# Cognitive Error Analysis

## in Accident and Incident Investigation in Safety-Critical Domains

**Thesis Submitted for the Degree of Doctor of Philosophy**
**(Ph.D.)**

**Daniela Karin Busse**

**Department of Computing Science**

UNIVERSITY
of
GLASGOW

September 2002

# Abstract

Major accidents and incidents, such as Three Mile Island or the Bristol Royal Infirmary Inquiry (a.k.a. the Bristol Baby Case, Kennedy et al., 2000) are often mediated by what is termed 'human error'. The analysis of these 'human errors' provides the basis for safety recommendations and thus has a crucial, albeit indirect impact on system design and the prevention of future accidents.

The analysis process often entails instances of erroneous behaviour (or its effects) being classified according to error taxonomies. These taxonomies might be grounded in psychological theory in order to provide more analytic power to the investigator. However, the classification process itself is often left unsupported when using these psychological error taxonomies as part of accident or incident investigation.

The reasoning process behind classification decisions is typically not being made explicit in the analysis process, and thus suffers from a lack of traceability. Human error classification that uses psychologically grounded taxonomies depends heavily on the expertise of the analyst. Although error categorizations are a useful basis for quantitative analysis as part of system safety management, they tend to elucidate mainly the "what" and "how" of an accident's causation, and not the "why".

The reasoning process that leads to the categorization often embodies several competing causal hypotheses being generated and tested by the analyst. This reasoning process itself, if documented, can provide valuable diagnostic insights into the complexity of an accident's cause. It can also embody important analytical information for the investigator when considering safety recommendations that are to address the system's weakness.

In this thesis, a cognitive error analysis framework is proposed as a tool to support accident and incident investigators in the analysis of "human error". A cognitive architecture is used to provide an analytic vocabulary that is grounded in

psychological theory. This cognitive error analysis approach provides a structured framework for expert analysts to reason about competing hypotheses on the role of ‘human error’ in an accident’s or incident’s causation.

By using the proposed cognitive error analysis approach, the rationale behind error analysis and classification can be made more transparent and be documented for future reference. The cognitive framework can provide a structured, grounded vocabulary for validating and reasoning about competing error explanations and safety recommendations.

These arguments will be illustrated in this thesis by means of case studies in human error analysis in accident and incident investigation. Retrospective analyses of accident and incident data drawn from aviation and medical work domains provide proof of concept and initial insights into the strengths and weaknesses of the cognitive error analysis approach that is proposed in this thesis.

The two core case studies presented here both concern the analysis of human error in incident investigation in intensive care units. The cognitive error analysis approach was applied both times in the context of real existing safety management in two Scottish hospitals (referred to as ‘the Edinburgh scheme’ and ‘the Glasgow scheme’ respectively). Both case studies were conducted in collaboration with the local medical and nursing staff.

A database of 10 years’ worth of medical incident data gathered in an Edinburgh Intensive Care Unit was analyzed using the proposed cognitive error analysis approach. In the second live case study, the error analysis approach was evaluated in the field by applying it to incident reporting data that was collected with a newly implemented incident reporting scheme in a Glasgow Neonatal Intensive Care Unit. The insights gained by analyzing the Edinburgh incident scheme were used to inform the design and implementation of the Glasgow incident scheme as part of the unit’s existing safety management. Since both were local incident reporting schemes, it was seen as an important factor for its success to take the local context and conditions into account while situating the cognitive error analysis approach as part of these hospitals' safety management strategies.

The evaluation of this incident reporting and analysis framework demonstrated the benefits of a structured, psychological "human error" analysis approach that centres on the human aspect of the incident, without isolating it from its context. It is argued that not only could the understanding of the underlying error mechanisms be improved for individual incidents, but the generation of safety recommendations could be supported, and these could then also be evaluated as to their impact on the human "in the loop". The resulting error analysis models could further be used as basis for comparing competing analyses, and also improve analysis traceability by documenting the analysis process and its resulting safety recommendations. Further work is needed in providing "best practices" for the application of the cognitive analytical framework. Further work is also needed in formalizing a way to situate the cognitive error analysis approach within the investigation of local work system factors in the search for the overall incident and accident causation.

## Thesis statement

This thesis aims at demonstrating the benefits of grounding the analysis of human error as part of incident and accident reporting in a cognitive theoretical framework. This will provide the means and the vocabulary to reason about alternative causal hypotheses while also acting as a tool to document and communicate the psychological analysis of human error and its resulting safety recommendations. This approach is proposed as complementing the analysis of human error data by means of error taxonomies grounded in psychological theory.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

I would like to thank my supervisors Prof. Chris W. Johnson and Prof. Stephen A. Brewster, for their careful, spirited, and generous guidance during the course of the research reported here.

The Glasgow Interactive Systems Group (GIST) has been a great source of inspiration, initiation, and support. Special thanks also goes to the Glasgow Accident Analysis Group.

I would also like to thank all the members of the Glasgow Yorkhill NICU for their participation in my studies, and their friendly welcome. I would also especially like to thank Drs Barbara Holland and David Wright for their enthusiasm and open collaboration.

I encountered many fellow researchers during the course of writing this thesis that supported and inspired me with their thoughts and discussion. Thanks to all of them, and especially to Phil Gray, Steve Draper, Richard Young, Lorraine Johnston, Penny Sanderson, and Phil Barnard.

This work has been sponsored by UK EPSRC grant No GR/L27800, and the MacRobertson Scholarship (Faculty of Physical Sciences, Glasgow University).

In memory of Billy Coutts, Bernard Weber, and Rodney Fuller.

Daniela K. Busse
Seattle, WA, U.S.A.
September 11th, 2002

# Declaration of Authorship

I hereby declare that I composed this thesis entirely myself and that it describes my own research.

Daniela K. Busse
Seattle, WA, U.S.A.
September 11th, 2002

# CHAPTER 1 INTRODUCTION

## MOTIVATION

Human involvement often plays a major role in the occurrence of accidents in safety critical systems such as in aviation, hospital systems, or nuclear power plants (Reason, 1990). Accident reports often resort to naming human error ("pilot error", "operator error") as the 'reason' why the accident happened. This 'blame' approach has frequently been criticized (op. cit.) but it still seems to prevail, such as illustrated by the all too frequent mention of it in news coverage of accidents and incidents (e.g. Kleinfield, 2001)

Identifying human involvement in an accident's causation does not necessarily mean that "human error" is the cause, or the only cause, of that accident. Human action (or inaction) is always stimulated by either the external context or an internal motivation (such as a set of expectations), it does not take place in a vacuum. Therefore, the identification of "human error" in an accident's causation should more often than not be the starting case for further investigation, rather than provide a convenient category where blame apparently has been found, and no need for more in-depth analysis exists. The label "human error" should not be equated with a stopping case for accident analysis.

The following case study of recent events illustrates this further. In the American Airlines crash of an Airbus A-300 in Queens, NY, on November 12th 2001 (Kleinfield, 2001), 260 passengers and all crew died when the plane plunged into a residential neighbourhood minutes after leaving Kennedy International Airport at

9:14 am for Santo Domingo (Dominican Republic). Eyewitnesses said they saw the burning engine separating from the plane and plummeting to the ground shortly before the crash.

The plane had broken into pieces in midair, with parts of the tail, most of the fuselage and both engines having broken off the plane and fallen to the ground. Although initially considered a material fault, later this was hypothesized as caused by the pilots' performance putting too much stress on the tail through the rudders "more than they were designed to handle". Thus, after initially concluding the accident was the result of a mechanical, the New York Times announced on January 5th 2002 that now the "crash inquiry is focusing on the pilots" (op. cit.).

By international regulation, a civil aircraft's tail is supposed to be able to withstand a force 50 percent stronger than the largest it is likely to ever encounter, and the A300 tail exceeded even that standard. However, certain rudder manoeuvres are known to temporarily increase a force that is seemingly larger than "likely ever to be encountered", and this had been known to aircraft manufacturers for a number of years. Crews at American Airline were trained in the mid-1990's to use the rudder to recover from "flight upsets", but Airbus, Boeing and the F.A.A. later warned against this practice, saying it could produce "dangerous stresses". American Airlines said it changed its training in 1999 to de-emphasize use of the rudder. However, the training's effectiveness might need to be evaluated, since evidence recovered from the plane's data recorder indicates that the pilots were using the rudder to try and stabilize the plane. This might have led to stress on the plane's tail that eventually led to the plane breaking apart. Furthermore, it might be argued that since it is known that these rudder manoeuvres create untenable material stress, pilot should be prohibited from carrying out these manoeuvres *through the system's design*, not through a "blame and train" approach. This, however, is likely to require further analysis of how pilots interact with the system in order to find an optimal technical design solution.

Implicating “pilot performance” in an accident’s causation still means that this human behaviour needs to be contextualized and analyzed in order to shed light on its origin and potential courses of action to improve future safety performance. Thus, identifying hazardous human involvement in accidents does not necessarily imply the identification of the actual ‘cause’ of the accident. The design of technology, task procedures, or organisational issues may be precursors to “human error”, for instance if they are not well matched to human capabilities and thus present an ‘unkind’ work environment that precipitates the occurrence of human error. Analysis of the human factor in an accident’s causation, therefore, cannot just focus on the human component alone, but must see it in relation and in interaction with the system which suffered the accident. This thesis builds on this insight and views errors and accidents as the result of a ‘mismatch’ between human capabilities and system design (see e.g. Rasmussen, 1982). It thus postulates viewing the human as ‘just another system component’ (with some set of known characteristics such as human cognitive capabilities and limitations). This human system component is seen as being in continuous interplay with the other system components, rather than as a malleable entity external to (and interfering with) the system. Thus, this ‘human factors’ approach to accident analysis (which corresponds to a ‘human computer interaction’ approach when dealing with computerized systems) stresses the need to consider the *interaction* between the human and the system in the accident’s causation, with view of determining potential mismatches that could be remedied through system re-design.

The understanding of human error and its relationship to the overall system is thus essential to further our understanding of accidents in complex systems. Insights into the causes and the course of accidents will help avoid similar events and situations in the future. For this goal it is necessary to more thoroughly understand why “human error"[1] occurred and to be able to offer a sound basis for safety recommendations from the analysis process.

## BACKGROUND

Human Error analysis in incident and accident investigation is often supported by error taxonomies or classification schemes (Maddox and Reason, 1996). These provide 'buckets' of common characteristics of error classes in which instances of assumed erroneous behaviour can be categorized. Some of these taxonomies aim at being grounded in some theoretical framework (e.g. a psychological theory of human action), others aim at being comprehensive in describing surface characteristics of incidents, and thus tend to be more domain-dependent (i.e. the behaviour and circumstances are not described generically, but in relation to their situated work context). These taxonomies also offer a terminology to describe, distinguish, and compare errors, or instances of erroneous behaviour.

However, these error categories can be ambiguous, vague, and overlapping (Busse, 1998). This is further exacerbated by error data often being under-specified and conflicting, requiring a high degree of interpretation on part of the analyst. For instance, a medical incident analyst might be presented with a description of a proximal incident cause that reads "forgot to check heart monitor". Given analysis

---

[1] "Human Error" is the term that is commonly used in contemporary work on the subject to denote human involvement in an accident's causation. It is an arguably controversial term, but it will be used throughout this dissertation (interchangeably with related terms such as the 'human factor' in accidents' causation, or 'human involvement'). The decision to use this term was made not only since the term has been coined by leaders of the field (e.g. Reason, 1990), but also because the juxtaposition of this term with the central thesis of this work (the relativity of 'human error') will further highlight the benefit of the suggested error analysis approach.

taxonomies such as the one used in the Australian Incident Monitoring Scheme (AIMS) (Runciman et al., 1993), the incident analyst will then have a choice of categories such as “inattention”, “fatigue”, “haste”, or “stress” to document the underlying ‘cause’ of the human error. Clearly, any choice of categories in this situation must be pure conjecture, relying solely on the “good judgement” of the analyst. Furthermore, such categories do not provide much substance or discriminatory power to help describe, distinguish, and compare errors in a meaningful way.

Thus, these error taxonomies often do not attempt to provide a ‘causal’ explanation as to why the error occurred. Instead, they list descriptive categories of what happened, and how. (I will return to this in more detail in Chapter 5 when evaluating the Edinburgh Incident Reporting Scheme).

There are more analytic taxonomies that ground the classification of erroneous behaviour in theoretical frameworks. For instance, Reason’s (1990) error taxonomy bases its error categorization scheme on generic underlying cognitive mechanisms, such as cognitive biases (e.g. it defines cognitive biases such as the “confirmation bias” which has been shown to hold for human decision-making: humans tend to form an initial hypothesis to understand a situation, and thereafter are *biased* towards only collecting evidence which further confirms this initial hypothesis[2]). But even these present in their categorization at most the result of a ‘causal’ analysis, but not how these conclusions were arrived at. Thus the path from the incident description to the categorization of the instance of human behaviour is a) based on the analysts ‘good judgement’ and b) not documented.

---

[2] It could be argued that this is one reason why “scientific” (i.e. unbiased and objective) decision-making does not come natural to humans, and why there needs to a rigorous method in place (such as the idea of ‘hypothesis falsification’) in order to avoid un-scientific, “common sense”, decision-making tainting analytic conclusions (e.g. Popper, 1934).

The rationale for choosing one error category over another is not documented, but left implicit. In this way, the reasoning behind the safety recommendations that result from the investigation process often cannot be traced back to the original error analysis. Safety recommendations provide the bridge between accidents and the prevention of future occurrences through advocating changes in the total system's design. Thus, the error analysis process will impact system redesign through its resulting recommendations, and would benefit from being documented.

The following list summarizes the weaknesses of current human error analysis approaches that this thesis will try to address.

In the case of error analysis by means of classification (common for incident reporting schemes):

- Incident and accident related error data is often underspecified and conflicting. Several layers of interpretation thus taint the eventual result of the incident's analysis. Especially in incident reporting, this is exacerbated by inadequate data collection tools and mechanisms, and poor staff awareness.
- Domain-dependent taxonomies are not generalizable (which prohibits e.g. cross-industry fertilisation, or even cross-department comparisons and exchange), they are inflexible and thus not change absorbent (cannot accommodate new technology, new procedures)
- Theoretical error taxonomies' categories can be ambiguous, vague, and overlapping, so that error classification does not show high reliability even given adequate data.
- Both domain-dependent and theoretical taxonomies can be pitched at different levels: the more detailed and low-level the classification the less

generalizability it affords, the more high-level and abstract the taxonomy, the less valuable information (understanding, precision in its specification, discriminatory value) it embodies and conveys.

- The classification process typically is not documented, nor is a justification of the classification provided. Classification decisions, especially in ambiguous category cases, can therefore seem arbitrary, and the taxonomies' reliability is further reduced.

In the case of error analysis without classification (the case of traditional accident reports):

- error analysis that is based mostly on 'common sense', i.e. that benefits from the analysts' expertise, but relies on implicit and highly individual knowledge
- Implicit use of error theories and taxonomies (such as 'slip') that have made their way into the expertise's knowledge base
- Imprecise and ill-defined use of error descriptors (such as 'mistake')
- The documentation of the analysis and conclusion finding proceeds in the form of the accident report. There are numerous problems associated with such a natural language based approach (Burns, 2000)
- There is a gap between the generation of safety recommendations and both taxonomical as well as natural language analysis data; intervention points for system redesign are not suggested or revealed by traditional analysis techniques. So far, the analyst's decision is left unsupported by the analysis methodology.
- In traditional analysis techniques there is no method of validating the chosen intervention points and the outcome of suggested interventions.

# THESIS GOALS

The principal goal of this thesis is to introduce an approach to support the human error analysis process that addresses its above-mentioned weaknesses. To this end, cognitive modelling techniques are employed as a tool to reason about competing interpretations of human error and its underlying cognitive processes. Thus, the goal for the approach that is put forward here is to guide error classification, while basing the error analysis in a cognitive-analytical framework rather than a behavioural-descriptive one. Furthermore, the reasoning process underlying the classification will be explicated and documented. An additional goal is for the cognitive error modelling approach to strengthen the bridge between the analysis of the 'causes' of human error and the resulting safety recommendations, and thus measures for system redesign. The cognitive error modelling approach thus also aims at providing support for the choice of system redesign decisions by grounding them in cognitive theory. In this way, the link between the analysis of human error and its implications for safety recommendations is made explicit and substantiated by an analysis that is situated in a cognitive theoretical framework.

This benefit of this error analysis approach will be illustrated throughout this dissertation by use of real-life case studies drawn from the domains of aviation and medicine. An incident reporting scheme was implemented from scratch in a Neonatal Intensive Care Unit addressing above-mentioned issues of data collection and staff awareness, and demonstrating the cognitive error modelling technique in the context of a clinical safety management framework.

# PROOF OF CONCEPT

The cognitive error modelling approach that is put forward in this thesis was piloted in an initial study that involved analysing user error when interacting with the Netscape internet browser (Busse and Johnson, 1998) (see Appendix A). The study confirmed the above-mentioned weaknesses of commonly used human error taxonomies, and the need for documenting the error analysis process. It was concluded that this approach of cognitive error modelling supported the analysis process. It offered a framework and method for evaluating competing error categorisations based on cognitive theory. Thus, interpretations of user behaviour and hypotheses as to its motivation could be expressed, compared and evaluated in a cognitive framework. The results also showed how design recommendations could be based on an in-depth cognitive analysis of user error, which thus provided a means for justifying and documenting design choices. Thus, this pilot work showed promising results for the cognitive error modelling framework to be applied to error analysis. The observations made in this investigation of user error refined the goals for this thesis.

# CONTRIBUTION TO THE FIELD

The core contributions to the field of the work presented here is the examination of the weaknesses of current error analysis practices in incident and accident investigation, and addressing these with the application of a novel cognitive error modelling approach in a revised incident reporting framework.

This dissertation provides theoretical justifications and practical, worked examples of how the analysis of human error in accidents and incidents can be supported by cognitive error modelling. It provides a framework for reasoning about alternative classifications and causal hypotheses. It also deals with the relationship between error analysis and the resulting safety recommendations. A goal of this work is to support the generation and corroboration of safety recommendations by grounding error analysis and its implication for system redesign in a cognitive theoretical framework. This thesis gives examples of how an error analysis that is supported by cognitive error modelling can be used to compare choices of safety recommendations and substantiate system redesign decisions.

A further goal is to support the practice of human error analysis in a real-life application. Thus, the analysis is not limited to the retrospective investigation of error as described in accident reports. It could be shown (see chapter 5) that a post-hoc cognitive analysis of error data in accident reports can be used to evaluate safety recommendations.

However, accident reports are rhetorical documents (Snowdon and Johnson, 1998), and can serve to lay out an argument that justifies the investigation’s conclusions more than portray the analysis process. Therefore, the cognitive error modelling approach was also applied to error data drawn from incidents reports. Incidents are near-miss accidents, where the chain of events leading to an accident was recognised in time and the accident was prevented from occurring. This means, the potential causation of accidents is investigated, rather than a post-hoc analysis of the causation of an accident that actually occurred. Incident reports are a means of documenting those near-miss system breakdowns and are authored solely by the person who detected the deviation and prevented the accident.

Thus, the error data drawn from near-miss incident reports provide less of a rhetoric argumentation but an (albeit subjective) description and interpretation of events leading to a near-miss system breakdown. This description presents raw error data with no other level of interpretation than that of the witness filtering the information. Therefore, one of the goals of this thesis is to propose a cognitive error modelling approach to support the analysis of error and the validation of safety recommendations when dealing with data drawn from incident reports.

An intermediate benefit of the work presented here also is to test the transferability of human error analysis concepts and techniques researched in safety-critical fields such as aviation and process control to the domain of medical (clinical) environments from a safety-critical field with high demands for error management techniques (Leape et al., 1991). A further goal addressed in this report, therefore, is to cover a range of characteristically different domains, such as aviation and medicine, in the application of our cognitive error modelling approach.

Subgoals also include a comparative evaluation of different cognitive architectures and their use as error analysis supporting expressive frameworks. The architectures considered here are EPIC and ICS (see Chapter 3).

## CAVEAT

This thesis does not present a novel cognitive framework or architecture – it draws on existing work in the field of cognitive psychology. The error analysis approach in this thesis is also does not supposed to replace existing error taxonomies, but to complement their value and support their application. Furthermore, the cognitive error analysis approach that is proposed in this thesis has not yet been cross-validated in the context of human error analysis in incident and accident investigation. Currently, it has only been applied to the field of safety-critical systems analysis within the Glasgow Accident Analysis Group. The approach has been, however, cross-validated in its original field of applications, Human-Computer Interaction (Buckingham-Shum et al., 1996).

# CHAPTER BREAKDOWN

Chapter 2 will introduce the state of the art of error analysis in accident and incident investigation, and point out weaknesses in current error analysis methods.

Chapter 3 suggests the use of cognitive modelling techniques to aid human error analysis in accident and incident reporting. It investigates and outlines the relative strengths and weaknesses of different cognitive modelling traditions when used to tackle human error analysis. One focus of this chapter is to provide a comparative evaluation of two architectures as to their suitability to support the error analysis process. After having identified one suitable cognitive architecture for error analysis, the chapter demonstrates a proof of concept of how this cognitive modelling technique can aid human error analysis from data drawn from an aviation accident report. A further implication of the analysis carried out in the work presented in chapter 3 will be shown to be the value of cognitive modelling in the grounding of safety recommendations.

In Chapter 4, it is shown how this cognitive error modelling technique can be used to complement conventional error analysis. This will be demonstrated by retrospectively re-analysing the human factor in the causation of selected aviation incidents and accidents.

In Chapter 5, the error modelling technique is taken to a real-life setting (rather than being applied retrospectively to an existing report). The established safety management process in a real-life clinical environment is investigated, and it is shown how it can benefit from using the cognitive error modelling approach for its error analysis process as put forward by this thesis. The application of cognitive modelling to error data drawn from a clinical incident reporting scheme is examined.

This chapter also again explicitly tackles the relationship of human error analysis and safety recommendations for system redesign.

In Chapter 6, the process of establishing an incident reporting scheme that is specially designed for contextual human error analysis is documented. Conventional treatment of human error analysis has been shown to be at times insufficient and at worst misleading, and thus alternatives are spelled out in this demonstration of a safety management approach that is geared towards a meaningful, contextual analysis of human error.

The final chapter (Chapter 7) provides a summary of our findings, conclusions, and suggestions for further research.

# CHAPTER 2 HUMAN ERROR ANALYSIS IN ACCIDENT AND INCIDENT INVESTIGATION

## DEFINITIONS OF TERMS

In the field of accident analysis, there has been much discussion as to the difference between terms such as ‘failure’, ‘accident’, ‘incident’, ‘occurrence’, or ‘adverse event’. This is partly since they are used differently in the various domains that concern themselves with accident analysis (aviation, medicine, process control to name but a few). In this chapter, I will give definitions and an introduction to these key concepts of this thesis. I will also present an overview of the current state of the art research and practice in human error analysis in incident and accident investigation in safety-critical domains. Throughout this overview, I will highlight the role of human involvement in an accident’s causal chain, and how this is dealt with in the relevant research and practice.

In this dissertation, a ‘failure’ refers to a system failure, in which there is a “non-performance or inability of the system or component to perform its intended function for a specified time under specified environmental conditions” (Leveson, 1995). A system failure does not necessarily lead to undesirable consequences. It might not even ever be detected. ‘System failure’ is often used implicitly to denote a purely mechanical failure, with no human involvement. However, a human interacting with the system is necessarily to be seen as one of its components. This dissertation will explicitly note human involvement (or lack of it) whenever it is relevant when talking about ‘system failures’.

An ‘accident’ is defined as “an undesired and unplanned [..] event that resulted in (at least) a specified level of loss” (op.cit.). A incident (referred to interchangeably as ‘critical incident’) is here defined as an event “that involves no loss (or only minor loss) but with potential for loss under different circumstances” (op.cit.).

Leveson does not explicitly refer to the medical domain in her definitions of accidents and incidents. As is often the case, terms take on different meanings in other domains, and so the term ‘accident’ takes on a different meaning in the medial domain. What is referred to as an “accident” in e.g. aviation is referred to in medicine as an ‘adverse event’. Thus, in this thesis the definitions of accidents and incidents will be modified to suit the appropriate terminology as this work progresses to discuss “human error” in the domain of medicine. Other terms relevant to the medical domain in this context are the concepts of “iatrogenic injury” or “iatrogenic events”. These refer to “treatment-caused injury” (sometimes commonly referred to as “doctor-caused injury”), and these terms are specific to the clinical field.

‘Human Error’ is here defined as any (expected or unexpected) human behaviour (including lack of action) that led to undesirable consequences of some specified level (i.e. not including violations*,* and not of terrorist intent, or sabotage) in the case of an accident, or that could have led to undesirable consequences of (at least) some specified level, in the case of an incident (as defined above). “Undesirable consequences” includes a specified level of loss (in the case of an accident). “Unexpected behaviour” here refers to both “unplanned” (i.e. unexpected by the individual) and “planned behaviour”, and it is also set in relation to other people’s expectations. The latter is of course subjective, and thus often brought in relation to rules laid down by a working community, such as standards and protocols, but also implicit ‘best practice’.

It can also refer to behaviour that was planned by the individual, but where the plan was either inappropriate (i.e. it was correct plan to achieve some goal, but the wrong situation (or goal) to apply it to), wrong, or incomplete (again, as measured against the (potential) negative consequences of that behaviour resulting from the plan). Crucially, ‘human error’ as used in the literature can also refer to *expected* human behaviour (not in relation to an inappropriate, wrong and incomplete plan) that nevertheless resulted in some level of loss, or had the potential to do so.

The example given in the introduction of the November accident of Flight 587 is a case in point. The pilots had been trained to use the rudder, and to their best knowledge the rudder was supposed to adhere to the international standard of withstanding any pressure level that could possibly occur during flight. But when the rudder failed to withstand the pressure of their manoeuvring, accident investigation turned to the pilots’ performance as the “cause” of the accident.

The above definition of ‘human error’ reflects one interpretation of error that is prevalent in the literature (e.g. Reason, 1990) and it will be this definition that is referred to in this dissertation. However, this definition can invite contradiction and misunderstandings, and this thesis aims at elucidating why this is so, and suggests ways of dealing with ‘human error’ to avoid these pitfalls.

For instance, expected ‘work according to rule’ is near equivalent of “strike” in any trade-union’s book, and implicit rules are still to a large part subjective and flexible. Often, standards contain inappropriate, wrong, or incomplete rules that are compensated for by humans. They may even contain contradictory rules, which the nurse or pilot must choose from without guidance and at their own risk. Plans often are trained, and expectations are set in advance, often without choice or control of the individual. Furthermore, the design of machines, devices, interfaces, and work contexts might be more error-inducing than helpful, with humans often compensating for their weaknesses (Kanse and van der Schaaf, 2000).

‘Human error’ can also be seen as inevitable part of human behavioural variability, and thus as the flipside of the coin that also enables humans to think and act flexibly and economically, to compensate for poor machine design, poor work protocols, and sub-optimal work conditions, to recover a system from near-miss incidents back into a healthy state, and also often enough to prevent accidents from occurring. This view thus holds that the variability of behaviour that includes ‘human error’ is inevitable, and thus the design of the work system must take this human variability into account, and be error-tolerant, ‘forgiving’, and aim at an optimal match with humans’ abilities.

The above definition of ‘human error’ will require further discussion throughout this dissertation, especially, as will be shown, when placed in the context of categorizing accident causes and instances of human behaviour involved in the chain of causation. This is exacerbated by the need for any definition of human error in the context of categorization to enable a “relevant” description (i.e. providing an analytic, constructive vocabulary), the distinction and comparison of instances of human behaviour over time, and across individuals, situations, locations and even domains.

The dissertation examines accidents and incidents in “safety-critical systems”. These are here defined to be systems whose incorrect function (i.e. failure) may have very serious consequences, such as loss of human life, severe injuries, large scale environmental damage, or considerable economical penalties. The safety-critical systems dealt with in this dissertation are the domains of aviation and clinical medicine (intensive care). Other domains that are defined as safety-critical and that the field of accident analysis deals with are e.g. process control (especially in nuclear power plants and chemical plants, but also manufacturing), the naval domain, automobile engineering and traffic, fire safety (internal and external) and associated crowd control. These are the fields that most of the work relevant to this thesis’ accident analysis aspects is informed by. In the following section, I will introduce the current state of the art in accident analysis research as the basis of this thesis.

# ACCIDENT ANALYSIS

Accident analysis should be an integral part of safety management in any safety-critical organisation or industry. Morone and Woodhouse (1986) identified “learning from experience”, which takes place thorough incident and accident analysis, as one of five vital strategies for risk regulation in safety critical domains, next to e.g. “protection against potential hazards”, and “prioritising potential risks”.

Next to the financial implications of the loss associated with accidents, there is also the moral obligation of accident prevention towards the employees that are at risk in safety-critical industries, towards anyone who might be impacted by the negative consequences of accidents, and also to society at large in terms of economic impact or, in the worst case, large-scale loss of life. Public scrutiny therefore rests on these industries, which will not only have to try their best to prevent (seemingly inevitable) accidents from happening, but also present a transparency of their decision-making processes concerning safety management, documenting any measures taken in the name of accident prevention, and their justification.

The International Civil Aviation Organization (ICAO) document of “International Standards and Recommended Practices: Aircraft Accident and Incident Investigation - Annex 13” (ICAO, 1994) details the standards and recommended practices of reporting, analyzing, and acting upon civil aviation accidents and incidents. I will use this standard as a baseline to illustrate the current state of the art in accident analysis practices in this chapter. In contrast, the later sections in this chapter will delve deeper into current accident analysis *research*.

Standards and Recommended Practices for Aircraft Accident Inquiries were first adopted by the ICAO Council in 1951 (op. cit.). The document details the duties and recommendations for every contracting state with respect to accident investigation, and especially, as is stressed in the document, the prompt dissemination of accident occurrences, and of the analysis results. The ICAO document clearly lays down that “the sole objective of the investigation of an accident or incident shall be the prevention of accidents and incidents.” It further stresses explicitly that “it is not the purpose of this activity to apportion blame or liability” (op. cit.). The ICOA recommended format of aviation accident and incident reports is taken as a blueprint (or at least as a framework) for accident reporting by ICOA’s contracting states. The following lists the sections and accident report format that is recommended by ICAO (see Figure 2-1).

It is noted that much diverse expertise is brought to populating these sections – meteorologists, metallurgists, fire safety experts, and of course aircraft specialists, and many others, all contribute to the analysis of these sets of factors. Under analysis, ICAO recommends to “analyse, as appropriate, only the information documented [under “Factual Information”] and which is relevant to the determination of conclusions and causes.” There is only one section out of 19 main sections of the report format’s body, however, that concerns itself with the human factor in the accidents causation, “personnel information”.

This section, “Personnel information”, includes the following, according to ICAO:

1. Pertinent information concerning each of the flight crew members, including: age, validity of licenses, ratings, mandatory checks, flying experience (total and on type), and relevant information on duty time.
2. Brief statement of qualifications and experience of other crew members.
3. Pertinent information regarding other personnel, such as air traffic services, maintenance, etc. when relevant.

These are all factors that impact the flight crew members’ *psychological* predisposition, and their psychological disposition (their mental abilities, decision-making powers, concentration levels etc.) at the time of the accident. However, it is unclear how these factors got to be included in this list of “factors influencing human cognition”, (do they all stem from one coherent psychological theory?), or why other psychological factors were left out.

On another level, however, it is, surprising that ‘human error’ does not warrant its own analytical section, given the fact that it is so often described as the major single cause of modern accidents (up to 80% of aviation accident and medical adverse events are frequently cited as the accident’s cause, see e.g. (Busse and Johnson, 1999)).

There seems to be a tendency to assume that if ‘the system’ ostensibly did not do wrong, there must have been a mistake by the people who operated the machines. This is corroborated by the observation which I will return to in later chapters, that accident causation taxonomies tend to focus on the classification of technical causes (material conditions), and environmental conditions, such as weather. The part of those taxonomies that relates to “human error” often serves as a bucket for instances of human behaviour that are associated with one or other non-human causal event in the accident chain for which no sufficiently satisfying technical explanation can be found (Hollnagel, 1983).

1. Title
2. Synopsis
3. Body
    a. Factual Information
        i. History of the flight
        ii. Injuries to persons
        iii. Damage to aircraft
        iv. Other damage
        v. Personnel information
        vi. Aircraft information
        vii. Meteorological information
        viii. Aids to navigation
        ix. Communications
        x. Aerodrome information
        xi. Flight recorders
        xii. Wreckage and impact information
        xiii. Medical and pathological information
        xiv. Fire
        xv. Survival aspects
        xvi. Tests and research
        xvii. Organizational and management information
        xviii. Additional information
        xix. Useful or effective investigation techniques
    b. Analysis
    c. Conclusions
    d. Safety Recommendations
4. Appendices

**Figure 2-1 -** ICAO recommended accident report format

There is typically no justification, however, why that instance of human behaviour is assumed to have *caused*[3] this accident chain[4], rather than just being associated (i.e. correlated) with it[5]. Moreover, the "analysis" of human factors that then takes place often is constrained to take place within the boundaries of this category, neglecting the interaction of human behaviour with the environmental context. Whether the human involvement in an accidents' chain of events is explicitly considered or not, it is often implied as the 'final' cause (or the "root cause") in the absence of any 'satisfying' technical or environmental factors. It has been noted that this decision often is a highly subjective one[6], and especially the interpretation and analysis of that instance of human behaviour that is reasoned to have 'caused' the accident is often implicitly influenced by background 'knowledge' and assumptions on part of the analyst[7]. This phenomenon has also been investigated empirically by Lekberg (1997), with comparable conclusions.

---

[3] See the second half of this Chapter (on 'Human Error') for a discussion and definition of the term 'cause'.

[4] Benner refers to the beginnings of accident analysis theory to explain this phenomenon: He states that "the statistical work of Greenwood and Woods in 1919 [..] suggested the "accident proneness" concept. Their work still influences some accident investigation [..]. Investigators still look for data in accidents that will support the idea that "conditions" such as attitudes, attentiveness and so forth "cause" accidents. This statistical work focused on static conditions and set the pattern for untold man years of research into "unsafe conditions" as causes of accidents" (Benner, 1975).

