Safety Science, Volume 49, Issue 6, July 2011, Pages 753-763

Human Reliability Analysis: a critique and review for managers

Simon French[1]
Manchester Business School,
University of Manchester,
Manchester, M15 6PB

Tim Bedford
Department of Management Science,
University of Strathclyde,
Glasgow, G1 1QE

Simon J.T. Pollard
The Collaborative Centre of Excellence
in Understanding and Managing
Natural and Environmental Risks,
Cranfield University,
Cranfield, MK43 0AL

Emma Soane
Employment Relations and
Organisational Behaviour Group,
Department of Management,
London School of Economics and
Political Science,
London, WC2A2AE

Abstract

In running our increasingly complex business systems, formal risk analyses and risk management techniques are becoming a more important part of a manager's tool-kit. Moreover, it is also becoming apparent that human behaviour is often a root or significant contributing cause of system failure. This latter observation is not novel; for more than 30 years it has been recognised that the role of human operations in safety critical systems is so important that they should be explicitly modelled as part of the risk assessment of plant operations. This has led to the development of a range of methods under the general heading of *human reliability analysis* (HRA) to account for the effects of human error in risk and reliability analysis. The modelling approaches used in HRA, however, tend to be focussed on easily describable sequential, generally low-level tasks, which are not the main source of systemic errors. Moreover, they focus on errors rather than the effects of all forms of human behaviour. In this paper we review and discuss HRA methodologies, arguing that there is a need for considerable further research and development before they meet the needs of modern risk and reliability analyses and are able to provide managers with the guidance they need to manage complex systems safely. We provide some suggestions for how work in this area should develop.

Keywords: Cynefin model of decision contexts; high reliability organisations; human reliability analysis (HRA); management of risk.

---

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

1. Introduction

Complex systems are never 100% reliable: they fail, sometimes catastrophically, more usually reparably. Perrow (1984, 1994) has argued that failures are an inevitable consequence of the increasing complexity of our systems. Whatever the case, inevitable or not, failures undoubtedly occur. Even in systems that appear to be largely technological rather than human, we find that in the majority of cases there is a human element involved. Maybe some erroneous or even malicious behaviour initiates the failure; maybe the human response to some event is insufficient to avoid system failure; or maybe the original design of the system did not anticipate a potential failure or unfavourable operating conditions.

Statistics show human error is implicated in (see also Hollnagel 1993):

- over 90% of failures in the nuclear industry (Reason 1990a), see also (United States Nuclear Regulatory Commission 2002);
- over 80% of failures in the chemical and petro-chemical industries (Kariuki and Lowe 2007);
- over 75% of marine casualties (Ren *et al.* 2008);
- over 70% of aviation accidents (Helmreich 2000);
- over 75% of failures in drinking water distribution and hygiene (Wu et al. 2009).

In addition to highly technological industries, there are other complex systems involving applications of technology in which we include complex mathematical modelling, software nd web-based systems. The growth of service industries with new business models implies an even greater dependence of businesses, organisations and even economies on reliable human interactions. For instance, recently human checks and balances failed to detect 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 sectors, indicating the growing interdependence of many disparate parts of the modern global economy. It also owes a lot to a loss of investors' confidence and trust, both of which inform human behaviour. These data indicate how vulnerable our systems are, even after many years of refinement and improvement; and how important an understanding of human behaviour is if we are to reduce the risk to systems. Another high profile example is the leak in the THORP plant at Sellafield (Thermal Oxide Reprocessing Plant) that was discovered in 2005: see (BNFL, 2005). This relatively modern plant had been designed to a high standard of safety, but information indicating 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 that the system was error-free and hence information that could suggest the contrary was ignored or dismissed.

Managerial processes are critical to successful operation of any complex system; and the quality of management processes depends on their understanding of the import and limitations of the results of analyses that are provided to them. In this article, we examine current and past approaches to human reliability analysis (HRA). We discuss its assumptions, limitations and potential in qualitative terms so that managers can better assess the value of the information that it provides them and so manage risks more effectively. We also suggest that further development of HRA methodologies should take more account of the managerial practices that could be applied to reduce the failures that occur at the interface of human behaviour and technology.

Managers understand human behaviour; good managers understand human behaviour extremely well. To bring out the best in a team one needs to know how each will respond to a request, an instruction, an incentive or a sanction. Yet only the most foolhardy and overconfident of managers would claim that they can predict human behaviour perfectly all the time – or even 95% of the time. The problem is that we often need to design systems with very high reliabilities, many times with overall failure rates of less than 1 in 10 million (i.e. 1 in 10^{-7}). To design and analyse such systems we need a deep understanding of human behaviour in *all* possible circumstances that may arise in their management and operation. And that is the challenge facing HRA. Our current understanding of human behaviour is not sufficiently comprehensive: worse, current HRA methodologies seldom use all the understanding that we do have.

Of course, there is a trivial mathematical answer to this. If we are to achieve an overall system reliability of 10^{-7}, we do not need humans to be perfectly reliable. We simply need to know how reliable they are and then ensure that we arrange and maintain sufficient safety barriers around the system to ensure that overall system failure probabilities are as low as required. Suppose we construct seven independent safety barriers perhaps some involving humans, some purely technological and suppose each has a probability of 1 in 10 of failing, then arranging them (conceptually) in sequence so that the whole system fails if and only if every one of the seven fails gives an overall probability of system failure of

$$\frac{1}{10}\times\frac{1}{10}\times\frac{1}{10}\times\frac{1}{10}\times\frac{1}{10}\times\frac{1}{10}\times\frac{1}{10}=10^{-7}.$$

The problem with this is that there are few barriers that are truly independent, most systems offer opportunities to 'bypass' these barriers. Moreover, human behaviour tends to introduce significant correlations and dependencies which invalidate such calculations, reducing the benefit that each extra

safety barrier brings; such problems with protective redundancy are well known (for example, Sagan 2004). So the simplistic calculation does not apply, and we shall argue that we have yet to develop sufficiently complex mathematical modelling techniques to describe human behaviour adequately for risk and reliability analyses.

In many ways 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 and the resulting numbers too much and fail to recognise the potential for unmodelled and possibly unanticipated behaviours – physical or human – to lead to overall system breakdown (cf. French and Niculae 2005). Broadly there are two ways in which such analyses 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. Such uses require quantitative modelling of human reliability; and overconfidence in these models can lead to overconfidence in the estimated probabilities and poor appreciation of the overall system 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 operated. 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.

