these residents including such parties as delivery vehicles.

The immediate cause of the incident was reported to be the fault of the driver of the vehicle, who failed to stop at the check point to observe for oncoming trains. In the incident report it was further disclosed that the authorised person who permitted the motorist to use the UWC did not give the motorist a briefing on how to use the crossing correctly i.e. in a safe manner.

The findings of the inquiry highlighted additional causal factors leading to the occurrence of the incident, these were:

- The gates on both sides of the crossing were found to have been left open for a lengthy period of time despite requiring to be closed when the crossing is not in use. The gates were unable to be closed due to the overgrown vegetation that had developed around them and thus rendered the gates usable; this was due to inadequate maintenance of the gates.
- The motorist involved in the collision had used the crossing for 36 years and 6 times in the week leading up to the incident; in this time the driver had never come across a train on this section of the track. Due to this past experience he, along with other authorised users, became accustomed to leaving the gates open and were thus of the expectation that a train would not come when the track was being crossed.

Further factors concerning the crossing, which were not immediately apparent, were also thought to contribute to the incident:

- the short warning time to alert a motorist of an oncoming train;
- the signs which warn a user to stop and read how to cross safely had poor visibility due to shrouding by foliage and vegetation and, moreover, there were problems with their wording;
- no telephone number was provided at the crossing for contacting a railway employee in the event of an accident.

In the subsequent enquiry it was also disclosed that letters had been sent to the authorised users of this gate several times, reminding them of the rules regarding gate closure. When questioned, they could not explain why the gates were left open. It is therefore probable that due to the low level of traffic the users became complacent about the safe use of the crossing. In summary, there was no physical breakdown that caused the accident, but a number of unwanted behaviours arose through poor information flows, an unchecked growth in poor practice and complacency.

4.3 Water

The provision of safe drinking water that has the trust of customers is the overarching objective of the water utility sector. The sourcing of raw water, its treatment, distribution and use by customers involves a complex array of processes, assets and procedures, all of which contribute to preventative risk management and to public health protection. A water distribution system for example, is an interconnected collection of sources, pipes, and hydraulic control elements (e.g., pumps, valves, regulators, and tanks) aimed at delivering water to consumers in prescribed quantities and at desired pressures. A typical water supply system is composed of water sources, raw water transmission pipes, water treatment plants, and water distribution networks. These subsystems expose a wide variety of risks for both natural and human-related influences since most of them are spatially diverse and accessible. Critically for water supply systems, in event of failures, customers will usually have been exposed to pathogens and/or chemicals for some time before the effects (usually waterborne disease) become evident. There is no opportunity for 'product recall' and often, many thousands of customers may have been exposed. Thus securing a culture of preventative risk management is critical to the provision of safe drinking water [62-65].

The experience with water systems has been that when failures occur, they may initially appear to be asset-centric (pipe bursts, filter breakthroughs etc.), but are frequently found also to have deep-seated causes, including human error. Consider six cases selected from Hrudey and Hrudey [66] (Table 3), representing fatal drinking water outbreaks in affluent countries over the past 20 years where human error was implicated. From the table, the most frequently occurred HRA factors contributing to the six outbreaks are *risk not recognised*, *poor system design and installation*, and *poor maintenance*. Each of them appears 4 times out of the 6 cases. Poor design, installation and maintenance are more or less associated with organisational influences in Reason's Swiss Cheese model. This suggests latent errors are major causes in the outbreaks. Water quality incidents are often triggered by major change – e.g., by extreme weather, livestock or wildlife faecal contamination – that presents a pathogenic challenge to the system under conditions at the edge or beyond its design parameters.

Given the significant consequences that can arise from water quality incidents, research has begun on the development and embedding of a risk management culture within the sector. But how do organisations to develop a risk management culture without having first to suffer a major accident? Weick and Sutcliffe [67] apply the concept of 'mindfulness' as one strategy to reduce the likelihood of accidents. Mindfulness is associated with a number of organisational characteristics: (i) preoccupation with failure and the root causes of it; (ii) reluctance to (over)simplify; (iii) sensitivity to operations; (iv) commitment to resilience; and (v) deference to expertise. For water utilities seeking to develop mindfulness [64, 68]:

- informed vigilance is actively promoted and rewarded;
- there exists an understanding of the entire system, its challenges and limitations is promoted and actively maintained;
- effective, real-time treatment process control, based on understanding critical capabilities and limitations of the technology, is the basic operating approach;
- fail-safe multi-barriers are actively identified and maintained at a level appropriate to the challenges facing the system;

Table 3. Summary of fatal drinking water outbreaks in affluent countries over the past 20 years (Adapted from [65])

Location & Time	Health Consequences	HRA Comments
Drumheller Alberta Canada, 1983 Feb	1326 confirmed cases of gastroenteritis, **2 deaths**	• vulnerable situation of sewage pump station upstream not recognized • failure of internal reporting of sewage spill to water operations • operating winter treatment without coagulation made system vulnerable
Cabool Missouri USA, 1989–1990 Dec–Jan	243 confirmed, 32 hospital admissions, **4 deaths**	• risks associated with water main break repair during extreme weather not recognized • poor sewerage systems maintenance exposing water distribution to risk • no treatment barrier in place
Milwaukee Wisconsin USA, 1993 Mar–Apr	285 confirmed cases, ∼ 4400 hospital admissions **50 deaths**	• risks associated with sewage contamination of water intake not recognized • apparently not aware of Cryptosporidium risk • failure to maintain optimum filtration performance • failure to recognize signal from consumer complaints
Gideon Missouri USA,1993 Nov–Dec	31 cases confirmed, 15 hospital admissions, **7 deaths**	• poor maintenance of water storage allowed faecal contamination • water quality management not based on good knowledge of system • no treatment barrier in place
Washington County Fair New York USA, 1999 Sept	161 confirmed cases, 71 hospital admissions, **2 deaths**	• not aware of risk to well from septic seepage field • allowed use of unchlorinated water from a shallow well • failure to consider that extreme drought of previous summer might affect water supply safety
Walkerton Ontario Canada, 2000 May	163 cases of *E. coli* confirmed, 65 hospital admissions, 27 cases of HUS, **7 deaths**	• ignored warnings about vulnerability of shallow well when first installed in 1978 • failed to adopt source protection recommendations at installation • regulator failed to implement policy requiring continuous chlorine residual monitors on vulnerable shallow wells • operators inadequately trained with no knowledge that contaminated water could kill consumers • failure to recognize that extreme weather and flooding could cause water contamination • failure to maintain chlorine residuals • failure to monitor chlorine residuals as required

- close calls are documented and used to train staff about how the system responded under stress and to identify what measures are needed to make such close calls less likely in future;

- operators, supervisors, lab personnel and management all understand that they are entrusted with protecting the public's health and are committed to honouring that responsibility above all else;
- operational personnel are afforded the status, training and remuneration commensurate with their responsibilities as guardians of the public's health;
- response capability and communication are improved, particularly as post 9-11 bioterrorism concerns are being addressed; and
- an overall continuous improvement, total quality management (TQM) mentality pervades the organisation.

It is clear that as with the incidents occurring in the nuclear and rail transport industries, the water industry is subject to arrange of threats arising from poor organisational culture and management practices.

5 The contexts of risk and reliability analyses

5.1 Introduction

There are many issues relating to the context in which human behaviour may contribute to a failure: e.g.

- the team and local management structures which set the local context in which the operators work;
- the organisational context – including strategic and economic imperatives – in which the teams and local management structures are embedded;
- the cultural context and – including misplaced trust in other safety barriers in the system – in which the operators find themselves;
- external influences on the operators, e.g. stresses from home life, tiredness;
- the lack of recent incidents leading to a growth of complacency.

All have been illustrated by the case studies discussed in the previous section. There are at least one further contextual issue that we should consider and on which we have been silent: the decision making activity in which the operators are engaged when the 'failure' occurs. It may seem a superfluous remark to make, but not all contexts in which human reliability is important are the same. Some concern operators performing standard tasks at the right time and place. Others require responses to novel, potentially catastrophic circumstances. Decision processes and reactions will vary accordingly. This means that the appropriate HRA methodology to assess the risks associated with the operator's behaviour may vary with the details of that context.

Hollnagel [26] also discusses the contexts of responses to incidents in his development of CREAM. However, his development and categorisation of contexts depends *both* on the underlying decision making activity *and* the various factors noted above which may affect human behaviour: see Section 5.4. We believe that there is some value in separating the contextual issues that relate to the decision to be made from those that influence the behaviour of the operator(s) engaged in that decision making. In Section

5.2 we introduce the Cynefin model of decision contexts and suggest how this may inform our discussion.

5.2 Cynefin

Cynefin is a conceptual framework developed by Snowden which, among other things, offers a categorisation of decision contexts [69]. The Cynefin model roughly divides decision contexts into four spaces: see Figure 2. In the *known space*, or the Realm of Scientific Knowledge, the relationships between cause and effect are well understood. All systems and behaviours can be fully modelled. The consequences of any course of action can be predicted with near certainty. In such contexts, decision making tends to take the form of recognising patterns and responding to them with well rehearsed actions. Klein [70] discusses such situations as recognition primed decision making; Snowden describes decision making in these cases as CATEGORISE AND RESPOND.

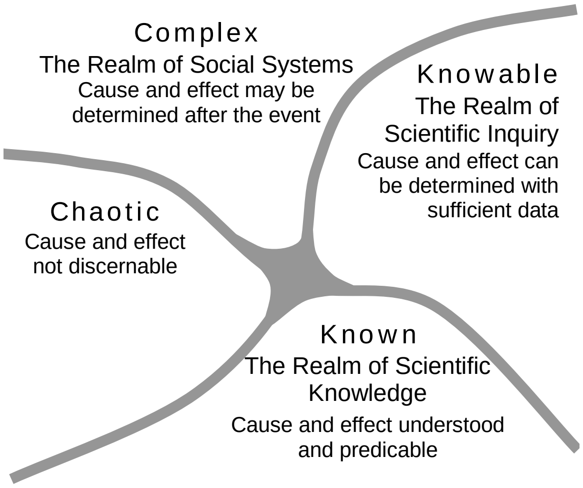

Complex
The Realm of Social Systems
Cause and effect may be determined after the event
Knowable
The Realm of Scientific Inquiry
Cause and effect can be determined with sufficient data
Chaotic
Cause and effect not discernable
Known
The Realm of Scientific Knowledge
Cause and effect understood and predicable

Figure 2: Cynefin

In the *knowable space*, the Realm of Scientific Inquiry, cause and effect relationships are generally understood, but for any specific decision there is a need to gather and analyse further data before the consequences of any course of action can be predicted with any certainty. Decision analysis and support will include the fitting and use of models to forecast the potential outcomes of actions with appropriate levels of uncertainty. This is the realm in which the standard methods of decision analysis as found in, say, Clemen and Reilly [71] apply. Snowden characterises decision making in this space as SENSE AND RESPOND.

In the *complex space*, often called the Realm of Social Systems though such complexity can arise in environmental, biological and other contexts, decision making situations involve many interacting causes and effects. Knowledge is at best qualitative: there are simply too many potential interactions to disentangle particular causes and effects. There are no precise quantitative models to predict system behaviours such as in the known and knowable spaces. Decision analysis is still possible, but its style will be broader, with less emphasis on details. Decision support will be more focused on exploring judgement and issues, and on developing broad strategies that are flexible enough to accommodate changes as the situation evolves. Analysis may begin and, perhaps, end with much more informal qualitative models, sometimes known under the general heading of soft modelling, soft OR or problem structuring methods [72-76]. If quantitative models are used, then they are simple, perhaps linear multi-attribute value models [77]. Snowden suggests that in these circumstances decision making will be more of the form: PROBE, SENSE, AND RESPOND.

Finally, in the chaotic space, situations involve events and behaviours beyond our current experience and there are no obvious candidates for cause and effect. Decision making cannot be based upon analysis because there are no concepts of how separate entities and predict their interactions. Decision makers will need to take probing actions and see what happens, until they can make some sort of sense of the situation, gradually drawing the context back into one of the other spaces. Snowden suggests that such decision making can be characterised as ACT, SENSE AND RESPOND. More prosaically, we might say 'trial and error' or even 'poke it and see what happens!'

The boundaries between the four spaces should not be taken as hard. The interpretation is much softer with recognition that there are no clear cut boundaries and, say, some contexts in the knowable space may well have a minority of characteristics more appropriate to the complex space.

The Cynefin framework provides a structure in which to articulate some concerns about the use if HRA in risk and reliability analysis and in relation to HRO studies.

- First generation HRA methodologies and arguably most of second and third generation ones focus on tasks that lie in the known or knowable spaces. Yet many of the perceived risks in modern systems arise because of their inherent complexity: *cf.* the normal accident theory of Perrow [13, 54]. In other words, we need be concerned with human behaviour as managers and operators strive to deal with events happening in the complex or even chaotic spaces. The Chernobyl Accident was initially managed as if it were in the known and knowable spaces, yet it was one of the most complex socio-technical accidents that have occurred [16, 78]. In the Three Mile Island Accident initially there was no conceptual understanding of the processes by which a hydrogen bubble might form and hence decision making in the first hours and days of handling the incident took place in the chaotic space.

- It is informative to read HRO studies from the perspective of Cynefin. For instance, Weick's classic paper [24] moves from discussions of how air traffic controllers manage flights in a highly reliable way – a repetitive task in the known/knowable spaces – and uses these to discuss how teams might react to complex events such as Bhopal, the decision to launch Challenger and the Three Mile Island Accident. It is far from clear that organisational practices that enable repetitive, intrinsically dangerous operations to be carried out safely can be used to develop organisational preparedness dealing with complex situations that bring many risks, some quite unanticipated.

Applying Cynefin to decision making has helped decision analysts recognise that different methodologies may be needed for decision contexts lying in the different spaces [31]. We believe that it can serve the same purpose in delineating when different HRA methodologies are appropriate.

Up until now we have discussed the implications of behavioural, organisational, and cultural contexts for human behaviour within a system and the consequences of this behaviour for the system's reliability. We have noted that that are many different types of activity required of the operator in contexts varying from the known to the complex and chaotic. In the next section we discuss the varying purposes that might be served by an HRA and which set the context for the analysis – as opposed to the context in which the system being analysed operates. We have already noted that an HRA may serve formative or summative objectives. In the next section we enlarge upon the implications of these.

5.3 Why perform HRA?

There are many reasons why one might undertake an HRA: e.g.

- In the design of a system one may be concerned with 'designing out' the potential for system failure. Part of this involves analysing how human behaviour may affect the system in its potential both to compromise its reliability and to avoid the threat of imminent failure.
- During licensing discussions between a government regulator and the system operator there may be a need to demonstrate that a system meets a safety target.
- Sometimes an organisation wants to restructure and change its reporting structures. In such circumstances, it may wish to understand how its organisational design may affect the reliability and safety of its systems; and in turn that understanding may inform the development of its safety culture.
- There may be a need to modify a system in which case there are needs to design the modification *and* the project to deliver the modification.
- There may be a need to choose which of several potential systems to purchase and the risk of system failure may be a potential differentiator between the options.

Risks occur throughout the life cycle: during construction or installation, during operation, during modification and during decommissioning. In event of failure at any stage there are risks associated with repair and recovery. In all cases HRA will inform risk management. Sometimes, it will be possible to consider controlling risk to within society norms; in the case of repair and recovery, there may be no choice but to take what normally would be considered excessive risk.