[5] The confusion between correlation and causation is common in "common sense" thinking, and has been pointed out especially by statisticians repeatedly (e.g. Blalock, 1972).

[6] See for instance, (Snowdon and Johnson, 1998), and Snowdon, 2002.

[7] "Investigators unconsciously base their investigative methods on methodologies adapted from their academic disciplines or previous work experience; this leads to highly individualized, personalized investigative methodologies" (Benner, 1981).

Furthermore, point 3d ("Safety Recommendations") only states "[a]s appropriate, briefly state any recommendations made for the purpose of accident prevention and any resultant corrective action. There is no recommendation of linking the "causal analysis of factual information" to the generation of safety recommendation, as would be necessary in order to trace back the justifications for certain recommendations, and in order to review the recommendations in case other factors come to light, and the analysis process is to be revised.

Of interest is the inclusion of point of 3.a.xix: "useful or effective investigation techniques". This section recommends that "when useful or effective investigation techniques have been used during the investigation" analysts should "briefly indicate the reason for using these techniques and refer [..] to the main features as well as describing the results under the appropriate subheadings [..]." This point is of interest since its frequent omission in ICAO inspired accident reports leads to the question whether the investigation techniques used are indeed hardly useful or effective enough to be mentioned[8].

The official accident investigation handbook by the US Department of Energy (DoE, 1997) similarly short-changes human error investigation methods. In chapter 7, it lists recommended analysis techniques to address critical areas of investigation. Under "advanced methods", fault tree techniques such as MORT (Management Oversight and Risk Tree, e.g. Knox and Eicher, 1980), but only under "other analytic techniques" can a mention of human factors analysis techniques be found. To quote, it states that "Human factors analysis identifies elements that influence task performance, focusing on operability, work environment, and management elements. Humans are often the weakest link in a system and can be the system component

---

[8] see for instance http://www.ntsb.gov/ntsb/query.asp, the US National Transportation and Safety Board's Aviation Accident Database and collection of aviation accident synopses (current January 8th, 2002).

most likely to fail. Often machines are not optimally designed for operators, thereby increasing the risk of error. High-stress situations can cause personnel fatigue and increase the likelihood of error and failure. Therefore, methods that focus on human factors are useful when human error is determined to be a direct or contributing cause of an accident.” But despite these insights, and the fact that 60-80 % of accidents are said to involve ‘human error’, the DoE handbook does not list or recommend any specific human factors investigation technique.

## Accident Causation

Accident analysis is informed, albeit often implicitly (Bibbings, 2001), by accident causation models.  In accident causation models, as is shown below, any system’s human component is central to the system description that forms the framework to describe the web of contributing factors that led to the accident. Human involvement in any accident is seen as inevitable when looked at from the perspective of ‘active and latent’ human error (see chapter 5 for an in-depth discussion). Whatever the focus of the investigation, accident causation models make clear that the accident’s causation needs to be viewed from several different levels of system components.

## Accident Causation: Causal Trees

Arguably one of the most common and intuitive approaches to the analysis of causal factors that led to the accident, next to the chronological timeline of events, is the “causal tree” family of methods.

An accident seldom happens in isolation: usually it is a concatenation of events. This idea initially gave rise to the concept of a ‘causal chain’  (e.g. Heinrich's Domino

Theory, 1936). However, this concept is mostly unhelpful in a meaningful analysis of the accident, since it is an oversimplification that ignores the interplay of causal and environmental factors (see also e.g. Perrow, 1984).

The reality is likely to be a complex web of interacting events, culminating in the accident. To grasp such a web of events while investigating the accident can be difficult. To present it in purely verbal form in an accident report, in such a way as to be comprehensible to even a well-informed reader, can be very difficult, if not impossible (Zotov, 1996).

Fault trees are a family of analysis and investigation methods that support the uncovering of the multitude of casual factors during accident analysis. Most fault tree methods are diagrammatic, and can facilitate the presentation of the culmination of events leading to an accident. They require the system, task, procedure, or component in question to be logically decomposed into functional elements. Once the relationships among these elements are identified, a diagram is developed that depicts the elements and their relationships. A typical fault tree consists of an outcome (either successful operation or a particular failure) shown at the top, or most general level, of the tree. Elements that contribute to the outcome are listed below it in the diagram with connections (branches) showing the logical relationship between the element and the outcome. Once an outcome is defined, then contributing factors are identified and placed into the diagram. The appearance of a fault tree is that many elements combine, in pyramid fashion, to produce a single outcome. The trick is to identify all of the underlying elements (and combinations of elements) that might combine to produce a particular outcome (Reason and Maddox, 1996).

Fault trees are used specifically in the field of Probabilistic Risk and Safety Analysis (e.g. Fullwood and Hall, 1988), and some variants of it concern themselves explicitly with ‘human error’ next to technical risks and faults.

For instance, the NTSB accident/incident database contains sections with factual elements (“who, what, where”), sequence of events, and a narrative description of the incident. The NTSB uses a multiple-cause classification system with allowance for up to seven “occurrences”, and several “findings” for each occurrence. Any of the findings may be prescribed as a cause or contributory factor. The probable cause statement is usually an integrated text. Each “finding” may, in turn, be divided into “subjects”. The subjects come from a coded list and refer to whether the finding was person-related, and whether it was a proximal cause or not (NTSB, 1998). Therefore, the database allows for a causal tree being constructed by ascribing the level of proximity of the findings to the reported occurrence. Thus, it can be distinguished between causal factors and conditions facilitating the occurrence, and suggested root causes. However, the available constructs largely identify who was involved in a finding but often stop short of an assessment of why the error was made. Even the probable cause statements are largely descriptive in nature, without reference to latent failures or human information processing. This information, however, is important in developing preventive strategies (Luxhoj et al., 1997).

It has thus been pointed out repeatedly in the recent past that accidents that presumably result from “human error” are better characterized as a complex interplay of technical, social, organisational, and managerial factors (e.g. Leveson and Turner, 1993). Thus, the design of the operating environment - displays, controls, and so on (Norman, 1983; 1990; Rizzo et al., 1996) - has profound implications for the ability of the system to be maintained effectively and safely. More progressive models of accident causation attempt to capture the broader view on human error, taking the ‘total system’ view. Two established models that take a broader view are Helmreich’s and Reason’s. Both aim at contextualising the human factors in accident causation, by placing the human and the ‘erroneous’ behaviour in relation to other factors, such as the environment, and also the technical system. These two models are introduced in the upcoming section.

## Helmreich’s Accident Causation Model

Helmreich envisaged e.g. aircraft crews as working within a series of environments, each of which might put pressure on the crew, and degrade their performance. This series of environments was visualised as concentric spheres of influence, each affecting those inside (see Figure 2-2). In this diagram, the innermost environment concerns matters among the crew such as communication, personality, and Crew Resource Management. The crew is affected by the physical environment: the aircraft, with its idiosyncrasies, defects, and performance characteristics; the weather, both local and general; and the aerodrome environment. Outside these is the organisation of the airline, which purchased and maintained the aircraft, trained the crews, and should support their actions. Surrounding all of these is the regulatory environment, in which regulatory action by e.g. Transport Canada (i.e. the relevant regulatory authority) should ensure safe standards of operation.

## Reason’s Accident Causation Model

To produce effective recommendations from accident analysis, the information collected and the conclusions reached must be analysed in a way that reveals the relationships between the individuals associated with the occurrence, and the design and characteristics of the systems within which those individuals operate.

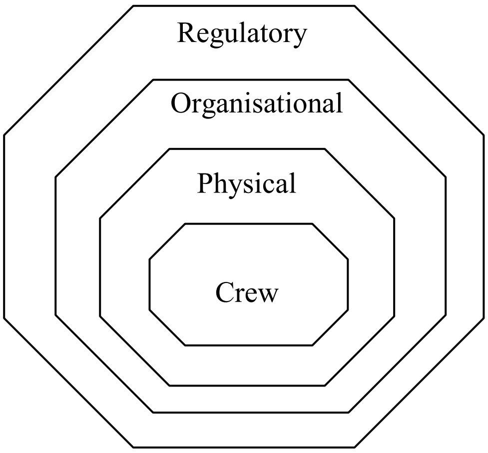


**Figure 2-2 -** Flightcrew Environments: Factors Influencing Behaviour

(Helmreich et al., 1990)

For the purposes of broad systems analysis, an analytical model researched and developed by James Reason can be used. The Reason accident causation model is becoming an industry standard, and has been recommended by ICAO (1993) for use in investigating the role of management policies and procedures in aircraft accidents and incidents (Zotov, 1996).

Central to Reason's approach is the concept of the 'organisational accident', in which latent failures arising mainly in the managerial and organisational spheres combine adversely with local triggering events (weather, location, etc.) and with the active failures of individuals at the 'sharp end' (errors and procedural violations) (Reason, 1990).

The relationship between these elements in the process of accident causation is shown in the accompanying diagram (Figure 2-3).

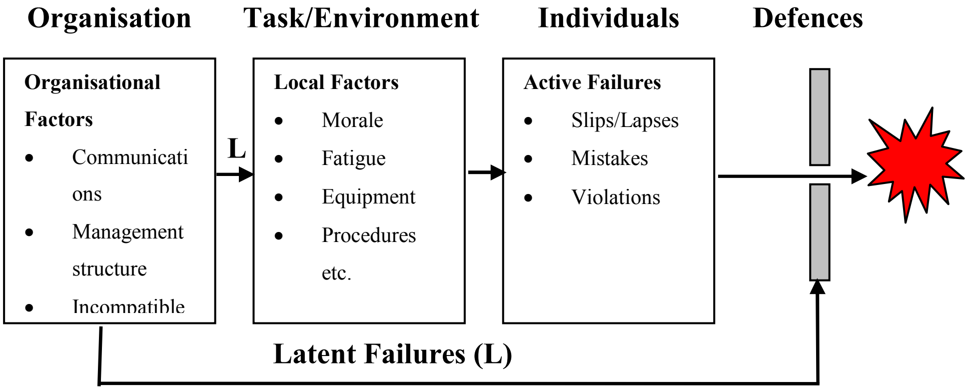


**Figure 2-3 -** 'Organisational Accidents' (Reason, 1997)

Thus in both these influential accident causation models, 'human error' as a causal factor is seen as intertwined with 'causes' of very diverging origins – from technological failure to managerial oversight, to organizational weakness or vulnerability (Johnson, 1999). These factors are demonstrated to not only stand in relation with the 'human error' but more drastically, the instance of 'erroneous' human behaviour would not have occurred were it not for the circumstances. Thus, the situational context can be seen as *defining* the causation of "human error".

## INCIDENT INVESTIGATION

Accident investigation is a form of post-hoc analysis, i.e. it is done after the fact. The underlying goal of any accident analysis is to prevent future accidents. Accident analysis is employed towards this goal (as stated in the above-mentioned ICAO reporting standard), since it is reasoned that an understanding of one accident's causation can lead to safety recommendations that improve on the system sufficiently to prevent its reoccurrence[9].

This pressing motivation (the goal to prevent future accidents) has led to the idea of "early warning systems" (Van der Schaaf et al., 1991). This refers to incident reporting schemes (or: "Critical Incident Reporting"), i.e. schemes that record *near-miss* accidents. These are seen as indications of accident potential within a system, and it is reasoned that if the thus identified weakness in the system is addressed as soon as a near-miss accident is identified, then actual accidents can be prevented from occurring. Not surprisingly, the concepts of causal analysis for accidents and incidents are closely related.

In many medical incident reporting schemes, in-depth analysis and a search for the root causes of adverse events does not take place (Leape, 1994; Chappell, 1996). In contrast, the formal investigation of adverse events in industry is an increasingly well-established concept. Studies of accidents in industry, transport and military spheres have led to a much broader understanding of accident causation, with less focus on the individual who makes the error, and more on pre-existing organisational factors.

---

[9] The shorter term view on the benefits of accident analysis also stresses the importance of *prompt dissemination* of the accident's occurrence, the identified circumstances and causation, and the "lessons learned", i.e. the safety recommendations. In the case of an aviation accident, or course the airline company needs to be notified promptly, as well as the manufacturer. But also other airlines that fly the same planes, or use the same procedures, or use the same runways etc. need to be notified. This *communication* of accident data is seen as crucial in the overall strategy of future accident prevention (ICAO, 1994).

For instance, Reason's organisational accident causation model (see Figure 2-3) shows how latent failures can pave the way for an accident to occur. To produce effective recommendations, the information collected must be analyzed in a way that reveals the relationships between the human error that occurred, the design, and characteristics of the systems.

The benefits of analysing a case using a formalised model are that it allows an analysis in a structured format based on theories of accident causation and human error. This type of analysis allows analysts not only to identify the active failures, which they are accustomed to do, but also the potentially more important latent failures which create the conditions in which people make errors. Stanhope et al. (1997) suggested that a more systematic approach dealing with a smaller number of cases in more depth is likely to yield greater dividends in understanding incident causation and generating action recommendation than the 'many' cases currently analysed quite briefly and hence less effectively.

Barnhart et al. (1975) present a seminal paper on NASA's accident investigation method. Combined with Reason's accident causation model (Eagle et al., 1992), it can be described as consisting of three steps:

"When":

- Assembly of facts and generation of timeline
- Embed event within task sequence

"What":

- Identification of active failures
    - Description of all behaviours
    - Root Cause Analysis, assembling proximal and distal causes and work conditions factors

“Why”

- Identification of latent failures, for instance:
    - “Information Processing Analysis”, e.g.:
        - sources of information available (latent)
        - what information perceived, used, and in what ways
        - which decisions made, listing all choices
        - whether actions matched decisions (latent and active)
        - consideration of managerial and organisational failures; follow-up of root cause analysis

Therefore, this model extends the above-described NTSB model by including the analysis of the “why” of the incident. The incident analysis proceeds from identification of ‘active failures’ and local working conditions that precipitate those to the identification of latent system failures. For instance, Reason (1990) has proposed a list of General Failure Types of organizations. They include: incompatible goals; organizational deficiencies; inadequate communications; poor planning; inadequate control and monitoring; design failures; inadequate defences; unsuitable materials; poor procedures; poor training; inadequate maintenance management; and inadequate regulation.

I will return to incident reporting in chapters 6 and 7 of this thesis (and in detail also in Appendices B and C).

# THE ROLE OF HUMAN ERROR

## Introduction and Definition of Terms

Next to accident causation models, accident analysis is typically informed just as implicitly by the analyst's model of the nature of human error[10]. The human factor is a central part in the above-mentioned accident causation models. The human operators can thus be seen as another "system component" that needs to be investigated in accident and incident analysis. Therefore, methods are needed that specialize in the analysis of the human contribution to an accident's causation.

There are many theories on the causation and underlying mechanisms of human error, just as there are as many on the general workings of the human mind. The human error models used in accident analysis, however, (or those found in the literature on accident analysis, even if they are not explicitly used in the process) are not solely influenced by psychological theory.

Human error research is traditionally not bound to any one particular discipline, but is distributed over several communities, such as psychology, reliability analysis, system, software and usability engineering, and cognitive science. Within each paradigm, the domain specific perspective dominates definition and scope of the term 'human error'. As we will see later, research into error focuses on those aspects of the topic that are relevant to the respective discipline or perspective.

The First Human Error Conference in Maine (see Reason, 1990) (and the preceding Three Mile Island accident) in 1980 brought into being a human error research community that started to pool the efforts scattered across disciplinary boundaries.

---

[10] There are some notable examples where the theories that were assumed to explain "erroneous" human behaviour were made explicit in accident reports (BASI, 1994).

The body of research has grown considerably since then, lead prominently by cognitive engineers and psychologists such as Jens Rasmussen, James Reason, and e.g. Veronique DeKeyser.

In order to review the field of error research, I will first draw attention to different views on what constitutes an 'error', and specifically, a 'human error'.

Within psychology, human error is typically viewed as a deviation from an individual's preliminary intention to reach a norm (Bes and Johnson, 1998). Error is seen in the context of cognitive processing and decision-making models (see for instance Ashcraft, 1994). For instance, Reason states that human error occurs because of either inappropriate planning (action specification), or inappropriate action execution on side of the individual (see Norman's Action Theory below).

In a system failure[11] context, the term 'human error' might be misleading (see for instance Leveson, 1995). An 'error', according to Reason (1990), is the consequence and not tantamount to 'erroneous action'. The term 'human error' is seen by many[12] as conveying oft-unjustified responsibility and blame in a system failure context.

---

[11] A Failure is defined to be the non-performance or inability of the system or component [including the human as a component] to perform its intended function for a specified time under specified environmental conditions (Leveson, 1995)

[12] Rasmussen (Rasmussen et al., 1987, see below), in his error taxonomy uses the descriptive phrase 'external mode of human malfunctioning' (inappropriate task performance) in order to avoid the term 'human error' "with its flavour of guilt".
Norman (1990) suggests that: "one major step would be to remove the term 'Human Error' from our vocabulary and re-evaluate our need to blame individuals". This is an attempt to shift away from blame, error, and cause. The search for cause, as Norman (1990) further suggests, ends the search for meaning, for once we have found the cause we can 'explain' it away. The idea of 'cause' can be seen as a legacy of the legal community (Hale et al., 1997).

Senders and Moray (1991) point out one useful distinction between system fault and human error. In the former the system did not provide the functionality necessary for performing certain tasks. In the latter, the system did provide the ‘tools’ (such as actions, operators and information) for carrying out a task, but the user failed to ‘use’ them. However, this distinction poses further questions.

For instance, Rasmussen et al. (1987) point out that, in accident analysis, the identification of an event as a human error depends entirely upon the stop rule applied for the explanatory search after the fact.

The backtracking of the causal chain will stop, according to Rasmussen et al. (op. cit.), when one or more changes are found which are familiar and therefore acceptable as explanations, and to which something can be done for correction. He states that the characteristics of a fault (which might include human error)[13] are:

1. It is the cause of deviation from a standard
2. It is found on the causal path backwards from this effect
3. It is accepted as a familiar and therefore reasonable explanation
4. A cure is known

This means, he says, that allocation of causes to people or technical parts in the system is a purely pragmatic question regarding the stop rule applied for analysis

---

Hollnagel (Hollnagel, 1991)argues the idea of ‘human error’ should be removed and replaced with a less emotive term like ‘erroneous action’.

[13] A fault is to be distinguished from failure (Leveson, 1995). Failure is defined to be the event (a behaviour) (i.e. something fails). All failures are faults. Not all faults are, or lead to, failures. Failures are basic abnormal occurrences. Faults are higher-order events.

after the fact. There is thus no defined ‘start’ of the causal chain involved in accidents, and this fact should be reflected in the error data taxonomy.

Events, according to Rasmussen et al. (op. cit.), should be considered occurrences of a man-machine mismatch, which can only be characterised by a multifaceted description (see below). Faults and errors cannot be defined objectively by considering the performance of humans or equipment in isolation.

Furthermore, since the system itself is built, programmed, and maintained by humans one can argue that faults are ultimately human error. Indeed, one school of thought seeks to ascribe the cause of all failures to human origins (Bignell and Fortune, 1984).

Reason picked up on this by pointing out, as noted above, that an error is not to be seen as the accident or failure itself, but as preceding the failure (Reason, 1990). ‘Error’ can lead to ‘active failure’ as well as ‘latent failure’.

Reason maintains that active failure is usually associated with the performance of ‘front-line’ operators (such as pilots and control room crews) and has an immediate impact upon the system. Latent failure is most often generated by those at the ‘blunt end’ of the system (designers, high-level decision-makers, construction crews, managers, etc.) and may lie dormant for a long time.

In this thesis I will concentrate on ‘human error’ leading to active failure. The analysis of these is most likely to benefit from cognitive modelling techniques. Latent failures are best analysed in ‘total system’ approaches, which go beyond individuals’ cognition and take for instance organisational aspects into account (see also Reason and Maddox, 1996; or Hale et al., 1997).

# THE NATURE OF HUMAN ERROR: MODELS AND TAXONOMIES

Given a more defined scope of ‘error’, the question ‘what is human error, and how does it come about?’ is still subject of discussion, and several theories and approaches have been put forward, each with their perspective on this issue, or sometimes building up on, or modifying, previous work (e.g. Reason, 1990). I will discuss some of the most influential approaches in the following sections.

Error classification’s distinction between varieties of human error are important because different error types have different underlying mechanisms (different psychological origins), occur in different parts of the system, and require different methods of error management, and remediation (Reason and Maddox, 1996).

Although a classification system is not likely to give insights into the underlying causes of error, Freitag (1997) notes that “the philosopher’s stone which we are all seeking is how to recognise the events we can learn from without actually having to analyse them all in detail first” (compromise between the generation of enough events to learn from, and avoiding swamping the analysis system with too much work which will cost more than its added value for improving management).

The distinction between varieties of human error plays a significant role in accident analysis because different error types have different underlying mechanisms (different psychological origins), occur in different parts of the system, and require different methods of error management and remediation (Reason and Maddox, 1996). The nature of ‘human error’ is still subject of discussion, and several theories and approaches have been put forward. I will discuss some of the most influential approaches in the following sections, notably Rasmussen’s Skill, Rule, Knowledge (SRK) framework, and Reason’s Generic Error Modelling System (GEMS).

## Behavioural Classification

Norman (1981) pointed out that error data examination reveals that errors can be categorised and fall into patterns. However, the categorisation and interpretation is theory dependent. There have been several attempts to categorise human error in terms of behaviouristic criteria (errors of commission, of omission, of substitution) (op. cit.).

Most notably, Hollnagel (1991) suggested focusing on the overt behaviour of the human leading to, or constituting an ‘error’.

Thus, Reason (1990) argues that Hollnagel’s error classification scheme is an example of a behaviour-based taxonomy of human error. It starts with the observable phenomena, such as errors of omission or commission, rather than cognitive theories. The observable phenomena, he states, make up the empirical basis for error classification. He referred to these behavioural descriptions as the *phenotype* of human error. The error *genotype* denotes the mental mechanism hypothesised to underlie the overt erroneous behaviour.

The main aim of this approach at the time was to develop a system that can recognise possible erroneous actions and alert the user of their occurrence. From this point of view, Hollnagel argues, it is in principle sufficient to be concerned only with the phenotype of erroneous actions. The purpose of his taxonomy of phenotypes of erroneous actions is to establish a basis for a working, computationally implemented, action classification system. It does not attempt to explain the erroneous actions.

Behaviourist classifications such as Hollnagel’s do not aid in understanding the underlying mechanism. Hollnagel points out that the phenotype is useful for computerised detection, but, since it does not provide any explanation, it cannot be used for correction and improvement. For that purpose a theory of error genotypes is needed. This will be a model description of human erroneous behaviour, which can be useful in clarifying how the user-system interaction takes place, and thus how it can be improved when necessary (op. cit.).

Moreover, behavioural classifications quickly become large and unwieldy. A complete error theory seems likely to require autonomous, subconscious processing, with intentions, past habits, thoughts, and memories all playing some role in corrupting the intended behaviour (Norman, 1981). Also, Rasmussen (1987) observes, that if the analysis is based only on the consideration of human error in terms of their external manifestation (such as omission, commission, and inappropriate timing) a priori error identification will be hindered by a combinatorial explosion.

## Conceptual Classification Schemes

At the conceptual level of classification, assumptions about the cognitive mechanisms involved in error production are the basis for human error taxonomies (Reason, 1990). In contrast to the behavioural level, these classifications are based more upon theoretical inference than on observable characteristics. Conceptual categorisation schemes seek to identify causal mechanisms underlying human error.

## Action Theory and Errors

Theories of action present core concepts of applied cognitive theory. They often form the basis for the conceptual error models introduced below. For instance, Norman’s Seven Stages of Action model (Figure 2-4) combines a simple cognitive processing model with action stages of human-computer interaction. It suggests how an intention is represented and acted upon. He articulates how the Gulf of Execution (stages of intention, specification of action sequence, and execution) and the Gulf of Evaluation (stages of perception, interpretation, and evaluation) bridge between the user’s goal and the world.

Several errors can arise along the line of these gulfs. Norman in particular distinguished between slips and mistakes. Mistakes were defined as inappropriate planning (action specification), whereas slips are inappropriate action execution. This distinction, along with the assumptions embodied in the above action theory play a major role in the following conceptual error models.

## Level of Performance and Human Error

One influential model of human decision making in complex environments is Rasmussen’s (1986) decision ladder (Figure 2-5) that distinguishes between skill-, rule-, and knowledge-based behaviour. As Reason (1990) and Sanderson and Harewood (1988) pointed out, Rasmussen’s Skill-Rule-Knowledge framework has become “a ‘market standard’ for the human reliability community the world over”.

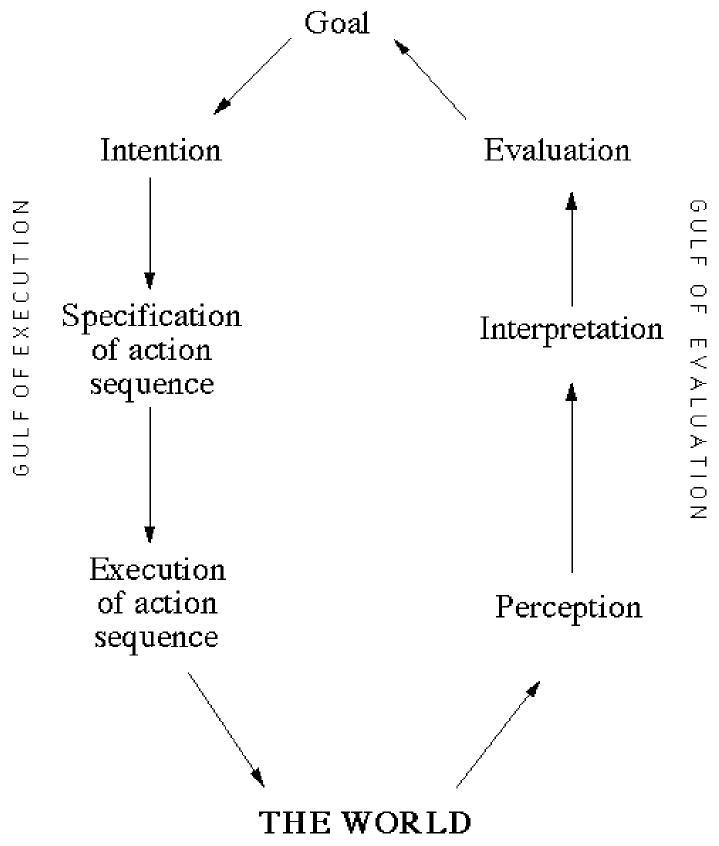


**Figure 2-4 -** Norman's (1988) Seven Stages of Action

*Levels of Performance: Skill, Rule, Knowledge (SRK)*

Skill-based behaviour represents sensory-motor performance during acts or activities that, after a statement of an intention, take place without conscious control as smooth, automated, and highly integrated patterns of behaviour.

In rule-based behaviour, the composition of a sequence of subroutines in a familiar work situation is typically consciously controlled by a stored rule or procedure. A rule may have been derived empirically during previous occasions, communicated from other persons' knowledge as an instruction, or it may be prepared on occasion by conscious problem solving or planning.

In general, skill-based performance flows without conscious attention and the actor will be unable to describe the information used to act. The higher level rule-based co-ordination in general is based on explicit know-how, and the rules used can be reported by the person, although the cues releasing a rule may not be explicitly known.

During unfamiliar situations, for which no rules for control are available from previous encounters, the control must move to a higher conceptual level, in which performance is goal controlled and knowledge based. The goal is explicitly formulated. Then a useful plan is developed. Different plans are considered and their effect tested against the goal, physically by trial and error, or conceptually by means of 'thought experiments'.

Rasmussen (1990) maintains that a very important aspect of the cognitive control to be captured by models of human behaviour is *the dynamic interaction* between the activities at the three levels. Only insights into this interaction would make cognitive models of human performance fit for complex environments, and thus for cognitive engineering; cognitive models derived from the cognitive science tradition suffers in this novel context from focusing on a well-bounded aspect of mental representation.

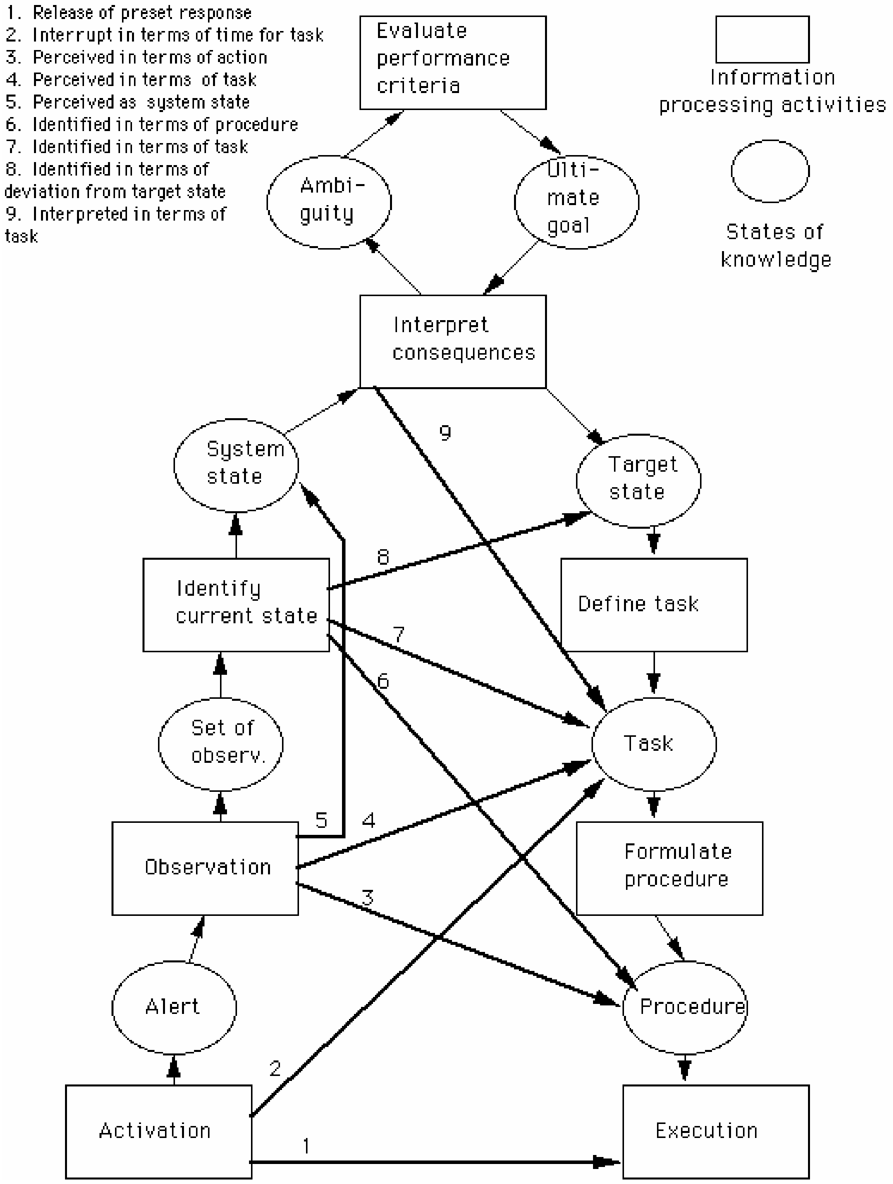


**Figure 2-5** Rasmussen's Decision Ladder (1987)

### *Errors as Human-Machine Mismatch*

Rasmussen et al. (1987) see human error as man-machine or man-task mismatch. In the case of systematic or frequent mismatches, the cause can typically be considered a design error. Occasional mismatches are either caused by variability on part of the system or the human, and will typically be considered component failures or human errors, respectively.

Rasmussen states that human variability is an important ingredient in adaptation and learning, and the ability to adapt to peculiarities in system performance and optimise interaction is the very reason for having people in the system. Human errors can in a way be considered to be unsuccessful experiments with unacceptable consequences. An 'unkind' work environment is then defined by the fact that it is not possible for a man to correct the effects of inappropriate variations in performance before they lead to unacceptable consequences.

### *Error Taxonomy*

Rasmussen et al. (1987) stress that a consistent taxonomy (see Figure 2-6) is necessary in order to extract data from reports on accidents and incidents. The analysis of this data is for instance useful for design of error-tolerant systems (see also Rasmussen and Vicente, 1989).

Rasmussen's 'External Mode of Malfunction' can be seen as a counterpart of Hollnagel's phenotype description. The SRK framework, however, also aims at relating mismatches to psychological mechanisms.