These uses are not independent – in designing, licensing and managing a system one inevitably iterates between the two – they do differ, however, fundamentally in philosophy. In summative analysis the world outside the system in question learns from the outcome of an analysis; in formative analysis the world inside the system learns from the process of analysis. In summative analysis the ideal is almost to be able to throw away the process and deal only with the outcome; in formative analysis the ideal is almost to throw away the outcome and draw only from the process. While we believe that HRA has a significant potential to be used more in formative ways; we are concerned at its current ability to fulfil a summative role, providing valid probabilities of sequences of failure events in which human behaviour plays a significant role. We believe that there is scope for considerable overconfidence in the summative power of HRA currently and that management, regulators and society in general need to appreciate this, lest they make poorly founded decisions on regulating, licensing and managing systems.

The four of us were part of a recent UK EPSRC funded multi-disciplinary project *Rethinking Human Reliability Analysis Methodologies* to survey and critique HRA methodologies (Adhikari *et al.* 2008). Our purpose in this paper is to draw out the relevant conclusions from this project for the management

community and, perhaps as well, for our political masters who create the regulatory context in which complex systems have to operate. Overall we believe that current practices in and uses of HRA are insufficient for the complexities of modern society. We argue that the summative outputs of risk and reliability analyses should be taken with the proverbial pinch of salt. But not all our conclusions will be negative. There is much to be gained from the formative use of HRA to shape management practices and culture within organisations and society which can lead to better, safer and less risky operations.

In the next section we briefly survey the historical development underlying concepts of HRA and its role in risk and reliability analyses. We reflect on the widely quoted Swiss Cheese Model (Reason 1990b), which seeks to offer a qualitative understanding of system failure – though we shall argue that it may actually lead to systematic misunderstandings! In Section 3 we turn to modern theories of human behaviour, particularly those related to judgement and decision. A key issue is that HRA focuses on human *errors*, whereas many systems failures may arise not just *despite*, but sometimes *because of* fully appropriate and rational behaviour on the part of those involved. Thus we need a broader understanding of human behaviour than that relating to human error. We also need to recognise that cultural, organisational, social and other contexts influence behaviour, perhaps correlating behaviour across a system, thus invalidating assumptions of independence commonly made in risk and reliability analyses. One of the flaws common to many current HRA methodologies is that they tend to focus on easily describable, sequential, generally low-level operational tasks. Yet the human behaviour that is implicated in many system failures may occur in other quite different contexts, maybe in developing higher level strategy or during the response to an unanticipated initiating failure event. In recent years there have been many studies of organisational forms which seem to be more resilient to system failures than might be expected and we discuss such studies of *high reliability organisations* (HROs) briefly in Section 4. Another flaw common to many current HRA methodologies is the lack of specification of the domain of applicability – hence making it difficult to select appropriate methods for a given problem. Therefore in Section 5, we use Snowden's Cynefin classification of decision contexts (Snowden 2002; Snowden and Boone 2007) to categorise different circumstances in which human behaviour may be involved in system failure. We believe that the use of Cynefin – or a similar categorisation of decision contexts – can help in delineating when different HRA methodologies are appropriate. Moreover, it points to areas in which we lack a really sound, appropriate HRA methodology. Our final two sections draw our discussion to a close, suggesting that:

- by drawing together current understandings from HRA with other domains of knowledge in behavioural, management and organisational theories, we can make better formative use of HRA in designing systems, process and the organisations that run these;

but that:

- the state of the art in quantitative HRA is too poor to make the summative assessments of risk and reliability that our regulators assume, and that society urgently needs to recognise this.

2. HRA methodologies and the Swiss cheese model

Reliability analysis and risk analysis are two subjects with a great deal of overlap (Aven 2003; Barlow and Proschan 1975; Bedford and Cooke 2001; Høyland and Rausand 1994; Melnick and Everitt 2008). The former is generally narrower in scope and tends to deal with engineered systems subject to repeated failures and the need for preventative maintenance policies to address these. Key concepts in reliability engineering include component availability, reliability and maintainability; mean times to, and between failure; the use of specific fault tree and failure mode tools; and the concepts of system redundancy. Reliability engineering owes a significant amount to advances in manufacturing engineering and the desire to improve production quality and optimise output (Lewis 1994). Risk analysis is a much broader term and tends to deal with more one-off failures that may write-off a system with concomitant impacts elsewhere. It is not necessarily restricted to technical systems and has developed into a broad interdisciplinary field with important inputs from the social sciences, alongside applied mathematics and decision science. But both reliability engineering and risk analysis are essentially are concerned with anticipating possible failures and assessing their likelihood. HRA specifically 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.

There are many reasons why one might undertake a risk or reliability analysis. In broad terms the first three items in our list relate to formative uses of risk and reliability analysis and the last two to summative uses[2].

1. The designers of a system 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.
2. 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 operating practices and safety culture.
3. 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.

[2] We make no claims that this list is exhaustive, just sufficient for our discussion.

4. 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. An assessment of the risks arising from human behaviour will be an integral part of this.

5. 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. Such differences may not be purely technical, since some systems may be more or less at risk from some human behaviours.

As a component of a full risk or reliability analysis, HRA may be used in any of these ways.

The origins of HRA lie in the early probabilistic risk assessments performed as part of the US nuclear energy development programme in the 1960s (Bedford and Cooke 2001; United States Nuclear Regulatory Commission 1975) Early first generation HRA methods such as the *Technique for Human Error Rate Prediction* (THERP) (Swain and Guttmann 1983) were very similar to those in other areas of reliability analysis: namely, the probability of a human error is assessed via a simple event tree analysis. The event tree simply listed an initiating event', which might be a system error reaching the human operator, and then considered a series of tasks that which had to be correctly carried out to prevent unwanted consequences. Essentially, in these early models, the human operator is treated as another component in the system. Hollnagel (1993) referred to this general approach as decomposition. A variety of other first generation methods have been developed with broadly similar features to THERP – the use of task analysis, use of nominal probabilities for task failure, adjustment factors to take account of different performance conditions, error factors and so on. The *Human Reliability Analysis Event Tree* method (HEART) (Williams 1985) is a good example of a method that aims to use many of the same features but in a simplified setting to give a more straightforward approach. Recognizing that many tasks have an associated time for completion, the *Human Cognitive Reliability* method (HCR) (Hannaman et al. 1984) modelled the time to successful completion. A wider review of these and many other methods is given in (Kirwan, 1994).