Thus there are many contexts and reasons for conducting an HRA. It would be surprising, therefore, if one HRA methodology served all. Yet, often reading the proponents of a given HRA methodology, there is a suggestion, perhaps implicit, that their methodology might serve all contexts. Equally critics of a methodology may point to particular faults without the recognition that for some contexts these may not be relevant. In the Section 6 we shall argue that within any large scale system risk or reliability analysis, there will be many points at which human behaviour may be a factor leading to system failure. Which HRA methodology is adopted at any one of these points will depend on the many contextual issues and purposes of the overall analysis discussed above. Thus the overall analysis may involve *several* HRA methodologies. But first we should note that Hollnagel's CREAM [26] methodology is to some extent based upon the same conception.

5.4 The importance of context in Hollnagel's CREAM methodology

Hollnagel [26] also recognises that there are many contextual issues that determine the appropriateness of a particular HRA. He suggests that the context should be categorised in various ways including four control modes: *opportunistic*, *tactical*, *strategic* and *scrambled*. At first sight these look rather similar to the four categories of decision contexts in Cynefin. However, there is one significant difference. Cynefin categorises the decision context without reference to behaviour of the human taking the decision. Hollnagel’s classification confounds these factors. Since it is perfectly possible for one individual to face up to the strategic response to incident in a calm focused frame of mind, while another may be distracted or in state close to panic, to confound these factors is to miss one of the key challenges of HRA.

Hollnagel’s CREAM methodology encompasses two analytical approaches: basic CREAM and extended CREAM. The former is quicker to apply, makes more approximations and presumably more suited when one needs rough guidance, e.g., on the relative risks resulting from human behaviour in different parts of a system; whereas the latter gives a more precise answer for circumstances when greater quantification is needed. Thus in this sense Hollnagel has begun to address the need for different HRA methods to meet the different possible purposes of such analyses. However, there are two issues that arise. Firstly, given the wide range of reasons for conducting an HRA given above, one may question whether *two* analytical approaches are sufficient. Second and more importantly, Bedford and Bayley [79] have tried applying both basic and extended CREAM to the same context and obtained inconsistent answers. It seems that the former method is *not* a consistent approximation to the latter.

6 The need for a portfolio of HRA methodologies

6.1 Behaviour not error

We have made the point several times that we should not think of the issue being one of assessing the risks of human ‘error’ or ‘failure’. Human reliability should focus on the interaction of a variety of human behaviours and the operation of the overall system. To do it job, HRA does not need to attribute pejorative terms such as ‘error’, ‘slip’ or ‘failure’ to a particular behaviour. It simply needs to ask what the humans are doing and can do at a particular point in the operation of the system. What behaviours might they exhibit and what effects might these behaviours have on the system?

In any particular application of HRA methods, there is a need to ensure that its focus is on a sufficiently wide grouping of the human part of the system. If one is examining the likelihood of a slow response to a signal on a computer screen then perhaps it is sufficient to consider the cognitive behaviour of an individual operator. If, however, the interest is in the likelihood that several signals will be ignored by several operators, then not only should the cognitive behaviour of one individual be considered but also the group dynamics of the team of operators and perhaps a wider range of organisational behaviours. Even when the focus is on a single operator there may be a need to consider management and other behaviours around him or her and perhaps training systems.

The focus should also be sufficiently wide in temporal terms. If an operator realises that his or her response to a signal was slow, he or she may reschedule immediate tasks or adopt some other strategy to recover and prevent a system failure. We rely on people in a system not just for their ability to perform many complex and sensitive tasks but also for their intelligence and ingenuity in resolving problems. Thus HRA should assess human behaviour and its effects from the initiating event or behaviour through to the point at which the overall system passes into a different state from its planned operation.

Perhaps we should refer not to human *reliability* analysis but to human *behaviour* analysis and its role in overall system risk and reliability studies. However, it is not the name that matters but the apparent quest for a single HRA method applicable to all tasks.

6.2 Context matters!

Our contention is that the variety of tasks that HRA is called upon to perform and the range of contexts in which it is applied are so great that it would be optimistic in the extreme to expect one method to be sufficient to meet these requirements. Hollnagel [26] recognised this, though his suggestion of two methods probably does not take us much further forward. What we believe is needed is a portfolio of HRA methods. The characteristics of each need to be well understood so that we can determine the appropriate contexts for its application and appreciate its accuracy.

First we need to recognise the context of the analysis itself. Why do we need to assess the risk or reliability of the system? In Section 5.3 we indicated a range of possible reasons for needing to perform such analyses. We need to recognise that some of these will need quantitative output, while other times more qualitative output may suffice. In qualitative cases, sometimes we need a ranking; but in the early stages of designing a system it may suffice to have an understanding of the human behavioural issues that may affect its performance and put it at risk of some system failure. If quantitative output is needed, then it will be necessary to have an understanding of the bounds on the numbers produced. How accurate are they? This will be particularly important in demonstrating that a safety target is reached – or maybe that they can never be reached. If Perrow's theory of normal accidents [13, 54] is to be believed, then some targets of, say, less than one failure in 10^7 years may be unachievable for some very complex systems; and the demonstration of this may well be based upon modelling limitations on the cognitive capacity of the operators or related behavioural issues.

We should note here that a good quantitative method may not necessarily provide the best qualitative information. 'Black Box' models [80], for instance, cannot by definition bring qualitative understanding. Even statistical prediction based upon measureable system and human behavioural factors may not bring much true qualitative understanding. Equally, good qualitative models may not develop easily into good quantitative models. Thus the purpose of the overall analysis does much to define the type of HRA method that should be adopted.

A system risk or reliability analysis should begin by exploring potential hazards and points at which failure may occur. Some of these will be clearly associated with human operations and interventions; but we should recognise that *all* failure modes potentially have human behavioural aspects once introduce considerations of response and recovery actions. Traditionally, task analysis is used to identify the points at which human activity can lead to system failure. Essentially one investigates what happens if an operator fails to complete an assigned task in the expected manner. However, while this is a necessary step in identifying possible ways in which human behaviour may be involved in system failure, it is not sufficient. Firstly, recovery actions introduce human behavioural aspects into all possible failure modes. Secondly and more importantly, task analysis focuses on closely related groups of operations – generally closely related not just conceptually, but also in terms of time and space. Yet organisational and cultural effects can correlate many aspects of behaviour in very different areas of operations. Remember our discussion of the Swiss Cheese Model (Section 2.1). Task analysis focuses attention on the potential 'holes in the cheese layers' whereas the wider behavioural issues can lead to 'movement of the cheese layers' becoming correlated. In design terms this means that the safety barriers that are built into the system, be they physical or organisational, may not be as independent as their designer think initially.

The appropriateness of any HRA method may depend on the decision context that is being assessed. Are we considering a repetitive task that an operator performs in the normal course of events? In which case we need modelling approaches that fit with behaviours in the known domain. Or are we looking at the response of an operator to something unexpected that may herald a departure of the system from its normal operating characteristics? In which case we need modelling behaviours for the knowable, complex or even chaotic domain. For repetitive events the key contextual pressures on operators that may modify their behaviour are likely to relate to complacency and organisational issues such as excessive workloads or requirements to work at the same task too long. External pressures and distractions such family problems or a national sporting event are more likely to affect behaviour in repetitive normal operations than in responding to the unexpected. In responding to events ranging from an indication of departure from normal operations to a full blown crisis, adrenaline, the importance of the matter, as well as cognitive interest are likely to focus the mind. So the operators' performance is more likely to be affected by issues such as cognitive overload, miscommunication between several operations and a range of behaviours that we commonly call panic! Organisational contexts that affect the operators' responses relate to, *inter alia*, the provision of training, including emergency simulations in a variety of scenarios, and the establishment of common mental models among response teams and, more generally, of supportive team behaviours. Thus to assess the appropriateness of any HRA method to a decision context, it is necessary to understand how each models such cognitive, behavioural and organisational effects and their strengths and weaknesses in doing so.

In summary, to choose an appropriate HRA method one needs to consider:

- the purpose of the analysis
- the decision context;
- the organisational, cultural, and managerial contexts.

6.3 Building a coherent composite analysis

Thus we would expect that several HRA methods would be used in a full systems risk or reliability analysis. At each point where human behaviour was an issue, an HRA appropriate for the task/decision context and for the purposes of the analysis would be chosen. The outputs of these would be drawn together into the overall systems risk or reliability analysis. The need to draw them together means that it is important that the HRA and other component risk and reliability modelling cohere in a number of respects.

- Datasets need to be common across all analyses; and the inputs and outputs of different analyses need to be compatible in many ways such as unit, accuracy and, most importantly, operational meaning. It is more than possible for models to involve parameters known by the same name, but to differ in terms their precise definitions within the models.

- Environmental parameters should represent the same context. The operation of the system needs to be explored within a consistent representation of the external world – and that includes not just the physical, but also the political, social and economic environment.
- The component analyses should be based upon compatible sets of assumptions. It would be clearly quite wrong to assume that an operator can work at some level of effectiveness for up to 90 min at one point of the analysis; and for up to 120 min at another point. But there are also many more subtle modelling assumptions made in any analysis and these need to be compatible across the whole systems analysis.
- The underlying methodological foundations should be compatible. For instance, it is far from clear that probabilistic methods can be safely combined with fuzzy or possibility methods [4, 81, 82]. In short, the different component methods drawn together into the full systems analysis should 'speak the same language'.

We have already noted that the basic and extended CREAM methods would seem to be incompatible in some of these respects [79].

6.4 Future research directions

How far are we from this conception of the role of HRA in systems risk or reliability analyses? A long way, probably.

First, let us assume for the present that we have available enough HRA methods for our needs. If we are to select from this portfolio an appropriate HRA method for particular part of the full analysis, we need to be able to compare them in terms their relative appropriateness for:

- the decision context,
- the organisational, cultural, and managerial contexts,
- the purpose, i.e. formative or summative, of the analysis and the accuracy needed from it.

We also need to compare the methods in terms of their compatibility and overall coherence.

While there are some comparative studies of some of the methods [e.g., 27, 83-85], there are too few to build full and consistent comparisons of the methods that are available. More are needed. Moreover, the comparisons need to recognise that each method may be appropriate some contexts but not others. Once this is done, we can then identify the gaps in our portfolio. For what contexts do we have no appropriate HRA methods and how might further methods be developed for these?

Second, the community needs to consider how the overall system risk or reliability analysis is pulled together. We have already noted that the Swiss Cheese model predisposes analysts to decompose systems into independent layers. We have also noted that human and organizational behaviours can become a common cause of failure in different parts of the systems, correlating these layers. The risk and reliability community need to recognise this and break away in some respects from the mindset engendered by the Swiss Cheese model: see also [86]. There needs to be a recognition that correlations due to behavioural and organisational issues can connect different subsystems in perhaps unexpected ways. Thus more than ever analysts need to challenge assumptions of independence between subsystems.

Third, the terminology of the discipline needs to change to recognise that failures can arise as a result of perfectly rational and reasonable behaviours. The HRA community needs to widen its focus from error to all human behaviour. In many ways the development of second and third generation HRA methods recognise this, but it seems very likely that there is a considerable way to go. To have an indication of how far, we need to explore and draw together the HRA, behavioural and cognitive sciences, and organisation theory literatures and explore the implications of each for the others.

7 Summary and Conclusion

Our project may not have led to a 'rethinking of human reliability analyses' at least not in the sense of seeing how a single all-embracing method might be developed. However, we have become acutely aware that a wide variety of contextual issues have not been fully appreciated and that these determine the characteristics needed in any HRA application. Moreover, we believe that the discipline has been dominated by too great a focus on human 'error' rather than a more balanced recognition that all human behaviour has potential to cause a system failure. In the previous section, we outlined three developments that we believe are needed:

(i) comparative analyses of the HRA methods that already exist and gap analyses to identify the need for further methods;

(ii) a greater recognition that human and organisational behaviour can act as a common cause, correlating failures in different subsystems or safety barriers;

(iii) a greater exploration of what can be learnt from the behavioural, cognitive and organisational sciences.

Until these steps are taken, we believe that it will be impossible to meet the need for society to understand and manage the risks brought by systems of ever-increasing complexity.

Acknowledgements

We are grateful for financial support from the UK EPSRC for funding a research project entitled *Rethinking Human Reliability Analysis Methodologies* (contract: EP/E017800/1). The following attended one or more of our workshops and offered very helpful contributions: Mark Bendig, Roland Bradshaw, Calvin Burns, Nicola Content, Andrew Evans, Ned Hickling, Andrew Howes, John Quigley, Osman Turan and Lesley Walls. Ronald Boring offered very helpful comments on a near final draft of the report.

Appendix A: List of Abbreviations and Acronyms

AEMA	action error mode analysis
APJ	absolute probability judgement
ATHEANA	a technique for human error analysis
CCA	cause-consequence analysis
CFP	cognitive failure probability (concept in CREAM)
COCOM	contextual control model
CPC	common performance condition (concept in CREAM)
CREAM	Cognitive reliability and error analysis method
EFC	error forcing conditions
EOC	errors of commission
EOO	errors of omission
EPC	error producing condition (concept in HEART)
FA	functional analysis
FMEA	failure modes and effects analysis
FTA	fault tree analysis
GBAS	ground based augmentation system
GEMS	generic error modelling system
HAZOP	hazard and operability analysis
HCR	human cognitive reliability
HEART	human error assessment and reduction technique
HEAT	human error action taxonomy (concept in CREAM)
HEP	human error probability
HFE	human failure event
HRA	human reliability analysis
HEART	human reliability analysis event tree
HRO	high reliability organisation
IDA	influence diagrams approach
JHEDI	justification of human error data information
LOCA	loss of coolant accident
MAUD	multi-attribute utility decomposition
MMI	man-machine interface

MTO	man-technology organisation
NGT	nominal group technique
NRC	nuclear regulatory commission
OR	operational research
ORCA	operator reliability calculation and assessment
PHEA	predictive human error analysis (concept in CREAM)
PRA	probability risk assessment
PSA	probabilistic safety assessment
PSF	performance shaping factor
RCA	root cause analysis
SARAH	systematic approach to the reliability assessment of humans
SHARP	systematic human action reliability procedure
SLI	success likelihood index
SLIM	success-likelihood index methodology
SRK	skill-rule-knowledge
STAHR	socio-technical assessment of human reliability
TESEO	Tecnica Empirica Stima Errori Operatori
THERP	technique for human error rate prediction
THORP	(Sellafield) thermal oxide reprocessing plant
TQM	total quality management
TRC	time reliability correlation
UA	unsafe action

Appendix B: A survey of HRA methods

Absolute Probability Judgement

Absolute Probability Judgement (APJ) is a technique used in the field of HRA for the purposes of evaluating the probability of a human error occurring throughout the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

APJ, which is also known as Direct Numerical Estimation [87], is based on the quantification of HEPs. Expert judgement is typically desirable for utilisation in HRA when there is little or no data with which to calculate HEPs, or when the data is unsuitable or difficult to understand. In theory, qualitative knowledge built through the experts' experience can be translated into quantitative data such as HEPs.

Required of the experts is a good level of both substantive experience (i.e. the expert must have a suitable level of knowledge of the problem domain) and normative experience (i.e. it must be possible for the expert, perhaps with the aid of a facilitator, to translate this knowledge explicitly into probabilities). If experts possess the required substantive knowledge but lack knowledge which is normative in nature, the experts may be trained or assisted in ensuring that the knowledge and expertise requiring to be captured is translated into the correct probabilities i.e. to ensure that it is an accurate representation of the experts' judgements.