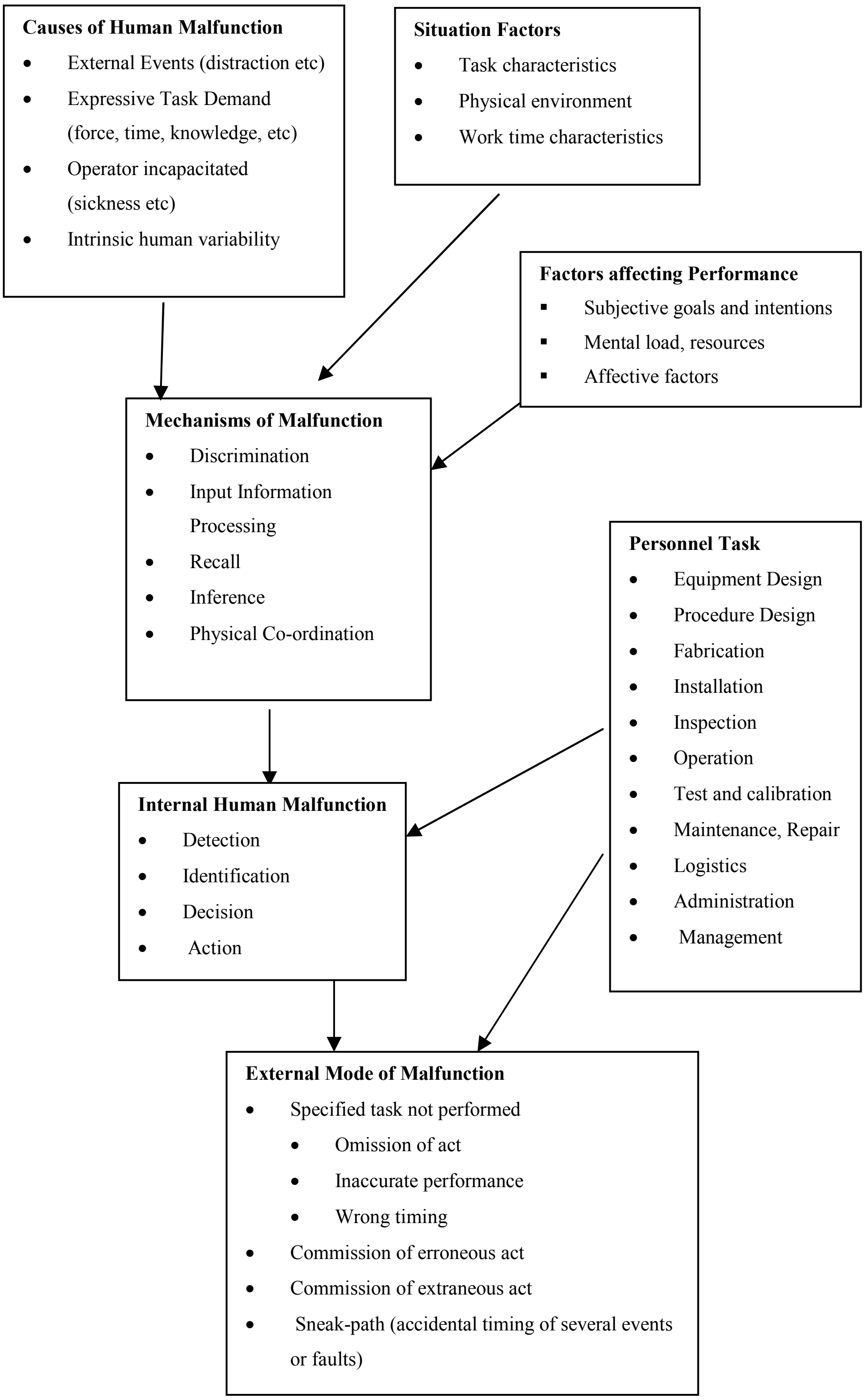


**Figure 2-6 -** Taxonomy for Description and Analysis of Events involving Human Malfunction (Rasmussen, 1982)

For this purpose, a cognitive task analysis[14] (within the SRK framework) is necessary (see also Chapter 3) to identify the element of the internal cognitive decision process which was affected, either by not being properly performed or by being improperly bypassed by a habitual short-cut. Rasmussen (1987) states, that for the more familiar routine task[15] it is possible for a knowledgeable expert to judge the internal mode of malfunction from a case study.

Consideration of the 'mechanisms of malfunction', in terms of the above described levels of performance and control, further characterises an event.

Rasmussen (op. cit.) points out, that human system interaction cannot be described adequately only considering the cognitive level. The dimensions 'Personnel task', 'Factors Affecting Performance (FAP)', and 'Situation Factors' are therefore also to be taken into account in the analysis of the event. Importantly, these dimensions, along with the dimension 'Causes of human malfunction' (similar to the FAP dimension, but with discrete events as opposed to continuing conditions), allow the causal backtracking to be continued "upstream from the human" (op. cit.). Thus it can help to bridge the gap between models of human error and system failure.

Rasmussen suggests a multifaceted description system rather than an exclusive, generic classification tree (see Figure 2-6). The described taxonomy is not intended to lead to a generic, hierarchical classification system. Complex scenarios can be identified (i.e. errors can be classified and analysed) directly from the underlying

---

[14] Rasmussen's work epitomises the introduction of Cognitive Task Analysis (CTA). The decision ladder (Figure 2-5) is used as a framework to carry out CTA. It defines a set of prototypical information processing activities and resulting states of knowledge, which are used for examining task strategies and cognitive functions (Bes and Johnson, 1998)

[15] this refers to the rule-based level of performance only.

psychological mechanisms. So human error (or ‘events involving human malfunction’) is categorised on the basis of the mechanisms at its origin.

### *Concluding Remarks on the SRK Model*

Rasmussen’s SRK framework can in particular account for rule-based errors. There is no detailed method to investigate the nature of occurring slips other than a one-way short cut on the decision ladder, and behaviourist identification on the overall framework level. Slips, however, make up a significant proportion of operator error (since expertise makes them possible in the first place), and are also much more readily predictable than for instance knowledge-based mistakes (Rasmussen, 1987).

SRK does not deal with knowledge-based mistakes. Rasmussen (op. cit.) notes that on the other levels, error mechanisms are described in terms of established, ‘normal’ action sequences (rules, procedures) in a rather behaviourist way (based on task analysis). On the knowledge level this is impossible. It is thus very difficult to characterise the mental data processing and the related mechanisms leading to mismatch. The sequence of arguments an operator will use during problem solving cannot be described in general terms (see also Reason, 1988). The goal to pursue must be explicitly considered, and the actual choice depends on very subjective and situation-dependent features. Rasmussen maintains that in present-day control rooms, the context in which operators make decisions at the knowledge level is far too unstructured to allow the development of a model of their problem-solving process, and hence, to identify typical error modes, except in very general terms.

This also leads to the observation that Rasmussen’s model, though aiming at being a generic cognitive performance model, has its origins and application in control rooms of for instance nuclear power plants (NPPs), and is limited in the sense that it is domain dependent. In NPPs’ control rooms, operators’ tasks are proceduralised to a high degree. Knowledge-level performance, as is needed when faced with novel situations, is to be avoided. The same holds for skill-based tasks, since monotony often related with automated task performance is highly error-prone, too. Thus, SRK focuses on the rule-based level of behaviour, having evolved in response to NPP requirements.

Furthermore, the representation of cognitive processing is pitched on a considerably high level, when compared to local theories of cognitive functioning[16]. Since not intended as such, the SRK model is something rather different from a psychological model. It emphasises *psychological products rather than processes* (Sanderson and Harwood, 1988). It is intended to be a “functional model to illustrate general aspects of the operator’s situation at a higher level as seen by the system designer” (Rasmussen, 1969, cited in op. cit.). This hinders a thorough analysis of the cognitive processing underlying the ‘external human malfunction’.

Furthermore, in a study of several Human Error Identification (HEI) techniques in 1992 (Kirwan, 1992), SRK obtained the lowest ratings on all but one category. Only its theoretical validity was judged moderately satisfactory, but on scales concerned with for instance comprehensiveness, consistency, validity, and acceptability (usage to date and availability of technique) it performed very poorly. However, Rasmussen’s SRK model paved the way for several other techniques, such as Reason’s, as described next.

## Cognitive Primitives, Cognitive Biases, and Human Error

Reason’s (1990) taxonomy of human error also represents a conceptual classification of error. It is predicated on assumptions about the cognitive mechanisms involved in error production. He locates basic human error tendencies within the general working

---

[16] As opposed to framework theories (Reason, 1988)

mechanism of the human cognitive system, paying particular attention to cognitive primitives and biases, rather than detailed information processing principles. His categorisation scheme is widely referred to in research into error modelling (Logan, 1984; Green, 1985; Rasmussen, 1985; Rouse and Morris, 1987; Woods and Roth, 1988; De Keyser, 1989; Rouse and Cody, 1989; Rasmussen, 1990; Rasmussen, 1990).

Rasmussen's SRK model is at the heart of Reason's model of human error tendencies. Reason's Generic Error Modelling System (GEMS) combines the SRK approach with the distinction of slips (and lapses) and mistakes introduced by Norman (1988). Norman (op. cit.) defined that slips result from automatic behaviour, mistakes from conscious decisions. Automatic behaviour means cognitive control is non-conscious, i.e. it takes place on the skill-based level of performance. Conscious decision-making relates to the knowledge-based level of cognitive control, as well as to the rule-based level (although to a lesser extent). In this way, Reason takes on board SRK performance levels, also associating them with levels of cognitive control as Rasmussen did in his model.

Thus, Reason bases his error classification on the definitions of skill-based slips and lapses on the one hand, and rule- and knowledge- based mistakes on the other. He defines slips and lapses to result in actions or states that deviate from the current intention due to execution and/or storage failures. Mistakes, on the other hand, result in actions that may run according to plan, but where the plan is inadequate to achieve its desired outcome (see also Norman's Action Theory, 1981, Figure 2-4).

Reason distinguishes between type and form of an error. Error types are related to the performance level (Rasmussen, 1983) at which the error occurs. On the skill-based level he lists *inattention* and *overattention* as major reasons for slips of action, on the rule-based level he indicates *misapplication of good rules* and *application of bad rules* as main forms of errors. On the knowledge based level he enumerates 10 different forms of mistakes and failures, namely *selectivity, workspace limitations,* *out of sight out of mind, confirmation bias, over-confidence, biased reviewing, illusory correlation, halo effects,* and *problems with causality and complexity*.

Error forms are basically dependent on two cognitive primitives, or biases: *similarity matching* and *frequency gambling*. Reason accumulated varied experimental results as evidence that these two processes are ‘computational primitives’ of the human cognitive system. These biases denote automatic retrieval processes, by which knowledge structures are located and their products delivered to consciousness or to the outside world. Reason postulates as a useful, working approximation of human information processing principles, that, *when cognitive operations are underspecified, they tend to default to contextually appropriate, high-frequency responses.* This approximation and the two primitives hold across all three levels of performance (see also chapter 4 and 5 for a more detailed description of some error types and forms).

## CONCLUSION

Although Reason’s model provides for the analysis of slips to a much greater extent than Rasmussen’s SRK, GEMS particularly assists the analyst in understanding the errors that may occur when the operator moves into the rule-based and knowledge-based behavioural domains (Kirwan, 1992). GEMS is notably comprehensive in its coverage of different types of cognitive failure.

However, the technique only gives limited guidance on which error shaping factors are likely to apply to the identified error forms. The classification of knowledge-based mistakes is also mainly assisted by heuristic “guidance”. Kirwan notes[17], that it

---

[17] Kirwan (1992) reviewed 12 Human Error Identification techniques in a seminal “Applied Ergonomics” paper. He measured the techniques against a set of criteria qualitatively and also by means of an empirical comparison study. The criteria

is very much left up to the analyst's insight and experience to ascribe particular error shaping factors to any individual task step, and then to propose measures by which these negative factors may be overcome. Furthermore, since guidance on how to choose the cognitive mechanisms underlying the categorised errors is ultimately relying on the assessor, GEMS is still largely treated as a taxonomy rather than an explanatory action and cognition model of human error.

Both SRK and GEMS are based on insights derived from cognitive theory, but do not present these in a systematic, model-oriented manner. They ultimately define static error classification systems, along with high level heuristic psychological knowledge. Although a classification system is not likely to give insights into the underlying causes of error, it offers the first step towards theory by identifying systematic patterns in otherwise unordered error data. Also, Freitag (1997) notes that "the philosopher's stone which we are all seeking is how to recognise the events we can learn from without actually having to analyse them all in detail first". A more observational error classification scheme can provide this kind of initial guidance. A more thorough, in-depth error analysis approach can then be applied to further investigate the classification choices and decisions (as proposed in this thesis).

In conclusion, the error taxonomies described provide for an error analysis that is pitched at a very high level in respect to the underlying cognitive error mechanisms. Typically, error categories are general, not necessarily discriminative, ambiguous, and the cognitive concepts used are partly loosely defined. A structured vocabulary, in which the error classes can be expressed as constrained by cognitive theory, will help to highlight error properties.

---

included comprehensiveness, consistency, and e.g. auditability. A follow-up evaluation was conducted in 1998 (Kirwan, 1998), with similar results for GEMS and SRK.

The possible mappings from category to underlying mechanisms can be examined by reasoning about the underlying processing within a cognitive architecture. In cases of mapping ambiguity, a systematic approach to the investigation of the relevant cognitive aspects will prove helpful. This also points to the issue of documentability, i.e. the degree to which the technique lends itself to auditable documentation. GEMS performed very poorly when rated on a scale for documentability (Kirwan, 1992). Complementary cognitive modelling of the error forms identified would provide an additional source of modelling decision documentation.

In this thesis, I detail how this approach of complementing Reason's GEMS taxonomy with a cognitive modelling approach can be put into practice. In the following chapter I will delineate the field of cognitive modelling and introduce a cognitive architecture that lends itself to cognitive error modelling, and was thus used in the combined approach.

One important implication from either accident causation model is that the specific analysis of "Human Error" is an essential concept in accident analysis. Furthermore, the analysis of the psychological side of Human Error has to be situated in accident causation models that take the interplay of all levels of system factors into account, ranging from individual 'active' error to team interactions, and the impact of governmental regulation policies. Thus far, accident analysis only often takes a 'common sense' approach to error analysis.

Reason points to several shortcomings of what he calls error management (EM) techniques. Though of proven value, these existing forms of EM have a number of limitations, particularly narrowness of focus. In brief, they tend to be "piecemeal rather than planned, reactive rather than proactive, fashion-driven rather than theory-driven" (Reason and Maddox, 1996). They also ignore the substantial developments that have occurred in the behavioural sciences over the past 20 years in understanding the nature, varieties, and causes of human error.

A thorough, grounded, i.e. theory-based, understanding of Human Error, its nature, its causes, and its relation to the rest of the 'total system' is needed. In order to fully comprehend an accident's causation, the focus cannot only be on the technical aspects of system failure, with 'human error' as the label that provides the stopping case for the investigation process. This will prohibit us to effectively prevent their re-occurrence. Human error models, as introduced in this chapter, when embedded in broader accident causation frameworks, can provide the tools for this more comprehensive accident and incident analysis. However, current methods for 'human error' analysis typically are "more of an art than a science" (Kirwan, 1992), and thus lack support for the human factors analyst in their application in accident and incident investigation.

# CHAPTER 3 COGNITIVE ERROR MODELLING

## INTRODUCTION

Bainbridge (1987) observed the “ironies of automation” more than a decade ago, and described it as a side effect of increasing automation in industry that operators’ tasks shifted towards *control*, with cognitive tasks (such as decision-making) predominating. A similar trend is noticeable in the corresponding forms of operator error.

These ‘cognitive errors’ demand a suitable taxonomy and method of analysis in order to be understood and explained thoroughly. An error analysis technique that is rooted in cognitive theory allows the analyst to gain an understanding of the processes underlying the error - the mechanisms of human cognition. This thesis argues that the use of a cognitive architecture as a structural framework for expressing the cognitive processing underlying operator error will lead beyond an observational approach to error modelling. Further research into cognitive error modelling and simulation is also discussed.

Human error plays a major role in the occurrence of accidents in safety critical systems such as those used in aviation, railway systems, or nuclear power plants (e.g. Hollnagel, 1991). Identifying human involvement in accidents is, however, not necessarily tantamount to identifying the actual 'cause' of the accident. The design of technology, task procedures, or organisational issues might not be well suited to human action and cognitive processing (Norman, 1993), and thus present an 'unkind' work environment. The understanding of the causes, principles and effects of human error is thus essential if we want to further our understanding of major accidents. Insights into the causes and course of accidents and incidents will help us to deal with them appropriately: avoiding similar events and situations in future, or by being prepared for their occurrence (for example through error-tolerant design, or emergency procedure based training). To attain this goal it is necessary to perform an accident analysis which identifies as many contributing factors as possible and to suggest actions for removing or blocking each contributing factor (Svenson and Sjöström, 1997). Human error makes up a substantial portion of the pool of contributing factors. Their identification is commonly guided by error taxonomies, some of which consider the origins of the classes of human error identified.

Research into human error has been dominated by work on 'error' classification schemes and 'error' models. Domain dependent error models and taxonomies cater for the instances of human erroneous behaviour that are typical for, or conditioned and postulated, by the domain at hand (e.g. SRK and proceduralised tasks in control rooms).

Domain-independent human error taxonomies can be seen as being based primarily, and to different extents, on generic theories of human cognition (such as Reason's taxonomy). Even those taxonomies that adhere more closely to cognitive theory in their explanation of human fallibility are pitched at a level too high to track the underlying cognitive processing undergone by the operator.

Examination of this processing could provide leads to the causes and 'inner workings' of the error. It is therefore useful to provide a framework to model the cognition that underlies human error. This will enable error analysts to benefit from error taxonomies' abstracting and simplifying effect on the wealth of error data, as well as from the more refined, structured, and detailed information gained by systematic cognitive modelling.

With increasing automation, the operator's role transformed to one of control rather than action (Bainbridge, 1987). With the attempt to decrease the user's (physical) workload, cognitive load increased inadvertently. Easily performed (physical or mental) tasks are transferred to the system's responsibility, while delegation, decision making and other general control processes stay within the operator's task space. This led to an increase in the demands posed on the mental capabilities of the human operator, and thus increases the occurrence of errors that originate in the constraints imposed by the human cognitive system.

Current error models, as noted above (and in later chapters in this thesis), do not provide an analysis detailed enough to trace the cognitive processing preceding or constituting human error. Our approach suggests utilising the cognitive 'vocabulary' that an information processing model offers to express error specifications.

However, most current cognitive modelling techniques focus on expert, error-free behaviour. Assumptions of 'idealised' task performance are built into the technique, hindering its adaptation to real-life behaviour of users and operators. Similar observations hold for current cognitive simulation architectures. Those cognitive architectures that do have the potential to incorporate error classification are typically either rather general (e.g. Interacting Cognitive Subsystems, ICS), or bound to a specific domain, such as nuclear plant operation (e.g. COSIMO, Cacciabue, 1998).

The goal of this thesis is to demonstrate the use of a cognitive architecture as a framework for expressing error classes using a cognitive vocabulary. The vocabulary of a cognitive architecture, like the syntax of a language, can help to constrain and guide the modelling process. The framework is structural in as much as it represents the constraints imposed by the structure inherent in the underlying cognitive theory (such as the stages in an information processing model).

Modelling erroneous behaviour within a cognitive architecture can thus provide a more in-depth description of the course the error took, and complement higher level description given by error categories. Non-executable cognitive models can present a medium for communicating the causes of error, and the processes and mechanisms underlying its production. This is demonstrated in this report by modelling user and operator error within the Interacting Cognitive Subsystems architecture (ICS, Barnard, 1993).

## DEFINITION OF TERMS

Several, quite different, research areas refer to the term “user modelling” when not necessarily referring to the same activity or product of that activity.

Newman and Lamming (1995) described various definitions of ‘User Model’ as follows:

- A user’s conceptual (or mental) model of a system (a mental picture a user has of a system and how it works).

- A model of the user held within the system software (a representation of the user’s abilities, limitations, beliefs and goals. May form the basis for automatic adaptation).

- A model or definition of the typical user referenced by the interface designer to aid his/her formulation of the system being built.

With the various theories of mental models, a whole range of terms has been introduced. The lack of a unified terminology is confusing, especially when different authors use different terms to describe the same type of model, and the same term is sometimes used for very different types of models (Nielsen, 1990).

In an attempt to bring some order into the great user model confusion, Nielsen (op. cit.) proposed a meta-model to classify models of user-system interaction. This meta-model encompasses a wider range of models than those mentioned here. His notation describes 'models' in terms of the relationships between 7 elements or participants of models:

- U – the user;
- D – the designer;
- C – the computer system;
- M – manuals and other documentation of C;
- T – the task performed by the user;
- W – the surrounding world in which U performs;
- R – the researcher looking at any of the above.

These elements can be combined, using a simple notation, to denote who holds a representation of what. For example, a user's model of a computer system is UC, the designer's model of a user's model is D(UC), and so on. This notation offers a parsimonious way of distinguishing between different models. The notation will be employed to clarify the exact nature of models involved in the theoretical and empirical work reviewed in this section.

The cognitive modelling approaches relevant to my work are the ones characterised by Newman and Lamming (1995) as defining "models or definitions of the typical user referenced by the interface designer to aid his/her formulation of the system being built". Thus, the term 'cognitive user modelling' will in this report refer to the user model as characterised by the relationships of RU, and also of R(DU). It describes the activity of capturing and communicating (diagrammatically or text-based) users' cognitive processes when performing some task, or what the designer perceives those processes to be. These can then be used to complement the analysis of error as provided by the describe taxonomies.

## User Modelling

Tognazzini (1991) suggests that designers should construct simple design models which:

- reflect users' tasks, rather than the underlying hardware and software of the system;
- correspond to users' experiences and expectations.

Similarly, Norman points out that "ideally, the model is based on the user's task, requirements and capabilities … (AAIB, 1989) must also consider the user's background, experience and the powers and limitations of the user's information processing mechanisms." (Norman, 1986).

To follow these recommendations, we need to include three sets of representations to the model-building process: users' tasks, general knowledge and experience, and information processing mechanisms.

Three types of methods that fulfil some or all of these requirements are briefly described and discussed below.

## Behavioural Task Analysis

The goal of behavioural task analysis is to study the capabilities and performance of human operators versus task demands necessary to achieve system goal(s) (Kirwan and Ainsworth, 1992). Task description techniques include charting and network techniques (including timelines, input-output diagrams, process charts, functional flow diagrams, information flow charts), decomposition methods, and operational sequence diagrams.

An almost universal part of this process is hierarchical decomposition, for example by Hierarchical Task Analysis (HTA). HTA is a broad framework originally articulated in Annett et al. (1971) and e.g. Kirwan and Ainsworth (1992).

Kirwan and his colleague (1992) investigated several diagrammatic task description methods. They stress how charting and network techniques (such as Petri Nets, or Functional Flow Diagrams) can present an integrative method, displaying the human embedded within the system context. However, they note in conclusion, that, as the cognitive task content increases, the mentioned chart representations become less satisfactory. Concurrency of mental operations, as well as processing complexity is difficult to capture by these techniques.

Thus, complex cognitive aspects of human performance are typically not addressed by this approach. It concerns itself exclusively with normative task sequences, in order to provide a system interface that meets the functional requirements of the user or operator. Behavioural task analysis, however, often represents the initial stage, or the prerequisite, for Cognitive Task Analysis, Cognitive User Modelling, and also Human Error Identification techniques.

## Cognitive Task Analysis

One typical, well-established model of calculational task analysis is Card, Moran and Newell's GOMS technique (Card et al., 1983).

GOMS stands for Goals, Operators, Methods, and Selection Rules. As many other cognitive task analysis models, it makes use of a model of mental processing in which the user achieves goals by solving subgoals in a divide-and-conquer fashion.

A goal is something the user wants to do; goals are often related hierarchically. Operators are actions that the user does, which may be external (perceptual or motor) or mental (e.g., make decision, store an item in working memory). Methods are a sequence of steps (operators) used to accomplish a goal. Methods can call submethods to accomplish subgoals. When multiple methods are possible to accomplish a goal, selection rules are used to choose an appropriate method. GOMS has been used as the basis for constructing production-rule simulations of users and for redesigning help systems.

Analysis of the GOMS goal structures can yield measures of performance. The stacking of a goal structure can be used to estimate short-term memory requirements. The model of the users' mental processes implied by this is highly idealised. However, GOMS was not intended for anything else but describing how experts perform a routine tasks. Furthermore, the GOMS (and associated Keystroke-level) model had been criticised (see for instance Sasse (1992)) by designers as being:

- too low-level;
- too limited in scope;
- focused on outdated technology;
- too difficult to apply.

A further cognitive task analysis technique is presented by the Cognitive Complexity Theory (CCT) approach introduced by Kieras and Polson (Kieras and Polson, 1985). It begins with the basic premises of goal decomposition from GOMS and enriches the model to provide more predictive power. CCT has two parallel descriptions: one of the user's goals and the other of the computer system (called a device in CCT). The description of the user's goals is based on a GOMS-like hierarchy, but is expressed primarily in production rules. CCT can represent more complex plans than the simple sequential hierarchies of GOMS. Concurrent plans can be expressed. CCT attempts to measure interface complexity by relating the number of production rules in the CCT description to the degree of complexity of the interface.

Knowles (1997) dealt with the CCT approach in her thesis on interface complexity. She identified major shortcomings:

- No underlying theory of knowledge sources and their interaction
- Restricted in application, by virtue of its reliance on GOMS to tasks involving no problem solving
- No empirical validation of proposed theory
- Over-reliance on quantitative components at the expense of qualitative aspects.

Especially the last point addresses a weakness most of the calculational CTA approaches share.

Carroll & Campbell (1989) suggest that, rather than reducing the psychological contribution to HCI to quantitative measurements and calculations, psychologists should ensure that the value of explanatory and conceptual HCI research is recognised and applied.

Application of quantitative methods to behavioural and cognitive phenomena which have not been sufficiently well described and understood can lead to the garbage in, garbage out problem well known in computer science. Carroll & Campbell (1989) are right to point out that the imposition of methodological strictures alone will not make HCI more creditable or applicable. It could have the opposite effect: a discipline can be strangled by methodology strictures that become more important than the theories they are supposed to serve (Sasse, 1997).

Theories of the user should ideally not restrict themselves to collecting observable quantitative data for mechanistic models, without trying to explain the underlying cognitive concepts. The trade-off between applicability and generality often causes explanatory and conceptual models to fall behind. The next section will discuss two approaches to the analysis of the cognition underlying more overt, observable phenomena: cognitive simulation, and approximate cognitive modelling. The latter is a more qualitatively oriented cognitive modelling approach, which nevertheless strives at achieving applicability.

## Cognitive Simulation

Human error models thus cannot convincingly model, classify, or predict knowledge-based mistakes, in spite of ongoing and well-developed research into problem solving within cognitive psychology. Bringing the insights gained in this field to bear on human error modelling would appear to be an appropriate contribution to the field.

Cognitive modelling techniques that do concern themselves with decision-making and problem-solving are to be found in the field of cognitive simulation.

Roth et al. (1992) define cognitive simulations as “runnable computer programs that represent models of human cognitive activities”. They maintain that their approach to cognitive simulation can be used to uncover the cognitive demands of a task, to identify where intention errors are likely to occur, and to point to improvements in the person-machine system.

Cognitive architectures employed in cognitive simulation tend to concentrate on higher order cognition – knowledge-based reasoning, problem-solving, or decision making (see for instance SOAR's reasoning by analogy mechanisms, (Newell, 1990)). Thus, executable cognitive models, or cognitive simulations, offer further means for exploring human error, beyond the approach taken in this report. Further research is suggested to the effect of exploring the potential of computationally implementing operator high-level thought processes in an error-oriented, domain-independent cognitive architecture.

As a first step, suitable architecture candidates need to be identified. Some cognitive architectures, as we have seen in earlier sections of this chapter, may lend themselves more easily to error modelling than others.

In order to choose an architecture, selection criteria need to be defined and explicitly stated. Current models can then be examined in the light of these criteria.

Grant (1996) investigated the range of cognitive theories, architectures, models, and simulations. He set about to find in how many significant ways works about cognition could differ. He discovered six dimensions along which to differentiate the approaches:

- Domain dependency
- Level of specification
- Coverage
- Correspondence with experimental or practical findings
- Parsimony

These dimensions will influence the choice of selection criteria that need to be adopted. Grant (op. cit.) classified Reason's GEMS (see Chapter 2) as "domain-independent but still cognitive", and as low level specified. He indicates its 'direction for possible future developments' as moving towards a coded, or executable, level of specification. This relates to my proposal for further research into linking an executable cognitive model with GEMS in order to investigate knowledge-based mistakes.

## Approximate Cognitive Modelling

As we have seen above, Cognitive Task Analysis looks at a system from the point of view of the user, and identifies aspects of the (interface or task) design that place heavy demands on the user's cognitive resources - their memory, attention, and so on.

This information allows the designer to focus upon the features that users will find hardest to learn, and where they are most likely to make errors. May and Barnard (1995b) state that "*by highlighting the sources of ambiguity, CTA helps designers to iterate towards design specifications that are cognitively straightforward, leaving users freer to concentrate on performing the tasks that the system supports rather than on using the interface itself.*"

This section concerns a rather *qualitative*, though systematic analysis of the mental processes in the light of cognitive theory.

For instance, Barnard's diagrammatic ICS framework examines users' mental load in terms of the resource demand and allocation issues that arise through parallel functioning. At its best, the linkage between the input and output of the black box cognitive processor with the environment is also described: reception of information as much as acting on the world.

Kirwan and Ainsworth (1992) noted in relation to task description methods, that graphic descriptions in this context can offer substantial benefits over purely textual approaches. The principal aim, so they state, of these techniques is to use a formal graphical representation of the task which is either easier to understand or significantly more concise than a textual description, or else to emphasise particular aspects of the task.

## ICS

ICS offers a diagrammatic approach to task description that focuses on complex cognitive processing of a human in interaction with his or her environment. It presents a distributed, parallel cognitive architecture approach, and is therefore well suited to represent concurrent, and overlapping mental tasks.

ICS is a technique reflecting a resource-based view of the multiple subsystems involved in cognition (Barnard, 1987). ICS provides a model of perception, cognition, and action. But unlike other cognitive user models, it is not intended to produce a description of the user in terms of sequences of actions. Rather, ICS provides a more holistic view of the user as an information processing system. The emphasis is on determining how easy particular sequences of actions become as they are made more automatic and proceduralised within the user (Young and Abowd, 1994, and see also a more detailed description in Chapters 4 and 5).

Barnard (1987) maintains that a CTA based on ICS framework can give qualitative information on users' mental processing involved in task performance. It does not force the complexity of the human mind into a calculational straightjacket. The representation constructed includes a specification of mental processes; procedural knowledge; the contents of episodic memory; and a characterisation of the way in which the cognitive mechanism is controlled during task execution. Pre-specified mappings from the contents of Task Models then predict aspects of user behaviour.

We conducted a feasibility study concerning the error modelling potential in ICS. This study can be found in the appendix in full detail. It was shown that ICS provides a good vocabulary to reason about error in interactive system users, and it also opens avenues for design implications dependent on the undertaken error analysis.

### EPIC

The second approximate cognitive architecture we investigated is EPIC – an engineering model of cognitive task performance. A comparison study of ICS and EPIC and their potential to contribute constructively to accident investigation was conducted. The results of this study will be discussed in the remainder of this chapter. In conclusion it can be said that the EPIC architecture and its focus on performance and timing in an idealised task setting proved less suited to human error analysis in the context of accident investigation in complex, safety-critical systems, than ICS.

## COGNITIVE MODEL COMPARISON

The investigation of pilots' cognitive processing plays an important role in the analysis of aviation accidents. Increasing cockpit automation led to an increase in the demands posed on the mental capabilities of the pilot, and thus increases the occurrence of errors that originate in the constraints imposed by the human cognitive system (Woods, 1987). Cognitive user modelling can capture and communicate an analytic view of pilots' cognitive processes when performing some task. Cognitive modelling can then also be used to enhance the analysis of pilot error by relating it to the constraints of the human cognitive system.

Cognitive modelling is employed by organisations such as NASA and the CAA to further their understanding of the cognitive processes that underlie pilot performance. One “classic” approach to cognitive modelling uses knowledge-based cognitive architectures, such as Soar (Laird et al., 1986) or ACT (Anderson, 1993). Such models focus on higher-level cognitive processes, such as reasoning and decision-making. However, sensory and physiological data often contribute to aviation accidents, but still are widely neglected in cognitive modelling approaches. In this chapter we will look into two cognitive modelling approaches that can also represent sensory and physiological data, EPIC (Executive Process/Interactive Control) (Kieras and Meyer, 1995) and ICS (Interacting Cognitive Subsystems, (Barnard, 1993).

EPIC is a knowledge-based simulation architecture. In other words, analysts can use it to simulate pilots’ cognitive processes and draw conclusions about the complexity and time demands of the cognitive processing required to carry out a task. EPIC also provides features that aim at integrating perceptual and motor processes in a holistic approach to cognition.

However, we argue that the investigator also needs a cognitive model which accounts for built-in constraints in the human cognitive system, such as memory limitations or problems with translating from one mental code to another (Bainbridge, 1993). These are central to making detailed recommendations based on cognitive analysis. Barnard (1987) provides a modular approach that does focus on these aspects of cognition, the Interacting Cognitive Subsystem (ICS) architecture. ICS offers a diagrammatic approach focusing on humans in interaction with their environment. Sensory as well as internal knowledge input into the cognitive system is integrated in the ICS approach.

The following sections demonstrate that pilots’ perception, cognition and action can be modelled within these two cognitive architectures. We argue that a holistic approach offers a deeper understanding of the pilot’s cognitive and physiological processes in aviation accidents. We conclude that EPIC lacks provision for retrospective modelling or non-expert behaviour, unless simple cognitive-behavioural processes are to be analysed. ICS offers these possibilities, but requires an analyst to be highly skilled in the complex psychological theory of the model.

In the remainder of this chapter, we first introduce the case study, an attempted night visual approach to a runway at Gatwick Airport. Then we review alternative cognitive modelling approaches for the representation and investigation of the pilots' cognitive processes. We look at EPIC, an executable human cognitive performance model. Then, we review Barnard's (1987) ICS technique, which combines a domain-independent view on human's cognitive processing capabilities. For each of the modelling approaches presented, a task model is constructed to show their scope. Also, their applicability to 'human error' modelling is tested by detailing 'real life' pilot error within each approach.

## THE CASE STUDY

The modelling of pilots' cognitive processes in this chapter will be illustrated by the events recorded in an Air Accident Investigation Branch (AAIB) incident report. This report is appropriate as a case study because it represents a relatively frequent occurrence (AAIB, 1989, p. 22). It also shows a variety of causal factors. These include higher cognitive mistakes as well as perceptual slips. The incident involved a Boeing 737 and a British Aerospace One-Eleven (BAC 1-11) at Gatwick Airport. The BAC 1-11 landed on taxiway 2 after making a night visual approach to runway 08L. A Boeing 737 had been ordered onto taxiway 2 just previously by the air traffic controller. The Boeing's commander attempted to turn off to the side after observing the landing lights of the BAC approaching. The manoeuvre immediately led to the aircraft's port main wheels leaving the paved surface. It bogged down in the soft ground partially blocking the taxiway with its left wing and rear fuselage. The BAC stopped only 190 metres short of the Boeing 737. There were no injuries.