Much of the discussion around these models focussed on the issue that errors of *omission* (failures to respond to events appropriately) were considered easier to model than errors of *commission*, i.e. inappropriate human actions. However, this simplistic dichotomy now appears too stark in light of our current, richer, qualitative understandings of human cognition, motivation and decision making, including the effects of stress, emotion, training, group interactions, organisational structures, cultures and so forth (Bazerman 2006; Bazerman 1999; French et al. 2009; Kahneman et al. 1982; Kahneman and Tversky 2000). Research in these fields has shown that there are systematic influences on decision making and behaviour that cannot be categorised simply as omissions or commissions: see Section 3 below. Human failure is far more complex than the failure of, say, a steel support beam or a

hard disk. To be fair, second[3] generation HRA methods (Barriere et al. 2000; Hollnagel 1993) attempted to incorporate contextual effects such as tiredness, stress and organisational culture on an operator's proneness to error; and third generation HRA methods (Boring 2007; Mosleh and Chang 2004) have sought to allow for the potential variation in response and recovery actions once an error chain has begun. Notwithstanding this, we argue that far more development is needed before any method takes account of all our current understandings of human behaviour.

Surveys of current HRA methodologies may be found in Adhikari *et al*. (2008), Forrester *et al*. (2006) and Hollnagel (1993, 1998). For other recent research and developments in HRA, see the special issue of the *Journal of Loss Prevention in the Process Industries* (2008, **21**, 225-343). Software reliability analysis also has a large literature (Courtois *et al.* 2000; Lyu 2005; Zhang and Pham 2000). Software engineering is largely an endeavour of human design and thus subject to all the risks that HRA seeks to explore and assess. To date, software reliability assessment has, by and large, also adopted a mechanistic or empirical modelling of human error similar in methodology to current quantitative HRA.

Reason (1990b) offered a metaphor for system failure involving human error likening failure processes to movements of slices of Swiss cheese relative to each other: see Figure 1. Essentially this suggested 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. Although it is clear from his writings that Reason knew the limitations of metaphors (Reason 1995, 1997), his readers have often interpreted the model too mechanistically. There has been a dominant tendency to imagine a *fixed* number of slices, sliding backwards and forwards *independently* of each other until a series of holes align. In safety studies one talks of the number of safety barriers (the multi-barrier concept) or layers between normal operation and system failure; and, in a sense, the slices of Swiss Cheese mirror these. Systems are designed with a set number of safety barriers and these barriers are intended to be independent: cf. the simplistic calculation of a failure rate of 1 in 10^7 above. But human behaviour can correlate the risks of failure of two or more barriers, and most systems also harbour the opportunity for the 'bypass' of these barriers. Human 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 and which may disrupt several safety barriers simultaneously. For instance, the Chernobyl Accident (International Atomic Energy Agency 1991; Marples 1997) 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 compromised several of the safety barriers simultaneously. Another potential unsafe behaviour is to discover an

[3] One should not take too chronological perspective on first, second and third generation HRA methods. Some of those developed earliest did make attempts to account for contextual effects (Adhikari *et al*. 2008).

indication of a ‘hole’ in one layer and to defer further investigation, relying on the ‘cover’ offered by other layers: such behaviour occurred during a recent leak of radioactivity at Sellafield (Adhikari *et al.* 2008). Hrudey et al. (2006) describe similar behaviour during the Walkerton drinking water tragedy in Ontario, where latent and active flaws left unaddressed exacerbated the impact of agricultural run-off infiltrating a town’s shallow groundwater supply. 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 – the latter is, of course, the principle of preventative risk management.

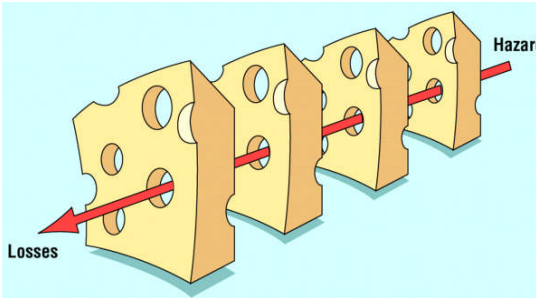

Figure 1 Reason's Swiss Cheese model (Reason 1990b)

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. Reason himself discusses similar criticisms (Reason 1995, 1997); but the simpler mechanistic thinking implicit in Figure 1 still pervades thinking in much of reliability engineering (Perneger 2005). The model visually 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. For example, it could be argued that the model struggles to fully represent the motives that might accompany deliberate violations of procedure, the creeping loss of vigilance with respect to a safety culture or the very real opportunities for the bypass of barriers in most technological systems.

We note that there is an established literature stemming from a range of work in France on the need to moderate reductionist, decomposable approaches to human reliability – or as they sometimes term it, ‘human factors of reliability’ – with an understanding of organisational, management and process contexts which can introduce dependencies (Fadier 2008; Fadier and Ciccotelli 1999; Fadier and De la Garza 2006; Leplat 1994).

3. Human behaviour and human error

Human behaviour is complex and often non rational. For instance, it seems sensible to use modern technological advances to make the physical components of 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 (Adams 1988; Hollnagel 1993). In this section we survey some recent findings from behavioural decision studies and consider how this area of theory and research can add to HRA. We do not focus on error behaviours per se, but take a more holistic approach. We do this for three reasons.

First, the error focus of HRA models may be too narrow (Hollnagel 1998, 2000a). Errors are just one of a range of behavioural products of a number of individual and organisational precursors; they are not a class of behaviours that are entirely distinct from other behaviours and thus 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 explicitly include human factors typically focus on cognitive aspects of decision making. Recent developments in the modelling of decision making emphasise the dual influences of cognition and emotion on decision outcomes (French *et al.* 2009; Loewenstein *et al.* 2001; Slovic *et al.* 2004). The integration of emotions and cognition models of decision making has improved the ability of such models to understand and predict behaviour (Phelps 2006). Furthermore, such an integrated approach is highly relevant to the risk-related decision making typically found within safety critical industries (Fenton-O'Creevy et al. 2008; Finucane et al. 2000).