Background

APJ is an expert judgement-based approach which involves using the beliefs of experts (e.g. front-line staff, process engineers etc.) to estimate HEPs. There are two primary forms of the technique; Group Methods and Single Expert Methods i.e. it can be done either as a group or as an individual exercise. Group methods tend to be the more popular and widely used as they are more robust and can be used to generate a consensus opinion. Moreover, within the context of use, it is unusual for a single individual to possess all the required information and expertise to be able to estimate solely, in an accurate manner, the human reliability in question. In the group approach, the outcome of aggregating individual knowledge and opinions is more reliable, in the sense that it allows for different expertises to be represented in the expert group.

APJ Methodologies

There are four main group methods by which APJ can be conducted.

- Aggregated Individual Method

  Utilising this method, experts make their estimates individually without actually meeting or discussing the task. The estimates are then aggregated by taking the geometric mean of the individual experts' estimates for each task (though in other expert judgement applications it is more common to take a weighted arithmetic average – possibly with the weights being determined by performance, see Cooke [82]. The major drawback to this method is that there is no shared expertise through the group; however a positive of this is that due to the individuality of the process, any conflict such as dominating personalities or conflicting personalities is avoided and the results are therefore free of any bias.

- Delphi Method

  Developed by Dalkey [88, 89], this method is very similar to the Aggregated Individual Method in that experts make their initial estimates in isolation. However following this stage, the experts are then shown the outcome that all other participants have arrived at and are then able to re-consider the estimates which they initially made. The re-estimates are then aggregated using the geometric mean. This allows for some information sharing, whilst avoiding most group-led biases; however their still remains the problem of a lack of discussion.

- Nominal Group Technique (NGT)

  This technique takes the Delphi method and introduces limited discussion/consultation between the experts. By this means, information-sharing is superior, and group domination is mitigated by having the experts separately come to their own conclusion before aggregating the HEP scores.

- Consensus Group Method

  This is the most group-centred approach and requires that the group must come to a consensus on the HEP estimates through discussion and mutual agreement. This method maximises knowledge-sharing and the exchange of ideas and also promotes equal opportunity to participate in discussion. However, it can also prove to be logistically awkward to co-ordinate as it requires that all experts be together in the same location in order for the discussion to take place. If the circumstance arises in which there is a deadlock or breakdown in group dynamics, it then becomes necessary to revert to one of the other group APJ methods.

APJ Procedure

1. Select subject matter experts

   The chosen experts must have a good working knowledge of the tasks which require to be assessed. The correct number of experts is dependent upon what seems most practicable, while considering any constraints such as spatial and financial availability. However, it should be noted that the larger the group the more likely problems are to arise.

2. Prepare task statement

   Task statements are a necessary component of the method; tasks are specified in detail. The more fuller the explanation of the task within the statement, the less likely it will be that the experts will resort to making individual guesses about the tasks. The statement should also ensure that any assumptions are clearly stated in an interpretable format for all experts to understand. The optimal level of detail will be governed by the nature of the task under consideration and the required use of the final HEP estimation.

3. Prepare response booklet

   These booklets detail the task statement and design of the scale to use in assessing error probability and by which experts can indicate their judgements [87]. The scale must be one which allows differences to be made apparent. The booklet also includes instructions, assumptions and sample items.

4. Develop instructions for subjects

   Instructions are required to specify to the experts the reasons for the session, otherwise they may guess such reasons which may cause bias in the resultant estimates of human reliability.

5. Obtain judgements

   Experts are required to reveal their judgements on each of the tasks; this can be done in a group or individually. If done by the former means, a facilitator is often used to prevent any bias and help overcome any problems.

6. Calculate inter-judge consistency

   This is a method by which the differences in the HEP estimates of individual experts can be compared; a statistical formulation is used for such purposes.

7. Aggregate individual estimates

   Where group consensus methods are not used, it is necessary to compute an aggregate for each of the individual estimates for each HEP.

8. Uncertainty bound estimation

   Calculated, or assessed, by using statistical approaches involving confidence ranges. (Note that these will usually not be confidence intervals in the usual sense of statistics, but will represent the expert group confidence in making their predictions).

Worked Example [87]

Context: In this example, APJ was utilised by Eurocontrol, at the experimental centre in Bretigny Paris, using a group consensus methodology.

Required Inputs: Each of the grades of staff included in the session took turns to provide estimates of the error probabilities, including ground staff, pilots and controllers. Prior to the beginning of the session, an introductory exercise was conducted to allow the participants to feel more comfortable with use of the technique; this involved an explanation to the background of the method and provided an overview of what the session would entail of. To increase familiarity of the method, exemplary templates were used to show how errors are estimated.

Method:

1. Initial task statements of the project were created leaving space for individual opinion of task estimates and additional assumptions the group may have collectively foregone.

2. A session was held in which the individual scenarios and tasks were accurately detailed to the experts
3. Experts, with this knowledge, were then able to enter individual estimations for all tasks under consideration
4. Discussion followed in which all participants were provided with the opportunity to express their opinion to the rest of the group
5. Facilitation was then used in order to reach a group consensus on the estimate values. Further discussion and amendment took place when necessary.

During the duration of the session it was revealed that the ease with which the experts were able to arrive at a consensus was low with regards to the differing estimates of the various HEP values. Discussions often changed individuals' thinking e.g. in the light of new information or interpretations, but this did not ease reaching an agreement. Due to this difficulty, it was therefore necessary to aggregate the individual estimates in order to calculate a geometric mean of these.

The following table displays a sample of the results obtained.

Table: Pilot APJ Session – extract of results

Potential Error {Code in Risk Model}	**Excluding PC test (unsound expertise discarded)**			
	Maximum	**Minimum**	**Range**	**Geometric Mean**
C1a	1.1E-03	2.0E-05	**55**	2.1E-04
C1b	2.5E-04	1.0E-05	**25**	3.5E-05
D1	1.0E-03	1.0E-04	**10**	4.3E-04
F1a	4.0E-04	1.0E-05	**40**	6.9E-05
F1b	1.0E-03	1.0E-04	**10**	4.0E-04
F1c	1.0E-03	1.0E-04	**10**	4.6E-04

In various cases, the range of figures separating the maximum and minimum values proved to be too large to allow to aggregated value to be accepted with confidence

These values are the events in the risk model which require to be quantified. There are 3 primary errors in the model that may occur:

- C1: Capturing false information about final approach path
- D1: Failure to maintain a/c on final approach path
- F1: Selecting wrong runway

There were various reasons which can explain the reasons why there was such a large difference in the estimates provided by the group: the group of experts was largely diverse and the experience of the individuals differed. Experience with Ground Based Augmentation System (GBAS) also showed differences. This process was a new experience for all of the experts participating in the process and there was only a single day, in which the session was taking place, to become familiar with its use and use it correctly. Of most significance was the fact that the detail of the assessments was very fine, which the staff were not used to. Experts also became confused about the way in which the assessment took place; errors were not considered on their own and were analysed as a group. This meant that the values estimated represented a contribution of the error on a system failure as opposed to a single contribution to system failure.

Results/Outcomes:

- Controllers and pilots provided good estimates for the errors and these have been used in some safety cases
- Participants highlighted their understanding of the importance of their participation in the process to provide expertise, as opposed to using external safety analysts instead i.e. their understood their role in carrying out a HRA of the system
- The experts were provided with a realistic representation of human performance within the system and therefore further safety requirements required to improve the safety and reduce the likelihood of the identified errors. This is particularly beneficial; for the future GBAS.

Lessons from the study:

Time is required to familiarise with the methodology and to understand what is needed to be done in the given context.

Experts are required to understand the circumstances in which HEPs are conditional.

There is a need for true experts to be included in the process and in significant number to allow for the necessary information to be gathered.

The use of existing information in the process is always helpful for the purposes of standardisation.

Advantages of APJ

- The method is relatively quick and straightforward to employ. With a greater degree of group discussion in use of the technique, there is more qualitative data that is produced; this can be considered as a useful by-product of the assessment [87].
- APJ is not restricted to or specialised for use in a particular field; it is easily applicable to an HRA on any industrial sector thus making it a generic technique for use in a wide range of potential applications [90].
- Useful suggestions may result from discussion as to ways in which a reduction in errors can be achieved [91].

Disadvantages of APJ

- APJ is prone to judgemental biases and group conflicts or problems. Selection of the correct group methodology or high-quality group facilitation may decrease the effect of these biases and increase the validity of the results [87].
- Locating suitable experts for the APJ exercise is a difficult stage of the process, more so due to the ambiguity with which the term ‘expert’ can be defined [90].

- Because there may be little or no empirical and/or quantitative reasoning underpinning the experts' estimates, it is difficult to be certain of the validity of the final HEPs i.e. there is no means by which guesses can be validated [87].

A Technique for Human Error Analysis (ATHEANA)

ATHEANA [92] is a technique used in the field of HRA for the purposes of evaluating the probability of a human error occurring throughout the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

ATHEANA is used in both retrospective analyses, that is in analysing incidents that have occurred, and in prospective analyses, that is in analysing incidents that could occur. After a series of studies of plant incidents, it was observed that the incidents occurred in a context where the combination of plant state, performance shaping factors and dependencies led, almost inevitably, to a human error. Hence the main underlying principle of ATHEANA is that error forcing conditions (EFCs) are described for non-nominal situations. Application of ATHEANA in probabilistic risk analysis hence requires an assessment of the possible EFCs and their likelihoods. The various drivers of an incident and the possible outcomes are categorised into one of the following groupings: organisational influences; performance shaping factors; error mechanisms; unsafe actions; human failure event; unacceptable outcome(s). The outcome provided by ATHEANA identifies various human actions within a system while also eliciting many contextual situations within this system, which influence whether the action will be carried out successfully or will lead to failure.

Background

ATHEANA is both a retrospective and prospective HRA methodology developed by the US nuclear industry regulatory commission in 2000. It was developed in the hope that certain types of human behaviour in nuclear plants and industries, which use similar processes, could be represented in a way in which they could be more easily understood. It seeks to provide a robust psychological framework to evaluate and identify PSFs - including organisational/environmental factors - which have driven incidents involving human factors, primarily with the intention of suggesting process improvement [92]. Essentially it is a method of representing complex accident reports within a standardised structure, which may be easier to understand and communicate.

ATHEANA methodology

The basic steps of the ATHEANA methodology are [93]:

1. Define and interpret the issue under consideration
2. Detail the required scope of analysis
3. Describe the Base case scenario for a given initating event, including the norm of operations within the environment, considering actions and procedures.
4. Define Human Failure Events (HFEs) and/or unsafe actions (UAs) which may affect the task in question

5. Identify potential vulnerabilities in the operators' knowledge base.
6. Search for deviations from the base case scenario for which UAs are likely.
7. Identify and evaluate complicating factors and links to PSFs.
8. Evaluate recovery potential.
9. Quantify HFE probability
10. Incorporate results into the PRA.

A schematic outline of the method is provided below [94].

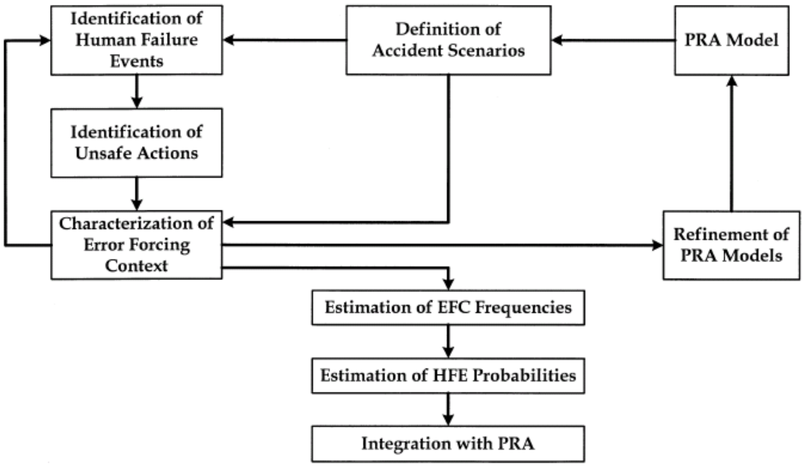

Identification of Human Failure Events
Definition of Accident Scenarios
PRA Model
Identification of Unsafe Actions
Characterization of Error Forcing Context
Refinement of PRA Models
Estimation of EFC Frequencies
Estimation of HFE Probabilities
Integration with PRA

Figure. Schematic outline of ATHEANA

The probability of a HFE in ATHEANA, given a particular initiator, is determined by summing over the different error forcing conditions associated to the HEF, taking account of the likelihood of unsafe actions given the EFC, and the likelihood of no recovery action given the EFC and the UA.

Advantages

- The most significant advantage of ATHEANA is that it provides a much richer and more holistic understanding of the context concerning the Human Factors known to be the cause of the incident, as compared with most first generation methods.
- It may also be suggested that carrying out the qualitative model structuring leads to the enhancement of understanding as it requires stakeholders and decision makers to consider and discuss the contributing aspects as part of the model-building procedure.
- It increases the guarantee that the key risks associated with the HFEs in question have been identified[11].
- Utilising the ATHEANA methodology, it is possible to estimate HEPs considering a variety of differing factors and combinations.

[11] http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1880/sr1880.pdf

- Compared to many other HRA quantification methods, ATHEANA allows for the consideration of a much wider range of performance shaping factors and also does not require that these be treated as independent. This is important as the method seeks to identify any interactions which affect the weighting of the factors of their influence on a situation.

Criticisms

- A peer review of the original version of ATHEANA [95], raised a number of criticisms, some of which could be argued to apply to the version described in Barriere *et al* [92]. Some significant critical points made by reviewers are that
  - the method is cumbersome and requires a large team;
  - the method is not described in sufficient detail that one could be sure that different teams would produce the same results;
  - the quantification method is weak.
- The taxonomic approach used by ATHEANA has been criticised: in particular the notion *error forcing* [96].

Cognitive Reliability and Error Analysis Method (CREAM)

CREAM [26] is a technique used in HRA for the purposes of evaluating the probability of a human error occurring throughout the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business; each technique has varying uses within different disciplines.

CREAM is a second generation HRA method. Compared to many other methods, it takes a very different approach to modelling human reliability. There are two versions of the technique, the basic and the extended version, both of which have in common two primary features; ability to identify the importance of human performance in a given context and a helpful cognitive model and associated framework, usable for both prospective and retrospective analysis. Prospective analysis allows likely human errors to be identified while retrospective analysis quantifies errors that have already occurred.

The concept of cognition is included in the model through use of four basic ‘control modes’ which identify differing levels of control that an operator has in a given context and the characteristics which highlight the occurrence of distinct conditions. The control modes which may occur are as follows:

- *Scrambled control*: the choice of the forthcoming action is unpredictable or haphazard. The situation in question may be portraying rapid alterations in unexpected ways thus eliminating the operator’s ability or opportunity to make deductions about the next action required.
- *Opportunistic control*: the next action is determined by superficial characteristics of the situation, possibly through habit or similarity matching. The situation is characterised by lack of planning and this may possibly be due to the lack of available time.
- *Tactical control*: performance typically follows planned procedures while some ad hoc deviations are still possible.
- *Strategic control*: plentiful time is available to consider actions to be taken in the light of wider objectives to be fulfilled and within the given context.

The particular control mode determines the level of reliability that can be expected in a particular setting and this is in turn determined by the collective characteristics of the relevant Common Performance Conditions (CPCs).

Background

CREAM was developed by Eric Hollnagel in 1998 following an analysis of the methods for HRA already in place. It is the most widely utilised second generation HRA technique and is based on 3 primary areas of work; task analysis, opportunities for reducing errors and possibility to consider human performance with regards to overall safety of a system.