The AAIB report lists as one causal factor that the BAC commander inaccurately interpreted the cues provided to him by the visual scene on the approach to runway 08L. He consequently landed on taxiway 2 believing it to be runway 08L. The other 'causes' describe various factors believed to have facilitated the cue misinterpretation. This includes the use of runway edge lighting as well as the taxiway green centreline. Communication between the BAC commander and the first officer, which might have facilitated the 'misjudgement', are also mentioned (see (Johnson, 1995) for a systems analysis approach to the incident that draws together the contributing factors).

## EPIC

Cognitive modelling techniques give us the opportunity to reason in more detail about the potential causes, or (cognitive) precedents, of the commander's misinterpretation. However, there exist a variety of modelling approaches to choose from. We will now investigate two techniques that each represent different traditions in cognitive modelling. By using the AAIB report as illustration, we will show how the two traditions serve different needs, and can delineate complementary perspectives on the data provided.

The two approaches we will assess in this chapter are based on diagrammatic cognitive modelling and on computational cognitive simulation.

As mentioned above, (Roth et al., 1992) define cognitive simulations as "runnable computer programs that represent models of human cognitive activities". Cognitive simulation approaches can potentially uncover the cognitive demands of a task, to identify where intention errors are likely to occur, and to point to improvements in the person-machine system. Cognitive processes are coded typically in production rules in the IF <CONDITION> THEN <ACTION> format.

EPIC is a computational performance simulation model. The “primary goal” in the development of EPIC has been “to account for human multiple-task performance in situations such as aircraft cockpit tasks” (Kieras et al., 1997). EPIC was designed to explicitly couple perceptual-motor and basic information processing mechanisms like those in the MHP (Card et al., 1983), ACT-R (Anderson, 1993), and SOAR (Laird et al., 1986). Although it is mainly demonstrated as a predictive performance model, it is also stated to achieve insights as an explanatory model of human behaviour and its underlying cognitive processes (Kieras and Meyer, 1995).

Figure 3-1 shows the overall architecture of an EPIC model interacting with a simulated system.

Human performance in a task is simulated in EPIC by programming the cognitive processor with production rules organised as methods for accomplishing task goals. The EPIC model is then run in interaction with a simulation of the external system and performs the same task as the human operator would. The model generates events (e.g. eye movements, keystrokes, vocal utterances) and the timing of these events is predictive of human performance.

These predictors also include various aspects of mental workload, such as how much information must be maintained in short-term memory. If the predicted performance is unsatisfactory, either in terms of overall system performance, or in comparison to another design, further examination of the simulated performance will reveal the cause of the performance limitation. For example, humans can perform many activities concurrently, but a poor interface design will limit the amount of overlapping that can be done. For example, requiring that the eyes be kept mostly on one location, thereby preventing a second, otherwise compatible task from being executed concurrently.

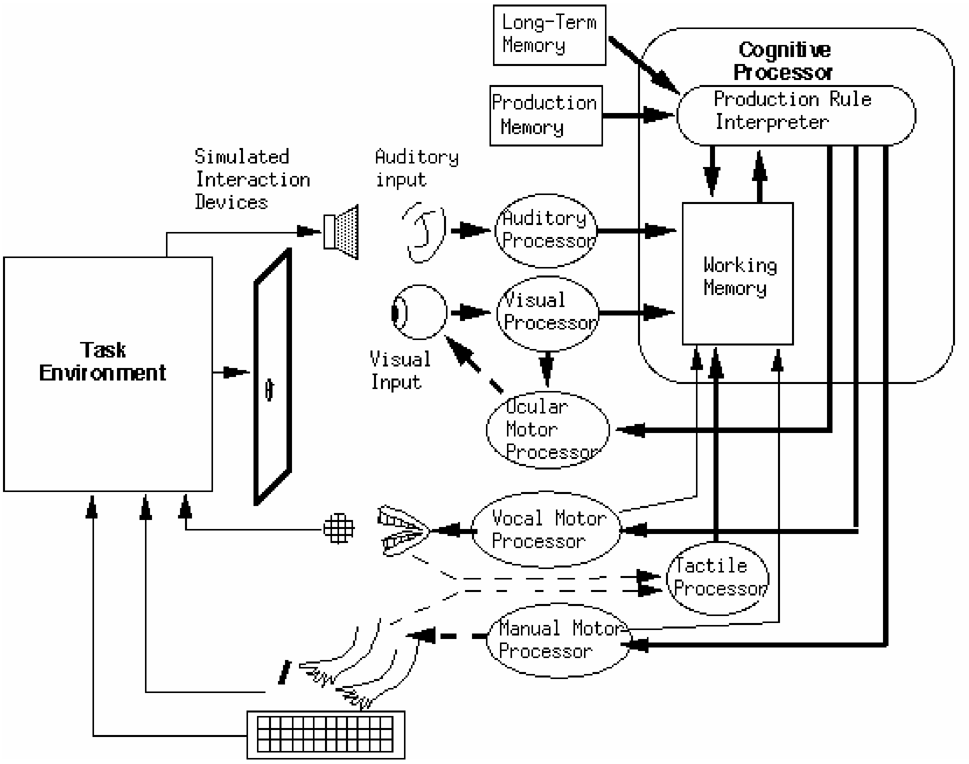


**Figure 3-1 -** The EPIC Cognitive Architecture

EPIC models are based on prior task analysis, using some established behavioural technique, such as Hierarchical Task Analysis (HTA, Kirwan and Ainsworth, 1992). This is then translated into production rules using cognitive task analysis. Note that the tasks mentioned in an HTA model derive from a normative task model, and thus presume the successful completion of each task, in other words, they assume error free goal achievement.

Since the accident arguably occurred due to non-successful completion of the “Locate Runway” task, this is the task we will concentrate on in our cognitive analysis.

## EPIC MODEL “LOCATING TARGET”

Figure 3-2 shows an EPIC model that, given parameters specifying the details of the visual situation, will result in runway identification.

The code shows the current state of the EPIC production system database, showing the various items present in the simulated human operator’s working memory.

The ‘pilot error’ from the Gatwick incident will thus be located in the matching process of the working memory (WM) target pattern and the pattern received in the visual field. This will show how, for example, underspecification of visual clues can lead to confusion between the taxiway and runway 08L.

The rule-based approach of EPIC forces the analyst to precisely specify the task parameters. Thus, the specification of the pilot’s mental model of the runway (WM TARGET-PATTERN) will be an assumed pattern of visual cues necessary for runway identification. This can then be run against the environmental input, namely the actual visual cues available to the pilot. Therefore, a mismatch can be identified, and a fine-grained analysis of the pilot’s visual perception is possible.

However, the pilot’s communication with the tower and his co-pilot are not included in the model. EPIC does not prompt the analyst to consider these environmental aspects. Visual perception is seen as a pure pattern matching process with no other modalities or task aspects interfering. Thus, important contributory factors to the

```
(IF-NOT-TARGET-THEN-SHIFT-VISUAL-FIELD
IF
 ((GOAL LAND PLANE
   (STEP LOCATE RUNWAY)
   (WM CURRENT-PATTERN IS ?OBJECT)
   (MOTOR OCULAR PROCESSOR FREE)
   (VISUAL ?OBJECT FEATURES ?NOTTARGET)
   (NOT(WM TARGET-PATTERN IS ?VAR)))
THEN
 ((DELDB(WM CURRENT-PATTERN IS ?OBJECT)
   (ADDDB(WM CURRENT-PATTERN IS ?NEXT-OBJECT))
   (SEND-TO-MOTOR OCULAR MOVE ?NEXT-OBJECT)))


(TARGET-IS-LOCATED-BEGIN-HOMING-IN
IF
 ((GOAL LAND PLANE
   (STEP LOCATE RUNWAY)
   (WM TARGET-PATTERN IS ?TARGET)
   (VISUAL ?TARGET-FIELD FEATURES ?TARGET)
   (WM PLANE IS ?PLANE-OBJECT)
   (MOTOR MANUAL PROCESSOR FREE)
THEN
 ((DELDB(WM STEP LOCATE RUNWAY)
   (ADDDB(WM STEP HOME-IN))
   (SEND-TO-MOTOR MANUAL PERFORM POINT ?PLANE-OBJECT ?TARGET-PATTERN)))
```

**Figure 3-2 –** EPIC Rules: Locating the Target

visual misperception are ignored.

The above two production rules show a high level representation of two steps in the goal hierarchy, both relating to the sub-task "Locate Runway". Although presented here in the form of production rules as used in EPIC models, these two rules only present a high level view. To enable their implementation in a computational EPIC model, every line in the rules' body has to be a fine-grain (atomic), low-level, perceptual or motor activity, This means, for an EPIC model, each line will need to be further decomposed into its component parts, down to specification of the angle of the current visual field (perceptual), and the current position of the pilot's hand when preparing to move (motor). This is difficult, if not impossible, to determine in the aftermath of an accident.

This low-level specification of units of behaviour and the provision of experimentally derived fixed timing parameters enables EPIC to provide the analyst with performance time predictions (Kieras et al., 1997) as well as real-time simulated pilot performance. However, the data necessary for this grain of analysis can typically not be derived from an accident report. EPIC modelling thus lends itself more to a priori modelling as part of a normative cognitive task analysis than to explanatory modelling of a specific, erroneously performed task instance.

Table 3-1 overleaf briefly summarises strengths and weaknesses of the application of EPIC modelling in Human Error analysis.

## ICS

Cognitive simulations, as most other cognitive modelling techniques, are typically intended to represent expert, error-free behaviour. ICS, however, is a general architecture, in which the cognitive basis of erroneous behaviour can be expressed without violating or changing architectural principles. Instead of aiming at simulating the human, ICS is a diagrammatic cognitive process-model, which enables the analyst to inspect information flow and resource demands in the human cognitive system.

| **Strengths** |
|---|
| • Detailed, forces precision<br>• Grounded in cognitive theory (albeit not very complex)<br>• Takes perceptual and motor processes into account<br>• Can generate timing of task performance |
| **Weaknesses** |
| • Requires expertise and time<br>• Does not model for interaction between central and peripheral cognitive processing<br>• Does only incorporate routine expert performance<br>• Very low level tasks only to be specified<br>• Focus is only on timing aspect, cognitive resource allocation is not covered<br>• Precise low-level data is post hoc not available |

**Table 3-1 -** The EPIC cognitive architecture reviewed

ICS represents a comprehensive account of human cognition, which has proved powerful in explaining cognitive phenomena such as the stability of users' mental models during dual task interference effects (Duke et al., 1995). It has been applied to real-life systems and tasks, such as cinematography (May and Barnard, 1995a). ICS provides a great level of detail in the representation of cognitive processes and also the inherent constraints these have to satisfy. ICS was designed to provide a theoretical framework in which to place user cognition. It attempts to "satisfy the need for applicable theory" (Barnard, 1987). ICS, therefore, bridges the gap between theory-oriented cognitive architectures and task-oriented cognitive user models ((Grant and Mayes, 1991); (Simon, 1988).

| | | |
|---|---|---|
| **Sensory subsystems:** | | |
| **VIS** | visual: | hue, contour etc. from the eyes |
| **AC** | acoustic: | pitch, rhythm etc. from the ears |
| **BS** | body-state: | proprioceptive feedback |
| **Effector subsystems:** | | |
| **ART** | articulatory: | subvocal rehearsal & speech |
| **LIM** | limb: | motion of limbs, eyes etc. |
| **Structural subsystems:** | | |
| **OBJ** | object: | mental imagery, shapes etc. |
| **MPL** | morphonolexical: | words, lexical forms |
| **Meaning subsystems:** | | |
| **PROP** | propositional: | semantic relationships |
| **IMPLIC** | implicational: | holistic meaning |

**Table 3-2** - The ICS Cognitive Subsystems

Cognition is represented in ICS as the flow of information between a number of different subsystems (see Table 3-2), and as the processing performed on this data. The nine subsystems cover both high level cognitive and low level perceptual and motor processors. Each of these can act in parallel. Each of the subsystems has associated with it a unique mental code in which it represents the information it receives and processes. It will transform its data output into the corresponding mental code of the subsequently receiving subsystems. Each subsystem can receive several input streams and achieve a blending of these data streams under certain circumstances as described below (May and Barnard, 1995a) (see also Appendix A for more details). Each subsystem also has at its disposal a local image store. This serves as an episodic memory buffer of infinite size. A copy of any input the subsystem receives will automatically be copied to the local image store, before being further processed.

## ICS MODEL OF ERROR-FREE VISUAL RUNWAY IDENTIFICATION

Figure 3-3 illustrates the error-free performance of a task of locating an object in the visual field (such as a runway).

This is modelled in ICS in terms of information flow between the subsystems, and thus the different resources that are employed. Visual information concerning the target arrives at the visual subsystem and is copied into the local store. It is transformed into object code (1). The propositional subsystem has generated a representation of the target of the location task (by conferring with its local buffer) and transforms this into object code (3). This is sent to the object subsystem, and can there be blended with the incoming structurally encoded visual information (2). The matching representation can be sent back to the propositional subsystem – the target has been located. During unsuccessful attempts to locate the target, visual information other than that concerning the target will not be able to blend with the object code data sent by the propositional subsystem. A loop between the two subsystems will be maintained until the representations match and the target, in this case the runway, is located.

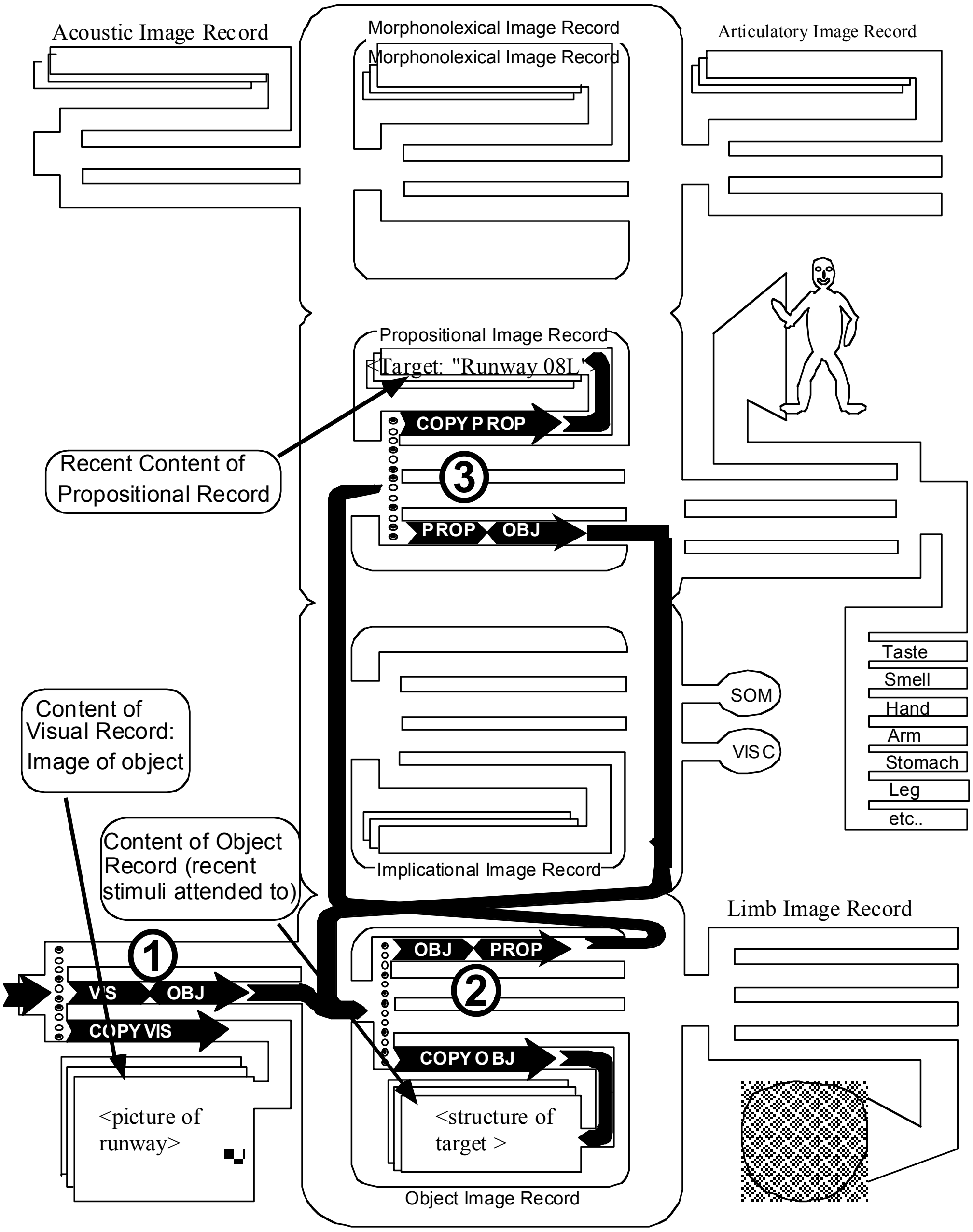


**Figure 3-3 -** Error Free Visual Identification of Runway

## FAULTY VISUAL IDENTIFICATION IN ICS

In a high-level, non-normative, task model, pilot’s communication with the tower and his co-pilot can now be included. The inclusion of task contributing factors such as tower communication, which in itself does not form an explicit ‘goal’ of the pilot’s task is specific to the subsequent ICS modelling. A conventional task analysis approach would not have considered this factor. ICS encourages the analyst not to treat perceptual and motor processes as peripheral, but as equal to cognitive processes. A task analysis prior to ICS modelling will therefore select the tasks to be included in the model prompted by the holistic process (rather than task) approach of ICS.

Thus, Figure 3-4 not only illustrates the visual perception configuration (1), but also how communication impinged on the higher-level processors (2). The consequence of the conflicting statements is increased central processor activity (3). The subsystems responsible for extracting the propositional and implicational meaning from the information presented to them by the peripheral processors form a loop between them in order to arrive at definite conclusions as to the further course of action to be taken. This loop takes up the processing resources from the peripheral subsystems, and also leads to reduced processing of novel information extracted from the environment.

Table 3-3 and Table 3-4 briefly summarise the strengths and weaknesses of the ICS approach to modelling the cognitive basis of erroneous action.

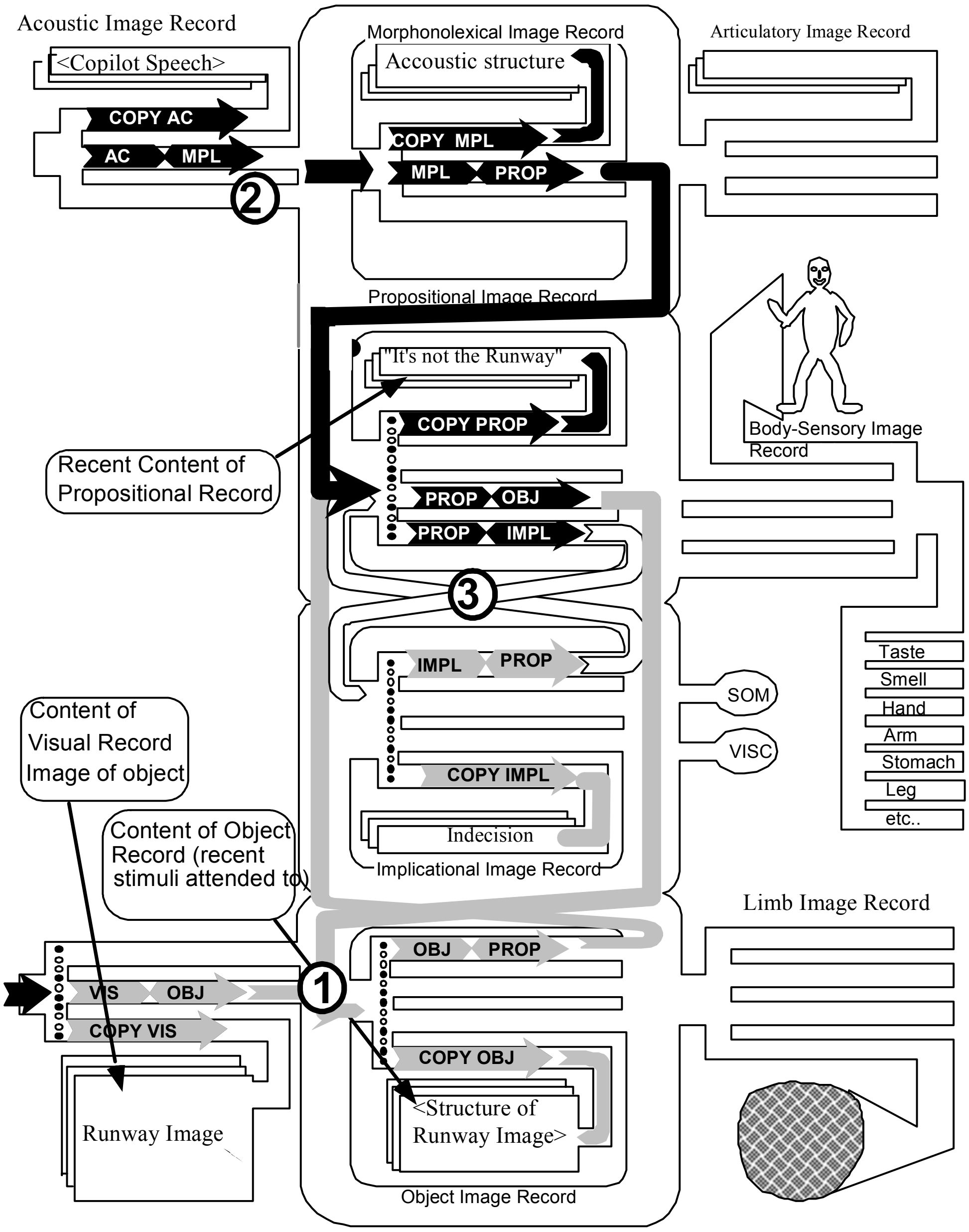


**Figure 3-4 -** Faulty Visual Identification of Runway

| Strengths |
|---|
| • Optionally supports low-level precision<br>• Clear link to (complex) cognitive theory<br>• Comprehensive in cognitive coverage<br>• Perceptual and motor processes are not only taken into account, but put on equal standing with cognition (holistic view)<br>• Interaction between all processors is made explicit<br>• Could be used to look at timing aspect, though not intended<br>• High explanatory power<br>• Easily incorporates non-expert, erroneous behaviour<br>• Assesses cognitive resource demand and allocation |

**Table 3-3 -** The ICS cognitive architecture reviewed (I)

| Weaknesses |
|---|
| • Weak on generating recommendations for error avoidance (though can act as a means to contextualise the effects)<br>• Requires expertise and time<br>• Understanding of underlying low level cognition is necessary. The analysis process has not much methodology support. |

**Table 3-4 -** The ICS cognitive architecture reviewed (II)

# RESULTS OF THE COGNITIVE MODEL COMPARISON

Cognitive models may help us to understand the causes of major accidents. Both EPIC and ICS highlight the perceptual and motor aspects of pilot performance. EPIC’s fine-grain analysis enables the analyst to identify mismatches in the required and actual visual cues. ICS explicitly shows the concurrency in cognitive processing. It also details the cognitive resources necessary for task completion. Thus, ‘pilot error’ can be examined in the light of the cognitive precursors of the actions that have taken place.

ICS is a rich and expressive modelling approach. Still, compared to EPIC, the grain of analysis is not predetermined by the architecture. Importantly, no assumptions regarding a certain level of performance (such as novice/expert) are made in ICS. Therefore, there are no obstacles to modelling erroneous behaviour directly. EPIC models are based on normative tasks and performance and have to be iteratively altered to pinpoint the trace when erroneous performance simulation has been induced. Only then can hypotheses about the cognitive precursor of the ‘error’ be formed and validated. EPIC architectural constraints also do not invite wide-scope inclusion of contributing factors, as does ICS. However, ICS as it stands now does not provide a clear modelling strategy and the outcome depends to a great part on the expertise of the analyst. Neither of the models provides an easy link between modelling outcome and action recommendations for future error prevention.

It has been suggested that the problem of the gap between research and application in the case of user models lies in the fact that much research has been based on the wrong type of psychology, an information processing view of psychology that is reductionist and context-free. An alternative, richer view of human behaviour that is holistic and contextualized is needed. ICS is a step towards this more holistic view of human cognition. Further research will need to concern itself with the integration of analysis approaches and explanation frameworks that look beyond the individual's cognition.

# CONCLUSION

Behavioural task analysis, as the term suggest, deals only with the behavioural analysis of users' tasks. The modified task analysis approach embodied in Cognitive Task Analysis (CTA) methods also attempts to capture the psychological task – goal hierarchy, and associated attributes such as information needs, interface complexity, or timing information. Typically, CTA approaches embody information on human information processing mechanisms implicitly, as a basis of the respective approach (see for instance GOMS). However, these cognitive principles are biased towards expert, error-free behaviour (Simon, 1988; Booth, 1991; Grant and Mayes, 1991). These assumptions are ingrained in the structure of those approaches, and thus major changes to the methods themselves would be necessary (if at all possible) in order to represent erroneous action. Finally, cognitive processing models, or cognitive architectures, such as ICS, present a framework in which both error-free as well as erroneous behaviour can be expressed. It embodies descriptions of the users' tasks, as well as user knowledge and general human information processing mechanisms.

Furthermore, the consideration of time is important for the analysis and description of human error (and human action more generally). Surprisingly, few of the existing action and error taxonomies include the aspect of time, but rather describe and classify human error on an atemporal (static) basis, i.e. a classification of past and observed events. Time may enter only in a few cases as one of the possible causes (i.e. incorrect timing of actions) (Hollnagel, 1991). Cognitive modelling based on the ICS framework can complement the identified error categories by providing temporal information.

ICS seems to be a potential candidate for the combined approach to error analysis proposed in this thesis. Most importantly, it lends itself to modelling the cognition underlying erroneous action, and is not dependent on cognitive models assuming ideal users in ideal situations. The cognitive error modelling approach was piloted in a study on how Barnard's ICS framework can be used to model human error in the context of interface design (see Appendix A).

In the remainder of this dissertation, the extent to which an ICS based cognitive modelling approach can complement a conceptual error taxonomy (such as Reason's) is examined, and its application to incident and accident analysis will be demonstrated.

# CHAPTER 4 SUPPORTING THE UNDERSTANDING AND CATEGORISATION OF HUMAN ERROR THROUGH COGNITIVE ANALYSIS

## INTRODUCTION

This thesis argues that error analysis and categorisation techniques in accident analysis will benefit from representing a cognition-based error model within a cognitive architecture, such as ICS. Using a cognitive framework to analyze the ‘erroneous’ instance of human behaviour can underpin any error categorization and support the reasoning process about alternative interpretations. It can also help document and communicate the analysis results and its rationale to the parties involved in the accident investigation as well as to fellow experts. Lastly, it can be used as a means to ground the error analysis in a theory of human thought. Throughout this chapter the above described critique on current human error analysis models in the context of accident analysis will be elucidated by further examples. The goal is to demonstrate how the cognitive error modelling approach suggested here can complement current error analysis models and support the meaningful analysis of human error. The suggested cognitive error analysis approach is shown to aid the analysis of and reasoning about human error and its potential causes. Thus a more complete understanding of human error in accidents can be achieved.

As described in Chapter 2 of this thesis, current human error analysis approaches in accident investigation leave a lot to be desired. The main shortcomings in accident investigation that were identified in this thesis concerned the use of "common sense" theories of human thought and action (i.e. not grounded in psychological action theories and empirical data)[18]. Thus, there might be an implicit use of personal theories on human error mechanism and accident causations being used, that might lead to the questionable, one-sided labelling of instances of human behaviour through subjective classification into "human error" categories. Furthermore, analysts (as humans prone to biases[19]) might jump to conclusions and link human error analysis and safety recommendations without referring to a grounded, theoretical understanding which can be communicated and validated in more than just a subjective, verbal, manner.

The nature and causes of failures due to human error remain relatively poor lyunderstood (O'Hare et al., 1994). Reason et al (Reason, 1990) maintained that 'one

---

[18] "Common sense" reasoning is typically characterized by its seemingly unproblematic and self-evident explanation of observed phenomena, explanations which often harbour contradictions and are often applied either randomly or they may fall prey of psychological biases (such as the confirmation bias) themselves (see e.g. Mills, 1970; Porter, 1981). "Scientific thought or method" tries to counteract these biases by grounding theories in empirical data or placing it within the scientific discourse. One critical distinction between the scientific method and 'common sense' is that scientific discourse highlights the necessity to unearth the assumptions upon which any explanations of an observed phenomenon are based. Thus, it aims at an 'explicit' reasoning process, i.e. transparent and open to criticism and further discourse.

[19] See also Johnson (2000) for analysts' biases, and, as mentioned previously, Lekberg (1997), and of course, Reason (1990).

of the applied psychologist's more pressing tasks is to provide accident investigators with a better classification of the possible varieties of human failure'. O'Hare and his colleagues analysed a database of aircraft accidents and incidents by applying two different error classification schemes. They stress that they only attempted to investigate *what* failed in each of these events, and not the 'mechanism of malfunction', meaning *how* it failed (Rasmussen, 1982). This, however, needs to be determined 'to trace the information processing failures associated with each event'. This chapter illustrates an approach with which a catalogue of theory-based error classifications can be translated into the framework of a cognitive architecture, thus bridging the gap experienced by O'Hare between mere categorisation of error and 'probing (Johnson, 1999) more deeply by means of theoretically based models of human information processing'.

As O'Hare also pointed out more recently in his review of the first 5 years of the International Journal of Aviation Psychology (O'Hare and Lawrence, 2000), the focus in the community of aviation researchers and practitioners could *still* be shown to be on a "blame and train" approach, rather than to focus on a meaningful, contextual analysis of "human error" with associated safety recommendations targeting the system on the whole, rather than e.g. an individual's memory (an approach that is not sufficient for the systematic prevention of accidents (e.g. see Busse and Wright, 2000; Westrum, 2001; or Johnson, 2002). Thus, the cognitive error analysis approach that is put forward here also aims at the generation and evaluation of safety recommendations, next to the initial analysis and categorization of the accident. The benefit of this approach will be demonstrated in this chapter, and in the case studies reported in the remainder of this thesis.

# CASE STUDY: THE TWIN SQUIRREL HELICOPTER ACCIDENT

On October 22 1996, five people were killed in an accident involving a helicopter AS 355F1 Twin Squirrel. The accident occurred when the helicopter was returning to London from a private landing site in Lancashire with one pilot and four passengers on board. The aircraft was being flown at night in visual contact with the ground when the pilot decided to climb to a higher altitude. During the climb he was deprived of external visual references and the aircraft adopted a steep nose-up attitude during which the air speed reduced below a minimum recommended speed for instrumental flight. This unintentional manoeuvre then developed into a fast, spiral descent. The helicopter did not recover from the dive and it crashed into a field on the outskirts of Middlewich, broke up and caught fire. The accident was not survivable (Air Accident Investigation Branch (AAIB), 1997).

The investigation identified six causal factors, one of which concerned the commander's workload in marginal weather conditions. Another one suggested that the commander may have been distracted at a critical time by the opening of a cabin door. Furthermore, the entire situation is underpinned by the pilot's disorientation and the inability to recover from it.

Examining the pilot's behaviour in the light of these causal factors can give rise to several interpretations.  Two of the possible viewpoints are discussed in detail as the section progresses. They refer to Reason's taxonomy of human error, and identify two rule-based failure modes underlying the pilot's inability to perform the appropriate recovery manoeuvres. On the one hand, as shown below, this could be put down to a rule-based mistake such as the 'First Exception' class of errors (described above). Alternatively, Reason identifies 'Information Overload' as a possible failure mode at rule-based level of performance. This could be seen as being a major contributing factor in the given accident sequence.

As can be seen, these two categorisations of pilot error are general in nature. We will show below how they can be complemented by an anlysis of the underlying cognition within the ICS framework. The more precise and detailed vocabulary offered by the ICS architecture can accommodate modelling to reach beyond surface characterisation of human error. We will illustrate this in the following section.

## ANALYSIS OF ERRORS IN TERMS OF THEIR UNDERLYING COGNITION

In the following section we show how errors leading up to the Middlewich accident can be categorised according to Reason's classification scheme and subsequently modelled in the ICS architecture. Thus, the relationship of these errors to the underlying cognitive mechanisms as proposed by Reason can be established.

At a crucial point in the run up to the accident, the pilot became disoriented after he lost external visual attitude reference, and in spite of several observed coping manoeuvres, he never recovered.

Attitude information was available through the main attitude indicator, and should have been confirmed by the standby attitude indicator. The latter, however, had most probably not been switched on at the beginning of the flight, and therefore showed erroneous indications. Furthermore, the pitch rate was sufficiently slow and steady for the commander not to be aware of the attitude change. He thus was faced with a mismatch of his expectations and two diverging indications on the standby and main attitude indicators. If both instruments had been giving much the same attitude information, the pilot may safely have assumed that he is suffering from an illusion.