Third, the use of high reliability systems designed and engineered to minimise 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. The *risk thermostat* model suggests there is a dynamic interaction between actors' perceptions and behaviours and their environment (Adams 1988; Wilde 1982, 1998). People will adjust their behaviour to be more or less risky, as appropriate for their preferences and their situation, perhaps relying on one safety system to protect them from the risk of failing to operate another. A high profile example is the leak in a modern plant at Sellafield mentioned previously. There was a belief that such a modern plant could not suffer from leaks or other failures. 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 at Sellafield (Adhikari *et al.* 2008; Board of Inquiry 2005). Marcus and Nichols (1999) discuss similar behaviours in which warning signs were not heeded and suggest that other priorities for limited resources make it too easy to drift towards what they term the 'safety border'.

Real human judgement and decision making is not as rational and analytic as one might wish. Since the early 1980s, psychologists have distinguished between two different forms of thinking (Chaiken *et al.* 1989)[4]:

[4] There is an unfortunate conflict of terminology here between our use of 'system' to mean the entire plant and processes which is at risk and 'systems of thinking' as referred to in the psychological literature. We use the phrases 'System 1 (or 2) *thinking*' to distinguish the latter.

- System 1 thinking, often referred to as ‘intuition’ or ‘gut reaction’ that involves a superficial analysis/interpretation of the relevant information based on much simpler forms of thinking on the fringes or outside of consciousness;
- System 2 thinking, characterised by conscious analytical thought that involves a detailed evaluation of a broad range of information, often based on a rule that is assumed to provide the ‘correct’ answer or solution.

While formal risk assessment techniques have the characteristics of System 2 thinking, system operators may use System 1 thinking in their day-to-day operations and responses to events. For example, a nuclear power plant is the outcome of considerable complex analysis, research and design, i.e. System 2 thinking. The operators of such a plant, however, do not typically engage in the same kind of analytical thinking as the system engineers and designers. The operators’ work comprises much more routine procedures and, where complex problems are faced, there is potential for operators to make them more manageable through system 1 heuristics. It has become common to refer to much of System 1 thinking as involving ‘heuristics and biases’, because of its deviation from the more rational, analytic System 2 thinking, though that terminology is as pejorative as the constant use of the term ‘human error’ in HRA which we reject in this article.

There is an extensive literature on decision making heuristics and biases (French *et al.* 2009; Kahneman *et al.* 1982; Kahneman and Tversky 2000). Numerous studies have demonstrated the existence of systematic and robust cognitive biases, and are well summarised by Bazerman (2006). For example, emotionally-laden or otherwise individually salient information is recalled easily and likely to be considered as significant to a decision when more objective evidence shows that other types of information are more important to a decision. The processes that drive such biases have not arisen without 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 by providing rapid mechanisms for recall of salient information and execution of choice behaviours (Goldstein and Gigerenzer 2002). However, such biases can be problematic. For example, Willman *et al*. (2001) and Fenton-O’Creevy *et al*. (2003) explored the dislocation between pure financial theories and the collective and individual behaviours of market traders. Their research showed that biases led to ineffective decision making and reduced performance.

It is of concern that very little use of this extensive, often empirically based literature has been made in developing HRA methodologies. Indeed, the mechanistic approach common to many such methodologies based on fault tree representations of human action assumes that the operators are using System 2 thinking when in all probability their intuitive responses and actions are guided by

System 1 thinking (Bargh *et al.* 1996)[5]. HRA methodologies should model the thinking and behaviours that are likely to occur rather than more rational, analytic actions and responses that one should like to think would occur.

In fairness to some current approaches to quantitative HRA, their proponents would not claim to be modelling actual behaviour, whether it be driven by System 1 or System 2 thinking; nor to be seeking a 'correct' answer to a quantitative problem. When risk analysis is used formatively, its purpose is to understand better systems and identify the key drivers of risk, rather than chase quantified estimates *per se*. Current HRA methods may help identify the key drivers relating to human behaviour, irrespective of what is going on inside people's heads and whatever organisational and environment contexts that surround them. However, such approaches do need data: and while there is generally no great problem in finding data relating to normal operations, appropriate data are – fortunately! – sparse in most contexts relating to serious system failures.

If we are to model actual behaviour in a variety of circumstances, then the concept of self-regulation may be needed. 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 (Lord and Levy 1994). 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 through activation and inhibition of decision making processes (Carver and Scheier 1981). Some of the decision processes take place at a subconscious level and never reach conscious deliberation, a process called automaticity (Bargh and Chartrand 1999). Thus, 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 (1908) introduced an inverted U model of the association between performance and arousal. More recent models of work performance show similar patterns: some effort and pressure can be effective, too much of either leads to burnout (Schaufeli and Bakker 2004). 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 potentially subject to the same processes of automaticity, flawed biases and self-regulation as individual decision making.

This recognition that human behaviour is complex and driven by a range of internal and external factors leads us to question the value of terminology such as 'error', 'slip' or 'failure' within HRA. Human errors and faults are socially defined events: a perfectly reasonable action to one person may

---

[5] Of course, one might hope that if operators have been subject to many training exercises, then their responses may be closer to those that would arise from system 2 thinking.

be an unreasonable failure to another (Hollnagel 2000b). 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 choice. We should focus more on human *behaviour* in individual, group and organisational contexts and recognise its potential involvement in system failure – without the pejorative judgement of whether that behaviour is aberrant in any sense. For example, in the Three Mile Island Incident (Commission on the Three Mile Island Accident 1979) 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 not 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. 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 within HRA theory. As we build and operate more and more complex systems, we should recognise that it is inevitable that we will encounter unanticipated events and conditions. Risk and reliability analyses need to take account of human responses to these and, although those responses may indeed lead to untoward outcomes, it is far from clear that they should be dubbed errors.

4. High reliability organisations

The past 20 years has seen several studies of *high reliability organisations* (HROs), which Roberts (1990) 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 (Weick 1987). 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 (Weick 1987). They emphasise a culture of sharing of learning and knowledge, of mental models: 'heedful inter-relating' (Weick and Roberts 1993), 'collective mindfulness' (Weick *et al.* 1999), 'extraordinarily dense' patterns of cooperative behaviour (La Porte 1996) and 'shared situation awareness' (Roth *et al.* 2006). Usually HROs apply a strategy of redundancy (Rochlin *et al.* 1987) with teams of operators 'watching each others backs'. As noted, it is suggested that teams share common mental models of both their internal organisational processes and the external world (Mathieu et al. 2000; Smith-Jentsch et al. 2005). Redundancy may increase complexity of operations as it makes the operations system less easily understood or opaque (Perrow 1984; Sagan 1993). However, redundancy also increases the probability or chance of getting adequate information to

solve probable dangers, consequently reducing the risks arising from complexity rather than increasing them. When necessary, HROs try to decentralise the authority of senior teams or management responsible for decision making. Rijpma (1997) 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 (1996) suggests the balance of these opposing effects can lie in the direction of higher reliability.