The aim of utilising this methodology is to assist an analyst to:

- identify work, actions or tasks within the system which necessitate or essentially depend on human thinking and which are therefore vulnerable to variations in their level of reliability;
- identify the surrounding conditions in which the cognition of these situations may be reduced and therefore determine what actions may lead to a probable risk;
- compile an evaluation from the assessment of the various outcomes of human performance and their effect on system safety – this can then be utilised as part of the Probability Risk Assessment (PRA);
- make suggestions as to how identified error producing conditions may be improved and therefore of how the system's reliability can be enhanced whilst also reducing risk.

Methodology

1. *Task Analysis*. The basic method adopted by the CREAM technique provides an immediate reliability interval based on an assessment of the given control mode, as highlighted by the figures provided in the table below. As can be seen by the contents of the table, each of the specified control modes has an individual reliability level. In the extended CREAM version, the control modes play the role of a weighting factor which scales a nominal failure probability associated to a given cognitive function failure. This version of CREAM is intended to be used for the purposes of a more in depth analysis of human interactions.

Control mode	**Reliability interval (probability of action failures)**
Strategic	0.5 E-5 < p < 1.0 E-2
Tactical	1.0 E-3 < p < 1.0 E-1
Opportunistic	1.0 E-2 < p < 0.5 E-0
Scrambled	1.0 E-1 < p < 1.0 E-0

Table Reliability intervals from [26].

2. *Context description*. The intention of the basic CREAM method is to use it as a screening technique with the aim of identifying processes which require a deeper level of analysis; this analysis may then be carried out by the extended CREAM method.

3. *Specification of Initiating Events*. When using the basic CREAM method, a task analysis is conducted prior to further assessment. CPCs are assessed according to the descriptors, given in the table below, in order to judge their expected effect on performance.

CPC name	**Level/descriptors**	**Expected effect on performance reliability**
Adequacy of organisation	Very efficient	Improved
	Efficient	Not significant
	Inefficient	Reduced
	Deficient	Reduced

CPC name	**Level/descriptors**	**Expected effect on performance reliability**
Working conditions	Advantageous	Improved
	Compatible	Not significant
	Incompatible	Reduced
Adequacy of MMI and operational support	Supportive	Improved
	Adequate	Not significant
	Tolerable	Not significant
	Inappropriate	Reduced
Availability of procedures/ plans	Appropriate	Improved
	Acceptable	Not significant
	Inappropriate	Reduced
Number of simultaneous goals	Fewer than capacity	Not significant
	Matching current capacity	Not significant
	More than capacity	Reduced
Available time	Adequate	Improved
	Temporarily inadequate	Not significant
	Continuously inadequate	Reduced
Time of day (circadian rhythm)	Day-time (adjusted)	Not significant
	Night-time (unadjusted)	Reduced
Adequacy of training and expertise	Adequate, high experience	Improved
	Adequate, limited experience	Not significant
	Inadequate	Reduced
Crew collaboration quality	Very efficient	Improved
	Efficient	Not significant
	Inefficient	Not significant
	Deficient	Reduced

Table Common Performance Conditions (CPCs)

4. *Error Prediction*. The assessments of the CPCs then require to be adjusted according to some specified rules in order to take account of synergistic effects. The matrix above would be considered in the context of the situation under assessment and by this means the previously considered initiating events are reviewed with respect to how they could potentially lead to the occurrence of an error. The rows of the matrix identify the possible outcomes while the columns show the precursors. The analyst then has the task of identifying the columns for which all the rows have been similarly classified into the same group according to the column headings.

Predicting the possible outcomes for each of the rows should be done until there are no remaining possible paths. Each of the identified errors requires to be noted along with the causes and the outcomes.

5. Finally, a simple count is performed of the number of CPCs that are causing an improvement in reliability and those which are reducing it. On the basis of this number the probable control mode is determined, by judging the region given in the graph as depicted below.

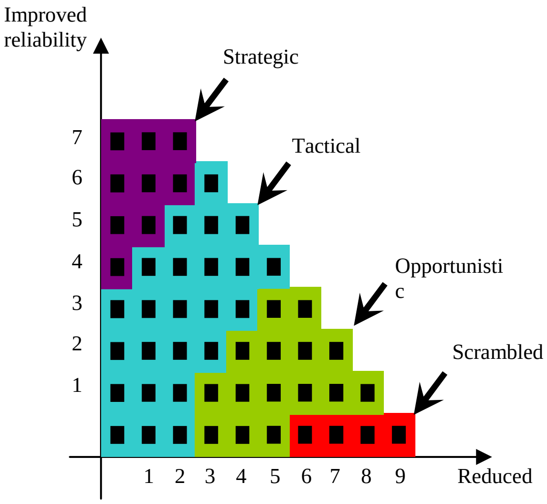


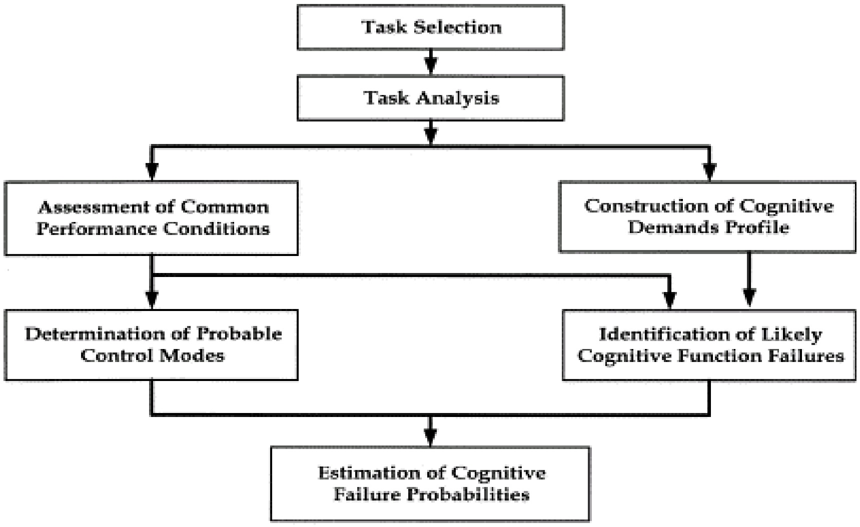
Figure **Allocation of probable control modes according to CPCs** – each section displays the relation between the CPC score and control modes. The x coordinate represents the number of reduced influence indexes and the y coordinate is the number of improved influence indexes.

6. The extended version of the CREAM methodology operates in a slightly different manner. Following the initial task analysis, a refinement is then provided in terms of the cognitive activities which are involved in the considered task (classified as co-ordinate, communicate, compare, etc). To these activities a Contextual Control Model (COCOM) function (observation, interpretation, planning and execution) is ascribed (following a table provided) so that a cognitive demand profile may be established.

7. Following this stage, the probable cognitive function failures are identified, based on a knowledge of the specific tasks, yet following a set of generic cognitive function failures associated to the COCOM functions. Each of these generic failures is associated with a nominal probability which is based on a table given in CREAM. However these probabilities are adjusted depending on the particular control mode.

Hence in the extended version of the CREAM methodology the control mode acts in the role of a Performance Shaping Factor with the task of performing adjustments to a nominal probability.

Provided below is a simplified diagrammatic representation of all the stages involved in the complete CREAM methodology.


Task Selection
Task Analysis
Assessment of Common Performance Conditions
Construction of Cognitive Demands Profile
Determination of Probable Control Modes
Identification of Likely Cognitive Function Failures
Estimation of Cognitive Failure Probabilities

Worked example

Context. The basic example that is provided below concerns the task of 'restarting a furnace following a system trip'.

The following figure illustrates the hierarchical task analysis carried out for the task.

O. Warm up furnace

Plan O: Do in order

O.1. Prepare plant and services

O.2. Start air blower

O.3. Start oil pump

O.4. Heat oil to 800^0 C

Plan O.1: Do in any order

O.1.1. Ensure plant is ready

O.1.2. Ensure gas-oil is available

O.1.3 Ensure O_2 analysis system is working

Plan O.4: Raise temperature to 800^0 C while monitoring O_2 and ΔT

O.4.1. Increase temperature controller as per chart

O.4.2. Monitor O_2

O.4.3. Monitor temperature

O.4.4. Switch furnace to automatic

The overall task is made up of four basic tasks, which are further completed by carrying out a number of sub-tasks. First level tasks are required to be performed in the given sequence while the tasks on sub level two can be carried out in any order. Finally the lowest level tasks are conducted as necessary in a repetitive manner.

Assumptions. In this example, a number of assumptions should be noted in order to aid understanding. It is assumed that the warm-up task does not have any procedural support nor is it one that has been trained in detail.

Required Inputs. From study of the task analysis, it is then possible to identify the necessary activities of the overall task that must be carried out.

This involves assessing the work conditions under which the task in question is performed. These are judged and rated on a scale as can be seen in the table provided below:

CPC name	Evaluation
Adequacy of Organisation	The quality of the support and resources provided by the organisation for the task or work being performed. This includes communication systems, Safety Management System, support for external activities etc.
Descriptors	*Very efficient/ Efficient/ **Inefficient**/ Deficient*
Working conditions	The conditions under which the work takes place, such as ambient lighting, glare on screens, noise from alarms, interruptions from the task etc.
Descriptors	*Advantageous**/Compatible/**Incompatible*
Adequacy of MMI and operational support	The quality of the MMI and/or specific operational support provided for operators. The MMI includes control panels, workstations, and operational support provided by specifically designed decision aids
Descriptors	*Supportive/Adequate**/Tolerable/**Inappropriate*
Availability of procedures/plans	The availability of prepared guidance for the work to be carried out, including operating/ emergency procedures, routines & familiar responses
Descriptors	*Appropriate/Acceptable**/Inappropriate***
Number of simultaneous goals	The number of tasks or goals operators must attend to. Since the number of goals is variable, this CPC applies to what is typical/characteristic for a situation.
Descriptors	*Fewer than capacity/**Matching current capacity/**More than capacity*
Available Time	The time available to complete the work; or the general level of time pressure for the task and the situation type. How well the task is synchronised to the process dynamics.
Descriptors	***Adequate/**Temporarily inadequate/Continuously inadequate*
Time of day (circadian rhythm)	The time at which the task is carried out, in particular whether the person is adjusted to the current time.
Descriptors	***Day-time(adjusted)**/Night-time (unadjusted)*
Adequacy of training and preparation	The level of readiness for the work as provided (by the organisation) through training and prior instruction. Includes familiarisation to new technology, refreshing old skills, etc. as well as the level of operational experience
Descriptors	*Adequate, high experience/ Adequate, limited experience/ **Inadequate***

CPC name	Evaluation
Adequacy of training and preparation	The quality of the collaboration between crew members, including the overlap between the official and unofficial structure, the level of trust, and the general social climate among crew members.
Descriptors	*Very efficient/**Efficient/** Inefficient/ Deficient*

Method. In order to calculate the combined CPC score, the assigned ratings of the CPCs are entered in the table as shown in step 3 of the methodology section. Using certain rules [26] an assessment is made as to whether it is necessary to adjust the CPCs. In this example this is not necessary. Therefore the combined CPC score for this example is [3, 5, 1]. This is interpreted as the CPCs pointing to reduced performance reliability, 4 CPCs indicate that there is no significant influence and one CPC suggests an improved performance reliability.

Result. By determining the most likely control mode for the example, the general action failure probability can also thus be identified. Referring to the graphical display in Figure 1, the result for this example is that the operator is expected to be in an opportunistic control mode. This adequately relates to the assumption provided earlier that the operator under consideration has only slight experience or training for the task and there is insufficient support for the operations involved in the task. It may therefore be suggested that the operator may task a 'try and test' approach, particularly for complicated tasks such as increasing the temperature under controlled conditions.

The last stage of the process is to determine the probability interval for the expected control mode, the opportunistic control mode. Referring to the table of reliability intervals on page 47, for this example the general action failure probability is within the range of 1.0 E-2 < p < 0.5 E-0. As this is not regarded as an acceptable there is no clear and justified reason to continue with the analysis being undertaken.

Advantages

- The technique allows for the direct quantification of HEP.
- It also allows the assessor using the CREAM method to specifically tailor the use of the technique to the contextual situation [97].
- The resultant model is highly integrate-able into the primary safety process in use.
- The technique uses the same principles for retrospective and predictive analyses [97].
- The approach is very concise, well structured and follows a well laid out system of procedure [97].

Criticisms

- The technique requires a high level of resource use, including lengthy time periods for completion [97].
- CREAM also requires an initial expertise in the field of human factors in order to use the technique successfully and may therefore appear rather complex for an inexperienced user [97].

- CREAM does not put forth potential means by which the identified errors can be reduced [97].
- The time required for application is very lengthy.

Human Cognitive Reliability Correlation (HCR)

Human Cognitive Reliability Correlation (HCR) is a technique used in the field of HRA for the purposes of evaluating the probability of a human error occurring throughout the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

HCR is based on the premise that an operator's likelihood of success or failure in a time-critical task is dependent on the cognitive process used to make the critical decisions that determine the outcome. Three PSFs – operator experience, stress level, and quality of operator/plant interface – also influence the average (median) time taken to perform the task. Combining these factors enables "response-time" curves to be calibrated and compared to the available time to perform the task. Using these curves, the analyst can then estimate the likelihood that an operator will take the correct action, as required by a given stimulus (e.g. pressure warning signal), within the available time window. The relationship between these normalised times and HEPs is based on simulator experimental data.

Background

HCR is a psychology/cognitive modelling approach to HRA developed by Hannaman *et al.* [98] in 1984. The method uses Rasmussen's idea of rule-based, skill-based, and knowledge-based decision making to determine the likelihood of failing a given task [99], as well as considering the PSFs of operator experience, stress and interface quality. The database underpinning this methodology was originally developed through the use of nuclear power-plant simulations due to a requirement for a method by which nuclear operating reliability could be quantified.

HCR Methodology

The HCR methodology is broken down into a sequence of steps as given below:

1. The first step is for the analyst to determine the situation in need of a human reliability assessment. It is then determined whether this situation is governed by rule-based, skill-based or knowledge-based decision making.
2. From the relevant literature, the appropriate HCR mathematical model or graphical curve is then selected.
3. The median response time to perform the task in question is thereafter determined. This is commonly done by expert judgement, operator interview or simulator experiment. In some literature, this time is denoted $T_{1/2}$ and sometimes referred to as the nominal response time.
4. The median $T_{1/2}$ is adjusted to make it specific to the situational context. This is done by means of the PSF coefficients K_1 (operator experience), K_2 (stress level) and K_3 (quality of operator/plant interface) given in the literature and using the following formula:

$$T_{1/2}adjusted = T_{1/2}\,nominal \times (1 + K_1)(1 + K_2)(1 + K_3)$$

Performance improving PSFs (e.g. worker experience, low stress) will take negative values resulting in quicker times, whilst performance inhibiting PSFs (e.g. poor interface) will increase this adjusted median time.

5. For the action being assessed, the time window (T) should then be calculated, which is the time in which the operator must take action to resolve correctly the situation.
6. To obtain the non-response probability, the time window (T) is divided by $T_{1/2}$, the median time. This gives the Normalised Time Value. The probability of non-response can then be found by referring to the HCR curve selected earlier.

This non-response probability may then be integrated into a fuller HRA; a complete HEP can only be reached in conjunction with other methods as non-response is not the sole source of human error.

Worked Example

The following example is taken from Human Factors in Reliability Group [87] in which Hannaman describes analysis of failure to SCRAM manually in a Westinghouse PWR.