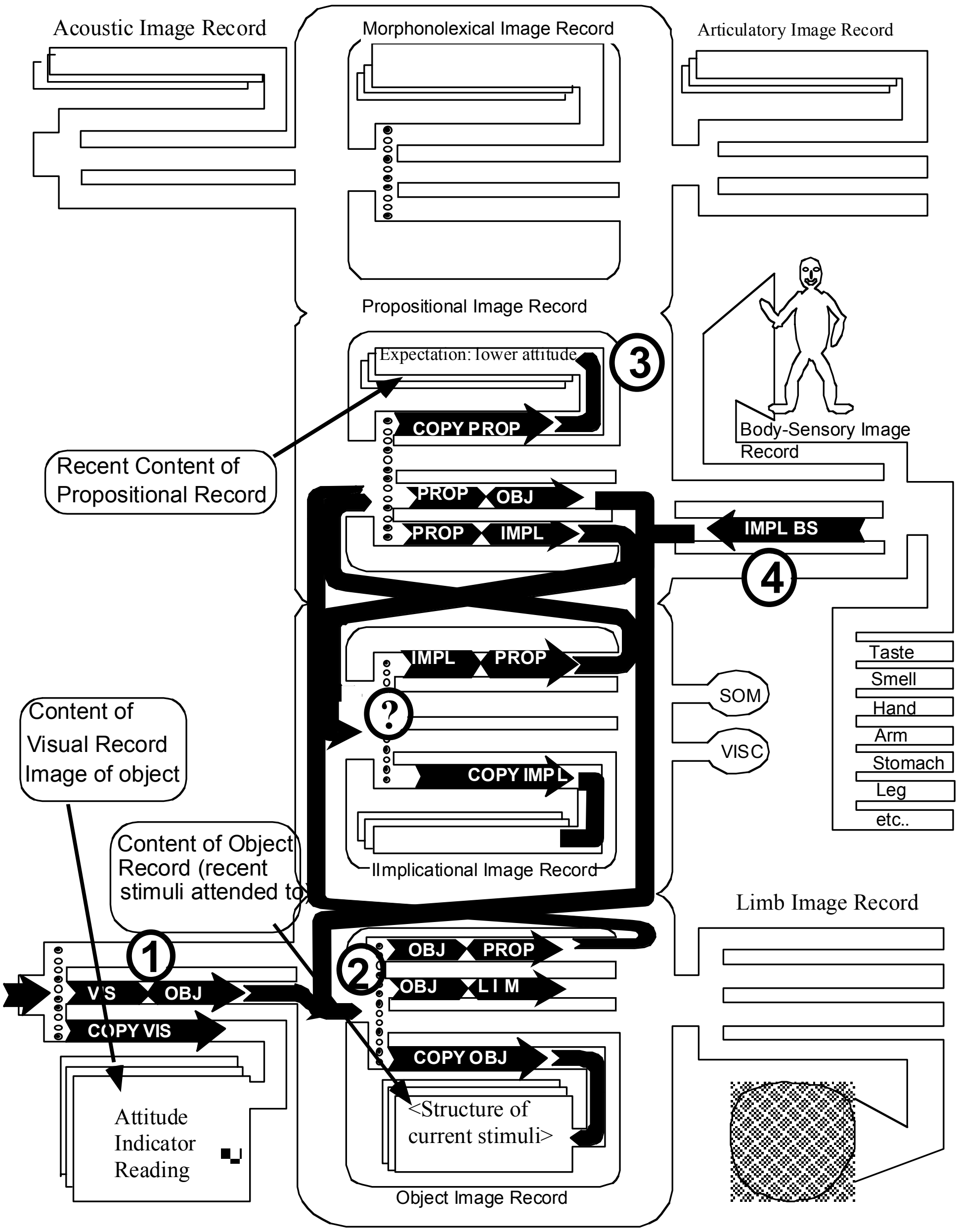


**Figure 4-1 -** Rule-based Mistake: First Exception

According to Reason’s taxonomy, the pilot’s cognition and resulting behaviour in this chain of events could be classed as a misapplication of a good rule (see above). Reason stresses the role of ‘first exceptions’ to a general rule, which are most likely to be overridden by ‘strong-but-wrong’ rules. The pilot’s instrument flying skills had not been formally examined since April 1992, and he had not been required to rehearse recoveries from unusual positions. His loss of orientation combined with facing a mismatch of sensory perception and instrument indication can be seen as the ‘first exception’ to the general rule when not experiencing a mismatch.

The underlying cognition can be modelled in ICS as shown in Figure 4-1.

The visual data is received at the visual subsystem (1), sent to the object subsystem for the recovery of a structural description (2), and finally interpreted by the propositional subsystem (3). The information is fed forward into the implicational subsystem, which interprets the data in the light of the current context. In the meantime, the implicational subsystem receives contradictory information from the body-sensory subsystem (4), which claims to sense no change in attitude. If, however, the propositional subsystem receives ambiguous structural information, and it proves unable to blend the incoming data streams, a selection process will take place, based on the rules available to it and their respective strengths. The feedback information received from the implicational subsystem also plays a guiding role in input and thus rule selection.

The choice of input stream taken by the propositional subsystem might fit in with the implicational interpretation of what is perceived, and thus stabilise in the cognitive system. If the assumption underlying the choice of what data is used to eliminate the ambiguity is wrong, however, the representation of what is thought to be perceived will also be incorrect. The wrong data will be favoured.
The modelling of the resulting rule-based mistake (as defined by Reason) in terms of a cognitive architecture sheds light on the underlying cognitive processes and hence the underlying causes of user error.

# REASONING ABOUT ALTERNATIVE ANALYSES OF ERROR CAUSES

Modelling this scenario in ICS showed how an instance of Reason's class of rule-based mistakes could be investigated at a more detailed level. This complements the more general categorisation of human error by Reason's taxonomy alone.

The above interpretation of causal factors represents one possible underlying cause of the described error. However, the same manifestation of user behaviour might also point towards a second, different underlying cognitive mechanism. Employing Reason's taxonomy, the commander not being aware of the attitude change can be classed as a rule-based mistake as modelled above. On the other hand, it could also be classed as a rule-based mistake as mediated by information overload.

Reason cites the abundance of information confronting the problem-solver in most real-life situations as one basis for rule-based mistakes. He states that this almost invariably exceeds the cognitive system's ability to apprehend all the signs present in a situation. Applied to our case study, the interplay of contradictory attitude information on the one hand, and the opening of the cabin door on the other can be seen as leading to cognitive information overload.

This scenario particularly lends itself to being expressed in the 'cognitive language' provided by ICS. The limitations of human cognition in the face of information overload, or cognitive strain, is built into ICS as the architectural constraint of subsystems not to processing simultaneous inputs which belong to distinct configurations. Using ICS can help to express the details of Reason's 'information overload' more precisely.

The problem-solving configuration described above remains, but now is supplemented by a second configuration, which describes the cognitive resources required when processing the opening of the cabin door (see Figure 4-2).

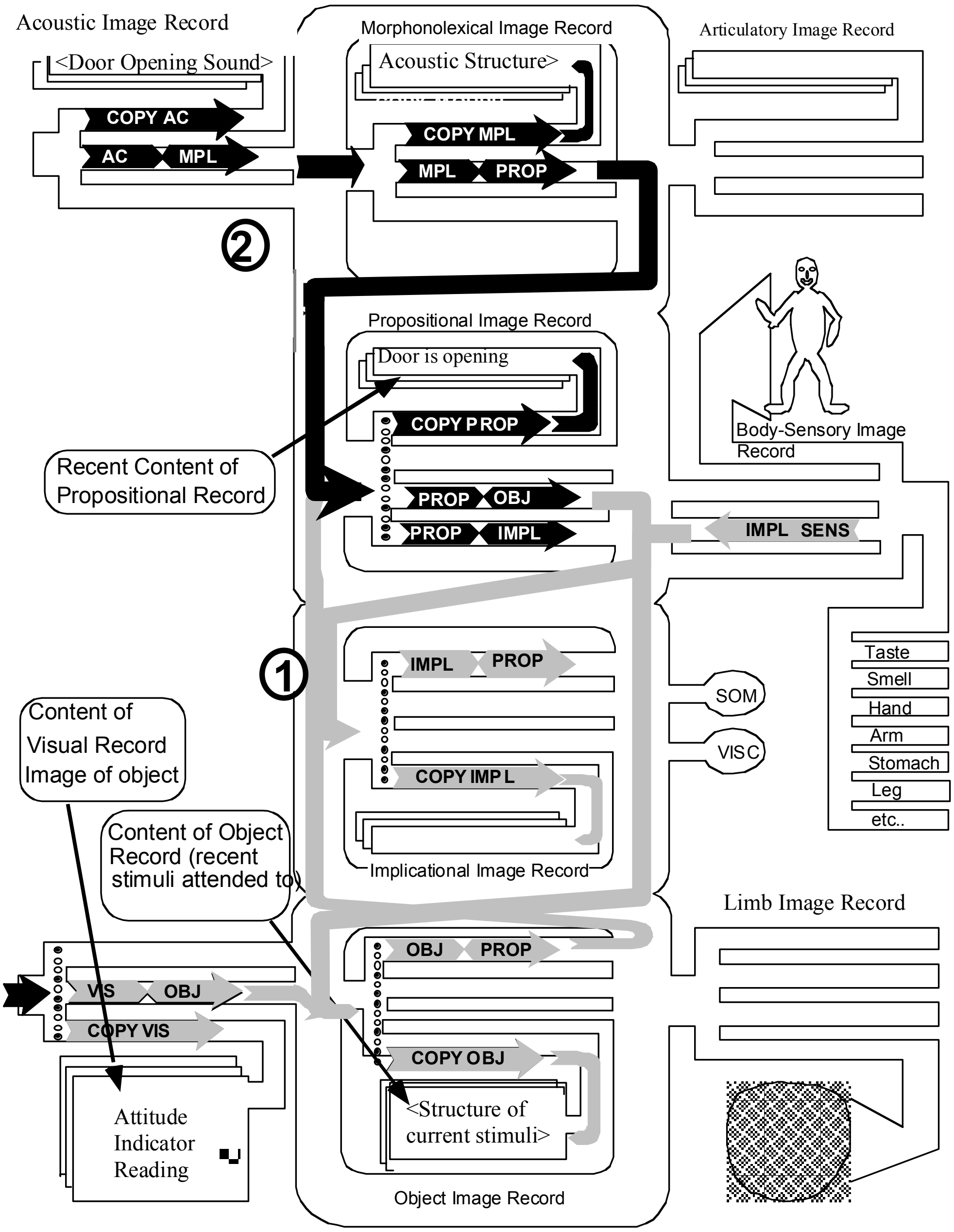


**Figure 4-2 -** Information Overload - Competing Configurations

The second configuration (2) originates from the input at the acoustic subsystem by the noise of the opening cabin door. This information demands access to the meaning subsystems, currently utilised by the first configuration (1). Since Principle 1 in ICS does not allow access to a process by more than one configuration at a time, the two configurations compete for the available cognitive processing resources. Thus, on the on the grounds of ICS Principle 1, cognitive overload is established.

Using ICS to model the underlying cognition of the error provides a means of further investigating the behaviour trace leading to an accident. Expressing the rationale for different interpretations within a cognitive framework facilitates their more precise communication and more detailed analysis. In that way, not only *what* failed in accidents, but also *how* and *why* it failed is examined and thus included in the investigation of human error.

# LINKING THE ANALYSIS OF HUMAN ERROR AND SAFETY RECOMMENDATIONS

## Introduction

The human error analysis process as part of accident and incident investigation often entails erroneous behaviour being classified according to error taxonomies. However, the error data to be classified is often under-specified and conflicting. Also, error categories can be ambiguous, vague, and overlapping. Error analysis that is rooted in cognitive theory allows the analyst to gain an understanding of the generic processes underlying the ‘error’, the mechanisms of human cognition. This thesis argues that the use of a cognitive architecture supports error analysis when used as a structural framework for expressing hypotheses about the cognitive origin of human error. By doing this, the rationale behind error classification can be explicated and documented. Cognitive frameworks can provide a vocabulary for reasoning about different possible explanations of the ‘error’, and a vocabulary for validating resulting safety recommendations. In the following sections of this chapter the link between the analysis of human error and safety recommendations is explored.

Examination of the cognitive processing underlying those classification instances could provide leads to the causes and ‘inner workings’ of the error mechanisms. We have, therefore, employed an analysis framework to model the cognition that underlies human error. This enables analysts to benefit from error taxonomies’ abstracting and simplifying effect on the wealth of error data, as well as from the more refined, structured, and detailed information gained by systematic cognitive modelling. Thus, the mappings from category to underlying mechanisms can be examined by reasoning about the underlying processing within a structural framework. This also supports the documentation of the error modelling process. Furthermore, safety recommendations that result from the error analysis can be validated by embedding reasoning about their impact in the existing cognitive error model.

## Case Study: The Gatwick BAC 1-11 Incident

The Air Accident Investigation Branch (AAIB) incident report on the Gatwick BAC 1-11 incident is used to illustrate the link between error analysis and safety recommendations. This incident report, as mentioned in the previous chapter, shows a variety of causal factors. These include higher cognitive mistakes as well as perceptual slips. The incident will be described in more detail in the following, and the cognitive description of human involvement in the incident (as modelled previously) will be shown to act as a backdrop against which possible safety recommendations can be contextualised and evaluated as to their effectiveness.

The incident involved a near-miss ground collision of a Boeing 737 and a British Aerospace One-Eleven (BAC 1-11) at Gatwick Airport. To remind the reader, the BAC 1-11 landed on taxiway 2 at 2123 hours after making a night visual approach to runway 08L. A Boeing 737 had been ordered onto taxiway 2 just previously by the air traffic controller. The Boeing's commander attempted to turn off to the side after observing the landing lights of the BAC approaching. This manoeuvre led to the aircraft's port main wheels leaving the paved surface. It bogged down in the soft ground partially blocking the taxiway with its left wing and rear fuselage. The BAC stopped only 190 metres short of the Boeing 737. There were no injuries.

The AAIB report lists as one causal factor that the BAC commander inaccurately interpreted the cues provided to him by the visual scene on the approach to runway 08L. He consequently landed on taxiway 2 believing it to be runway 08L. The other 'causes' describe various factors believed to have facilitated the cue misinterpretation. This includes the use of both white runway edge lighting as well as bi-directional green taxiway centreline lighting for runway 08L. Communication between the BAC commander and the first officer, which might have facilitated the 'misjudgement', are also mentioned.

From March 1988 to November 1988, there were major night-time reconstructions of runway 08R under way, the main runway at Gatwick. As a result, runway 08R was closed routinely at 2100 hours and runway 08L, which had been serving as taxiway during the day, took over as the main runway. 08L was also known as the 'emergency runway', and featured visual ground (lighting) aids that could signal both its status as a runway or a taxiway.

Taxiway 2 is situated parallel and next to 08L. In order to speed up traffic flow, National Air Traffic Services (NATS) lifted the prohibition of the use of taxiway 2 while aircraft were taking-off or landing on runway 08L during the time of runway 08R reconstruction. This made it possible for the BAC to approach runway 08L for landing while the nearby taxiway 2 was being utilised by the Boeing 737.

The crew of the Boeing saw the plane approaching and attempted to leave the runway, whereas it was not possible for the crew of the BAC to identify obstruction of the runway by another aircraft. Maximum Braking and/or maximum reverse thrust were not used for the whole of the landing run by the BAC, which suggests that the crew was unaware that they had landed on the (obstructed) taxiway until well into the landing roll (AAIB, 1989, p.18).

The prime 'cause' of the incident was identified by the AAIB report to be the "fundamental error by the crew" when they "convinced themselves that taxiway 2, with standard green taxiway centreline lighting, was runway 08L and landed on it" (AAIB, 1989, p.17). The report's conclusion focuses primarily on the ambiguity of the visual ground aids as the main contributing factor for the 'pilot error'.

Accordingly, the safety recommendations list changes to the ground lighting system and updating of the Aeronautical Information Publication (AIP) manual that detailed the visual ground aids provided at Gatwick's main runway. In the following sections we will demonstrate how the ICS cognitive architecture can be employed to reason about the underlying mechanisms of the cited 'pilot error', and also how resulting safety recommendations can be embedded in the resulting cognitive analysis.

## Contextualising Safety Recommendations within the ICS Model

The Gatwick incident report focuses primarily on the inadequacy of the visual ground aids in its treatment of the incident's causes and of the given safety recommendations. This view of the priorities in the investigation's findings can be illustrated and weighed up in an ICS model as shown in Figure 4-2.

Figure 4-3 below details the cognitive processing relevant to the runway identification specific to this case. In the diagram, (1) depicts the conflict between incoming visual information and the pilot's expectation as derived from his mental model of the situation (2). Thus, there is an unsuccessful attempt at matching the propositional and the visual input. Again, visual information other than that concerning the target will not be able to blend with the data originating in the propositional subsystem. A loop between the two subsystems will be maintained until the representations match. Such a match is possible only if either the correct runway is in the field of vision and provides sufficient discriminatory visual cues, or, if this is not the case, then only if the propositional information is underspecified or incorrect.

According to the investigator's judgement on the relevant processes, the faulty identification was caused predominantly by insufficiently discriminatory visual input by means of ambiguous visual ground aids. Thus, the focal point in the ICS model lies in this case at (1), with the visual input wrongly blending with the propositional specification. However, given the ICS architectural constraints, the model also shows how the propositional subsystem, conferring with the implicational processor, has generated a representation of the target (2) and passes this expectation to the object subsystem. This explicates the importance of anticipation and top-down information in human visual processing, and in the pilot's processing of the information available to him. Mirroring the report's interpretation, the 'cause' of the 'pilot error' is to be found in the visual information provided to the pilot. However, the ICS architecture draws attention to the fact that the pilot's expectation also played a significant role. The incident report emphasises that the BAC crew were aware of the use of 08L as emergency runway while not being certain as to which of the light patterns constituted runway 08L as opposed to taxiway 2. Crucially, also the crew's uncertainty about whether the switch to the emergency runway had already taken place can be viewed as a major factor influencing the interpretation of the visual cues available. For instance, one factor dismissed by the report was the radio information transmitted up to 2109 hours to crew stating runway 08R as the current runway without warning the crew of the impending change of runway.

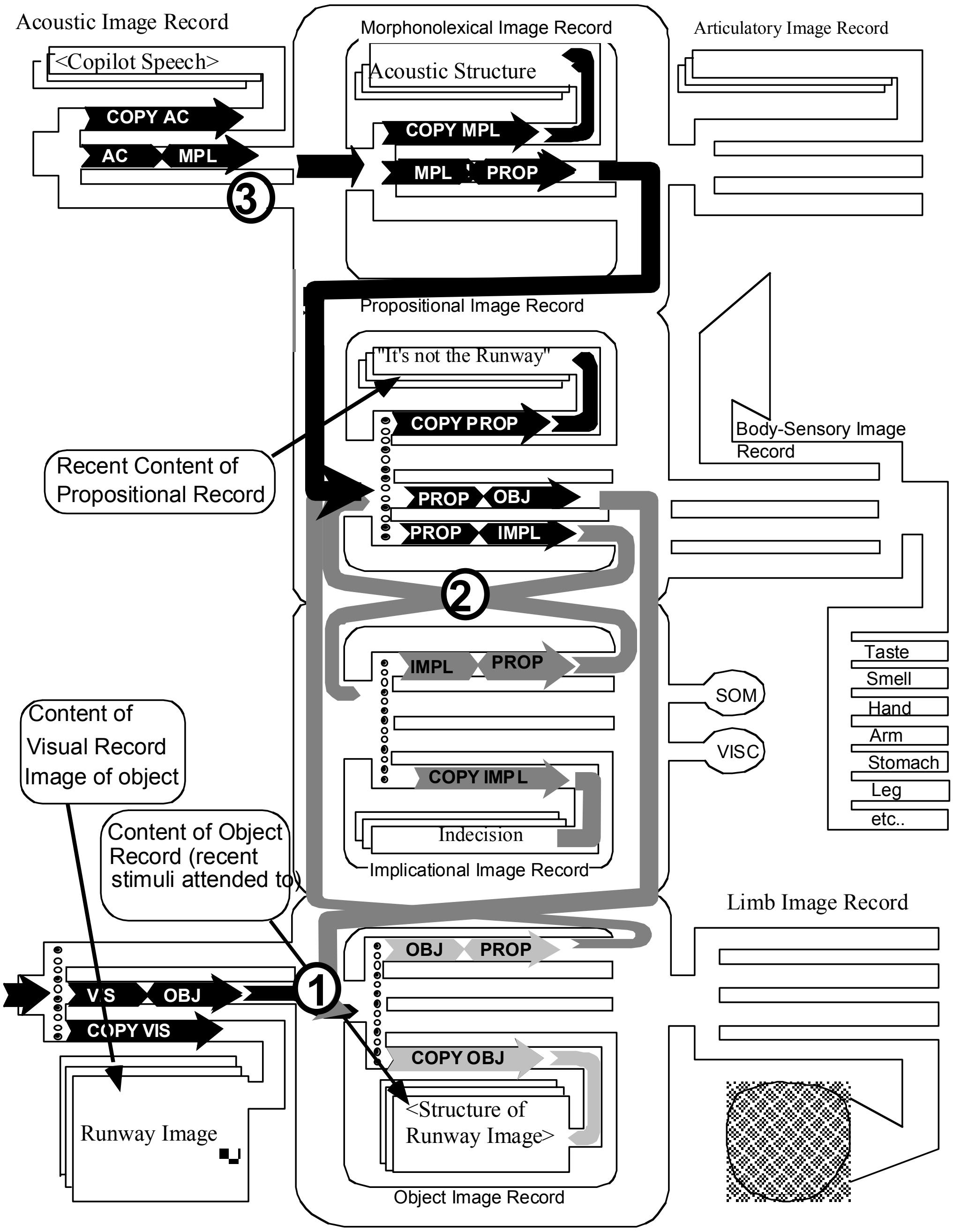


**Figure 4-3 –** The safety recommendations contextualised

The broadcast message at 2109 informed the crew that the runway in use was 08L, but did not explicitly draw attention to the change. This fact was dismissed by the investigation as unimportant since the pilot had responded to subsequent pilot-tower communication, repeating clearance to runway 08L. In Figure 4-2, the propositional expectation is shown as being informed by auditory/morphonolexical as well as implicational input (3). Thus, different hypotheses as to the cognitive processing underlying ‘human error’ can be reasoned about and documented by utilising a cognitive architecture such as ICS.

The effects of safety recommendations that result from the error analysis process can now be embedded and thus evaluated within the cognitive model. The above described approach to error analysis therefore also provides a ‘field’ (or backdrop) on which proposed safety recommendations and their effect on cognitive processing capabilities can be validated. In the given case study, safety recommendations that target visual ground aids only rely on an undisturbed propositional specification. However, there is ample evidence in the AAIB report that points to the contrary.

## CONCLUSION

Human error has been recognised as a predominant factor in aviation mishaps. O’Hare and colleagues (O'Hare et al., 1994) cite estimates of the proportion of mishaps due to human error as ranging between 60% and 80%.

Describing human cognitive errors occurring in the run-up to accidents in detail, or analysing in terms of underlying psychological factors is difficult, since typically, error taxonomies do not provide the necessary depth of analysis. Expressing such error classes within the framework of a cognitive model will allow us to investigate and reason about their underlying psychological causes. A conceptual, systematic technique for categorisation of errors is a prerequisite.

It has been argued that cognitive modelling may help us to understand the causes of accidents and incidents. Furthermore, architectures such as ICS can provide a ‘tool for thought’ and for reasoning about competing explanatory hypotheses on the causes and underlying cognitive mechanisms of ‘human error’. They can provide theoretical grounding and a documentation facility for such speculations, and a backdrop against which effects of safety recommendations can be validated.

ICS explicitly shows the concurrency in cognitive processing. It also details the cognitive resources necessary for task completion. Thus, ‘pilot error’ can be examined in the light of the cognitive precursors of the actions that have taken place. Although ICS is a rich and expressive modelling approach, the grain of analysis is not predetermined by the architecture. Importantly, no assumptions regarding a certain level of performance (such as novice/expert) are made in ICS. Therefore, there are no obstacles to modelling erroneous behaviour directly. ICS architectural and structural constraints invite the wide-scope inclusion of contributing factors, such as interacting with the environment through auditory, visual or other sensory channels. However, ICS in its current form does only partly provide a modelling strategy, and the analysis outcome depends ultimately still on the expertise and modelling skill of the analyst. Although the potential impact of safety recommendations can be assessed, ICS models do not provide an easy link between modelling outcome and action recommendations for future error prevention. However, as was shown in this chap, an ICS analysis can provide the theoretical framework in which the effect of potential safety recommendations can be evaluated.

Thus, this chapter showed the adoption of Reason’s error taxonomy and Barnard’s ICS for the systematic representation of operator error within a theoretical cognitive framework. Operator error in accidents can be described more precisely by linking it to its underlying cognition. Analysis can reach beyond surface categorisation, and it is made possible to reason about the actual causes of error. As a consequence, this approach paves the way for ergonomic design that takes full advantage of the insights expressed in cognitive theory.

Embedding human error modelling into a cognitive theoretical framework helps to express analysts' understanding of the error sources. Communication of their reasoning, based on expertise and experience, has been illustrated in this chapter by using Reason's taxonomy and ICS, as well as by the error analysis provided by existing aviation incident reports. Although the potential impact of safety recommendations can be assessed, ICS models do not provide an *easy* link between modelling outcome and action recommendations for future error prevention. However, as was shown in this chapter, an ICS analysis can provide the theoretical framework in which the effect of potential safety recommendations can be evaluated.

The table below summarizes the key points of critique on current error analysis approaches in accident investigation and provides an overview whether the suggested cognitive error modelling approach would contribute to the solution of the problem.

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Analysis of human behaviour only in terms of "error" (no room for non-erroneous behaviour) | The cognitive error analysis approach embodies erroneous as well as error-free behaviour and thought processes. Error is seen as the other side of the coin of an otherwise efficient thought, emotion, and action apparatus. |

**Table 4-1** - Benefits of the proposed approach (I)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Analysis of 'human error' only in terms of isolated behaviour | Human behaviour is analyzed in its interaction with external situation and events. |
| There is a gap between describing human behaviour as implicated in the accident's causal chain, to assigning a label to it through categorization | Cognitive error analysis models the behavioural description and its underlying cognition in a theoretical framework. The framework can be used to reason about competing categorizations. This justification process, the rationale, is thus traceable and documented. |
| Verbal description of human behaviour as implicated in an accident's causation. | The framework supports diagrammatic reasoning, which has been shown to aid understanding and communication of the analysis results (Moran and Carroll, 1996).<br>The symbolic implementation of the model (May et al., 1993; Barnard, 1988) will aid precision, flexibility, standardization, and predictive power. |

**Table 4-2** - Benefits of the Proposed Approach (II)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Human error predictions based on implicit assumptions | The cognitive architecture allows 'what-if' scenarios to be executed, and presents a framework to reason about potential mental and behavioural consequences of cognitive precursors. |
| Taxonomical labels alone often lack explanatory power | Modelling the cognition underlying the error's categorization contextualizes its causal processes, and thus adds explanatory power grounded in a theoretical framework |
| Taxonomical labels do not support the generation of safety recommendations, and might even mislead the analyst (such as in the case of "reminder statements") | The cognitive error model can help to identify intervention points for future safety measures in the causation processes. It can also be used to validate (or reject) suggested safety recommendations, by grounding both (the error analysis and the simulation of safety measures' impacts) in a common theoretical framework. |

**Table 4-3 -** Benefits of the Proposed Approach (III)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Taxonomic labels do not provide any support for evaluating safety recommendations. | The cognitive error model can be used to compare competing recommendations, help predict their outcomes, and test the compatibility of recommendations with analysis evidence and results. |
| Error categorizations do not sufficiently take team factors into account. | It is possible to model interactions between several individuals in the proposed cognitive error framework, and also the interplay of team members' communication and their knowledge and mental processes. |
| Taxonomies are finite, they cannot be exhaustive, and typically cover only a limited range of errors. | Since the proposed cognitive error analysis approach is grounded in a generic psychological theory, the range of behaviour and cognition that can be described is much wider than a restricted enumerations of categories |

**Table 4-4 -** Benefits of the Proposed Approach (IV)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Taxonomies often list mental and emotional states (such as stress, fatigue) as causal categories, but fail in contextualizing the human's mental state in terms of its impact on cognitive processes or action in a situated work activity. | The cognitive error framework can model emotions (Teasdale and Barnard, 1993) and can thus contextualise the human actor's mental state in terms of their interactions with their environment. |
| Seemingly 'conclusive' labelling of instances of human behaviour, rather than starting point for human error analysis. | Encourages in-depth analysis of the contextualized cognitive precursors defining the human involvement in the accident's causation. |
| Common sense analysis, based on implicit assumption on theories of accident causation as well as of the human psyche. | Encourages use of appropriate expertise on human thought, decision-making, and action processes. |
| Varying levels of "Goodness of Fit" to the accident investigation process and overall safety-management strategy. | The next two chapters in this thesis will evaluate the feasibility of the cognitive error analysis approach in two real-life clinical safety management strategies. |

**Table 4-5 -** Benefits of the Proposed Approach (V)

In the remaining chapters of this thesis, the suggested cognitive error analysis approach will be validated in real-life case studies in the medical domain: one retrospective analysis of an incident reporting scheme, and its associated analysis of “human error”, that had been run in an adult Intensive Care Unit for over 10 years, and one case study in which a full incident reporting system, including the suggested error analysis approach, was implemented and evaluated from scratch as part of existing safety management in a Neonatal Intensive Care Unit.

# CHAPTER 5 RETROSPECTIVE COGNITIVE ERROR ANALYSIS IN AN INCIDENT REPORTING SCHEME IN ADULT INTENSIVE CARE

## INTRODUCTION

As noted in Chapter 2, accident investigation is only one part of a safety management strategy that aims at preventing loss through system failure of any kind. One other prominent method is incident reporting. Incidents are typically defined as near-miss (or minor) accidents. The concepts of accidents and incidents are thus closely related. Van der Schaaf summarized their relationship in a diagram as shown in Figure 5-1.

This diagram (Figure 5-1) illustrates the core concepts in safety management – the factors that might influence an accident's causation, and the role that system defences and recovery (typically accomplished by a human) play in preventing an accident from occurring.

Heinrich (1936) studied the relationship of accidents and incidents in further detail in the early half of the last century, and noted the overwhelming rarity of actual accidents (with associated major loss). He identified a ratio of 300:1 of accidents and incidents, coining what is now known as the "Heinrich ratio". The diagram shown in

Figure 5-2 illustrates this ratio – The Heinrich ratio has also become known as the “iceberg model”.

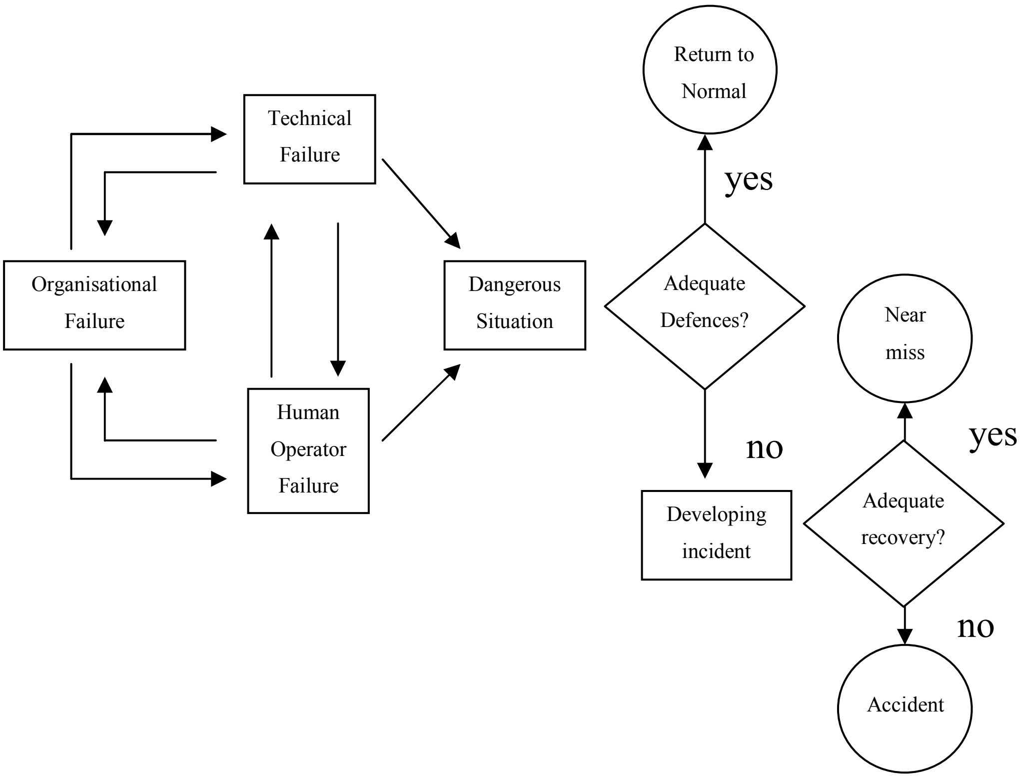


**Figure 5-1 -** The Relationship of Incidents and Accidents

(Van der Schaaf et al., 1991)

One assumption that is conveyed by the iceberg diagram is that accidents and incidents share the same root causes. It is also an accepted assumption that incidents are much more frequent than accidents, and that they thus lend themselves much more readily to large-scale statistics analysis. There is just not enough data on accidents to make this worthwhile, especially given that accidents themselves are often assumed to be a one-off occurrence, whose causation is rooted in an unrepeatable web of unique causation. Whereas incidents are assumed to cluster around failure potential in safety-critical systems, and they are assumed to give clues to "accidents waiting to happen" (Reason, 1990).

Collecting incident data in a database provides a much bigger sample size that enables the drawing of conclusions and generalizations based on statistical analysis that would not be possible with one-off accident occurrences. The need for causal classifications becomes even stronger in the context of incident analysis, since its statistical analysis is dependent on quantitative measures (rather than qualitative descriptions). This need overlaps e.g. with Probabilistic Risk and Safety Analysis (PSA, see e.g. Apostolakis, 1991) and the related Human Reliability Analysis (HRA), which also rely on quantifications for their results – and thus on distinct classifications of 'human error'. However, as is argued in this thesis, a human error classification approach that does not explicate its analysis process and assumptions bears the risk of leading to meaningless, subjective and untraceable classifications, with error categories functioning as empty labels of instances of human behaviour rather than carrying any real semantics. Furthermore, incident classifications may also help in prioritizing future analysis efforts. Thus, an efficient, but grounded, method for error analysis and categorization is needed.