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 (Perrow 1994; Sagan 1993, 1994). A second criticism relates to the absence of validation for the empirical studies underpinning HRO theory (Clarke 1993; Perrow 1994; Sagan 1993). Critics argue that the context of some of the most important HRO studies, e.g. on the flight decks of aircraft carriers, is misleading, with evidence of safety only 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. But the HRO work has given us an insight into the way in which error and failure is managed by social organisations, and how collective, rather than individual, phenomena like collective mindfulness (Weick *et al.* 1999) are what produce reliability in the face of supposedly unreliable individuals and unreliable technologies. The emphasis of HRA on individuals and on atomised tasks therefore misses the probability that collective actions and behaviours might lead to or avert system failure.

There would seem to be considerable potential for formative uses of HRA to influence the development of HRO theory, at least in so far as it can be applied in system and organisational design; and vice versa, complementing the work of, e.g., Grabrowski and Roberts (1999).

5. Decision contexts

There is a further aspect of context that HRA should consider: decision context. The judgements and decisions needed of humans in a system can vary from those needed to perform mundane repetitive operational tasks through more complex circumstances in which information needs to be sought and evaluated to identify appropriate actions to the ability to react to and deal with unknown and unanticipated. Decision processes will vary accordingly. Design decisions can inadvertently introduce further risks to the system that arise from limitations inherent in human foresight. This means that the appropriate HRA methodology to assess the risks associated with the human decision making behaviour may vary with the details of that context.

Cynefin is a conceptual framework developed by Snowden which, among other things, offers a categorisation of decision contexts (Snowden 2002; Snowden and Boone 2007). 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 (1993) discusses such situations as recognition primed decision making. 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 making can be proceduralised with clear guidance decided *a priori*. 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. Before decisions can be made, it is necessary to think widely, explore issues, frame the problem and develop broad strategies that are flexible enough to accommodate changes as the situation evolves. Much judgement and expertise will be needed in making the decision itself. 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. 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.

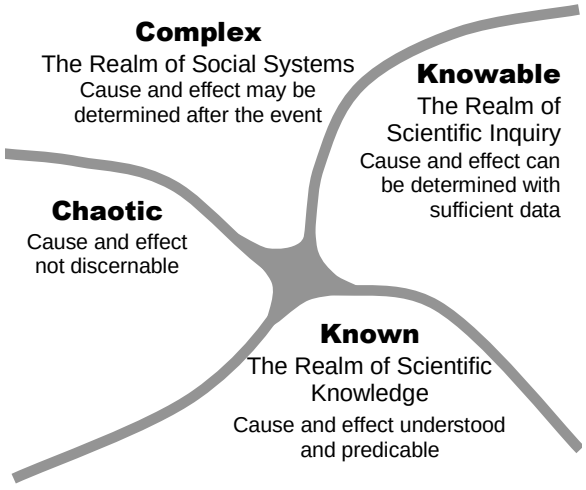

Figure 2: Cynefin

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 repetitive, operational tasks that lie in the known or, perhaps, knowable spaces. Yet many of the perceived risks in modern systems arise because of their inherent complexity (Perrow 1984, 1994). 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 (French and Niculae 2005). 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 (1987) discussion 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. (For discussions of the tension between operational risk management practice, and incident preparedness and management, see, e.g., Jalba et al. 2009; Pollard et al. 2009).

The appropriateness of any HRA methodology may depend on the context that is being assessed. As is the case with all risk methodologies, the characteristics of the risk and the availability of data to support the application of specific tools and techniques has a forceful influence on their feasible use. Are we considering a repetitive task that an operator performs in the normal course of events? In this 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 an unanticipated departure of the system from its normal operating characteristics? In this 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.

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 methodology to be sufficient to meet these requirements. Hollnagel (1998) recognised this, though his suggestion of two methods probably does not take us much further forward, particularly as his basic method is more of a screening method for his extended approach rather than appropriate to a different set of circumstances. 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. It is also important to work out a way of integrating them so that we do not perpetuate the fallacy of thinking tasks can be divided up and broken down, and methods can be selected in isolation.

6. Toward an extended model of HRA

Summative HRA and related approaches emphasise quantification and prediction. While cognitive understanding of people and cultural perspectives on organisations are acknowledged, the gulf between these and quantitative risk models is generally considered too significant to be bridged. Yet the conjoining of these approaches could yield a superior model of safety critical organisations and the people working within them. In the short term, exploring the interfaces between HRA and behavioural, organisational and related studies is likely to benefit formative analyses to support the design and operation of complex systems. The barriers to the quantification needed in summative analyses are currently too substantial – we do not have sufficiently developed and validated models of behaviour and organisations to provide the precision needed. Moreover, progress in improving and developing quantitative HRA methods is likely to proceed most quickly in relation to tasks and activities falling in the known and, perhaps, knowable Cynefin spaces. Successful quantitative modelling of such tasks and activities depends on having sufficient data to develop and validate models. For systems that are long established or straightforward developments thereof, we are likely to have useful data. For novel systems we might generate such data by involving operators in simulations of component tasks and activities in known and knowable spaces. Ĉepin (2008) has suggested as much, though without the language of Cynefin. Ĉepin's modelling, as might be expected from our discussion, is focused on probabilistic assessments of errors of omission and commission. His proposed development focuses on manufactured situations with tight parameters. While additional data gathered within such a paradigm would add considerable utility to HRA models, there remains the issue of scope. The approach cannot be easily extended to tasks and activities in the other spaces. By definition, in the complex space we have neither sufficient qualitative understanding

nor relevant data to develop quantitative HRA models that predict individual, group and organisational human behaviour and how these may impact overall system reliability and safety.