<u>Context.</u> The example concerns a model in which failures occurs to SCRAM manually in a Westinghouse PWR. The primary task to be carried out involves inserting control rods into the core. This can be further broken down into two sub-tasks which involve namely detection and action, which are in turn based upon recognising and identifying an automatic trip failure

<u>Assumptions.</u> Given that there exists the assumption that there is simply one option in the procedures and that within training procedures optional actions are disregarded, the likelihood that a reactor trip failure will be incorrectly diagnosed is minimal.

It is also assumed that the behaviour of the operating crew under consideration is skill-based; the reactor trip event which takes place is not part of a routine, however the behaviour adopted by the crew when the event is taking place is nevertheless recognised. Moreover, there are well set procedures which determine how the event should be conducted and these are assumed comprehended and practised to required standards in training sessions.

The average time taken by the crew to complete the task is 25 seconds. The average completion times for the respective subtasks are set as 10 seconds for detection of the failure and 15 seconds for taking subsequent action to remedy the situation.

<u>Method.</u> The PSFs (K factor) judged to influence the situation are assessed to be in the following categories:

- operator experience is “well trained”
- stress level is “potential emergency”
- quality of interface is “good”

The various K factors are assigned the following values:

$$K_1 = 0.0$$

$K_2 = 0.28$

$K_3 = 0.0$

Referring to the equation in Step 4 above, the product is therefore equal to the value of 1.28. In response, the average tasks times are altered from 10 and 15 seconds to 12.8 and 19.2 seconds respectively. Given that the PSFs are identical for both of the given sub-tasks, it is therefore possible to sum the median response times to give a total of 32 seconds, adjusting the figure for stress, compared to a previous total of 25 seconds.

The time window (T) to perform the task as part of the overall system is given as 79 seconds. This time is derived from a study conducted by Westinghouse in which it was discovered that the crew had approximately 79 seconds to complete the task of inserting the control rod to the reactor and then to shut the reactor down in order to inhibit over-pressuring within the main operating system.

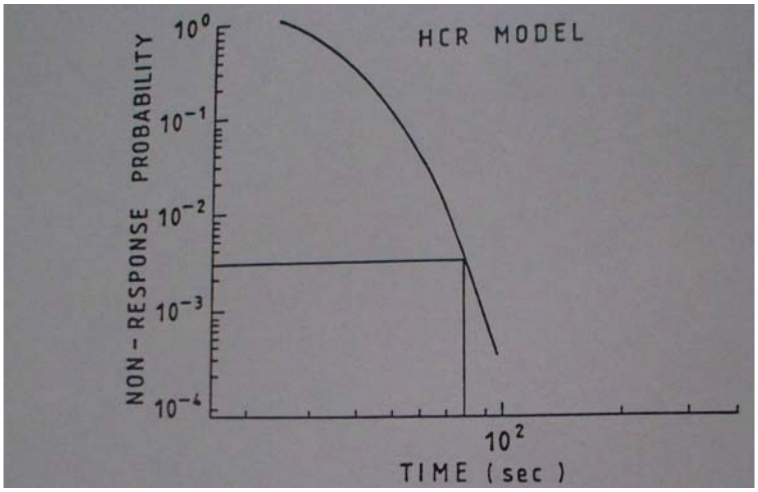
Results/Outcome. Consulting the graphical curve central to the technique, the normalised time for the task can thus be established. It is determined by the division of 79 seconds and 32 seconds, giving a result of 2.47 seconds. Identifying this point on the abscissa (the HCR curve model) provides a non response probability of 2.9 x 10^{-3}; this can also be checked for validation utilising the formula:-

$P_{RT}\,(79) = exp - [\ (79/32) - 0.7 \ /\ 0.407]^{\ 1.2}$

$P_{RT}\,(79) = 2.9 \times 10^{\ -3} / \ demand$

where $P_{RT}\,(T)$ equals the probability of non success within the system time window T.

Provided below is the graphical solution for the assessment using the HCR technique:



Advantages of HCR

- The approach explicitly models the time-dependent nature of HRA [87].
- It is a fairly quick technique to carry out and has a relative ease of use [87].
- The three modes of decision-making, knowledge-based, skill-based and rule-based are all modelled [87].

Disadvantages of HCR

- The HEP produced by HCR is not complete; it calculates the probability that a system operator will fail to diagnose and process information, make a decision and act within the time available. It does not give any regard to misdiagnoses or rule violations [87].
- The same probability curves are used to model non-detection and slow response failures. These are very different processes, and it is unlikely that identical curves could model their behaviour. Furthermore, it is uncertain as to whether such curves could be applied to situations in which detection failures or processing difficulties are the primary dominating factors of influence [87].
- The rules for judging Knowledge-based, Skill-based and Rule-based behaviour are not exhaustive. Assigning the wrong behaviour to a task can mean differences of up to two orders of magnitude in the HEP [87].
- The method is very sensitive to changes in the estimate of the median time. Therefore, this estimate must be very accurate otherwise the estimation in the HEP will suffer as a consequence [87].
- It is highly resource intensive to collect all the required data for the HCR methodology, particularly due to the necessity of evaluation for all new situations which require an assessment [87].
- There is no sense of output from the model that indicates in any way of how human reliability could be adjusted to allow for improvement or optimisation to meet required goals of performance [87].
- Only three PSFs are included in the methodology; there are several other PSFs that could affect performance which are unaccounted for.
- The model is relatively insensitive to PSF changes as opposed to, for example, time parameter changes [87].
- As the HCR correlation was originally developed for use within the nuclear industry, it is not immediately applicable to situations out-with this domain [87].

Human Error Assessment and Reduction Technique (HEART)

Human Error Assessment and Reduction Technique (HEART) is a technique used in the field of HRA for the purposes of evaluating the probability of a human error occurring throughout the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

The HEART method is based upon the principle that every time a task is performed there is a possibility of failure and that the probability of this is affected by one or more EPCs – for instance: distraction, tiredness, cramped conditions etc. – to varying degrees. Factors which have a significant effect on performance are of greatest interest. These conditions can then be applied to a "best-case-scenario" estimate of the failure probability under ideal conditions then to obtain a final error chance. This figure assists in communication of error likelihoods with the wider risk analysis or safety case. By forcing consideration of the EPCs potentially affecting a given procedure, HEART also provides a means of considering the impact of possible risk reduction measures.

Background

HEART was developed by Williams in 1986 [100]. It is a first generation HRA technique, and is still widely used throughout the UK. The method essentially takes into consideration a range of important factors which may negatively affect human performance of a task. Each of these factors is then independently quantified to obtain an overall HEP, depending on each of the factors.

HEART Methodology

1. The first stage of the process is to identify the full range of sub-tasks that a system operator would be required to complete within a given task.
2. Once this task description has been constructed a nominal human unreliability score for the particular task is then determined, usually by consulting local experts. Based around this calculated point, a 5th – 95th percentile confidence range is established.
3. The EPCs, which are potentially relevant for the given situation, are then considered and the extent to which each EPC applies to the task in question is discussed and agreed, again with local experts.
4. A final estimate of the HEP is then calculated using the EPC scores.

Worked Example [91]

Context. A reliability engineer has the task of assessing the probability of a plant operator failing to carry out the task of isolating a plant bypass route as required by procedure. However, the operator is fairly inexperienced in fulfilling this task and therefore does not always follow the correct procedure; the individual is therefore unaware of the hazards created when the task is carried out incorrectly.

Assumptions. There are various assumptions that should be considered in the context of the situation:

- the operator is working a shift in which he is in his 7th hour.
- there is talk circulating the plant that it is due to close down
- it is possible for the operator's work to be checked at any time
- local management aim to keep the plant open despite a desperate need for re-vamping and maintenance work; if the plant is closed down for a short period, if the problems are unattended, there is a risk that it may remain closed permanently.

Method. A representation of this situation using the HEART methodology would be done as follows:

- From the relevant tables it can be established that the type of task in this situation is of the type (F) which is defined as 'Restore or shift a system to original or new state following procedures, with some checking'. This task type has the proposed nominal human unreliability value of 0.03.
- Other factors to be included in the calculation are provided in the table below:

FACTOR	TOTAL HEART EFFECT	ASSESSED PROPORTION OF EFFECT	ASSESSED EFFECT
Inexperience	× 3	0.4	(3.0-1) × 0.4 + 1 = 1.8
Opposite technique	× 6	1.0	(6.0-1) × 1.0 + 1 = 6.0
Risk Misperception	× 4	0.8	(4.0-1) × 0.8 + 1 = 3.4
Conflict of Objectives	× 2.5	0.8	(2.5-1) × 0.8 + 1 = 2.2
Low Morale	× 1.2	0.6	(1.2-1) × 0.6 + 1 = 1.12

Result. The final calculation for the normal likelihood of failure can therefore be formulated as:

$$0.003 \times 1.8 \times 6.0 \times 3.4 \times 2.2 \times 1.12 = 0.27$$

Advantages of HEART

- HEART is very quick and straightforward to use [87].
- The technique provides the user with useful suggestions as to how to reduce the occurrence of errors[12].
- It provides ready linkage between Ergonomics and Process Design, with reliability improvement measures being a direct conclusion which can be drawn from the assessment procedure.

[12] http://www2.hf.faa.gov/workbenchtools/default.aspx?rPage=Tooldetails&toolID=110

- It is highly flexible and applicable in a wide-range of areas which contributes to the popularity of its use [87].

Disadvantages of HEART

- The main criticism of the HEART technique is that the EPC data has never been fully released and it is therefore not possible to review fully the validity of Williams EPC data base. Kirwan has dome some empirical validation on HEART and found that it had “a reasonable level of accuracy” but was not necessarily better or worse than the other techniques in the study [27, 83, 84]. Further theoretical validation is thus required [91].
- HEART relies to a high extent on expert opinion, first in the point probabilities of human error, and also in the assessed proportion of EPC effect. The final HEPs are therefore sensitive to both optimistic and pessimistic assessors
- The interdependence of EPCs is not modelled in this methodology, with the HEPs being multiplied directly. This assumption of independence does not necessarily hold in a real situation [91].

Influence Diagrams Approach (IDA)

Influence Diagrams Approach (IDA) is a technique used in the field of HRA for the purpose of evaluating the probability of a human error occurring during the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

An Influence Diagram (ID) is essentially a graphical representation of the probabilistic dependencies between PSFs, the factors which influence probability of failure in the performance of a task. The approach originates from the field of decision analysis and uses expert judgement to formulate and often to quantify the models. The role of the ID is both to depict these influences and the nature of the interrelationships in a comprehensible format, and to provide a method of calculating failure probability. In this way, the diagram may be used to represent the shared beliefs of a group of experts on the outcome of a particular human action and the factors that may or may not influence that outcome. For each of the identified influences quantitative values are calculated, which are then used to derive final HEP estimates.

Background

IDA is a decision analysis based framework which is developed through eliciting expert judgement, usually in group workshops. Unlike other first generation HRA, IDA explicitly considers the inter-dependency of operator and organisational PSFs. The IDA approach was first outlined by Howard and Matheson [101, 102], and then developed specifically for the nuclear industry by Phillips *et al* [103].

IDA Methodology

The IDA methodology is conducted in a series of 10 steps as follows:

1. Describe all relevant conditioning events

   Experts who posses sufficient knowledge of the situation under evaluation form a group. The chosen individuals include a range of experts - typically those with first hand experience in the operational context under consideration – such as plant supervisors, reliability assessors, human factor specialists and designers. The group collectively assesses and gradually develops a representation of the most significant influences which will affect the success of the situation. The resultant diagram is useful in that it identifies both immediate and implied influences of the considered factors with regards their effect on the situation under assessment and upon one another.

2. Refine the target event definition

   The event which is the basis of the assessment must be defined as tightly as possible.

3. Balance of Evidence

The next stage is to select a middle-level event in the situation and using each of the bottom level influences, assess the weight of evidence, also known as the 'balance of evidence'; this represents expert analysis of the likelihood that a specific state of influence or combination of the various influences is existent within the considered situation.

4. Assess the weight of evidence for this middle-level influence, which is conditional on bottom-level influences

5. Repeat 3 and 4 for the remaining middle-level and bottom-level influences

   These three steps are conducted in the aim of determining the extent to which the influences exist in the process, alone and in different combinations, and their conditional effects.

6. Assess probabilities of target event conditional on middle-level influences

7. Calculate the unconditional probability of target event and unconditional weight of evidence of middle-level influences. For the various combinations of influences that have been considered, the experts identify direct estimates of the likelihood of either success or failure.

8. Compare these results to the holistic judgements of HEPs by the assessors. Revise if necessary to reduce discrepancies.

   At this stage the probabilities derived from the use of the technique are compared to holistic estimates from the experts, which have been derived through an APJ process. Discrepancies are discussed and resolved within the group as required.

9. Repeat above steps until assessors are finished refining their judgements

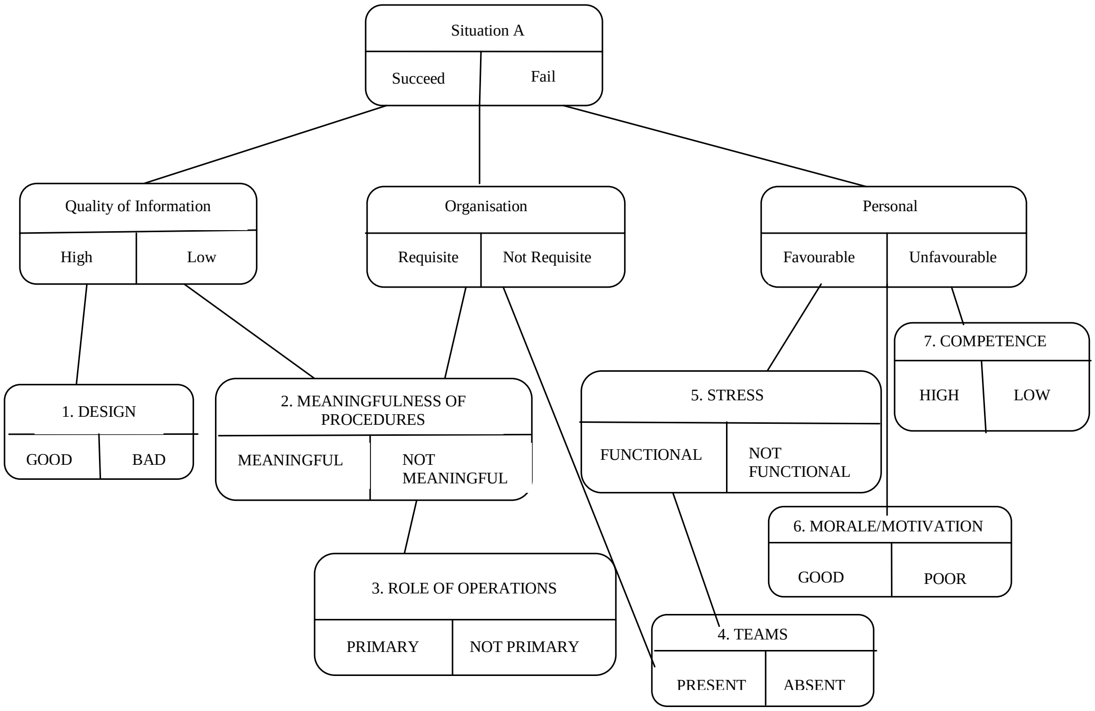
   The above steps are iterated, in which all experts share opinions, highlight new aspects to the problem and revise the initially made assessments of the situation. The process is deemed complete when all participants reach a consensus that any misgivings about the discrepancies are resolved.

10. Perform sensitivity analyses

    If individual experts remain to be unsure of the discrepancies about the assessments which have been made, then sensitivity analysis can be used to determine the extent to which individual influence assessments affect the target event HEP. Conducting a cost-benefit analysis is also possible at this stage of the process.