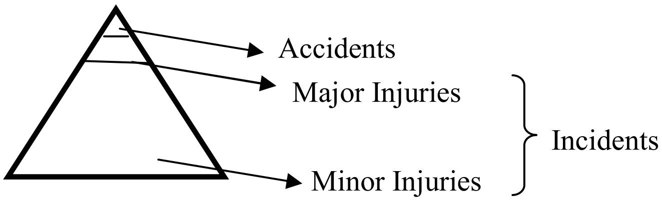


**Figure 5-2 -** Heinrich Ratio: Incidents and Accidents

This chapter will further demonstrate the suggested cognitive error analysis approach by means of a retrospective analysis of an incident reporting scheme and its collected data. Since standardized reporting forms were used in the incident reporting scheme, the data obtained adhered to a certain format. The incident data used had been collected (and pre-analyzed) over the last 10 years. I will first introduce the field of incident reporting and human error analysis as it pertains specifically to the medical domain, before I describe how the cognitive error analysis approach was applied to the incident data in order to shed light on underlying causes, and possibly remedies. A detailed overview of medical incident reporting and an analysis of the Edinburgh Incident Scheme are available in Appendix B.

## IDENTIFICATION AND ANALYSIS OF INCIDENTS IN COMPLEX, MEDICAL ENVIRONMENTS

Medical risk management is often seen as lagging behind other safety-critical industries, where there has been considerable research into safety and accident causation models. Accident analysis models used in, for instance, aviation and process control recognise the importance of formalised root cause analysis, and the multi-level nature of incident causation. Latent factors, such as management and organisational issues, are stressed as underlying structural precursors for incident occurrence. Also, the constraints of the human cognitive system and their relationship to task performance are taken into account. These considerations are reflected in the use of incident reporting schemes, the analysis of the collected incident data, and the generation of remedial action recommendations. In this chapter, we will illustrate how these concepts can be applied in a clinical setting. An incident reporting scheme implemented at an Edinburgh Intensive Care Unit will serve as a case study.

## Clinical Adverse Events

In 1990, the Harvard Medical Practice Study (Harvard University, 1990) investigated the occurrence of patient injury caused by treatment - so-called adverse events. It found that nearly 4% of patients suffered an injury that prolonged their hospital stay or resulted in measurable disability. Leape (1994) pointed out that, if these rates are typical of the US, then 180000 people die each year partly as a result of iatrogenic ('doctor-caused') injury. Since most of the precursors to iatrogenic injuries are perceived to be 'Human Error' (Runciman et al., 1993), the possibility of negligence causes great concern. This is mirrored in the litigious climate in the US (Leape et al., 1991), and led to a considerable increase in interest in the causes of adverse events (Bogner, 1994). The cost of adverse events is high; not only in human suffering, but also in compensation claims and the need for prolonged treatment of afflicted patients.

## Learning from Adverse Events

In the UK, the drive towards clinical effectiveness (Dawson et al., 1998) has led investigators to concentrate on increasing the quality of care while lowering the costs associated with our current health care system. In the course of the clinical effectiveness program, the idea of clinical audits has gained in strength – a "professionally led initiative which seeks to improve the outcome of patient care as a result of clinicians examining their practice and modifying it appropriately" (Wedderburn, 1998). However, the data collected for audits concerns itself only with factors peripheral to adverse events – cost effectiveness being the focus of the investigation (see e.g. Van der Schaaf, 1996). Formal analysis of the causes of adverse events does not take place. Rather, a variety of committees and commissions are typically set up locally, meeting regularly to review any cases of iatrogenic injury that had occurred and had been brought to attention by the staff involved (Leape, 1994). The reporting as well as the analysis of these events, however, are subject to local convention, and thus express the self-regulating policy of the health care community.

In some cases, incident reporting schemes are in place. This concerns “near miss” adverse events, i.e. cases in which iatrogenic injury was likely to have occurred, but the hazardous situation could be recovered from successfully. However, it was noted that even under the clinical reporting schemes, in-depth analysis and search for root causes of adverse events does not take place (Leape, 1994).

## Human Error Analysis in Aviation and Process Control Revisited

In safety-critical domains other than health care, accidents (i.e., like clinical adverse events, non-intended events which lead to negative outcomes and loss) have received a great deal of attention. Research into their causes and prevention has made considerable advances. Accident causes formerly described solely as ‘Human Error’ have come under close scrutiny, notably with the work of Rasmussen (e.g. Rasmussen et al., 1987), Reason (Reason, 1990), Hollnagel (1991), and Hale (Hale et al., 1997), as also described in Chapter 2. They have systematically analysed cognitive mechanisms underlying the various phenotypes of human error. Also, latent contributing factors are taken into account. These concern organisational as well as managerial influences on the course of the accident. Thus, there was a shift from ‘blaming the human’ (such as the oft-cited ‘pilot error’) to the insight that error invariably occurs in complex systems. The aim now is to create error-tolerant systems that absorb errors through ‘system defences’ and provide redundancy and possibilities for error recovery. The move away from the blame culture also made possible the introduction of institutionalised, anonymous, and non-punitive incident reporting schemes. Subsequent detailed comparison and analysis of the identified root causes is carried out by organisations such as the US National Transportation Safety Board (NTSB), or the US Federal Aviation Administration (FAA) and NASA. In the clinical domain the argument has also been put forward for the use of incident reporting to complement post-hoc accident investigation with the inherent problems for scant information, altered perception and outcome bias (Runciman et al., 1993). Furthermore, analysis of the reported events should extend beyond the investigation of proximal causes to include latent system failures. Thus, the use of incident reporting schemes and theories on accident causation are the two main contributions so far to safety applications in the medical domain.

This chapter looks at incident management practices in a clinical setting, as compared to the approaches in other safety-critical domains. The main issues we will investigate in this chapter concern the problems associated with incident reporting, categorisation and subsequent analysis. We will illustrate existing risk management in medicine by an incident reporting scheme employed in an Edinburgh Intensive Care Unit (ICU). First, we will introduce current applications of the Critical Incident Technique (Flanagan, 1954) in medicine and give an outline of the Edinburgh implementation. Then incident reporting, categorisation, and analysis with reference to the Edinburgh scheme will be described in turn, mirroring the stages of the incident investigation process. A later section will then investigate the generation of action recommendations from the prior analyses. In each section, we will compare and contrast theory and methodology 'lessons' from safety-critical domains such as aviation with the implementation of the Edinburgh incident reporting scheme.

## The Edinburgh ICU Incident Reporting Scheme

The Edinburgh incident reporting scheme was set up in an adult intensive care unit in 1989. It has been maintained by Dr David Wright, who is an anaesthetist and one of the ICU consultants. The unit has 8 beds at its disposal, and there are roughly 3 medical staff, one consultant, and up to 8 nurses per shift on the ward. Equipment in an ICU ranges from monitors displaying life sign data, such as heart rate and intra-cranial pressure (ICP), to drug administration equipment, automatic breathing machines, and oxygen humidifier masks. Patient management involves tracking and transcribing monitored vital data, laying and maintaining lines such as endotracheal tubes, and chest drains, and handling equipment, such as three-way taps for drug administration, ventilators and defibrillators.

Incidents reported over ten years (see Wright, 1999) fell mainly in four task domains: relating to ventilation, vascular lines, drug administration, and a miscellaneous group.

The incident scheme employed reporting forms that encouraged staff to describe the event in narrative form, as well as noting contributing factors, detection factors, grade of staff involved in the event and that of the reporting staff.

One crucial factor in the implementation of the scheme was its anonymity. Dr Wright, in his role of the scheme manager, was the only person who had access to the completed forms. The collected data was coded into categories and summarised. It was then collated into frequency tables. Information that may identify staff was removed, and action recommendations were proposed. The results of this initial analysis were then iterated over by Dr Wright and the Senior Nurse of the unit. Together, the data was again inspected and final revisions of action recommendations were carried out. The findings and recommendations were disseminated regularly among the staff of the unit, and thus an effective feedback loop was created.

For a more detailed description of the implementation and findings of the incident reporting scheme, see Wright et al. (1991), Wright (1999), Busse and Johnson, (1999), and Appendix B.

There are several definitions of what constitutes an ‘incident’. Incidents might be considered adverse events only, near miss events only, or both. In the Edinburgh study, staff are asked to report ‘critical incidents’, which are defined as any occurrence that might have led (if not discovered in time) or did lead, to an undesirable outcome. In consequence, each recorded incident:

1. was caused by an error made by a member of staff, or by a failure of equipment
2. can be described in detail by a person who was involved in or who observed the incident
3. occurred while the patient was under our care, though […] it could be patients in transit
4. was clearly preventable

## The Edinburgh Categorisation Scheme

In the Edinburgh study, information drawn from the incident reports were categorised into ‘causes’, ‘contributory factors’, and ‘detection factors’ (see Wright, 1999). The categories were arrived at through informal coding of the narrative incident data. This bottom-up approach led to a domain-specific, behavioural categorisation scheme.

‘Causes’ offers the subcategories of Human Error and Equipment Failure. Any incident that has some degree of human involvement is considered a Human Error. Furthermore, the human error incidents are classified as to the various task and equipment domains these refer to, such as “vascular lines related”, “drugs-administration-related”, or “ventilator-related”. Thus, the categorisation mainly labels the incidents without providing a step towards causal analysis. Rather, it points to where in the patient management task sequence the incident occurred. ‘Cause’ here refers to the task domain of the proximal causal factor. Mostly, the actual proximal ‘cause’ of the incident cannot be inferred from this categorisation per se. A summary of the narrative description of the occurrence is referred to in order to reconstruct the proximal cause.

However, the categorisation of the contributing factors sheds some light on more distal, and less domain-dependent, ‘causes’. Both, the task/equipment domain and the contributory factors together can be seen as representing pointers as to in which task sequence, and where in the task sequence, the underlying problems that led to the incident might be found. Thus, they can focus further enquiry.

The initial categories (Wright et al., 1991) were:

- Inexperience with equipment
- Shortage of trained staff
- Night time
- Fatigue
- Poor Equipment Design
- Unit Busy
- Agency nurse
- Lack of Suitable equipment
- Failure to check equipment
- Failure to perform hourly check
- Poor Communication
- Thoughtlessness

The ‘contributing factors’ categorisation scheme evolved since the time of creation, and nearly doubled from 12 categories to 23. This was the result of the ongoing iterative coding of the collected incident data, and of experience with the reporting scheme. The added categories are:

- Presence of students/teaching

- Too many people present
- Poor visibility/position of equipment
- Grossly obese patient
- Turning the patient
- Patient inadequately sedated
- Lines not properly sutured into place
- Intracranial Pressure Monitor not properly secured
- Endotracheal tube not properly secured
- Chest drain tube not properly secured
- Nasogastric tube not properly secured

In the initial version of the taxonomy, mainly so-called Performance Shaping Factors (see also Rasmussen, (1982)) are listed as contributing causes, such as Fatigue, Unit Busy, and Night Time. Poor Communication can also be considered a performance shaping factor. The one factor notably not a PSF is 'Thoughtlessness'. The only factor that clearly denotes a latent failure is 'Poor Equipment Design'.

Thus, the initial categorisation scheme concerned itself mostly with factors that created the situation precipitating the incident. However, the refined categories are increasingly task and domain specific, and do not denote generic Performance Shaping Factors. Instead, a behavioural and task domain dependent taxonomy is introduced, see especially the last seven factors above. These categories can provide the basis for descriptive statistics on the relative frequency of occurrences and for denoting trends in the distribution and combination of incidents. However, analysis of the underlying causes of the incident is not facilitated.

The evolution of the categorisation scheme itself provides valuable information. The novel factors were created by filtering them out of the incident data reported and analysed over the years. Such a bottom-up approach establishes and clarifies problem areas within both the task domain and the handling of the provided equipment. Thus, insights into task characteristics and performance can be gained.

This can be put to use, for instance, for providing training focus, while offering strong empirical support. For instance, action recommendations in the period form August 1995 to August 1998 pay heed to the recurring problem of dislodged endotracheal tubes. Initially, reminders are repeatedly publicised about this common problem. Reasons for dislodgement are given. Then, a list is devised that summarises “reasons for endotracheal tubes coming out”. This is disseminated, and later ‘suggested actions’ recommend to revise this list and publicise it further.

In comparison of the revised categorisation scheme with the categorisation approaches discussed above it can be noted that a linear causal chain can still be constructed by dividing the data according to, for instance, Rasmussen’s Taxonomy for Description and Analysis of Events involving Human Malfunction (see Figure 2-6). For instance, contributing factors on the AIMS-ICU form are divided into ‘system-based factors’, which detail work condition factors as well as latent failures, and ‘human factors’, which note factors that impact on the human cognitive processing levels. There is a trade-off, however, since the provision of fixed categories lessens the flexibility of the data reported, and might stifle creativity for staff attempting to explain how and why the incident occurred.

## Multi-causal Classification

In the classification of data into the contributing causes categories, combinations of factors are allowed, and are noted frequently in the data sample. The data analysis might thus also be based on noting the frequency or likelihood of certain factors correlating. Trends could be established, and conclusions drawn that are justified by a richer data set than only noting the occurrence of single factors. To illustrate this, we took two data samples (see Table 5-1 and Table 5-2), one sample covering the first categorisation interval, January and February 1989 (sample89), and the other covering a more recent interval from May to November 1998 (sample98). Both samples cover 25 incident reports.

| **'Cause' Occurrence '89** | | **'Contributing Factors' Occurrence '89** | | **'Detection' Occurrence '89** | |
|---|---|---|---|---|---|
| 'Ventilator': | **10** | Poor Communication: | **14** | D1 Regular Checking: | **11** |
| 'Vascular line': | **6** | Poor Equip. Design: | **11** | D2 Alarms: | **11** |
| 'Miscellaneous': | **5** | Inexperience with Equipment: | **5** | D3 Experienced Staff: | **8** |
| 'Disposable Equipment': | **4** | Lack of Suitable Equipment: | **4** | D5 Patient Noticed: | **1** |
| 'Drug-administration': | **3** | Night Time: | **3** | | |
| 'Non-disp. Equipment': | **2** | Fatigue: | **3** | | |
| | | Unit Busy: | **2** | | |
| | | Failure to Perform Hourly Check: | **2** | | |
| | | Thoughtlessness: | **2** | | |

**Table 5-1** - Causal Categorisation Sample '89

In sample98, the predominant factors are 'Thoughtlessness' (10 occurrences), 'Poor Communication' (9 occurrences), and 'Inexperience with Equipment' (5 occurrences). In contrast, in sample89, factors 'Poor Communication' (14 occurrences), 'Poor Equipment Design' (11 occurrences), 'Inexperience with equipment' (6 occurrences), and 'Lack of suitable Equipment' (4 occurrences) were most implicated in incidents.

| **‘Cause’ Occurrence ‘98** | **‘Contributing Factors’ Occurrence ‘98** | **‘Detection’ Occurrence ‘98** |
|---|---|---|
| ‘Drug-administration’: **10**<br>‘Ventilator’: **8**<br>‘Vascular line’: **4**<br>‘Miscellaneous’: **4**<br>‘Non-disp. Equipment’: **1** | Thoughtlessness: **11**<br>Poor Communication: **8**<br>Inexperience with Equipment: **4**<br>Night Time: **3**<br>Failure to Check Equipment: **3**<br>Failure to Perform Hourly Check: **2**<br>Endotrach. Tube Not Properly Sutured: **2**<br>Poor Equipment Design: **1**<br>Patient Inadequately Sedated: **1**<br>Turning the Patient: **1** | D1 Regular Checking: **9**<br>D3 Experienced Staff: **8**<br>D2 Alarms: **2**<br>D4 Unfamiliar Noise: **1**<br>D5 Patient Noticed: **1**<br>D7 Handover Check: **1** |

**Table 5-2 -** Causal Categorisation Sample ‘98

A closer look at the data reveals that in the 1989 interval, half of all ‘Poor Equipment Design’ incidents and one third of ‘Poor Communication’ are not single factor categorisations, but are placed in combinations. ‘Poor Equipment Design’ is predominantly (four out of five incidents) paired with ‘Lack of Suitable Equipment’. ‘Poor Communication’ is combined with a variety of factors, such as ‘Fatigue’, ‘Thoughtlessness’, and ‘Unit Busy’. This use of combinatorial categorisation embeds behavioural and person factors (e.g. Failure to Check Equipment, Thoughtlessness) within latent system and work condition factors such as Poor Equipment Design, Fatigue, and Poor Communication.

Also, in sample 98, ‘Poor Communication’ is paired with other factors in 6 out of 9 incidents. ‘Thoughtlessness’ is shown in combination with other factors (such as Inexperience with Equipment) in 4 out of 10 incidents. ‘Inexperience with Equipment’ is only ever mentioned in combination. In sample 89, ‘Inexperience with Equipment’ is left as sole contributory factor only twice out of a total of six incidents. Three times it is mentioned in combination with ‘Poor Equipment Design’ and ‘Lack of Suitable Equipment’, respectively. Again, this shows how factors that warrant further explanation can be placed in context by considering the multi-combinatorial categorisation. It also shows, however, that descriptive statistics neglecting this facet of analysis shed a slightly misleading light on the collected incident data.

## Incident Detection

The crucial role of detection factors is being recognised in the Edinburgh scheme. This is reflected on the incident form, as well as in the conclusions that are drawn form the data (see Table 5-1 and Table 5-2).

Not only needs to be observed that although humans cause incidents, it is also humans who detect it and either remedy the consequences or prohibit the course of the incident to proceed. Staff is encouraged specifically to note which factors are believed to have aided detection. This is not only the data that can with significant confidence be assumed to be the reporter’s own experience, but also that what ultimately can assist in finding ways of reducing the number of incidents and consequently, accidents.

The detection factor taxonomy evolved alongside the iterative development of the contributory factors taxonomy. Initially it consisted of the categories ‘Repeated Regular Checking’, ‘Presence of Alarms on Equipment’, ‘Presence of Experienced

Staff', 'Hearing Unfamiliar Noise', 'Patient Noticed', and 'Relative Noticed'. A task specific factor 'Having Lines or Three Way Tap Visible' was added, as well as the factor 'Handover Check'. These added factors point to possible system improvements to facilitate detection. They can be actively influenced by system factors such as work procedures, design, training, and staffing levels.

This is pointed out in the initial presentation of the incident scheme (Wright et al., 1991). It states that "regular checking by experienced staff is critical in detecting errors, but this may be adversely affected by nurse staffing policies where agency staff are commonly used or where little time is available for handovers".

The iteration over the collected incident data thus clarified two more detection facilitating conditions. The importance of handover checks to make up for contributing causes 'Failure to Check Equipment' and 'Failure to Perform Hourly Check' is pointed out. Without iterative revision and coding of the data categorisation, these factors might have gone neglected. A formalised framework of analysis, such as those used in the aviation domain, can aid the recognition of detection factors and the generation of suggested actions (see also later sections in this chapter).

## THE EDINBURGH STUDY: INCIDENT ANALYSIS

The narrative given by the reporting staff on the incident report form provides the first level of interpretation of what happened. In the case of the person reporting the incident not being the same as the one having 'caused' it, the reporter provides a second-level interpretation of the events.

The narrative, together with the contributory and detection factors mentioned, typically lays out a timeline of the events. Otherwise, this is inferred when analysing the data. One method of establishing a task-related timeline is to embed the erroneous task event into a sequential, high-level task model. This is partly carried out by classifying occurrences according to task aspects, as shown in the examples below.

The classification of events involves informal (and non-documented) analysis followed by the above noted categorisation into ‘causes’, contributory, and detection factors. This can be seen as representing an informal root cause analysis process (see below). However, to repeat, the categorised data often seems to present behavioural descriptions or proximal ‘causes’.

Following Reason’s accident causation and analysis model (see Figure 2-3), Table 5-3 shows a tentative classification of the provided ‘contributing factors’ categories into latent failure types (distal causal factor), work conditions failure types (distal causal factor), and active failures (proximal causal factor). The latter constitute in our case task and behaviour oriented categories, rather than the error types based on cognitive theory as suggested by Reason. This classification can be compared to Rasmussen’s event description scheme (see Figure 2-6), where latent factors and work conditions are mirrored in ‘Personnel Task’, and  the ‘Causes of Human Malfunction’, ‘Situation Factors’, and ‘Factors Affecting Performance’ respectively.

| Contributing Factor Categories | Failure Type |
|---|---|
| 1. Inexperience with equipment | 1. Distal: Training, Protocol |
| 2. Shortage of trained staff | 2. Distal: Management, shift arrangement |
| 3. Night time | 3. Distal: e.g. Fatigue/Staffing level |
| 4. Fatigue | 4. Distal: Shift |
| 5. Poor Equipment Design | 5. Distal: System Design |
| 6. Unit Busy | 6. Distal – refined by #13 and #14 |
| 7. Agency nurse | 7. Distal: Contract work, training |
| 8. Lack of Suitable equipment | 8. Distal: management of equipment maintenance |
| 9. Failure to check equipment | 9. Proximal, behavioural |
| 10. Failure to perform hourly check | 10. “ |
| 11. Poor Communication | 11. Distal - Team Communication or Training |
| 12. Thoughtlessness | 12. Proximal |
| 13. Presence of students/teaching | 13. Distal (refining #14) |
| 14. Too many people present | 14. Distal (refining #6) |
| 15. Poor visibility/position of equipment | 15. Distal – detection facilitating factor |
| 16. Grossly obese patient | 16. Patient characteristic |
| 17. Turning the patient | 17. Proximal – Task/behavioural |
| 18. Patient inadequately sedated | 18. Proximal “ |
| 19. Lines not properly sutured into place | 19. Proximal “ |
| 20. ICP monitor not properly secured | 20. Proximal “ |
| 21. Endotracheal tube not properly secured | 21. Proximal “ |
| 22. Chest drain tube not properly secured | 22. Proximal “ |
| 23. Nasogastric tube not properly secured | 23. Proximal “ |

**Table 5-3** – Failure Type Categorisation

In order to be able to categorise the proximal failure types into Reason's cognitive Error Types, more detailed incident data is required than could be accessed from the samples. There is no one to one relationship between contributing factors and their underlying cognitive mechanisms (see for instance Busse and Johnson (1998), and a more in-depth analysis of the error types is required.

Despite informal causal analysis during the categorization process, the Edinburgh study does currently not proceed much beyond the "what" phase within the above mentioned analysis model. However, given the contributing factors classification above, root cause analysis can be used to reflect the variety of levels in the incident causation tree. The consideration of latent and work condition factors draws attention to the deficiency of single-cause categorisation. Multi-causal categorisation can be used to reconstruct a possible root cause analysis as an example.

For instance, the above-mentioned combination of factors 'inexperience with equipment', 'poor equipment design', and 'lack of suitable equipment' can be illustrated in a causal tree as shown in Figure 5-3. The incident (e.g. ventilator related) was 'caused' by staff 'inexperience with equipment'. This is then hypothesised to be mediated by 'lack of suitable equipment', which in turn pointed to 'poor equipment design'. Thus, a causal chain is established, formalised, and documented.

It could also be argued that 'lack of suitable equipment' contributed directly to the occurrence of the incident. This hypothesis, again, has different implication for potential system redesign. Thus, this kind of analysis, taking several levels of causation into account, can aid precise and structured reasoning about the incident occurrence. Factors to be considered in lower levels of causation are work conditions and latent system failure types, for instance Training, and alternative contributing factors why a failure to check equipment occurred.

A structured, formalised analysis framework is also necessary to prevent hindsight from biasing error analysis. Attribution of error is a social and psychological judgement process rather than a matter of objective fact. Hindsight view is fundamentally flawed because it does not reflect the situation confronting the practitioners at the scene. Thus, rather than being a causal category, human error should be seen as representing a symptom, and a starting point for investigation (Woods et al., 1994).

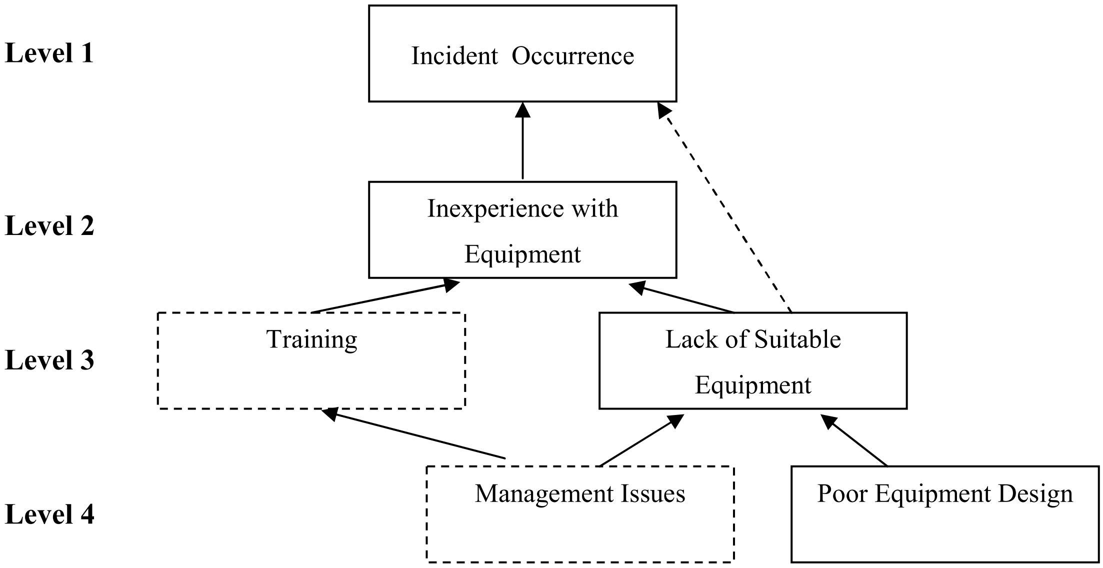


**Figure 5-3** - Example of Incident Root Cause Analysis

# ACTION RECOMMENDATION

In order to arrive at sound and relevant action recommendations, a systematic and structured way of bridging the result of the analysis process to remedial measures is needed. The documentation of this process plays an important role, for instance to allow monitoring of the effect of the measure.

In industry domains such as aviation and process control, cognitive analysis of error occurrences is often used to point towards remedial actions. For instance, Reece et al. (Reece and Hill, 1995) investigated human error in radiation exposure events. The proximal cause to the incident was situated in the task sequence and, additionally, cognitive failure analysis was carried out. Then the relationship between them was analysed, and thus it could be identified:

- where training might be most effective
- where equipment interface enhancements may be most appropriate
- where job aids might help performance (e.g. checklists).

In a similar vein, van der Schaaf (1996) proposed the Eindhoven Classification Scheme for classifying events and identifying incident causes in process control. The main categories represent Technical Factors, Organisational Factors, and Human Error categorised according to Rasmussen's Skills Rules and Knowledge (SRK) framework. The translation into proposals for effective, preventive, and corrective action can then be guided by means of a proposed Classification/Action Matrix. Thus, the action categories relate back to the SRK error types, and include Equipment, Procedures, Information & Communication, Training, and Motivation.

This shows how error categorisation, when done according to cognitive level of performance and latent factors, can provide the basis for sound, structured, and theory-based remedial recommendations. Without error categories being based on sound psychological theory, systematic and relevant action recommendation generation is not possible.

| **‘Cause’** | **Contributing Factors** | **Detection Factors** | **Action Recommendations** |
|---|---|---|---|
| Ventilator **10**<br>Vasc. Line **6**<br>Misc. **5**<br>Disp. Equip. **4**<br>Drug-admin. **3**<br>Non-disp. Equip. **2** | Thoughtlessness: **14**<br>Poor Equip. Design: **11**<br>Inexperience with Equipment: **5**<br>Lack of Suitable Equipment: **4**<br>Night Time: **3**<br>Fatigue: **3**<br>Unit Busy: **2**<br>Failure to Perform Hourly Check: **2**<br>Thoughtlessness **2** | Reg. Checking: **11**<br>Alarms: **11**<br>Exp. Staff: **8**<br>Pat. Noticed: **1** | **4** Ventilator<br>**3** Vascular Line<br>**2** Drugs<br>**1** Miscellaneous<br>**1** Equipment |

**Table 5-4** - Action Recommendations for the reporting period Jan/Feb 1989

| **Incident ‘Cause’ 98** | **Contributing Factors 98** | **Detection Factors 98** | **Action Recommendations 98** |
|---|---|---|---|
| Drug-admin. **10**<br>Ventilator **8**<br>Vasc. Line **4**<br>Misc. **4**<br>Non-disp. Equip**1** | Thoughtlessness **11**<br>Poor Communication **8**<br>Inexperience with Equipment: **4**<br>Night Time: **3**<br>Failure to Check Equipment: **3**<br>Failure to Perform Hourly Check: **2**<br>End. Tube not Properly Secured: **2**<br>Poor Equip. Design: **1**<br>Patient Inadequately Sedated: **1**<br>Turning the patient: **1** | Reg. Checking: **9**<br>Exp. Staff: **8**<br>Alarms: **2**<br>Unfamiliar Noise: **1**<br>Patient Noticed: **1**<br>Handover Check: **1** | **3** Ventilator<br>**2** Vascular line<br>**1** Drugs<br>**2** Miscellaneous |

**Table 5-5 -** Action Recommendations for the reporting period May/Nov 1998

## The Edinburgh Study: Action Recommendations

In the Edinburgh study, the incident data was categorised and summarised by the scheme manager. Action recommendations were arrived at in an iterative process, whereby the scheme manager suggested remedial actions and presented those together with the summary data to the senior nurse of the ICU. Together, the data was discussed and the rationale for the action recommendations reviewed. This led to a final version of suggested actions for each incident analysis period. Table 5-4 and Table 5-5 show a categorisation of the suggested actions for our two samples (sample89 and sample98). The revised action recommendations are listed below.

**Revised classification (Sample 89):**

- **3 "**Remind Staff…"
- **2** change equipment
- **2** create protocol for equipment use
- **2** review protocol for equipment use
- **1** create protocol for equipment maintenance
- **1** review equipment

**Revised Classification (Sample 98):**

- **4** "Remind Staff…"
- **1** Training viz new equipment
- **1** Equipment maintenance (management)
- **1** Create protocol for equipment use
- **1** Review procedure viz home patients' safety

First, we related the recommendations back to the initial 'cause' categories, with the scheme manager's assistance. Then we re-interpreted the suggested actions in the light of system safety design concepts, such as presented by van der Schaaf or Reason. In sample89, Thoughtlessness and Poor Equipment Design featured most often as contributing factors. This is mirrored in the action recommendations falling in the 'remind staff…', 'change equipment' and 'create protocol for equipment use'. Entries under 'remind staff…' typically are in the form of a reminder statement, drawing attention to problematic task or equipment characteristics, for instance "Remind all staff of the importance of careful, correct use of 3-way taps on central venous and arterial lines" (February 1989). 'Change equipment' is represented by recommendations such as "Particular sort of disposable ventilator tubing used on trial should no longer be used". 'Create protocol for equipment use' mentioned for instance "Consider use of small Graseby syringe drivers with smaller volumes of solution".

In the period of May to November 1998, a marked increase in reminder statements can be noted. Following inspection of recommendation data, the dissemination of reminder statements were noted to be the single most often suggested action. In the period August 1995 to August November 1998, 82 "Remind Staff…" statements out of a total number of 111 recommendations could be noted. The 29 other recommendations concerned procedure creation or change suggestions (e.g. "produce guidelines for care of arterial lines - particularly for femoral artery lines post coiling" ), or were equipment related (e.g. "Obtain spare helium cylinder for aortic pump to be kept in ICU").

Reminder statements as potential error prevention mechanism have come into disrepute in domains such as aviation (Reason, 1990). Rather than further burdening operators' and pilots' memory capacity, indirect safety methods such as reduced complexity, standardisation, proceduralisation, and work aids such as checklists have been introduced.

However, on closer inspection, the nature of the Edinburgh reminder statements proves to be interesting. Reminders seem to target either very common, but still error-prone, details of tasks and practices, or problem points that occur very infrequently, or otherwise problematic parts or uses of procedures.

Distinguishing thus between types of recommendation, a link to Rasmussen's SRK framework can be created. Skill (S) level performance concerns automatic behaviour routines, such as the common but still error prone details of tasks. Rule (R) level performance concerns the conscious but practiced following of procedures and protocols, and Knowledge-based (K) performance relates to potentially effortful, fully conscious problem solving and decision making. In aviation and process control, it has been realised that performance on the rule-based level is the least error-prone. Therefore, design methodologies such as Ecological Interface Design (EID, Rasmussen and Vicente, 1989) emphasise proceduralised tasks, and ensure that task features that relate to S or K level performance are assisted accordingly. For instance, K based performance can be supported by careful information design.

The Eindhoven classification/action matrix (Van der Schaaf, 1996) is also based on a SRK style cognitive classification of error. It details, as described above, that R-based error is best targeted with Training measures, K-based error with improvements in the Information & Communication domain, and S-based error with change in equipment. Thus, this is at odds with the Edinburgh results of the recommendation generation. Instead of reacting with reminder statements indiscriminately of cognitive performance level, these could be taken into account when suggesting remedial actions. The categorisation of error according to cognitive mechanisms will also further the understanding of performance problems.

We have illustrated how methods and insights from safety-critical domains other than medicine can be applied in a clinical setting. This concerns the use of incident schemes, as well as the application of accident causation models in the analysis of incidents, and in the generation of action recommendations. These recognise the importance of latent failure as part of the causal chain of an incident, which is reflected in the incident analysis process.

Incident investigation schemes often neglect formalised, in-depth analysis of single incidents in favour of a quantitative surface analysis. Also, the crucial role of detection factors and the need to support those is often underestimated. The Edinburgh incident scheme caters for those in the data collection process as well as in the generation of action recommendations. Thus, the analysis process and its results of the Edinburgh study showed how not only theoretical 'top-down' approaches can inform incident analysis, but also how practical incident avoidance can be supported by a 'bottom-up', detailed (albeit non-formalised) analysis process.

In the remainder of this chapter, we will demonstrate the benefit of the suggested cognitive error analysis approach for further in-depth analysis of incidents' human factors, as well as for the evaluation of safety recommendations.