Thus we believe that the dominant HRA paradigm, suited as it is for the known and knowable spaces, needs to be complemented by paradigms developed specifically for the complex space. In achieving this, we will need to move away from many systems engineering approaches in which hazards are purportedly designed out of a system. Complex systems involve some human activity if only in their design and hence are susceptible to some risk arising from human behaviour. Such systems engineering approaches may work in the known or knowable spaces – and there the question is moot. Even the simplest systems in the known space need to be designed and that is a human activity. Moreover the risk homeostasis model suggests that there can be an over-reliance the safety promised by the system and a concomitant increase in overall risk. But in the complex space, we have no such hopes that current approaches can design hazards out of the system.

It will not be easy to develop models for new HRA paradigms suitable for the complex space. For instance, organisational behavioural studies can be useful in identifying the individual, cultural and organisational factors relevant to system safety; but they do not lend themselves to simple quantification; indeed, it is the very nature of the complex space that quantification is difficult if not impossible on current knowledge. We are also limited both by the availability of data from real incidents and from the generalisability of laboratory based studies. Some commentators, notably Le Coze (2005), consider the question of whether current forms of organisations can lend themselves to effective modelling. Furthermore, linear models of cause and effect cannot be simply applied (Morin 1977). Le Coze provides a useful analysis of organisational theories, and their limitations. He proposes that approaches from complexity theory[6] (Morin and Lemoigne 1999; Prigogine 1994; Simon 1996) could assist in integrating methodologies. One of the contributions of complexity theory is the guiding philosophy that complex problems cannot be meaningfully decomposed and retain utility since the whole is greater than the sum of its parts. Le Coze concludes by emphasizing the need for holistic approaches to organisations with additional data from both organisational events and empirical studies.

Another family of approaches that might lead to a broadened conceptualisation of HRA in the complex space are the socio-technical (Mumford 2000). For instance, Reiman and Oedewald (2007) propose that safe and effective organisations can arise only when there is integration of organisational culture and organisational activities. Their model includes a range of qualitative and quantitative methods designed to elicit descriptions of the cultural features and the organisational core tasks resulting in a

---

[6] There are differences between complexity theory and Snowdon's notion of the complex space in Cynefin, but there are also similarities and these link Le Coze's and our arguments.

thorough understanding of alternative ways to approach organisational thinking, strengths and weaknesses of practices and opportunities to create dialogue regarding the effectiveness of work. Reiman and Oedewald's paper represents a useful contribution to the development of the field of socio-technical systems and their potential links with more quantitative approaches to error and risk. What they do not encompass are individual approaches to understanding organisational behaviour. Their work represents another step in the right direction but there is a long way to go before human activities and behaviours in complex space can be modelled sufficiently for quantitative HRA.

None of the above addresses activities and behaviours which might arise if through some unanticipated event the system 'moves into' the chaotic space: e.g. the unanticipated formation of a hydrogen bubble in the Three Mile Island Incident (Commission on the Three Mile Island Accident 1979; Niculae 2005). By definition these characteristics cannot be represented in a model – certainly not in anything other than a schematic manner – simply because the chaotic space represents that part of our environment that we do not understand yet and so cannot predict.

So to take stock: current quantitative HRA methodologies seem applicable to behaviours and activities in the known and knowable spaces. There are the barest hints of how some quantitative models might be developed to predict the impacts of human activities and behaviours in the complex spaces; and, by definition, it is logically inconceivable that we can develop quantitative models for the chaotic space. Thus it is not currently possible to perform summative risk and reliability analyses for any system in which human behaviour and activity can enter the complex or chaotic spaces. Governments and regulators should be concerned because this accounts for the majority of the technological systems currently being operated and commissioned. This does not mean that they are unreliable or unsafe; only that we cannot assure their reliability or safety to within some negligibly small probability. But there are ways forward.

Firstly and most immediately, we can look to formative uses of HRA, the behavioural and organisational sciences and many other related disciplines to inform the design of organisational and management structures and the establishment of appropriate safety cultures to improve the systems that we have and are designing. This will not be easy because the imperatives that drive this approach fly in the face of the dominant reductionist thinking in risk and reliability communities. One cannot simply decompose systems into smaller subsystems, focus on these in turn and expect these to represent the total system, because culture, organisational structures and other drivers of human behaviour correlate actions, judgements and decision making in the different subsystems. Modern perspectives on risk demand a systemic rather than an atomised perspective of the technical, human and organisational features of systems. Further, because many systems have shared, and arguably, often fragmented responsibilities for management and risk management (e.g. flood defence, social care, biosecurity in the food chain), one needs to take a more holistic perspective. The conceptualisation offered by Cynefin may again give us a way forward. The simple visual

categorisation of different decision contexts has proven very successful in one of the author's experiences in helping in problem formulation and issue structuring (Franco et al. 2006, 2007; Mingers and Rosenhead 2004; Rosenhead and Mingers 2001). The managers who decide on the choice of managerial system, its components and operational processes could map these onto a Cynefin diagram. The discussions and deliberations that would occur as they undertook this would naturally surface many issues that their design and management decisions would need to address. In other words, we propose a careful use and reflection upon a Cynefin mapping would augment current hazard identification procedures and make clearer some of the issues relating to human behaviour that management will face in operating the system. When they identify that issue although important lies in the known or knowable space they can look to current HRA – or, preferably, somewhat enhanced – models to guide their thinking and planning. But when discussion identifies an issue as lying in the complex space then they will to rely much more on judgement and put into place management processes that can deal with behaviours more subtly than seeking to police against 'slips, errors, and omissions'.

We also believe that in time it will be possible to develop better quantitative HRA methodologies to give wider assurance at the summative level. But it is unlikely that this will lead to single methodology. Rather we will need a multi-faceted approach that combines empirically validated HRA models for the known and knowable spaces with more judgementally based methods for the complex space. The Cynefin model suggests a broad framework with which to categorise the human tasks and activities in system to determine which form of HRA modelling would be most appropriate; but it is only a broad framework. To develop this methodology it will probably need extending to recognise, among other things:

- whether the human behaviours and activities take place at the individual, group, organisational level;
- the wider organisational context – including strategic and economic imperatives – in which the teams and local management structures are embedded;
- the team and local management structures which set the local context in which the operators work;
- the cultural context and – including misplaced trust in other safety barriers in the system – in which the operators find themselves;
- external influences, particularly those arising from larger external and societal pressures;
- the historical context, including perhaps the lack of recent incidents leading to a growth of complacency.