Example

The diagram below depicts an influence diagram which can be applied to any human reliability assessment [87].



This diagram was originally developed for use in the HRA of a scenario within the setting of a nuclear power situation. The diagram depicts the direct influences of each of the factors on the situation under consideration as well as providing as indication as to the way in which some of the factors affect each other. There are 7 first level influences on the outcome of the high level task, numbered 1 to 7. Each of these describes an aspect of the task under assessment, which is modelled as being in one of two states.

- The design of the task is judged to be either good or bad
- The meaningfulness of the procedures involved in the completion of the task are simply meaningful or not meaningful
- Operators either possess a role in the task that is or is not of primary importance
- For the purposes of completing the considered task, they may or not be a formation of teams of individuals
- The stress levels associated with the task can affect performance and render individuals either functional or not functional
- The surrounding work ethic and environment in which the task takes place will provide either a good level of morale or a poor motivation level
- Competence of the individuals who are responsible for carrying out the task is either of a high level or a low level

Differing combinations of these first level influences affect the state of those on the second level.

- The quality of information, which can either be classed as good or bad, is dependent upon the meaningfulness of the procedures of the task and the task design.

- The organisation, whether it is assessed as either requisite or not requisite, is determined by the role of operations functions in completing the task, the meaningfulness of the procedures and whether or not teams are formed to complete the task
- The personal aspect of the task can be judged as either favourable for successful completion or unfavourable. The way in which this is assessed is dependent on competence level of the concerned individuals, stress levels present, morale/motivation levels of the individuals and whether or not teams are formed to complete the task.

By assessing the state of the second level influences, the quality of information, organisation and personal factors, the overall likelihood of either success or failure of the task can be calculated by means of conditional probability calculations.

Advantages of IDA

- Dependence between PSFs is explicitly acknowledged and modelled [87]
- It can be used at any task "level", i.e. it can be used in a strategic overview or in a very fine breakdown of a task element [87].
- Data requirements are low [87].
- PSFs are precisely defined and their influence is explored in depth [87].
- PSFs and other influence creating error producing conditions are prioritised and if desired, the less significant ones may be ignored
- Sensitivity analysis is possible with use of this technique [87].
- It is possible to generate high amounts of qualitative data through the group discussion process

Disadvantages of IDA

- Building IDAs is highly resource-intensive in terms of organising and supporting an extensive group session involving a suitable range of experts [87].
- Eliciting HEPs requires further research with regards to their accuracy and justification [87].

Success Likelihood Index Method (SLIM)

Success Likelihood Index Method (SLIM) is a technique used in the HRA field for the purpose of evaluating the probability of a human error occurring during the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

SLIM is a decision-analytic approach to HRA which uses expert judgement to quantify PSFs; factors concerning the individuals, environment or task, which have the potential to either positively or negatively affect performance e.g. available task time. Such factors are used to derive a Success Likelihood Index (SLI), a form of preference index, which is calibrated against existing data to derive a final HEP. Significant PSFs for the context under study are selected by experts.

The technique consists of two modules: Multi-Attribute Utility Decomposition (MAUD) which scales the relative success likelihood in performing a range of tasks; and Systematic Approach to the Reliability Assessment of Humans (SARAH) which calibrates these success scores with tasks with known HEP values to provide an overall figure.

Background

SLIM was developed by Embrey *et al* [104] for use within the US nuclear industry. By use of this method, relative success likelihoods are established for a range of tasks, and then calibrated using a logarithmic transformation.

SLIM Methodology

The SLIM methodology breaks down into ten steps of which steps 1-7 are involved in SLIM-MAUD and 8-10 are SLIM-SARAH.

1. Definition of situations and subsets

   Upon selection of a relevant panel of experts who will carry out the assessment, these individuals are provided with as fully detailed a task description as possible, the group of individuals designated to perform each task and further factors which are likely to influence the success of each of these. An in-depth description is a critical aspect of the procedure in order to ensure that all members of the assessment group share a common understanding of the given task. This may be further advanced through a group discussion prior to the commencement of the panel session to ascertain of consensus. Following this discussion, the tasks under consideration are then classified depending upon the homogeneity of the PSFs that have an effect on each. Subsets are thus defined by those tasks which have common specific PSFs and possibly also by their weighting within a certain sub-group; this weighting is only an approximation at this stage of the process.

2. Elicitation of PSFs

   Random sets of 3 tasks are presented to experts from which they are required to compare one against the other two and subsequently identify an aspect in which the highlighted task differs from the remaining two; this dissimilarity should be a characteristic which affects the probability of successful task completion. The experts are then asked to highlight the low and high end-points of the identified PSF, i.e. the optimality of the PSF in the context of the given task. For example the PSF may be Time Pressure and therefore the end points of the scale would perhaps be "High level of pressure" to "Low level of pressure". Other possible PSFs may be stress levels, task complexity or degree of teamwork required. The purpose of this stage is to identify those PSFs which are most prevalent in affecting the tasks as opposed to eliciting all the possible influencing factors.

3. Rating the Tasks on the PSFs

   The endpoints of each individual PSF, as identified by the expert, are then assigned the values 1 and 9 on a linear scale. Using this scale, the expert is required to assign to each task a rating, between the two end points, which accurately reflects, using their judgement, the conditions occurring in the task in question. It is optimal to consider each factor in turn so that the judgements made are independent from the influence of other factors which otherwise may affect opinion.

4. Ideal Point Elicitation and Scaling Calculations

   The "ideal" rating for each PSF is then selected on the scale constructed. The ideal is the point at which the PSF least degrades performance – for instance both low and high time pressure may contribute to increasing the chance of failure. The MAUD software then rescales all other ratings made on the scale in terms of their distance from this ideal point, with the closest being assigned as a 1 and the furthest from this point as a 0. This is done for all PSFs until the experts are agreed that the list of PSFs is exhausted and that all the scale positions identified are correctly positioned.

5. Independence Checks

   Using the figures which represent the relative importance of each task and their rating on the relevant scale, these are multiplied to produce a SLI figure for each task. To improve the validity of the process it is necessary to confirm that each of the scales in use are independent to ensure no overlap or double counting in the overall calculation of the index.

   To help carry out this validation task, MAUD software checks for correlations between the experts' scoring on the different scales; if the scale ratings indicate a high correlation, the experts are consulted to reveal whether they agree in their meanings of the ratings on the two scales which are showing similarities. If this situation occurs, the experts are asked to define a new scale which will be a combination of the meaning of the two individually correlated scales. If the correlation is not significant then the scales are treated as independent; in this case, the concerned facilitator is required to make an informed decision as to whether or not the PSFs showing similarities are actually similar and should therefore ensure that a strong justification is explainable for the final decision.

6. Weighting Procedure

   This stage of the process concentrates on eliciting the emphasis required to be reflected in the weights on each of the PSFs in terms of the influence on the success of a task. This is done by enquiring, with the experts, the likelihood of success between pairs of tasks while considering two previously identified PSFs. By noting where the experts' opinion is changed, the weighting of the effect of each PSF on the task success can thus be inferred. To enhance the accuracy of the outcome, this stage should be carried out in an iterative manner.

7. Calculation of the SLI

   The SLI for each task is deduced using the following formula:

   $$SLI_j = \sum_{i=1}^{x} (R_{ij}W_i)$$

   where

   - SLI_j is the SLI for task j
   - W_i is the importance weight for the i^{th} PSF
   - R_{ij} is the scaled rating of task j on the i^{th} PSF
   - x represents the number of PSFs considered.

   These SLIs are estimates of the probability with which different types of error may occur.

8. Conversion of SLIs to probabilities

   The SLIs previously calculated require to be transformed to HEPs as they are only relative measures of the likelihood of success of each of the considered tasks.

   The relationship

   $$Log\ P = a\ SLI + b$$

   is assumed to exist between SLIs and HEPs. P is the probability of success and a and b are constants; a and b are calculated from the SLIs of two tasks where the HEP has already been established.

9. Uncertainty Bound Analysis

   Uncertainty bounds can be estimated using expert judgement methods such as Absolute Probability Judgement (APJ).

10. Use of SLIM-SARAH for Cost-Effectiveness Analyses

   As SLIM evaluates HEPs as a function of the PSFs, considered to be the major drivers in human reliability, it is possible to perform sensitivity analysis by modifying the scores of the PSFs. By considering the PSFs which may be altered, the degree to which they can be changed and the importance of the PSFs, it is possible to conduct a cost-benefit analysis to determine how worthwhile suggested improvements may be i.e. what-if analysis, the optimal means by which the calculated HEPs can be reduced.

Worked Example

The following example, based on [105], provides a good illustration of how the SLIM methodology is used in practice in the field of HRA.

Context. In this context an operator is responsible for the task of de-coupling a filling hose from a chemical road tanker. There exists the possibility that the operator may forget to close a valve located upstream of the filling hose, which is a crucial part of the procedure; if overlooked, this could result in adverse consequences, of greater effect to the operator in control. The primary human error of concern in this situation is 'failure to close V0204 prior to decoupling filling hose'. The decoupling operation required to be conducted is a fairly easy task to carry out and does not require to be completed in conjunction with any further tasks; therefore is failure occurs it will have a catastrophic impact as opposed to displaying effects in a gradual manner.

Required Inputs. This technique also requires an 'expert panel' to carry out the HRA; the panel would be made up of for example two operators possessing approximately 10 years experience of the system, a human factors analyst and a reliability analyst who has knowledge of the system and possesses a degree of experience of operation.

The panel of experts is requested to determine a set of PSFs which are applicable to the task in question within the context of the wider system; of these, the experts are then required to propose those PSFs, of the identified, which are the most important in the circumstances of the scenario.

For this example, it is assumed that the panel put forth 5 main PSFs for consideration, which are believed to have the greatest effect on human performance of the task: training, procedures, feedback, perceived risk and time pressure.

Method. PSF rating. Considering the situation within the context of the task under assessment, the panel are asked to provide further possible human errors which may occur that have the potential of affecting performance e.g. mis-setting or ignoring an alarm. For each of these, the experts are required to establish the degree to which each is either optimal or sub-optimal for the task under assessment, working on a scale from 1 to 9, with the latter being the optimal rating. For the 3 human errors which have been identified, the ratings decided for each are provided below:

Errors	**training**	**procedures**	**feedback**	**perceived risk**	**time**
VO2O4 open	6	5	2	9	6
Alarm mis-set	5	3	2	7	4
Alarm ignored	4	5	7	7	2

PSF weighting. Were each of the identified human errors of equal importance, it would then be possible to obtain the summation of each row of ratings and come to the conclusion that the row with the lowest total rating- in this case it would be alarm mis-set- was the most probable to occur. In this context, as is most often the case, the experts are in agreement that the PSFs given above are not of equal weighting. Perceived risk and feedback are deemed to be of greatest importance, twice as much as training and procedures, which are considered to be one and a half times more important than the factor of time. The time factor is of considered of minimal importance in this context as the task is routine and is therefore not limited by time.

The importance of each factor can be observed through the allocated weighting, as provided below. Note that they have been normalised to sum to unity.

PSF	Importance
Perceived risk	0.30
Feedback	0.30
Training	0.15
Procedures	0.15
Time	0.10
SUM	1.00

Using the figures for the scaled weighting of the PSFs and the weighting of their importance, it is now possible to calculate the SLI for the task under assessment.

Weighting	PSFs	V0101	Alarm mis-set	Alarm ignored
0.30	Feedback	0.60	0.60	2.10
0.30	Perc'd Risk	2.70	2.10	2.10
0.15	Training	0.90	0.75	0.60
0.15	Procedures	0.75	0.45	0.75
0.10	Time	0.60	0.40	0.20
	SLI (total)	5.55	4.30	5.75

From the results of the calculations, as the SLI for 'alarm mis-set' is the lowest, this suggests that this is the most probable error to occur throughout the completion of the task.

However these SLI figures are not yet in the form of probabilities; they are only indications as to the likelihood by which the various errors may occur. The SLIs determine the order in which the errors are most probable to occur; they do not delineate the absolute probabilities of the PSFs. To convert the SLIs to HEPs, the SLI figures require to first be standardised; this can be done using the following formulation.

$$Log_{10}(HEP) = a.SLI + b$$

Result. If the two tasks for which the HEPs are known are incorporated in the task set which is undergoing quantification then the equation parameters can be determined by using the method of simultaneous equations; using the result of this the unknown HEP values can thus be quantified. In the example provided, were two additional tasks to be assessed e.g. A and B, which had HEP values of 0.5 and 10^{-4} respectively and SLIs respectively of 4.00 and 6.00, respectively, then the formulation would be:

$$Log(HEP) = -1.85SLI + 7.1$$

The final HEP values would thus be determined as

V0204 = 0.0007

Alarm mis-set = 0.14

Alarm ignored = 0.0003

Advantages of SLIM

- The method has a sound basis in decision theory and a reasonably high level of theoretical validity [87].

- The technique is highly visible and auditable and is also sophisticated and well developed [87].
- Sensitivity analysis is relatively straightforward to execute
- SLIM can be used to evaluate HEPs for discrete tasks as well as at a higher, more holistic level.
- It is useful in allowing comprehensive cost benefit evaluations to be carried out [87].

Disadvantages of SLIM

- Extensive use of expert judgement is required [87].
- The method by which PSFs are selected is somewhat subjective and is considered to be unsuitable as the question is posed as to the means by which the PSFs suggested by the experts are judged to be reliable; why are the judges expected to have in depth knowledge of the affecting factors and details surrounding these [91].
- SLIM is relatively resource-intensive to carry out compared to some other HRA methods.
- The validity of the logarithmic transformation has not been established i.e. it requires some empirical and experimental justification.
- The absolute HEPs depend on the two calibration events used to establish the linear scaling parameters. Hence errors in these two HEPs will induce systematic errors in the other calculated HEPs.

Tecnica Empirica Stima Errori Operatori (TESEO)

Tecnica Empirica Stima Errori Operatori (TESEO) is a technique used in the field of HRA for the purposes of evaluating the probability of a human error occurring throughout the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

This is a time based model which describes the probability of a system operator's failure as a multiplicative function of 5 main factors. These factors are as follows:

1. K_1: The type of task to be executed
2. K_2: The time available to the operator to complete the task
3. K_3: The operator's level of experience/characteristics
4. K_4: The operator's state of mind
5. K_5: The environmental and ergonomic conditions prevalent

Using these figures, an overall HEP can be calculated with the formulation provided below:

$$K_1 \times K_2 \times K_3 \times K_4 \times K_5$$

The specific value of each of the above functions can be obtained by consulting standard tables that take account of the method in which the HEP is derived.

Background

Developed in 1980 by Bello and Columbari [106], TESEO created with the intention of using it for the purpose of conducting HRA of process industries. The methodology is relatively straightforward and is easy to use but is also limited; it is useful for quick overview HRA assessments as opposed to those which are highly detailed and in-depth. Within the field of HRA, there is a lack of theoretical foundation of the technique as is widely acknowledged throughout.

TESEO Methodology

When putting this technique into practice, it is necessary for the designated HRA assessor to thoroughly consider the task requiring assessment and therefore also consider the value for K_n that applies in the context. Once this value has been decided upon, the tables, previously mentioned, are then consulted from which a related value for each of the identified factors is found in order to allow the HEP to be calculated.

Worked Example

Provided below is an example of how TESEO methodology can be used in practice; each of the stages of the process described above are worked through in order.

Context. An operator works on a production transfer line which operates between two tanks. His role is to ensure the correct product is selected for transfer from one tanker to the other; this can be done by operation of the relevant valves which are located remotely. The essential valves must be opened to allow the task to be carried out. The operator possesses average experience in fulfilling this role. The individual is situated in a control room which has a relatively noisy environment and poor lighting. There is a time window of 5 minutes to carry out the required task.