## COGNITIVE ANALYSIS OF HUMAN ERROR IN ICU INCIDENT REPORTS

As described above, 'human error' is often assumed to be the prime 'cause' of incidents and accidents in clinical systems. This chapter investigates incidents in an adult intensive care unit (ICU). Our human error analysis approach stresses the importance of taking cognitive factors into account. The case study presents data drawn from an incident reporting scheme that has been running for over ten years. An in-depth analysis of example cases is carried out which considers human cognitive constraints during task performance. The genesis of erroneous action can thus be considered in relation to the underlying cognition. We embed work practice and the problems encountered in a holistic cognitive perspective that recognizes the importance of physical and visual input into human cognitive processing. Also, the cognitive analysis can provide pointers to constraints of humans' performance in context. It does not suffice to only consider the behavioural aspect of 'Human Error'. We argue that understanding of 'Human Error' is limited unless full credit is given to the impact that the characteristics of the human cognitive system has on task performance.

Instead of only investigating *what happened* in each incident, much can be gained from understanding the underlying *why* of the event. The cognitive 'mechanism of malfunction' (O'Hare et al., 1994) can be traced by using cognitive architectures. These provide the basis for cognitive models, which strive to represent some aspects of people's understanding, knowledge, or cognitive processing when performing some task. These models can, therefore, contribute to our understanding of the cognitive limitations interacting with task performance, for example the effects of cognitive load on performance (Barnard, 1993).

The distinction between varieties of human error according to their cognitive origin plays a significant role in accident analysis because they require different methods of error management and remediation (Maddox and Reason, 1996). In this thesis a cognitive architecture is used as a vehicle for expressing not only expert task performance but also the more realistic error-prone thought and action sequences processed by the human operator. By doing this, the error modelling capability implicit in the cognitive architecture is made the focus of inquiry into the underlying cognition of user performance. Such explicit modelling of erroneous performance can thus help to communicate user cognition analyses, and to ground incident analysis and subsequent action recommendation in a cognitive theoretical framework.

Interacting Cognitive Subsystems (ICS; Barnard, 1993) is used to illustrate the modelling of human error within a cognitive architecture. Errors resulting in medical incidents are rarely described in such detail, or even analysed in terms of underlying psychological factors (Busse and Johnson, 1999). Expressing human error within a cognitive model will allow us to investigate and reason about their underlying psychological causes. The model is thus used as a tool for reasoning about human error on a further, more detailed level.

As stated above, in the Edinburgh scheme, analysis of the underlying cognition of those proximal causal factors of the incident was not facilitated. In analysing the scheme, we took two data samples (Table 5-1 and Table 5-2), one sample covering the first categorisation interval, January and February 1989 (sample89), and the other covering a more recent interval from May to November 1998 (sample98). Both samples cover 25 incident reports.

In the classification of data into the contributing cause categories, combinations of factors are allowed, and are noted frequently in the data sample. For instance, in sample98, one of the predominant factors is 'Thoughtlessness' (10 occurrences). Looking at the combinations, however, shows that 'Thoughtlessness' is paired with other factors (such as Inexperience with Equipment) in 4 out of 10 incidents.

One common problem identified in the Edinburgh study is the endotracheal tube coming back out through the larynx. Over time, incident analysis resulted in a list of factors that led the tube to come out. Some of those now constitute 'contributory factors' in the incident data categorisation. One factor is that the tube was not properly secured. Another is that the patient was inadequately sedated, which led to the patient being able to pull out the tube. Judging the appropriate level of sedation is difficult, since oversedating the patient can lead to vital signs being disrupted, while under-sedation increases the possibility of the endotracheal tube being pulled out, which can itself be life threatening.

The tube can also be dislodged when turning the patient. Care must be taken to ensure the tube is long enough to be securely located in the larynx, while being short enough to prohibit it being caught in, for instance, the near-by intra-cranial pressure monitor during the turning manoeuvre. Often, several lines need to be monitored while turning the patient. Thus, dislodging the tube during turning can be exacerbated by not perceiving the relationship of the patient's position and the various tubes connecting measurement and drug administration equipment to the patient.

Thus, the scenario described above can be modelled in ICS as shown in Figure 5-4. The visual data, the patient and the position of the lines in relation to the surrounding environment, is received at the visual subsystem (1), sent to the object subsystem for the recovery of a structural description (2), and finally interpreted by the propositional subsystem (3). A loop is entered in order to maintain a stable cognition. The resulting interpretation on the propositional level of the success of the turning strategy influences the further view of the object. If the visual information perceived is inadequate, for instance, the position of the endotracheal tube is neglected, an inappropriate turning strategy will be chosen, and this inappropriate information is sent to limb subsystem (4) to initiate the motor movements.

This details a perspective on the cause of a dislodged endotracheal tube that emphasises the perception and interpretation of visual information. By modelling the underlying mechanisms of the causes of the incident during turning the patient within ICS, we can shed some light on the processes that are fundamental to the production of the incident as mediated by human error.

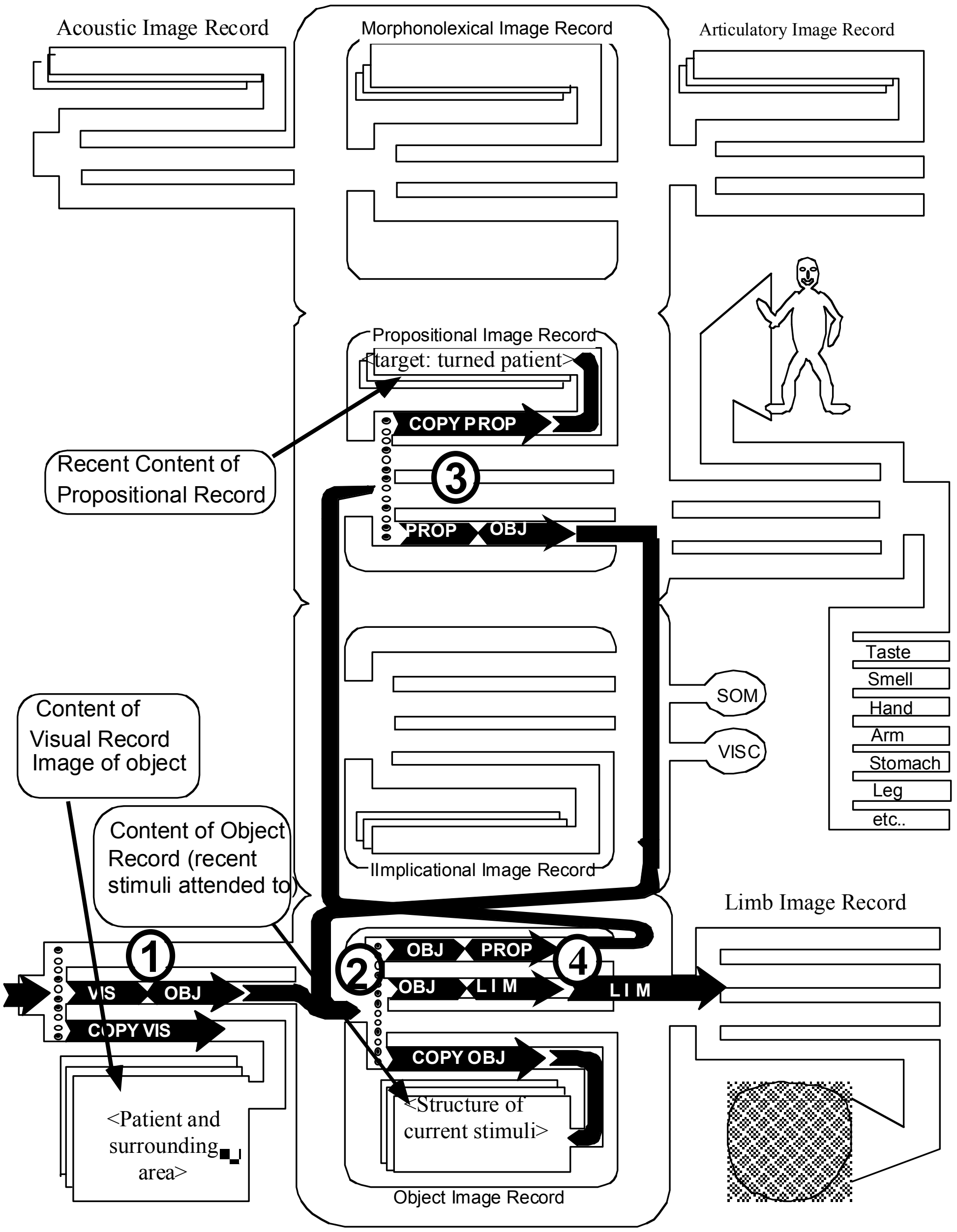


**Figure 5-4** - Lack of Visual Information

Contrasted to the categorisation scheme (table 1), this analysis highlights the relationship between two separate ‘contributory factors’: ‘Poor Visibility of Equipment’ and ‘Turning the Patient’. It is worth noting that these factors are not placed in multi-factor categorisation, but appear as single contributory factors to endotracheal tube incidents in the sample data.

## REASONING ABOUT ALTERNATIVE ERROR CAUSES

The third factor listed as a contributory factor for dislodged endotracheal tube incidents is listed as ‘Endotracheal Tube Not Properly Secured’. Endotracheal tubes are typically secured in the patient’s larynx by means of a cuff (situated on the lower end of the tube) being pumped with air until it sits firmly in the larynx. The failure to secure the tube suggests a ‘human error’ based on skill-level performance. Attentional resources are only minimally required and the action can be at least partly automated. Figure 5-5 presents an ICS model that details skill-based error leading to a dislodged endotracheal tube.

As can be seen in Figure 5-5, minimal resources are required to perform the task of placing the tube. The fixing of the tube in the larynx by pumping air into the cuff is carried out on a skill-based level, not requiring knowledge-based processing as would be provided by the implicational subsystem. Instead, body state information from the proprioceptive subsystem (1) is sufficient to enable the propositional subsystem (2) to interpret the state of the tube and to send motor movement information to the limb subsystem (3).

However, the task problem might not be based on failures on the skill-base level of performance, but alternatively on failures on a higher level of human cognition. Using ICS, alternative hypotheses as the underlying cognition of the dislodged endotracheal tube class of incidents can be investigated.

For instance, securing the endotracheal tube via the air cuff was taken as a skill-based task, with the procedural skill readily available. However, this is not always the case.

Not only does the level of staff experience play a significant role, but so do possible exceptions to the rule. For instance, in certain cases, the air cuff on the endotracheal tube is of a different make. Usually, when staff noted that the tube is not properly secured, the first measure is to re-secure the tube by pumping air into the cuff. With the different make of cuff, however, this has a counterproductive effect. In this specific case, pumping air into the cuff will force the cuff, and therefore the endotracheal tube, even further out of the larynx. This special case needs to be considered by staff when re-securing the tube.

Since it is an exception, a mistake on the rule-based level of performance is likely. It is one of the predispositions of the human cognitive system to opt for a well-known and practised rule even when faced with the exceptional situations. Reason (1990) calls this a misapplication of a ‘Strong but Wrong’ rule. This cognitive mechanisms comes into play especially in emergencies, since knowledge-level performance tends to be attenuated during high attentional requirements (Rasmussen et al., 1987). The human cognitive system then tends to fall back on well-practised strategies and procedures. Thus, in an environment such as an ICU, where emergencies are part of prototypical work situations, remembering the precise demands associated with rare, exceptional tasks is a resource intensive requirement. This can be modelled in ICS as shown in Figure 5-6.

Figure 5-6 demonstrates that the processing of an exceptional case involves more cognitive complexity, and requires increased cognitive resources. The body state information (1), being passed on to the propositional subsystem (via a PIP loop passed the implicational subsystem) (2), now only presents a small aspect of the cognitive demands posed upon the human. The decision to be taken as to what make of tube is involved, and how the body state information can be interpreted needs to draw on implicational input (3). This hypothesis, again, has different implications for potential system redesign. Thus, this kind of analysis, taking the impact of cognitive constraints into account, can aid precise and structured reasoning about the incident occurrence.

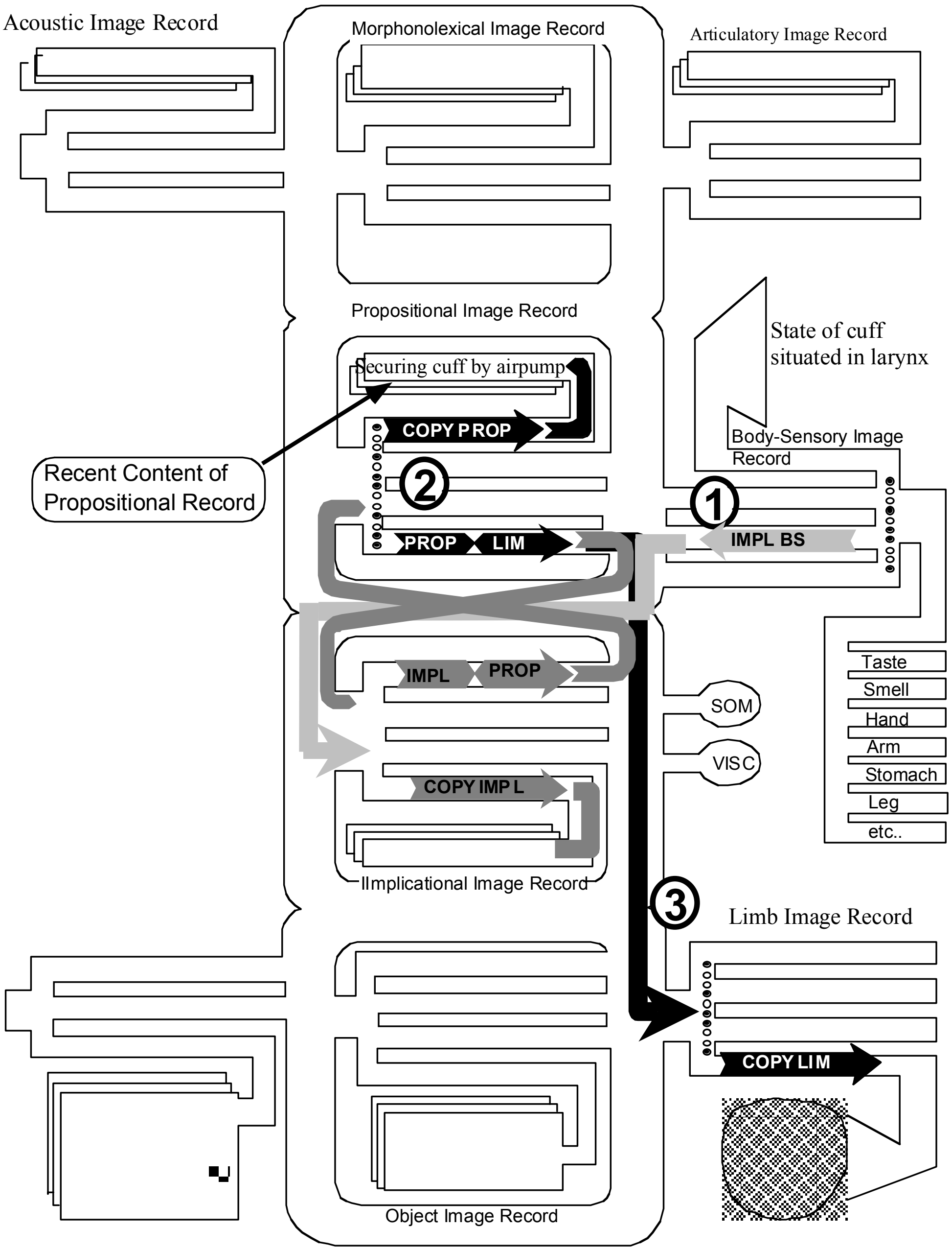


**Figure 5-5** - Skill-based Error

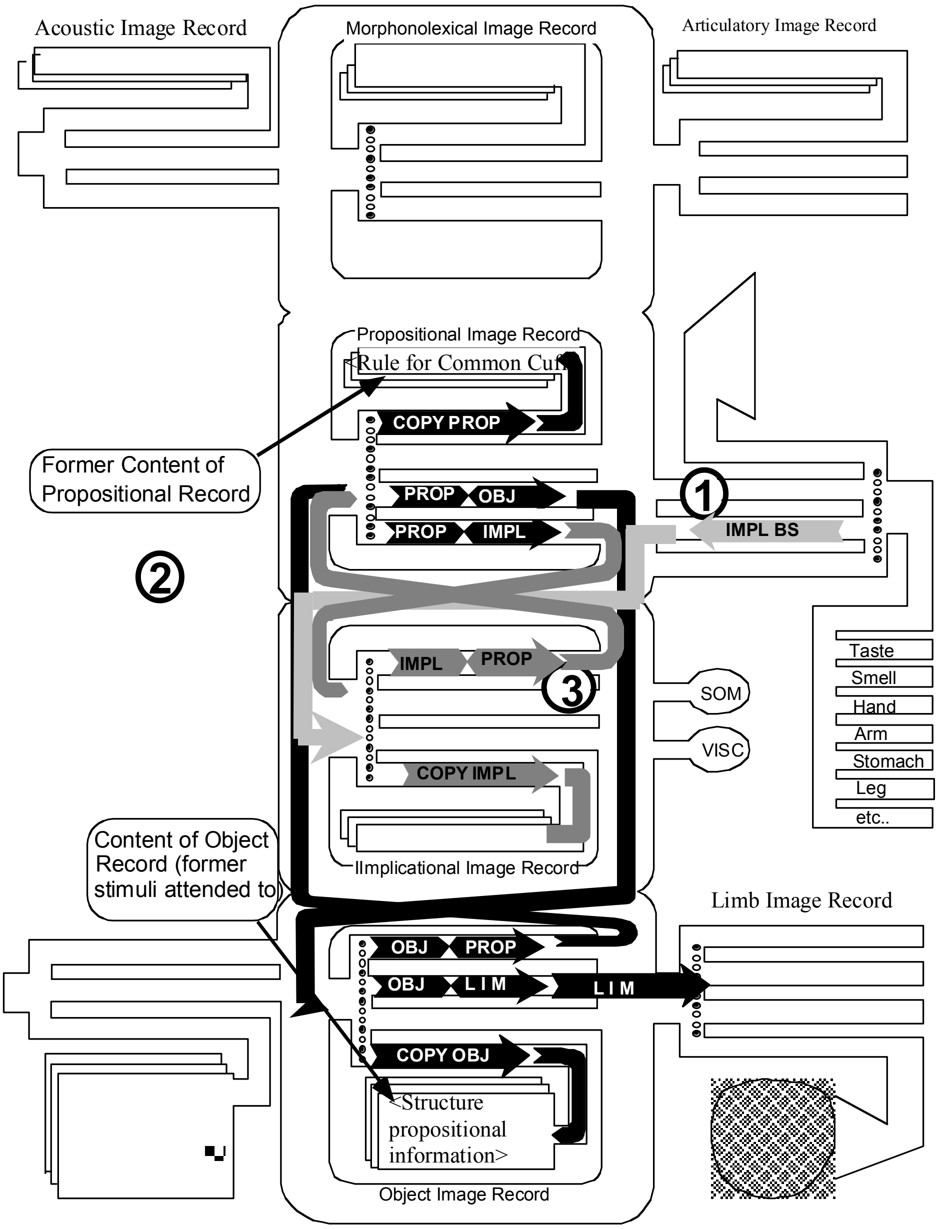


**Figure 5-6** - Higher Level Cognition

The examples elaborated above show clearly how different cognitive mechanisms might be implicated in the same overt task performance problem. This multi-way relationship between cause and error might go undetected if systematic error modelling within a cognitive architecture does not take place, this helps analysts to explicitly consider the detailed causes of task performance problems.

Human Error leading to a dislodged endotracheal tube might be grounded in varying cognitive processes, and not stem from one kind of cognitive mechanism alone. Unless these two different causes are considered, an analysis might misdiagnose an important problem in the task. Using a cognitive architecture to reason about the potential underlying cognitive error production processes allows work system and equipment designers to investigate the detected task problem in a systematic way.

Often, error categorization systems do not proceed much beyond the "what" phase within the above mentioned analysis model. A structured, formalised analysis framework also helps prevent 'hindsight' from biasing error analysis. Attribution of error is a social and psychological judgement process rather than a matter of objective fact. Hindsight view is fundamentally flawed because it does not reflect the situation confronting the practitioners at the scene. Thus, rather than being a causal category, human error should be seen as representing a symptom, and a starting point for investigation (Woods et al., 1994).

Often, especially in the case of clinical staff trained for taking on responsibility for their actions (Berwick, 1998), error analysis can be tainted by human cognitive mechanisms such as the 'Fundamental Attribution Error'. According to this, humans are significantly more likely to attribute error occurrence to situational aspects when the error was 'committed' by themselves. However, when looking for reasons why others were involved in 'error', one is most likely to blame the person rather than the situational aspects. This is another reason why an analysis framework is needed that aids objectivity in interpreting the actions of others.

## ACTION RECOMMENDATION

In order to arrive at sound and relevant action recommendations, a systematic and structured way of bridging the result of the analysis process to remedial measures is needed. The documentation of this process also plays an important role, for instance, to allow monitoring of the effect of the measure.

In industry domains such as aviation and process control, cognitive analysis of error occurrences is often used to point towards remedial actions. For instance, Reece et al. (Reece and Hill, 1995) investigated human error in radiation exposure events. The proximal cause of the incident was situated in the task sequence and, additionally, cognitive failure analysis was carried out. Then the relationship between them was analysed, and thus it could be identified: where training might be most effective; where equipment interface enhancements may be most appropriate; and where job aids (e.g. checklists) might help performance.

This shows how error categorisation, when done according to cognitive level of performance and latent factors, can provide the basis for structured remedial recommendations, rooted in theory. Without error categories being based on sound psychological theory, systematic and relevant action recommendation generation is often not possible.

In the Edinburgh study, the incident data was categorised and summarised by the scheme manager. Action recommendations were arrived at in an iterative process, whereby the scheme manager suggested remedial actions and presented those along with the summary data to the senior nurse of the ICU. The data was discussed and the rationale for the action recommendations reviewed. This led to a final version of suggested actions for each incident analysis period. Table 5-4 and Table 5-5 show a categorisation of the suggested actions for our two samples (sample89 and sample98).

We categorised the suggested actions in the light of system safety design concepts, such as those presented by Reason (1997). Entries under ‘remind staff’ typically are in the form of a reminder statement, drawing attention to problematic task or equipment characteristics, for instance “Remind all staff of the importance of careful, correct use of 3-way taps on central venous and arterial lines” (February 1989). ‘Change equipment’ is represented by recommendations such as “Particular sort of disposable ventilator tubing used on trial should no longer be used”. ‘Create protocol for equipment use’ mentioned for instance “Consider use of small Graseby syringe drivers with smaller volumes of solution”.

In the period of May to November 1998, a marked increase in reminder statements can be noted. Following inspection of recommendation data, the dissemination of reminder statements was noted to be the single most often suggested action. In the period August 1995 to November1998, there were 82 “Remind Staff…” statements out of a total number of 111 recommendations.

Instead of reacting with reminder statements indiscriminately of cognitive performance level, these could be taken into account when suggesting remedial actions. The categorisation of error according to cognitive mechanisms can also further the understanding of performance problems.

## Using ICS to detail risk situations and arrive at action recommendations

The Edinburgh Study also notes incident detection factors, which is often neglected in other reporting systems. It has been suggested that provisions for incident detection and recovery provide more effective safety measures than an approach solely targeting accident prevention or avoidance (Rasmussen and Vicente, 1989; Reason, 1990).

However, even if detection factors are noted, they are typically not being analysed in depth. The analysis should include system factors as well as cognitive aspects of the task and work environment. Unless we can monitor those adverse situations which are, and those which are not, reported, we can have little confidence in the accuracy of the system.

The detection factor taxonomy evolved alongside the iterative development of the contributory factors taxonomy. The factors added over time are ‘Having Lines or Three Way Tap Visible’, and ‘Handover Check’. The iteration over the collected incident data thus clarified the importance of handover checks to make up for contributing causes ‘Failure to Check Equipment’ and ‘Failure to Perform Hourly Check’. Without iterative revision and coding of the data categorisation, these factors might have gone neglected. A formalised, cognitive analysis of the incidents can aid the recognition of detection factors and the generation of suggested actions.

For instance, one recurring incident concerned the use of three-way taps. These are used to feed, for instance, two different drugs to the patient via one intravenous line. On changing one of the drug syringe drivers, the corresponding line connecting to the three-way tap is turned off. After the drug change, staff must remember to return the tap settings back to allow both drugs to run.

Currently, this procedure is being supported by the memory aid **T.A.P.** – an acronym for **T**ap **A**ligned **P**roperly. Thus, this problem initially prompted ‘Reminder’ statements in the recommended actions summaries. It was then modified, through an iterative analysis process, to recommend keeping the three-way tap visible, to facilitate staff recognition that the tap was still left in the turned off position. The suggestion to counteract three-way taps being left in the incorrect setting evolved over time. An analysis framework, such as ICS, can point towards weak points in human cognitive task performance such as this. The problem of three-way tap (in)visibility is detailed in ICS in Figure 5-7.

The ICS model illustrates how the process of changing the drug takes over the implicational and propositional subsystem processes. Higher-level, semantic cognition is involved, since active problem-solving is required for the task. The procedure to be followed prescribes the returning of the tap to the correct setting. However, relying solely on the semantic subsystems in remembering all steps in the procedure is insufficient, especially when inexperienced staff are involved (one of the main contributing factors of incidents, see Table 5-2).

By providing prompts for the next step in the procedure, task performance can be pulled onto the rule-based level of performance, rather than requiring knowledge-based problem-solving (Rasmussen and Vicente, 1989). Thus, cueing information can be fed into the system via the peripheral subsystems. This principle is often realised, for instance, by providing alarms on equipment. In ICS terms, the peripheral cueing information is in this case being fed in via the acoustic subsystem, for instance by using alarms. Alternatively, and often overlooked, information can also be provided via the visual subsystem.

In the task at hand, the visual subsystem has been focused on observing the status of the syringe driver task. After completion of the task, the resources of the visual subsystem are freed to take in additional status information. The visibility of the three-way tap comes to take a crucial cueing role. After processing the goal hierarchy for selecting a link, the cognitive system can shift its focus back onto the visible three-way tap. Only if the procedural step is remembered, will the propositional subsystem signal anticipation of the three-way tap via internal input into the object subsystem. Thus, the visibility of the tap is crucial to change the propositional representation.

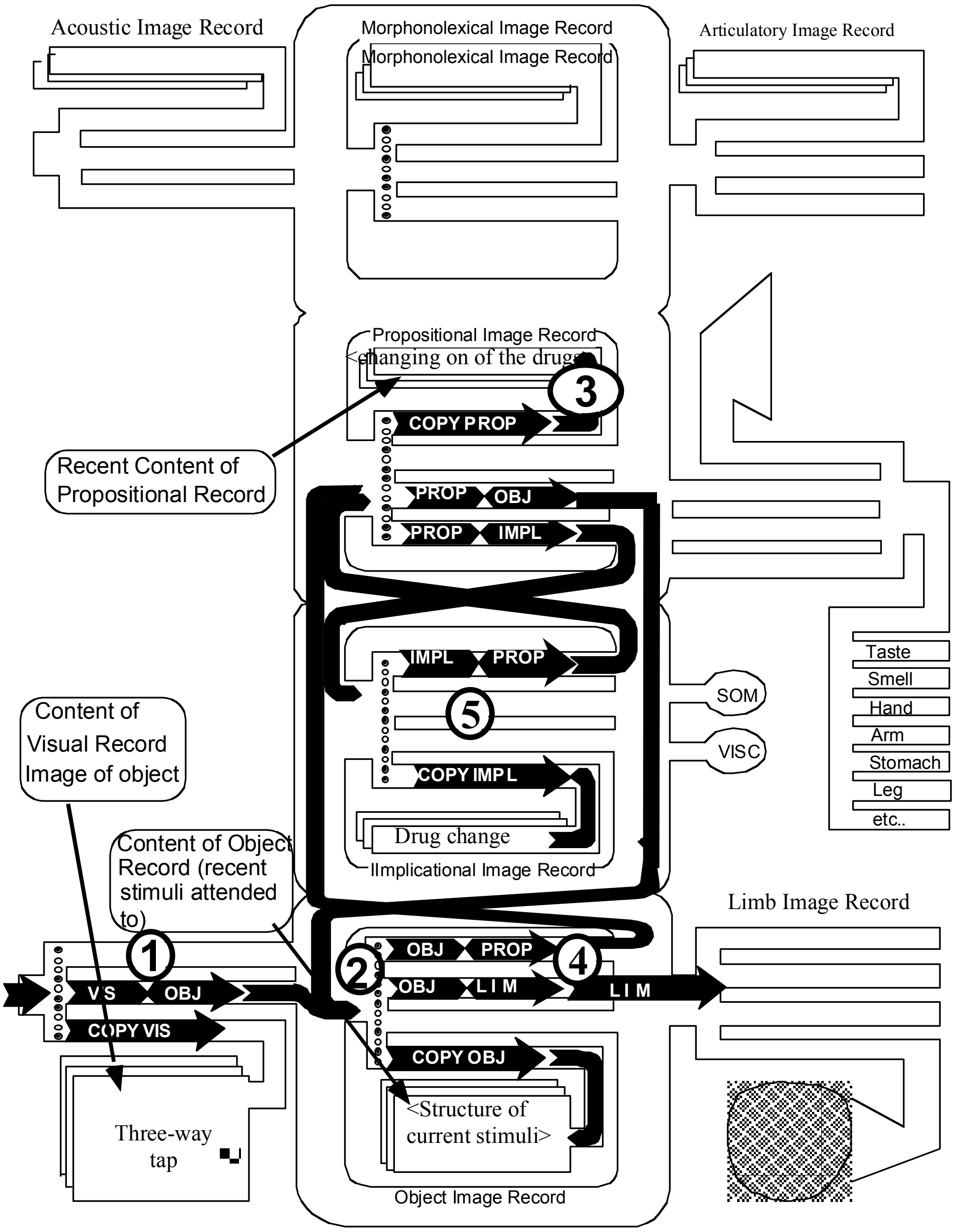


**Figure 5-7 -** Three-way tap visibility as Task Cue

Using ICS to model the underlying cognition of the error provides means of investigating the behaviour trace leading to an incident. Expressing the rationale for different interpretations within a cognitive framework facilitates more precise communication and more detailed analysis. In that way, not only *what* failed in incidents, but also *how* and *why* it failed is examined in an investigation of human error.

## CONCLUSIONS

There is a need in medicine to recognise the inevitability of error and adverse events (Leape, 1994). Safety culture that takes this into account in clinical system design is still lacking (Leape et al., 1998). There have been some notable exceptions in the recent past where incident reporting schemes were implemented and the identified incidents analysed, such as Runciman et al. (Runciman et al., 1993) and (Battles et al., 1998). However, these focus on the data collection process and somewhat neglect the in-depth analysis, with which valuable insight into incident causation and remedy can be achieved.

Incident investigation schemes often neglect formalised, in-depth analysis of single incidents in favour of a quantitative surface analysis. Also, the crucial role of detection factors is often underestimated. The Edinburgh incident scheme represents those factors in the data collection process, as well as in the generation of action recommendations. However, this process has not been formalised, and is not based on insights gained from cognitive theory.

We have illustrated in this chapter how cognitive modelling can be applied to focus more narrowly on the psychological precursors of the human actions leading to incidents, by means of retrospective analysis of incidents that had occurred in an adult intensive care unit during the last decade. In the following chapter, we address the weaknesses identified here by implementing and evaluating an incident reporting scheme from scratch in a Neonatal Intensive Care Unit (NICU), to which we will further demonstrated the benefit of the suggested cognitive error analysis approach, as situated in ongoing clinical safety management.

# CHAPTER 6 IMPLEMENTATION OF CRITICAL INCIDENT REPORTING IN A NEONATAL INTENSIVE CARE UNIT

## INTRODUCTION

The occurrence of medical adverse events is a growing cause for concern worldwide. Critical incident reporting schemes have recently been suggested as an effective means to tackle the problem of medical adverse events. There are few comprehensive frameworks that accommodate the specific requirements of local settings as well as generic issues in incident reporting. The local setting radically influences a scheme's successful implementation and maintenance. Issues that impact the overall success of incident reporting schemes concern the format of data collection, and especially, a meaningful data analysis. This chapter reports on the introduction of a critical incident reporting scheme to a neonatal intensive care unit (NICU). Issues concerning the implementation and maintenance of the reporting scheme are discussed. Incident analysis is described in terms of the process and the results of incident categorization. The implications of such a viewpoint are considered.

Critical incident reporting schemes have been cited as a major safety tool to combat human error and adverse events in medicine. Typically, a distinction can be drawn between standardized, national schemes of broad scope but less depth, and smaller, local schemes. Local schemes permit close scrutiny of in situ adverse events. Their implementation can also be fine-tuned to the local culture and conditions, a prerequisite for successful delivery of the scheme. This is at the expense of the benefits of statistical evaluation of the collected data that a more extensive, standardized data set offers. Both types of scheme operate by presenting employees with a data collection form, which prompts for a number of characteristics of the problem description. The nature of the form varies, but there are some fundamentals common to most schemes: general circumstances of the adverse events need first to be established (e.g. at what time did the incident occur?), as well as general facts about the reporters themselves (e.g. how experienced they are). Typically, this is followed by asking for a narrative description of the incident, and includes questions about the presumed ‘causes’ and contributing factors, and how the incident was detected. The question on detection factors is not typically included, but is clearly valuable for future incident avoidance. This data can then be used to instigate further, more in-depth investigation of the incident. It can also be categorized and archived for statistical purposes, if the data set permits. Major national safety schemes are NASA’s Aviation Safety Reporting System (ASRS), and its UK equivalent CHIRP (Confidential Human Factors Incident Reporting Programme). An example for a local incident reporting scheme that has been implemented in healthcare and maintained for over 10 years is described in (Busse and Wright, 2000).