In Adhikari *et al.*(2008) we outline a programme of research and benchmarking that may help us develop such a multi-faceted portfolio of HRA methodologies that may eventually provide much better summative guidance on the risks inherent in complex systems.

None of this will be easy and it will only be possible if we can break the current mechanistic paradigms that permeate the risk and reliability communities. We need to move on from the Swiss Cheese model.

7. Conclusion: a message for managers

The key point that we have been trying to convey in this paper is the current dislocation between the mechanistic reductionist assumptions on which current HRA methodologies are primarily built and our current understandings of human and organisational behaviour. We must bring these into better register. Managers, regulators, politicians and the public need to beware of this lest they believe the numbers that are sometimes touted about the safety of our systems. This should not be read as a manifesto for Luddism. We are not against the development of more and more complex systems, providing that they bring benefits, of course. Nor are we against risk *per se*. Rather we are concerned at the prevalence of overconfidence in our ability to assess the risks that arise from human behaviour. We need to take the numbers with that 'pinch of salt', recognising that when we build complex systems our uncertainty is greater than the raw numbers suggest and we need to monitor and watch for the unanticipated. As is often the case with the application of risk and reliability tools, the valuable insight comes from a systemic and often qualitative understanding of which systems features 'drive' the risk, rather than from the risk estimates *per se*.

We in the research community have much to do. But so does the management community. It is too easy to trust the assurances of current risk and reliability analyses which promise that the chance of an untoward event is small, to believe in the cumulative effect of 'independent' safety barriers and to manage the subsystems separately unaware of the interconnections between them that organisational culture and human behaviour bring. Human reliability has too long been treated as something that relates to individuals. It needs to be seen and managed at the organisational level. The key question is not how likely is an individual's behaviour is to impact a system, but how well the organisational structures around and within that system enable the system to run safely and reliability, and how well they will recover if an untoward event threatens or happens.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (Contract number: EP/E017800/1). We are grateful to our co-investigators and colleagues on this: Sondipon Adhikari, Clare Bayley, Jerry Busby, Andrew Cliffe, Geeta Devgun, Moetaz Eid, Ritesh Keshvala,

David Tracy and Shaomin Wu. We are also grateful for many helpful discussions with Ronald Boring, Roger Cooke and John Maule.

References

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

Adhikari, S., C. Bayley, T. Bedford, J.S. Busby, A. Cliffe, G. Devgun, M. Eid, S. French, R. Keshvala, S. Pollard, E. Soane, D. Tracy, S. Wu. 2008. Human Reliability Analysis: A Review and Critique. Manchester Business School, Booth Street West, Manchester M15 6PB.

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

Bargh, J.A., T.L. Chartrand. 1999. The unbearable automaticity of being. *American Psychologist* **54** 462-479.

Bargh, J.A., M. Chen, L. Burrows. 1996. Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* **71** 230-244.

Barlow, R.E., F. Proschan. 1975. *Statistical Theory of Reliability and Life Testing*. Holt, Reinhart and Winston, New York.

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

Bazerman, M. 2006. *Managerial Decision Making*, 6th ed. John Wiley and Sons, New York.

Bazerman, M.H. 1999. Reviews on decision making. *Administrative Science Quarterly* **44**(1) 176-180.

Bedford, T., R. Cooke. 2001. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, Cambridge.

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

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

Carver, C.S., M.F. Scheier. 1981. *Attention and Self-Regulation: a Control Theory Approach to Human Behavior*. Springer Verlag, New York.

Ĉepin, M. 2008. Importance of human contribution within the human reliability analysis (IJS-HRA). *Journal of Loss Prevention in the Process Industries* **21**(3) 268-276.

Chaiken, S., A. Liberman, A.H. Eagly. 1989. Heuristic and systematic information processing within and beyond the persuasion context. J.S. Uleman, J.A. Bargh, eds. *Unintended Thought*. Guilford, New York, 212-252.

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

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

Courtois, P.-J., B. Littlewood, L. Strigini, D. Wright, N. Fenton, M. Neil. 2000. Bayesian belief networks for safety assessment of computer-based systems. E. Gelenbe, ed. *System Performance Evaluation: Methodologies and Applications*. CRC Press, 349-363.

Fadier, E. 2008. Editorial of the Special Issue: Design Process and Human Factors Integration. *Cognition, Technology and Work* **10**(1) 1-5.

Fadier, E., J. Ciccotelli. 1999. How to integrate safety in design: methods and models. *Human Factors and Ergonomics in Manufacturing & Service Industries* **9**(4) 367-379.

Fadier, E., C. De la Garza. 2006. Safety design: towards a new philosophy. *Safety Science* **44** 55–73.

Fenton-O'Creevy, M., N. Nicholson, E. Soane, P. Willman. 2003. Trading on illusions: unrealistic perceptions of control and trading performance. *Journal of Occupational and Organizational Psychology* **76**(1) 53-68.

Fenton-O'Creevy, M., E. Soane, N. Nicholson, P. Willman. 2008. Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders *Academy of Management Conference*, Anaheim, California. .

Finucane, M.L., A. Alhakami, P. Slovic, S.M. Johnson. 2000. The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making* **13** 1-17.

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

Franco, A., D. Shaw, M. Westcombe. 2006. Problem Structuring Methods I *Journal of the Operational Research Society*, 757-878.

Franco, A., D. Shaw, M. Westcombe. 2007. Problem Structuring Methods II *Journal of the Operational Research Society*, 545- 682.

French, S., A.J. Maule, K.N. Papamichail. 2009. *Decision Behaviour, Analysis and Support*. Cambridge University Press, Cambridge.

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

Goldstein, D.G., G. Gigerenzer. 2002. Models of ecological rationality: the recognition heuristic. *Psychological Review* **109**(1) 75-90.

Grabowski, M., K.H. Roberts. 1999. Risk mitigation in virtual organizations. *Organization Science* **10** 704-721.

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

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

Hollnagel, E. 1993. *Human Reliability Analysis: Context and Control*. Academic Press, London.

Hollnagel, E. 1998. *Cognitive Reliability and Error Analysis Method – CREAM*. Elsevier Science, Oxford.

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

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

Høyland, A., M. Rausand. 1994. System Reliability Theory. John Wiley and Sons, New York.

Hrudey, S.E., E.J. Hrudey, J.W.A. Charrois, S.J.T. Pollard. 2006. A 'Swiss cheese' model analysis of the risk management failures in the fatal Walkerton outbreak. *IWA world water congress and exhibition*, Beijing, China.