Method. The figures for the HEP calculation, obtained from the relevant tables, are given as follows:

- The type of task to be executed: $K_1 = 0.01$
- Time available to complete the task: $K_2 = 0.5$
- Level of experience: $K_3 = 1$
- Operator's state of mind: $K_4 = 1$
- Environmental and ergonomic conditions: $K_5 = 10$

The calculation for the final HEP figure is therefore calculated as:

$$K_1 \times K_2 \times K_3 \times K_4 \times K_5$$

$$= 0.01 \times 0.5 \times 1 \times 1 \times 10$$

$$= 0.05$$

Result. Given the result of this calculation, it can be deduced that were the control room notified of the valves' positions and if the microclimate was better, K_5 would be unity, and therefore the HEP would be 0.005, representing an improvement of 1 order of magnitude.

Advantages of TESEO

- The technique of TESEO is typically quick and straightforward in comparison to other HRA tools, not only in producing a final result, but also in sensitivity analysis e.g. it is useful in identifying the effects improvements in human factors will have on the overall human reliability of a task. It is widely applicable to various control room designs or with procedures with varying characteristics [87].

Disadvantages of TESEO

- There is limited work published with regards to the theoretical foundations of this technique, in particular relating to the justification of the five factor methodology [87]. Regardless of the situation, it remains to be assumed that these 5 factors are suffice for an accurate assessment of human performance; as no other factors are considered, this suggests that to solely use these 5 factors to adequately describe the full range of error producing conditions fails to be highly realistic. Further to this, the values of $K_{1\text{-}5}$ are unsubstantiated and the suggested multiplicative relationship has no sufficient theoretical or empirical evidence for justification purposes.

Technique for Human Error Rate Prediction (THERP)

Technique for Human Error Rate Prediction (THERP) is a technique used in the field of HRA for the purposes of evaluating the probability of a human error occurring throughout the completion of a specific task. From such analyses measures can then be taken to reduce the likelihood of errors occurring within a system and therefore lead to an improvement in the overall levels of safety. There exist three primary reasons for conducting an HRA; error identification, error quantification and error reduction. HRA techniques have been utilised in a range of industries including healthcare, engineering, nuclear, transportation and business sector; each technique has varying uses within different disciplines.

THERP models HEPs using an event-tree approach in a similar way to an engineering risk assessment, but also accounts for PSFs that may influence these probabilities. The probabilities for the human reliability analysis event tree (HRAET), which is the primary tool for assessment, are calculated from the database developed by the authors Swain and Guttman [107]. The resultant tree portrays a step by step account of the stages involved in a task, in a logical order. The technique is known as a total methodology as it simultaneously manages a number of different activities including task analysis, error identification, represented in form of HRAET and HEP quantification.

Background

The Technique for Human Error Rate Prediction (THERP) is a first generation methodology which means that its procedures follow the way conventional reliability analysis models a machine [26]. The technique was developed in the Sandia Laboratories for the US Nuclear Regulatory Commission [107]. Its primary author is Swain, who developed the THERP methodology gradually over a lengthy period of time [91]. THERP relies on a large human reliability database containing HEPs which is based upon both plant data and expert judgements. The technique was the first approach in HRA to come into broad use and is still widely used in a range of applications even beyond its original nuclear setting.

THERP Methodology

The methodology for the THERP technique is broken down into 5 main stages:

1. Define the system failures of interest

   These failures include functions of the system in which human error has a greater likelihood of influencing the probability of a fault, and those which are of interest to the risk assessor; operations in which there may be no interest include those which are not operationally critical or those for which there already exist safety counter measures.

2. List and analyse the related human operations, and identify human errors that can occur and relevant human error recovery modes

   This stage of the process necessitates a comprehensive task and human error analysis. The task analysis lists and sequences the discrete elements and information required by task operators. For each step of the task, possible occurring errors which may transpire are considered by the analyst and precisely defined. The possible errors are then considered by the analyst, for each task step. The opportunity for error recovery must also be considered as this, if achieved, has the potential to reduce drastically error probability for a task.

The tasks and associated outcomes are input to an HRAET in order to provide a graphical representation of a task's procedure. The trees' compatibility with conventional event-tree methodology i.e. including binary decision points at the end of each node, allows it to be evaluated mathematically.

An event tree visually displays all events which occur within a system. It starts off with what is known as an initiating event, and then branches are developed as various intermediate events are added. The event tree thus shows a number of different paths each of which has an associated end state or consequence.

3. Assess the relevant error probabilities

   HEPs for each sub-task are entered into the tree; it is necessary for all failure branches to have a probability otherwise the system will fail to provide a final answer. HRAETs provide the function of breaking down the primary operator tasks into finer steps which are represented in the form of successes and failures. This tree indicates the order in which the events occur and also considers likely failures that may occur at each of the represented branches. The degree to which each high level task is broken down into lower level tasks is dependent on the availability of HEPs for the successive individual branches. The HEPs may be derived from a range of sources such as: the THERP database; simulation data; historical accident data; expert judgement. PSFs should be incorporated into these HEP calculations; the primary source of guidance for this is the THERP handbook. However the analyst must use their own discretion when deciding the extent to which each of the factors applies to the task

4. Estimate the effects of human error on the system failure events

   With the completion of the HRA the human contribution to failure can then be assessed in comparison with the results of the overall reliability analysis. This can be completed by inserting the HEPs into the full system's fault event tree which allows human factors to be considered within the context of the full system.

5. Recommend changes to the system and recalculate the system failure probabilities

   Once the human factor contribution is known, sensitivity analysis can be used to identify how certain risks may be improved in the reduction of HEPs. Error recovery paths may be incorporated into the event tree as this will aid the assessor when considering the possible approaches by which the identified errors can be reduced.

Worked Example

Context. The following example illustrates how the THERP methodology can be used in practice in the calculation of HEPs. It is used to determine the HEP for establishing air based ventilation using emergency purge ventilation equipment on In-Tank Precipitation (ITP) processing tanks 48 and 49 after failure of the nitrogen purge system following a seismic event.

Assumptions. In order for the final HEP calculation to be valid, the following assumptions are required:

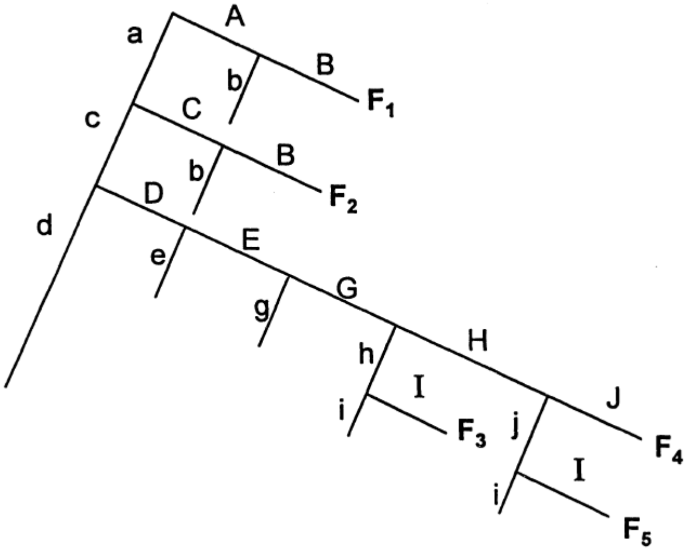
1. There exists a seismic event initiator which leads to the establishment of air based ventilation on the ITP processing tanks 48 and 49
2. It is assumed that both on and offsite power is unavailable within the context and therefore control actions which are performed by the operator are done so locally, on the tank top
3. The time available for operations personnel to establish air based ventilation by use of the emergency purge ventilation, following the occurrence of the seismic event, is a duration of 3 days
4. There is a necessity for an ITP equipment status monitoring procedure to be developed to allow for a consistent method to be adopted for the purposes of evaluating the ITP equipment and component status and selected process parameters for the period of an accident condition
5. Assumed response times exist for initial diagnosis of the event and for the placement of emergency purge ventilation equipment on the tank top. The former is 10 hours while the latter is 4 hours.
6. The In-Tank Precipitation Process has associated Operational Safety Requirements (OSR) which will identify the precise conditions under which the emergency purge ventilation equipment should be hooked up to the riser
7. The "Tank 48 System" Standard Operating Procedure has certain conditions and actions which must be included for correct completion to be performed.
8. A vital component of the emergency purge ventilation equipment unit is a flow indicator; this is required in the event of the emergency purge ventilation equipment being hooked up incorrectly as it would allow for a recovery action
9. The personnel available to perform the necessary tasks all possess the required skills
10. Throughout the installation of the emergency purge ventilation equipment, carried out by maintenance personnel, a tank operator must be present to monitor this process.

Method. An initial task analysis was carried out on the off normal procedure and standard operating procedure. This allowed for the operator to align and then initiate the emergency purge ventilation equipment given the loss of the ventilation system.

Thereafter, each individual task was analysed from which it was then possible to assign error probabilities and error factors to events which represented operator responses.

- A number of the HEPs were adjusted to take account of various PSFs which had been identified
- Upon assessment of characteristics of the task and behaviour of the crew, recovery probabilities were deciphered. Such probabilities are influenced by such factors as task familiarity, alarms and independent checking
- Once error probabilities were decided upon for the individual tasks, event trees were then constructed from which calculation formulations were derived. The probability

of failure was obtained through the multiplication of each of the failure probabilities along the path under consideration.



HRA event tree for align and start emergency purge ventilation equipment on In-Tank Precipitation Tank 48 or 49 after a seismic event

The summation of each of the failure path probabilities provided the total failure path probability (FT)

Results.

Task A: Diagnosis, HEP 6.0E-4 EF=30

Task B: Visual Inspection performed shiftly, recovery factor HEP=0.001 EF=3

Task C: Initiate standard operating procedure HEP= .003 EF=3

Task D: Maintainer hook-up emergency purge ventilation equipment HEP=.003 EF=3

Task E: Maintainer 2 hook-up emergency purge, recovery factor CHEP=0.5 EF=2

Task G: Tank operator instructing /verifying hook-up, recovery factor CHEP=0.5 Lower bound = .015 Upper bound = 0.15

Task H: Read Flow Indicator, Recovery Factor CHEP = .15 Lower bound = .04 Upper bound = .5

Task I: Diagnosis HEP= 1.0E-5 EF=30

Task J: Analyse LFL Using portable LFL Analyser, Recovery Factor CHEP= 0.5 Lower bound = .015 Upper bound =.15

From the various figures and workings, it can be determined that the HEP for establishing air based ventilation using the emergency purge ventilation equipment on In-tank Precipitation processing tanks 48 and 49 after a failure of the nitrogen purge system following a seismic event is 4.2 E-6. This numerical value is judged to be a median value on the lognormal scale. However, it should be noted that this result is only valid given that all the previously stated assumptions are implemented.

Advantages of THERP

- It is possible to use THERP at all stages of design. Furthermore THERP is not restricted to the assessment of designs already in place and due to the level of detail in the analysis it can be specifically tailored to the requirements of a particular assessment [87].
- THERP is compatible with PRA; the methodology of the technique means that it can be readily integrated with fault tree reliability methodologies [87].
- The THERP process is transparent, structured and provides a logical review of the human factors considered in a risk assessment; this allows the results to be examined in a straightforward manner and assumptions to be challenged [87].
- The technique can be utilised within a wide range of differing human reliability domains and has a high degree of face validity [87].
- It is a unique methodology in the way that it highlights error recovery and it also quantitatively models a dependency relation between the various actions or errors.

Disadvantages of THERP

- THERP analysis is very resource intensive, and may require a large amount of effort to produce reliable HEP values. This can be controlled by ensuring an accurate assessment of the level of work required in the analysis of each stage [87].
- The technique does not lend itself to system improvement. Compared to some other HRA tools such as HEART, THERP is a relatively unsophisticated tool as the range of PSFs considered is generally low and the underlying psychological causes of errors are not identified.
- With regards to the consistency of the technique, large discrepancies have been found in practice with regards to different analysts assessment of the risk associated with the same tasks. Such discrepancies may have arisen from either the process mapping of the tasks in question or in the estimation of the HEPs associated with each of the tasks through the use of THERP tables compared to, for example, expert judgement or the application of PSFs [27, 83, 84].
- The methodology fails to provide guidance to the assessor in how to model the impact of PSFs and the influence of the situation on the errors being assessed.
- The THERP HRAETs implicitly assume that each sub-task's HEP is independent from all others i.e. the HRAET does not update itself in the event that an operator takes a sub-optimal route through the task path. This is reinforced by the HEP being merely reduced by the chance of recovery from a mistake, rather than by introducing alternative (i.e. sub-optimal) "success" routes into the event-tree, which could allow for Bayesian updating of subsequent HEPs.
- THERP is a "first generation" HRA tool, and in common with other such tools has been criticised for not taking adequate account of context [26].

References

[1] Barlow, R.E. andProschan, F. (1975) *Statistical Theory of Reliability and Life Testing*. Holt, Reinhart and Winston, New York.

[2] Høyland, A. andRausand, M. (1994) *System Reliability Theory*. John Wiley and Sons, New York.

[3] Aven, T. (2003) *Foundation of Risk Analysis: a Knowledge and Decision Oriented Perspective*. John Wiley and Sons, Chichester.

[4] Bedford, T. andCooke, R. (2001) *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, Cambridge.

[5] Melnick, E.L. andEveritt, B.S. (Eds) (2008) *Encyclopedia of Quantitative Risk Analysis and Assessment*. John Wiley and Sons, Chichester.

[6] Reason, J. (1990) The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society of London*. **B327**(1241) 475-484.

[7] Kariuki, S.G. andLowe, K. (2007) Integrating human factors into process analysis. *Reliability Engineering and System Safety*. **92** 1764-1773.

[8] Ren, J., Jenkinson, I., Wang, J., Xu, D.L. and Yang, J.B. (2008) A methodology to model causal relationships in offshore safety assessment focusing on human and organisational factors. *Journal of Safety Research*. **39** 87-100.

[9] Helmreich, R.L. (2000) On error management: lessons from aviation. *British Medical Journal*. **320**(7237) 781–785.

[10] Hollnagel, E. (1993) *Human Reliability Analysis: Context and Control*. Academic Press, London.

[11] Wu, S. andPollard, S.J.T., *Human reliability analysis has a role in preventing drinking water incidents* 2008, School of Applied Science, Cranfield University, Cranfield, MK43 0AL.

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

[13] Perrow, C. (1984) *Normal accidents: living with high-risk technologies*. Basic Books, New York.

[14] Adams, J. (1988) Risk homeostasis and the purpose of safety regulation. Ergonomics. *Ergonomics*. **31**(4) 407 - 428.

[15] Hollnagel, E. (2000) Looking for errors of omission and commission or The Hunting of the Snark revisited. *Reliability Engineering and System Safety*. **68** 135–145.

[16] French, S. andNiculae, C. (2005) Believe in the Model: Mishandle the Emergency. *Journal of Homeland Security and Emergency Management*. **2**(1).

[17] Kolaczkowski, A., Forester, J.A., Lois, E. and Cooper, S., *NUREG-1792: Good practices for implementing human reliability analysis*. 2005, US Nuclear Regulatory Commission: Washington, DC.

[18] United States Nuclear Regulatory Commission, *Reactor safety study: an assessment of the accident risks in US commercial nuclear power plants*. 1975.

[19] Reason, J. (1990) Human error: models and management. *British Medical Journal*. **320**(7237) 768-770.

[20] Reason, J. (1987) Cognitive Aids in Process Environments: Prostheses or Tools? *International Journal of Man-Machine Studies*. **27** 463-470.

[21] Reason, J. (1990) *Human Error*. Cambridge University Press, Cambridge.

[22] Janis, I.L. (1982) *Groupthink: Psychological Studies of Policy Decisions and Fiascos*. Houghton Mifflin, Boston.

[23] Roberts, K.H. (1990) Some characteristics of one type of high reliability organisation. *Organization Science*. **1**(2) 160-176.

[24] Weick, K.E. (1987) Organisational culture as a source of high reliability. *California Management Review*. **29** 112-127.

[25] Doughty, E. (1990) Human reliability analysis - where shouldst thou turn? *Reliability Engineering and System Safety*. **29**(3) 283-299.

[26] Hollnagel, E. (1998) *Cognitive Reliability and Error Analysis Method – CREAM*. Elsevier Science, Oxford.

[27] Kirwan, B. (1996) The validation of three human reliability quantification techniques - THERP, HEART, JHEDI: Part I -- technique descriptions and validation issues. *Applied Ergonomics*. **27**(6) 359-373.

[28] Forester, J.A., Kolaczkowski, A., Lois, E. and Kelly, D., *NUREG-1842: Evaluation of human reliability analysis methods against good practices.* 2006, US Nuclear Regulatory Commission: Washington, DC.

[29] Boring, R.L. (2007) *Dynamic human reliability analysis: benefits and challenges of simulating human performance*. in *European Safety and Reliability Conference (ESREL 2007)*, INL/CON-07-12773, Idaho National Laboratory.

[30] Hollnagel, E. (2000) Looking for errors of omission and commission or The Hunting of the Snark revisited *Reliability Engineering and System Safety*. **68** 135-145.

[31] French, S., Maule, A.J. and Papamichail, K.N. (2008) *Decision Making: Behaviour, Analysis and Support*. Cambridge University Press, Cambridge.

[32] Loewenstein, G., Weber, E.U., Hsee, C.K. and Welch, N. (2001) Risk as feelings. *Psychological Bulletin*. **127**(2) 267-286.

[33] Janis, I.L. andMann, L. (1976) Coping with decision conflict. *American Scientist*. **64**(6) 657-667.

[34] Adams, J. (1985) *Risk and Freedom: The Record of Road Safety Regulation*. Transport Publishing Projects, London.

[35] Lord, R.G. andLevy, P.E. (1994) Moving from cognition to action – a control theory perspective. *Applied Psychology - An International Review (Psychologie appliquee - Revue Internationale)*. **43**(3) 335-398.

[36] Carver, C.S. andScheier, M.F. (1981) *Attention and Self-Regulation: a Control Theory Approach to Human Behavior*. Springer Verlag, New York.

[37] Bargh, J.A. andChartrand, T.L. (1999) The unbearable automaticity of being. *American Psychologist*. **54** 462-479.

[38] Schaufeli, W.B. andBakker, A.B. (2004) Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study. *Journal of Organisational Behavior*. **25** 293-315.

[39] Kahneman, D., Slovic, P. and Tversky, A. (Eds) (1982) *Judgement under Uncertainty*. Cambridge University Press, Cambridge.

[40] Kahneman, D. andTversky, A. (Eds) (2000) *Choices, Values and Frames*. Cambridge University Press, Cambridge.

[41] Bazerman, M. (2006) *Managerial Decision Making*. 6th. John Wiley and Sons, New York.

[42] Goldstein, D.G. andGigerenzer, G. (2002) Models of ecological rationality: the recognition heuristic. *Psychological review*. **109**(1) 75-90.

[43] Weick, K.E. andRoberts, K.H. (1993) Collective mind in organizations: heedful interrelating on flight decks. *Administrative Science Quarterly*. **38** 357-381.

[44] Weick, K.E., Sutcliffe, K.M. and Obstfield, D. (1999) Organizing for high reliability: processes of collective mindfulness. *Research in Organizational Behavior*. **21** 81-123.

[45] La Porte, T.R. (1996) High reliability organizations: unlikely, demanding and at risk. *Journal of Contingencies and Crisis Management*. **4** 60-71.

[46] Roth, E.M., Multer, J. and Raslear, T. (2006) Shared situation awareness as a contributor to high reliability performance in railroad operations. *Organization Studies*. **27** 967-987.

[47] Rochlin, G.I., La Porte, T.R. and Roberts, K.H. (1987) The self-designing high reliability organization: aircraft carrier operations at sea. *Naval War College Review*. **40** 76-90.

[48] Sagan, S.D. (1993) *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton University Press, Princeton, NJ.

[49] La Porte, T.R. andConsolini, P.M. (1991) Working in practice but not in theory: theoretical challenges of 'High Reliability Organizations'. *Journal of Public Administration Research and Theory*. **1** 19-47.

[50] Rijpma, J.A. (1997) Complexity, tight-coupling and reliability: connecting normal accidents theory and high reliability theory. *Journal of Contingencies and Crisis Management,*. **5**(1).

[51] Roberts, K.H., Stout, S.K. and Halpern, J.J. (1994) Decision dynamics in two high reliability organizations. *Management Science*. **40** 614-624.

[52] Schulman, P.R. (1993) Negotiated order of organizational reliability. *Administration and Society*. **25** 356-372.

[53] Sagan, S.D. (1994) Toward a political theory of organizational reliability. *Journal of Contingencies and Crisis Management*. **2** 228-240.

[54] Perrow, C. (1994) The limits of safety: the enhancement of a theory of accidents. *Journal of Contingencies and Crisis Management*. **2** 212-220.

[55] Clarke, L. (1993) Drs Pangloss and Strangelove meet organizational theory: high reliability organizations and nuclear weapons accidents. *Sociological Forum*. **8** 675-689.

[56] Commission on the Three Mile Island Accident, *Report of The President's Commission on the Accident at Three Miles Island*. 1979, US GPO: Washington DC.

[57] International Atomic Energy Agency, *The International Chernobyl Project: Technical Report*. 1991, IAEA: Vienna.

[58] Marples, D.R. (1997) *Nuclear Power in the Former USSR: Historical and Contemporary Perspectives*, in *In Nuclear energy and security in the former Soviet Union*, Marples, D.R. and Young, M.J. (Eds). Westview Press.

[59] Dörner, D. (1996) *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*. Metropolitan Books, New York.

[60] Board of Inquiry, *Fractured pipe with loss of primary containment in the THORP feed clarification cell.* 2005, British Nuclear Fuels Limited.

[61] Rail Accident Investigation Branch (RAIB), *Progress Report: Derailment at Grayrigg, Cumbria, 23 February 2007* 2007, Department of Transport.

[62] MacGillivray, B.H., Hamilton, P.D., Strutt, J.E. and Pollard, S.J.T. (2006) Risk analysis strategies in the water utility sector: an inventory of applications for better and more credible decision making. *Critical Review of Environmental Science and Technology*. **36** 85-139.

[63] Hrudey, S.E., Hrudey, E. and Pollard, S.J.T. (2006) Risk management for assuring safe drinking water. *Environment International*. **32** 948-995.

[64] MacGillivray, B.H., Sharp, J.V., Strutt, J.E., Hamilton, P.D. and Pollard, S.J.T. (2007) Benchmarking risk management within the international water utility sector. Part II: a survey of eight water utilities. *Journal of Risk Research*. **10**(1) 105-123.

[65] Pollard, S.J.T., Strutt, J.E., MacGillivray, Hamilton, P.D. and Hrudey, S.E. (2004) Risk analysis and management in the water utility sector – a review of drivers, tools and techniques. *Transactions of the Institute of Chemical Engineers, Part B Process Safety and Environmental Protection*. **82**(B6) 453-462.

[66] Hrudey, S.E. andHrudey, E. (2004) *Safe Drinking Water — Lessons from Recent Outbreaks in Affluent Nations*. IWA Publishing, London.

[67] Weick, K.E. andSutcliffe, K. (2001) *Managing the Unexpected: Assuring High Performance in an Age of Complexity*. Academy of Management Journal. Jossey Bass, San Francisco.

[68] Pollard, S.J.T., Strutt, J.E., MacGillivray, B.H., Sharp, J.V., Hrudey, S.E. and Hamilton, P.D. (2006) *Risk management capabilities – towards mindfulness for the international water utility sector*, in *Water Contamination Emergencies: Enhancing our Response*, Thompson, K.C. and Gray, J. (Eds). Royal Society of Chemistry Publishing, Cambridge. 70-80.

[69] Snowden, D. (2002) Complex acts of knowing - paradox and descriptive self-awareness. *Journal of Knowledge Management*. **6** 100-111.

[70] Klein, G. (1993) *A recognition primed decision model (RPM) of rapid decision making*, in *Decision Making in Action: Models and Method*, Klein, G. (Ed. Ablex.

[71] Clemen, R.T. andReilly, T. (1996) *Making Hard Decisions with Decision Tools*. 2nd Edition. Duxbury, Thomson Learning, Pacific Grove, CA.

[72] Rosenhead, J. andMingers, J. (Eds) (2001) *Rational Analysis for a Problematic World Revisited*. John Wiley and Sons, Chichester.

[73] Mingers, J. andRosenhead, J. (2004) Problem Structuring Methods in Action. *European Journal of Operational Research*. **152** 530-554.

[74] Franco, A., Shaw, D. and Westcombe, M., *Problem Structuring Methods I*, in *Journal of the Operational Research Society*. 2006. p. 757-878.

[75] Franco, A., Shaw, D. and Westcombe, M., *Problem Structuring Methods II*, in *Journal of the Operational Research Society*. 2007. p. 545- 682.

[76] Pidd, M. (Ed (2004) *Systems Modelling: Theory and Practice*. John Wiley and Sons, Chichester.

[77] Belton, V. andStewart, T.J. (2002) *Multiple Criteria Decision Analysis: an Integrated Approach*. Kluwer Academic Press, Boston.

[78] Niculae, C., *A socio-technical perspective on the use of RODOS in nuclear emergency management*, in *Manchester Business School*. 2005, The University of Manchester.

[79] Bedford, T. andBayley, C., *Sensitivity analysis of the CREAM method for human reliability*. 2008, Department of Management Science, University of Strathclyde: Galsgow, G1 1QE.

[80] Mitchell, G. (1993) *The practice of operational research*. John Wiley and Sons Ltd, Chichester.

[81] French, S. (1984) *Fuzzy decision analysis: some criticisms*, in *Fuzzy Sets and Decision Analysis*, Zimmermann, H.J., Zadeh, L.A. and Gaines, B.R. (Eds). North Holland, Amsterdam.

[82] Cooke, R.M. (1991) *Experts in Uncertainty*. Oxford University Press, Oxford.

[83] Kirwan, B. (1997) The validation of three human reliability quantification techniques - THERP, HEART, JHEDI: Part II - Results of validation exercise*. Applied Ergonomics*. **28**(1) 17-25.

[84] Kirwan, B. (1997) The validation of three human reliability quantification techniques - THERP, HEART, JHEDI: Part III -- practical aspects of the usage of the techniques*. Applied Ergonomics*. **28**(1) 27-39.

[85] Boring, R.L., Hendrickson, S.M.L., Forester, J.A., Tran, T.Q. and Lois, E., *SAND2008-2619: Issues in Benchmarking Human Reliability Analysis Methods: A Literature Review*. 2008, Sandia National Laboratories: Albuquerque.

[86] Perneger, T.V. (2005) The Swiss cheese model of safety incidents: are their holes in the metaphor*. BMC Health Services Research*. **5** 71-77.

[87] Humphreys, P.C., *Human Reliability Assessor's Guide*. 1995, Human Factors in Reliability Group, SRD Association.

[88] Dalkey, N. andHelmer, O. (1963) An experimental application of the Delphi method to the use of experts*. Management Science*. **9**(3) 458-467.

[89] Linstone, H.A. andTuroff, M. (1978) *The Delphi Method: Techniques and Applications*. Addison-Wesley, London.

[90] Eurocontrol Experimental Centre, *Review of Techniques to Support the EATMP Safety Assessment Methodology*. 2004, EuroControl.

[91] Kirwan, B. (1994) *Practical Guide to Human Reliability Assessment*. Taylor and Francis (CRC Press), London.

[92] Barriere, M., Bley, D., Cooper, S., Forester, J., Kolaczkowski, A., Luckas, W., Parry, G., Ramey-Smith, A., Thompson, C., Whitehead, D. and Wreathall, J., *NUREG-1624: Technical basis and implementation guidelines for a technique for human event analysis (ATHEANA)*. 2000, US Nuclear Regulatory Commission.

[93] Forester, J., Bley, D., Cooper, S., Lois, E., Siu, N., Kolaczkowski, A. and Wreathall, J. (2004) Expert elicitation approach for performing ATHEANA quantification. . *. Reliability Engineering and System Safety*. **83** (2) 207-220.

[94] Kim, I.S. (2001) Human reliability analysis design review*. Annals of Nuclear Energy*. **28** 1069-1081.

[95] Forester, J., Ramey-Smith, A., Bley, D., Kolaczkowski, A., Cooper, S. and Wreathall, J., *SAND--98-1928C: Discussion of comments from a peer review of a technique for human event analysis (ATHEANA),* . 1998, Sandia Laboratory

[96] Doughty, E. (1998) Human errors of commission revisited: an evaluation of the ATHEANA approach*. Reliability Engineering and System Safety*. **60**(1) 71-82.

[97] Salmon, P., Stanton, N.A. and Walker, G., *Humans Factors Design Methods Review*. 2003, Defence Technology Centre.

[98] Hannaman, G.W., Spurgin, A.J. and Lukic, Y.D., *Human cognitive reliability model for PRA analysis. Draft Report NUS-4531, EPRI Project RP2170-3*. 1984, Electric Power and Research Institute: Palo Alto, CA.

[99] Rasmussen, J. (1983) Skills, rules, knowledge; signals, signs and symbols and other distinctions in human performance models*. IEEE Transactions on Systems, Man and Cybernetics*. **SMC-13**(3).

[100] Williams, J.C. (1985) *HEART – A proposed method for achieving high reliability in process operation by means of human factors engineering technology*. in *Proceedings of a Symposium on the Achievement of Reliability in Operating Plant, Safety and Reliability Society*. NEC, Birmingham.

[101] Howard, R.A. andMatheson, J.E. (2005) Influence diagrams*. Decision Analysis*. **2**(3) 127-143.

[102] Howard, R.A. andMatheson, J.E. (2005) Influence diagrams retrospective*. Decision Analysis*. **2**(3) 144-147.

[103] Phillips, L.D., Humphreys, P.C., Embrey, D.E. and Selby, D. (1990) *A Socio-technical approach to assessing human reliability*, in *Influence Diagrams, Belief Nets and Decision Analysis*, Oliver, R.M. and Smith, J.Q. (Eds). John Wiley and Sons, Chichester.

[104] Embrey, D.E., Humphreys, P.C., Rosa, E.A., Kirwan, B. and Rea, K., *SLIM-MAUD: An approach to assessing human error probabilities using structured expert judgement. NUREG/CR-3518*. 1984, US Nuclear Regulatory Commission: Washington DC.

[105] Wilson, J.R. andCorlett, E.N. (1995) *Evaluation of Human Work: a Practical Ergonomics Methodology* Taylor and Francis.

[106] Bello, G.C. andColumbari, C. (1980) The human factors in risk analyses of process plants: the control room operator model, TESEO*. Reliability Engineering*. **1** 3-14.

[107] Swain, A.D. andGuttmann, H.E., *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*. 1983, NUREG/CR-1278, USNRC.