International Atomic Energy Agency. 1991. The International Chernobyl Project: Technical Report. IAEA, Vienna.

Jalba, D., N. Cromar, S. Pollard, J.W.A. Charrois, R. Bradshaw, E. Hrudey. 2009. Safe drinking water: critical components of effective inter-agency relationships. *Environment International* (in press - doi:10.1016/j.envint.2009.1009.1007).

Kahneman, D., P. Slovic, A. Tversky, eds. 1982. *Judgement under Uncertainty*. Cambridge University Press, Cambridge.

Kahneman, D., A. Tversky, eds. 2000. *Choices, Values and Frames*. Cambridge University Press, Cambridge.

Kariuki, S.G., K. Lowe. 2007. Integrating human factors into process analysis. *Reliability Engineering and System Safety* **92** 1764-1773.

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

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

Le Coze, J.-C. 2005. Are organisations too complex to be integrated in technical risk assessment and current safety auditing? *Safety Science* **43**(8) 613-638.

Leplat, J. 1994. Collective dimensions of reliability: some lines of research. *European Work and Organizational Ppsychologist* **4**(3) 271-295.

Lewis, E.E. 1994. *Introduction to Reliability Engineering*. Johne Wiley and Sons, Chichester.

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

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

Lyu, M.R. 2005. *Handbook of Software Reliability Engineering*. IEEE Computer Society Press and McGraw-Hill Publishing Company.

Marcus, A., M.L. Nichols. 1999. On the edge: heeding the warnings of unusual events. *Organisazational Science* **10**(4) 482-499.

Marples, D.R. 1997. Nuclear power in the former USSR: historical and contemporary perspectives D.R. Marples, M.J. Young, eds. *Nuclear Energy and Security in the Former Soviet Union*. Westview Press.

Mathieu, J.E., T.S. Heffner, G.F. Goodwin, E. Salas, J.A. Cannon-Bowers. 2000. The influence of shared mental models on team process and performance. *Journal of Applied Psychology* **85**(2) 273-283.

Melnick, E.L., B.S. Everitt, eds. 2008. *Encyclopedia of Quantitative Risk Analysis and Assessment*. John Wiley and Sons, Chichester.

Mingers, J., J. Rosenhead. 2004. Problem Structuring Methods in Action. *European Journal of Operational Research* **152** 530-554.

Morin, E. 1977. La method – tome I, La nature de la nature. Ed du seuil (coll point), Paris (The method – Vol I, the nature of nature).

Morin, E., J.L. Lemoigne. 1999. L'intelligence de la complexité. L'Harmattan (The intelligence of complexity)

Mosleh, A., Y.H. Chang. 2004. Model-based human reliability analysis: prospects and reliability. *Reliability Engineering and System Safety* **83** 241-253.

Mumford, E. 2000. A socio-technical approach to systems design. *Requirements Engineering* **5** 125-133.

Niculae, C. 2005. A socio-technical perspective on the use of RODOS in nuclear emergency management, The University of Manchester.

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

Perrow, C. 1984. *Normal accidents: living with high-risk technologies*. Basic Books, New York.

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

Phelps, E.A. 2006. Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology* **57** 27-53.

Pollard, S., R. Bradshaw, D. Tranfield, J.W.A. Charrois, N. Cromar, D. Jalba. 2009. Developing a risk management culture — 'mindfulness' in the international water utility sector (Report TC3184). Water Research Foundation, Denver, CO.

Prigogine, I. 1994. Les lois de chaos. Flammarion (The laws of chaos).

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

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

Reason, J. 1995. Understanding adverse events: human factors. *Quality Health Care* **4** 80-89.

Reason, J. 1997. *Managing the Risks of Organisational Accidents*. Ashgate, Aldershot, UK.

Reiman, T., P. Oedewald. 2007. Assessment of complex socio-technical systems – theorerical issues concerning the use of organisational culture and organisational core task concepts. *Safety Science* **45** 745-768.

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

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).

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

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

Rosenhead, J., J. Mingers, eds. 2001. *Rational Analysis for a Problematic World Revisited*. John Wiley and Sons, Chichester.

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

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

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

Sagan, S.D. 2004. The problem of redundancy problem [sic]: why more nuclear security forces may produce less nuclear security. *Risk Analysis* **24** 935-946.

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

Simon, H. 1996. *The Sciences of the Artificial*. MIT Press.

Slovic, P., M.L. Finucane, E. Peters, D.G. MacGregor. 2004. Risk as analysis and risk as feelings: some thoughts about affect, reason, risk and rationality. *Risk Analysis* **24**(2) 311-322.

Smith-Jentsch, K.A., J.E. Mathieu, K. Kraiger. 2005. Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology* **90**(3) 523-525.

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

Snowden, D., M. Boone. 2007. A leader's framework for decision making. *Harvard Business Review* 68-76.

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

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

United States Nuclear Regulatory Commission. 2002. Review of Findings for Human Performance Contribution to Risk in Operating Events (NUREG/CR-6753). US GPO, Washington, DC.

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

Weick, K.E., K.H. Roberts. 1993. Collective mind in organizations: heedful interrelating on flight decks. *Administrative Science Quarterly* **38** 357-381.

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

Wilde, G.J.S. 1982. The theory of risk homeostasis: implications for safety and health. *Risk Analysis* **2** 209-225.

Wilde, G.J.S. 1998. Risk homeostasis theory: an overview. *Injury Prevention* **4** 89-91.

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

Willman, P., M. Fenton-OCreevy, N. Nicholson, E. Soane. 2001. Knowing the risks: theory and practice in financial market trading. *Human Relations* **54**(1) 887-910.

Wu, S., S.E. Hrudey, S. French, T. Bedford, E. Soane, S.J.T. Pollard. 2009. Human reliability analysis has a role in preventing drinking water incidents *Water Research*(in press).

Yerkes, R.M., J.D. Dodson. 1908. The relation of strength of stimulus to rapidity of habit-formation. . *Journal of Comparative Neurological Psychology* **18** 459-482.

Zhang, X., H. Pham. 2000. An analysis of factors affecting software reliability. *Journal of Systems and Software* **50**(1) 43-56

Human reliability analysis: A critique and review for managers

French, Simon

2011-07-01T00:00:00Z