This chapter offers a report on the prerequisites to successfully running an incident reporting scheme. The implementation that is described here attempted to localize an incident reporting scheme by taking contextual factors such as safety culture into account. It had staff actively participate in the conception of the scheme and the design of the reporting form. This is argued to be crucial in achieving long-term staff participation, and is argued to optimize the scheme’s efficiency. The scheme that is described here also attempted to incorporate standardization issues that are posed by schemes implemented on a grander, e.g. national, scale. The experiences with this standardization approach will be discussed. This chapter thus presents a matrix of issues that still need to be addressed in any safety-critical domain that employs critical incident reporting schemes as part of ongoing safety management.

## INCIDENT REPORTING

In the UK, a Department of Health report (2000) revealed that as many as 850,000 adverse incidents are happening in UK hospitals each year. This, in terms of litigation and the extra care needed by victims, added up to a £2bn bill. (BBC News, 2001). There is an urgent need to make patient safety one of the highest priorities. The BBC also cites fears that the medical community is "complacent" about the toll of accidents, and notes that to date the National Health Service "did not even collect figures on the number of medical accidents".

Until recently, evidence of medical incidents (or near-misses) was mostly anecdotal. In the case their existence was acknowledged, the data typically did not leave a hospital's boundaries. This lack of distribution of incident data can lead to the replication of similar, preventable incidents across hospitals. This not only concerns, e.g. faulty or badly designed equipment which might lead to deadly consequences, it also concerns badly designed work procedures that might be in place in hospitals across the country, the safety threat of which might only be recognized locally and sporadically. Similarly, it concerns drugs that might have similar sounding names, but that have very different effects on a patient's condition. In order to prevent incidents needlessly repeating themselves, incident data must be recorded, analyzed, and then made available for distribution. In industries such as process control for chemical plants and power stations, incident reporting schemes have often been used as 'early warning schemes'. This has yet to translate fully to the medical domain.

Obstacles to the implementation of safety measures in medicine as established in aviation also lie in the differences between the two work domains. Accidents in aviation are comparatively infrequent but very visible, receiving high media attention, and often involve massive loss of life (Helmreich, 2000). In contrast, accidents in medicine typically only involve not more than one patient (or member of staff), with less or no media coverage, with news about adverse events often not leaving a hospital's boundaries.

Also, the type of standardized, unified safety management measures implemented in aviation often cannot translate to the less standardized, less regulated, and thus less clear-cut work environment that medicine presents. Doctors and clinical staff often learn 'on the job' to a large extent, in contrast to aviation or nuclear power plant operation (a domain that also has a long-term history of use of safety management measures such as incident reporting). Medicine, described recently again as a "humbling art and a complex team activity" (Berger, 2001), deals with humans, whose conditions and responses are typically less predictable than an aircrafts' (Helmreich, 2000).

Errors may be particularly difficult to recognize in health care because variations in an individual's response to treatment is expected. In addition, medical professionals may not recognize that a particular product or procedure may have contributed to or caused the problem because the patient is already ill, the product is not expected to work perfectly at all times, or the event appears unrelated to the product or procedure (QuIC, 2000).

Studies suggest, for instance, that uncertainty about the most effective diagnostic and therapeutic approaches is pervasive (Macias-Chapula, 1997). One area in medicine that resembles more closely the more proceduralized and well-defined and thus more predictable task space of aviation is often cited to be anaesthesia. Correspondingly, safety measures such as in-depth error analysis (e.g. Gaba et al., 1987) and incident reporting (e.g. Runciman et al., 1993) have been applied to anaesthesia prior to a more wide-spread adoption in other areas in medicine. The work of Runciman and his colleagues forms the basis of the incident analysis method described in this chapter as applied to neonatal intensive care.

The tendency to lay blame on staff involved in the incident rather than e.g. error-prone equipment design (Busse and Johnson, 1999) further prohibits the use of incident reporting as a constructive safety measure in medicine. Thus, incidents might not necessarily be perceived by staff to be ‘accidents waiting to happen’ (Reason, 1990). Incidents might be seen as mere task characteristics, with mistakes seen as human fallibility, and with incident detection and recovery taken for granted. Incident reporters might also not be aware of ‘upstream precursors’ to the incident, such as underlying system faults (‘upstream’ since in systemic incident analysis, multiple layers of incident causation are assumed, with systems factors being the lowest layer).

Staff might not acknowledge the significance of local workplace factors. For instance, if staff have been accustomed to working with substandard equipment, they may not report this as a contributing factor since they see it as the ‘normal’ work context; if they habitually perform a task that should have been supervised but was not, they may not recognize the lack of supervision as a problem (Reason, 1997). This tendency, and the associated ‘work-around’ culture in medicine, emphasizes the need for explicit scrutiny of potential upstream precursors (i.e. system factors) in incident reporting and analysis.

System factors might be organizational in nature, such as the notorious under-staffing in healthcare, with its associated stress on hospital staff, and the known increase in error-prone behaviour of individuals under stress. There might also be a lack of end-user consideration when choosing and procuring equipment, and also, for instance, neglect of training requirements on part of the organization purchasing the equipment (Jeffcott and Johnson, 2001). Accidents can often be traced back to equipment design that induces ‘human error’. It is rarely the case that accidents are caused by one single point of failure, and failure histories can usually be traced back through several layers of causation. While it is important for hospitals to prioritize error-tolerant and usable design and functionality in equipment procurement, there is often no empirical or analytical data available evaluating the system at hand in those terms. Thus, given difficult to use equipment design that seems obvious on scrutiny, accidents often seem preventable in retrospect. This seems particularly unforgivable in safety-critical work environments such as surgery or intensive care units.

Introducing incident reporting into hospital wards is the first step towards recording information on incidents' nature and frequency. Summary data can then be used for trend analysis to identify systematic sources of error, and to prompt more in-depth analysis of potential causes. However, in order to base valid and relevant conclusions on this frequency counts, the classification of incidents clearly needs to be meaningful.

For instance, in current incident studies, most of the incidents' precursors are perceived to be 'human error' (e.g. Runciman et al., 1993). There are doubts, however, how meaningful this category, and its implications, really are. Often, the fact that an incident does not fit a category such as 'equipment failure' alone is seen as justifying labelling the incident as 'human error' (see also e.g. Rasmussen et al., 1981). Such categorization might provide an initial filtering of immediately attributable equipment faults, but does not tell us much about how to prevent future instances of such 'human error'.

Furthermore, as soon as poor equipment design is considered as an instance of equipment failure, there is no telling as to what constitutes error-inducing design (such as similarly named drug containers) and what constitutes human error (mistaking the drug containers). Thus, the artificial distinction between equipment failure and human error (more meaningfully described by Rasmussen as 'Human-Machine Mismatch') is cemented and perpetuated by such a classification.

## Critical Incident Reporting Form

**The Incident**

Description of what happened:

(please also answer the questions overleaf in case of **Drug Error**)

What factors contributed to the incident?

What factors minimised the incident?

**The Circumstances**

Date: Time: Place:

What procedure was being carried out?

What monitoring was being used?

Did the equipment alarm?

If equipment failure give details of equipment:

**Personnel**

Grade of relevant responsible staff: Grade of staff discovering the incident:

Were you involved in the incident?

**Outcome**

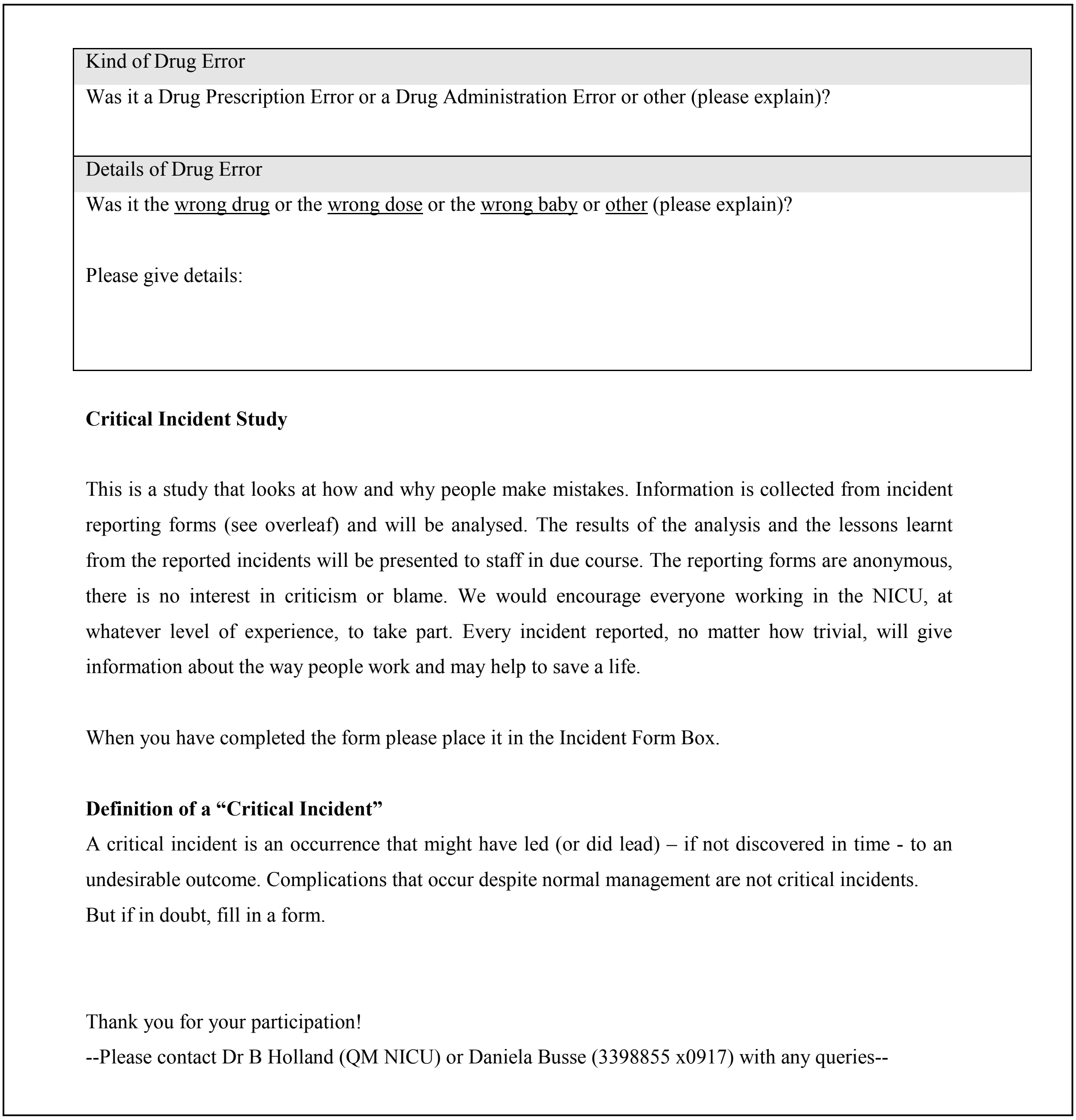
What happened to the patient?

What is the severity of potential outcome for the patient?

**Prevention**

How might such incidents be avoided in the future?

**Figure 6-1 -** NICU Incident Reporting Form (I)



| |
|---|
| Kind of Drug Error |
| Was it a Drug Prescription Error or a Drug Administration Error or other (please explain)? |
| Details of Drug Error |
| Was it the wrong drug or the wrong dose or the wrong baby or other (please explain)?<br>Please give details: |

**Critical Incident Study**

This is a study that looks at how and why people make mistakes. Information is collected from incident reporting forms (see overleaf) and will be analysed. The results of the analysis and the lessons learnt from the reported incidents will be presented to staff in due course. The reporting forms are anonymous, there is no interest in criticism or blame. We would encourage everyone working in the NICU, at whatever level of experience, to take part. Every incident reported, no matter how trivial, will give information about the way people work and may help to save a life.

When you have completed the form please place it in the Incident Form Box.

**Definition of a "Critical Incident"**
A critical incident is an occurrence that might have led (or did lead) – if not discovered in time - to an undesirable outcome. Complications that occur despite normal management are not critical incidents.
But if in doubt, fill in a form.

Thank you for your participation!
--Please contact Dr B Holland (QM NICU) or Daniela Busse (3398855 x0917) with any queries--

**Figure 6-2** - NICU Incident Reporting Form (II)

This clearly poses a very real problem for the validity of incident data and its analysis. However, this subjectivity in the incident classification process, and the resulting spurious precision of trend analyses based on the data, is not sufficiently recognized as would seem necessary for any wider distribution of incident data beyond the local setting.

As we know, an example of notable attempts at creating grounded and relevant categorizations schemes is reported in Runciman et al. (1993) who studied anaesthesia incidents in Australian hospitals as part of the Australian Incident Monitoring Study (AIMS). This study was subsequently extended to also investigate intensive care unit (ICU) incidents. The AIMS categorization scheme presents an integrated summary of previous categorization schemes, and has had substantial impact on future ones. For the study reported in this chapter, the AIMS-ICU categorization scheme was utilized to analyze incident data that had been collected in a neonatal intensive care unit. In the following sections, the outcome of this process is reported

## THE NICU INCIDENT REPORTING SCHEME

In the Neonatal Intensive Care Unit (NICU) in which the study took place, current safety management included informal checks, communication and consultation with fellow members of staff, morbidity and mortality meetings, and an adverse events reporting scheme, which addresses incidents that in fact resulted in harm to the patient or to staff and that legally require investigation. Near-miss adverse events that do not require legal investigation, but that could also lead to harm to the patient or staff, were dealt with on a local and immediate basis. They were not documented or kept track of, and distribution of known sources of error in the system was at best infrequent. There was demand to complement the existing safety management measures with a critical incident reporting scheme.

## Critical Incident Definition

A ‘critical incident’ was here defined as follows: “A critical incident is an occurrence that might have led (or did lead) – if not discovered in time - to an undesirable outcome. Complications that occur despite normal management are not critical incidents.” Staff that participated in the study were also asked to fill in an incident reporting form “if in doubt”. This reflected the intention to collect rich, qualitative data, rather than data that would be fit for exact statistical analysis.

## Set Up

A ‘critical incident’ thus includes near-misses as well as actual adverse events. Reporting schemes involve staff reporting critical incidents using the provided reporting forms on a voluntary and anonymous basis. The incident reports are regularly analyzed and categorized. The main aim of the analysis is to identify factors contributing to the causation of incidents that may be rectified. Accordingly, similar incidents are hoped to be avoided in future.

Van der Schaaf (Van der Schaaf et al., 1991) suggested a model of incident data collection and analysis which was considered for the NICU scheme’s implementation. Van der Schaaf (op. cit.) provided a convenient summary of the steps that need to be present in a successful event[20] analysis system, as summarized in Figure 6-3. They represent the inputs to the system (1-3), the way these are processed (4-6) and the output and monitoring which allows the recommendations (7).

---

[20] Hale et al. (1997) define an event to be a deviation in an activity or technology which leads towards unwanted, negative consequences.

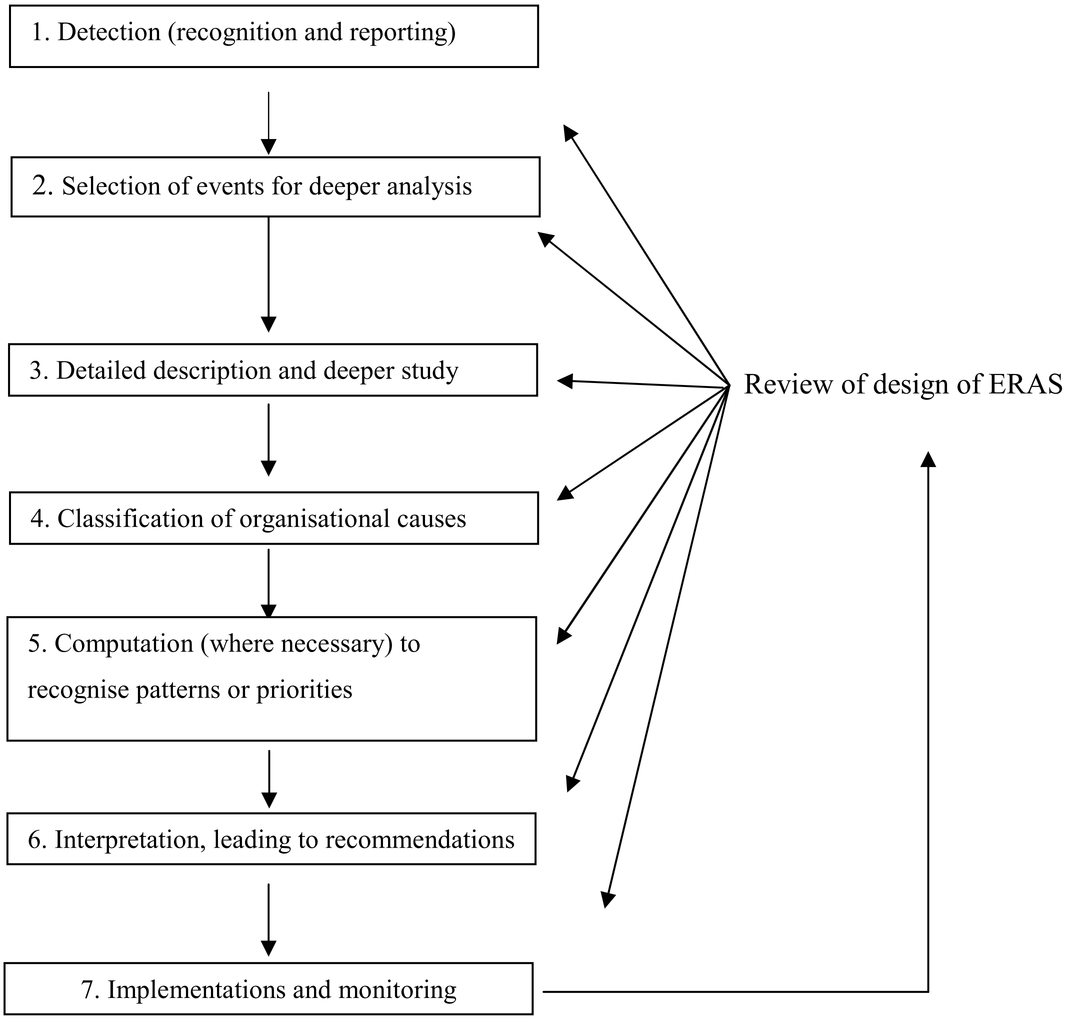


**Figure 6-3** - The Seven Modules of an ERAS - Event Reporting and Analysis System (Van der Schaaf et al., 1991)

In the NICU scheme, however, it was proposed to use the AIMS causal classification taxonomy on every incident that was detected and reported. Only subsequent meta-analysis of the classified incidents, and any specific observations that were noted during this first classification system would then lead to further in-depth analysis of the incident. This way it was to be evaluated whether the benefits of both approaches

– the quick and efficient surface categorisation approach to classify *all* data and to prepare the data for meta-analysis, as well as the in-depth cognitive error analysis applied to individual incidents – could be captured in one methodology. Therefore, the scheme's set up was suggested as illustrated in Figure 6-4 below.

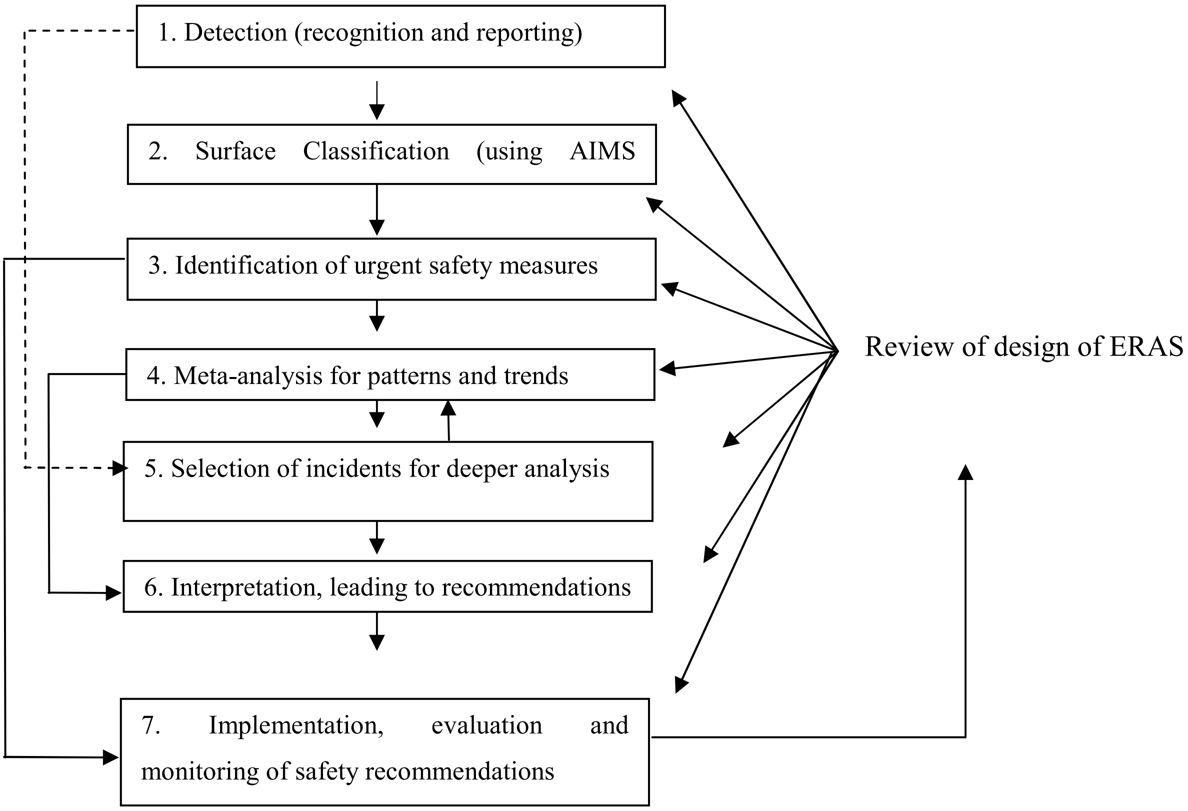


**Figure 6-4 -** NICU incident analysis process

This approach aims at maximising the benefit from a short-hand surface classification system such as the one used by the AIMS project (as described earlier), as well from a further in-depth analysis that reaches beyond surface classification and informs interpretation and safety recommendations.

This is the approach taken in the here described case study of incident reporting in Neonatal Intensive Care. The aim was to validate the suggested incident reporting model, including the cognitive error analysis approach. The implementation and evaluation of the approach is described in the remainder of this chapter (with respect to previously identified weaknesses in incident reporting), as is the in-depth analysis of sample incidents that occurred in the time of the study, using the cognitive error analysis approach.

## The Incident Reporting Form

The incident form was developed iteratively, and evaluated by means of a questionnaire survey of the unit staff (Busse, 2000). The current form covers the following questions: the first section asked for a "description of what happened"; 'Drug Confusion Error' is treated as a category distinct from other types of critical incident on the form, due to its known frequency (Bogner, 1994). This separate treatment allowed for more specific data to be gathered on drug errors. Other questions related to what factors contributed to the incident, and which factors minimized it. The next section covers details on the circumstances: which procedure was being carried out, which monitoring was being used, and which equipment failed (if any). A question on the presence of alarms was added to the form, since it was felt that incidents discovered through alarm sounding fell in a sufficiently distinct category of incident circumstances. This was then validated by the data that was collected. One section of questions touched on the personnel that were involved in the incident and its detection. The data was collected on an anonymous basis, so the reporting staff was not asked to provide contact details. However, experience levels and job titles were covered in the personnel section. Another section noted the estimated and actual outcome of the incident to the patient (or in terms of other costs). The final section provided an open-ended question regarding suggestions for improvements by the reporting staff - future prevention of similar incident being the primary goal of the incident reporting scheme.

## 3.4. Incident Analysis using AIMS

The AIMS-ICU analysis scheme was used. AIMS used a reporting form that consisted primarily of given categories which were to be ticked off by the reporting staff (see Figure 6-5). This way of recording staff reports could potentially lead to decreased analysis time (since the reporters essentially did the categorization themselves). It could also be argued that this decreased the degree of indirection in the analysis process – categorization based on subjective second-hand interpretation of gathered data could be replaced by the reporter's own interpretation of the actual events. In the NICU study, the form that was used was specifically developed to address the local needs. However, the collected data was subsequently analyzed by assigning it to categories as listed in the AIMS. The breakdown of the results is shown in the following section.

## RESULTS

As can be seen in Figure 6-5, 14 causes of incidents were classified as system factors, whereas only 11 where classified has human factors. This is in contrast to findings in comparable studies, where up to 80% of causal factors were classified as human error. The most commonly attributed subgroup of causal factors was 'equipment' with 7 occurrences, followed by 'physical environment/infrastructure' (4) and 'knowledge-based errors' (4). Furthermore, the categories 'rule-based errors', 'skill-based errors', 'technical errors', and 'work practices' were also all represented in the results. All causal factors could be categorized, and 'other system factors' and 'other human factors' were not assigned in this study.

SYSTEM-BASED FACTORS

| **Physical environment / infrastructure** | **#** |
|---|---|
| Lack of space / room | |
| Lack of facility | |
| Excessive noise | |
| High unit activity level | 4 |
| Staff mealtime | |
| Handover / ward round | |
| Lack of support staff | |
| **Equipment (including monitors)** | |
| Unavailable equipment | |
| Inadequate equipment | |
| Poor design | 4 |
| Poor maintenance | |
| Equipment failure | 3 |
| Inadequate inservice | |
| **Work Practices / Policies / Protocols** | |
| Communication problem | 1 |
| Inadequate assistance | |
| Lack of supervision | |
| Inadequate training | |
| Inadequate protocol | |
| Insufficient staff | 1 |
| Unable to contact staff | |
| Inapprop. staff / patient allocation | 1 |

HUMAN FACTORS

| **Knowledge-based error** | **#** |
|---|---|
| Lack or faulty knowledge | 1 |
| Error of: | |
| Judgement | 1 |
| Problem recognition / anticipation | |
| Diagnosis | |
| Treatment decision | |
| Use of investigation procedures | |
| Timing of investigation procedures | |
| Omitting intended treatment | 1 |
| Incorrect charting | 1 |
| Incorrect prescription | |
| Incorrect interpretation of information | |
| Information not sought | |
| Information not available | |
| **Rule-based error** | |
| Patient assessment inadequate | |
| Patient preparation inadequate | |
| Failure to check equipment | |
| Misuse of equipment | |
| Unfamiliar equipment | |
| Unfamiliar environment | |
| Unfamiliar patient | |
| Failure to follow protocol | 2 |
| Labelling error | |
| Calculation error | |
| **Skill-based error** | |
| Distraction / inattention | 1 |
| Fatigue | |
| Haste | |
| Stress | 1 |
| **Technical error** | |
| Fault of technique | |
| Inexperience | 1 |
| Uncooperative patient | 1 |
| Difficult patient body habitus | 1 |
| Patient physiological factors | |

**Figure 6-5** - Incident Categorization and Analysis with the AIMS Classification

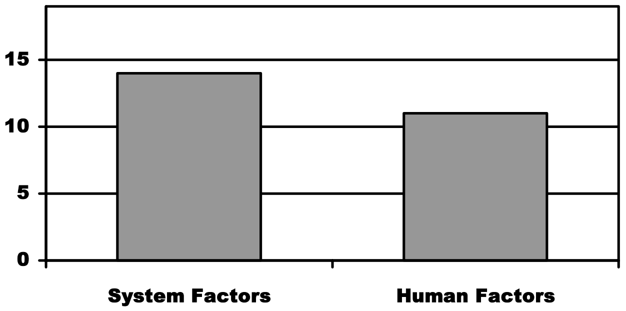


**Figure 6-6 -** Distribution of System vs. Human Factors contributing to NICU incidents

However, it was found that for most incidents, multiple categorizations were necessary. There were mostly several causal factors per incident, for instance the categorizations listed above show at least two categorizations of causal factors per incident. The combination of factors proved to provide a more meaningful picture of the incident's causation and its potential future prevention, than single categorizations. This confirms previous findings (Busse and Wright, 2000). The most striking finding in this study was arguably the difficulty of arriving at a meaningful classification of incidents. The nature of a meaningful classification has not yet been sufficiently discussed, let alone been operationalized, in the current discourse on incident reporting. Early work in the process control domain has covered substantial ground in delineating a meaningful analysis of Human Error (Rasmussen, 1982), but such work has still be addressed in incident categorization and analysis. The suggested cognitive error analysis approach aims at presenting a step in this direction.

# COGNITIVE ERROR ANALYSIS IN THE NICU SCHEME

In this section, some sample NICU incidents are introduced that had been collected by the incident reporting scheme, and the analysis process is demonstrated by juxtaposing the in-depth cognitive error modelling approach with the AIMS classification that was arrived at by a collaboration of the 'human error expert', the clinical consultant, and one representative nurse. By doing this, the benefit of the cognitive error analysis approach is further illustrated. The remainder of this chapter then reports on the evaluation of the incident reporting scheme with a questionnaire survey of implicated hospital staff.

## Analysis Case 1:

Incident Number 2 that was recorded in the incident reporting scheme concerned a drug error. The record entries read (in sum): "drug error due to heavy workload, negative consequences minimized by vigilance of staff". Measure for "Future Avoidance" that were suggested were "Vigilance of Staff" and "Lighter Workload".

The AIMS categorization of this incident comprised the following causal factors:

- High unit activity level
- Omitting intended treatment
- Distraction/inattention
- Stress

Clearly, the information given in the incident reporting form is not very informative as to the "causal context" in which this incident occurred. And clearly, the AIMS classification, although being able to reflect all the core points that were raised by the incident report, does not aid a meaningful understanding of the incident's causation or potentially effective safety recommendations, nor does it promise any meaningful (quantitative) meta-analysis of a set of incidents that might share the causal factors as listed by the AIMS classification. Interestingly, all but one of the causal factors that were "identified" from the incident record using AIMS were 'human factors', and as such impacted the distribution of 'human factors' versus 'system factors' (Figure 6-6 above).

As a response, and also in the knowledge of the magnitude of drug error occurrences reported elsewhere (Webster et al., 2001), the incident reporting form was thus changed to enable future elicitation of decisive situational factors that led to "drug error" (see Figure 6-1 and Figure 6-2 above, illustrating the revised form).

There is a fundamental contradiction in the classification of the causal and the minimizing factors in incidents such as these. On the one hand, stress and distraction are assumed to have led to crucial staff oversights with respect to their drug administration task. On the other hand, however, staff *vigilance,* and thus alertness (which can be understood as an opposite of distraction), have prevented this incident from becoming a full-blown adverse event. Both 'stress' and the state of 'vigilance' refer to psychological states. It is assumed that "high workload" led to nurses to go into one mental state (i.e. 'stress', with the underlying assumption that stress 'causes' oversight), but with the same high workload present, the opposite mental state (vigilance) was possible, too. Thus, the classification system is insufficient, since its categories do not provide differentiation power to discriminate between two instances of two different classes. Neither does the classification system shed any light on the "causes" of the incident that might be addressed and remedied by safety recommendations. By implication, it does not support bridging the gap between analytical classification of incidents and the generation of safety recommendations to address these.

This incident classification thus raises questions as to exactly which external or internal factors influenced the nurses' mental state in such a way that it led to a drug error (given that unit activity level is high by norm, not by exception), and also as to exactly which aspects of the nurses' task performance were impacted by these mental states and why?

Were the nurses overloaded with alarms, for instance? (A common problem in intensive care (Stanton, 1994)). Or was their mental capacity to pay attention to their drug administration maybe taken up by higher-level decision-making that was necessitated by the task context? Were there any assumptions or expectations present that conflicted with how the situation actually unfolded, and thus slowed down the comprehension or mental processing of further relevant data? What was the state of the nurses' existing knowledge and expectations, and what 'processing input' did these have to compete with? Since these questions concern the psychological underpinnings of this incident's causation, it will be necessary to enlist a cognitive vocabulary for its further investigation, and the use of a cognitive error framework suggests itself.

Thus, these questions illustrate the kind of information that would shed further light on the underlying cognitive causes of nurses' "high workload"-related oversight on the one hand, and to their vigilance and prevention behaviour on the other. ICS can support the formulation of these questions, since, using the cognitive error analysis approach, the actual cause of events is scrutinized within a cognitive framework. Figure 6-7 below shows which aspects of the ICS model would be implicated in investigating this incident further, relating to the questions posed previously.

Thus, employing a cognitive architecture such as ICS to further investigate incidents such as Case 1 will help scrutinizing the psychological, cognitive, underpinnings of the incident's causation. It will certainly highlight the insufficiencies of categorical labels such as the 'human error' classification suggested elsewhere with respect to providing a stepping stone to an understanding of the incident that goes beyond surface classification.

Case 1 presents a crucial example for the weakness of existing analysis approaches of drug administration approaches. Although their incidence is well-known (Webster et al., 2001), it was observed that "there seems to be no definite strategy for the elimination of drug error" (Merry and Peck, 1995) although it was also observed that "error is inherent in drug administration in anaesthesia, as it is in any complex human endeavour".

The New Zealand Green Lane Hospital (Auckland) instead found their own way of tackling the ubiquitous drug administration error by adapting the work environment (rather than human fallibility) to the task (Merry et al., 2001). They noted that "conventional methods of injectable drug administration makes little use of technology to support manual checking and are idiosyncratic and relatively error-prone" (op. cit.). They further noted, in line with my previous reasoning, that "most anaesthetists, if not all, make drug administration errors at some stage during their career (Apostolakis, 1991) no amount of good intent or harsh deterrence will stop them" (Merry and Webster, 1996). They therefore proceeded to develop a drug administration system to combat the potential for human error that has now been successfully used for over 2 years, and that is in daily use in Green Lane hospital. The core components of this system were the aspects that tackled the system's support for the anaesthetists' *cognitive capacities* during drug administration.

## Analysis Case 2:

In another incident that wasn't considered strictly technical failure, an intravenous infusion into a neonate's hand became interstitial (dislodged). Contact of the intravenous injection fluid with the baby's skin might have resulted in tissue burns, but was discovered in time, and thus the damage was contained and limited to a swollen arm.

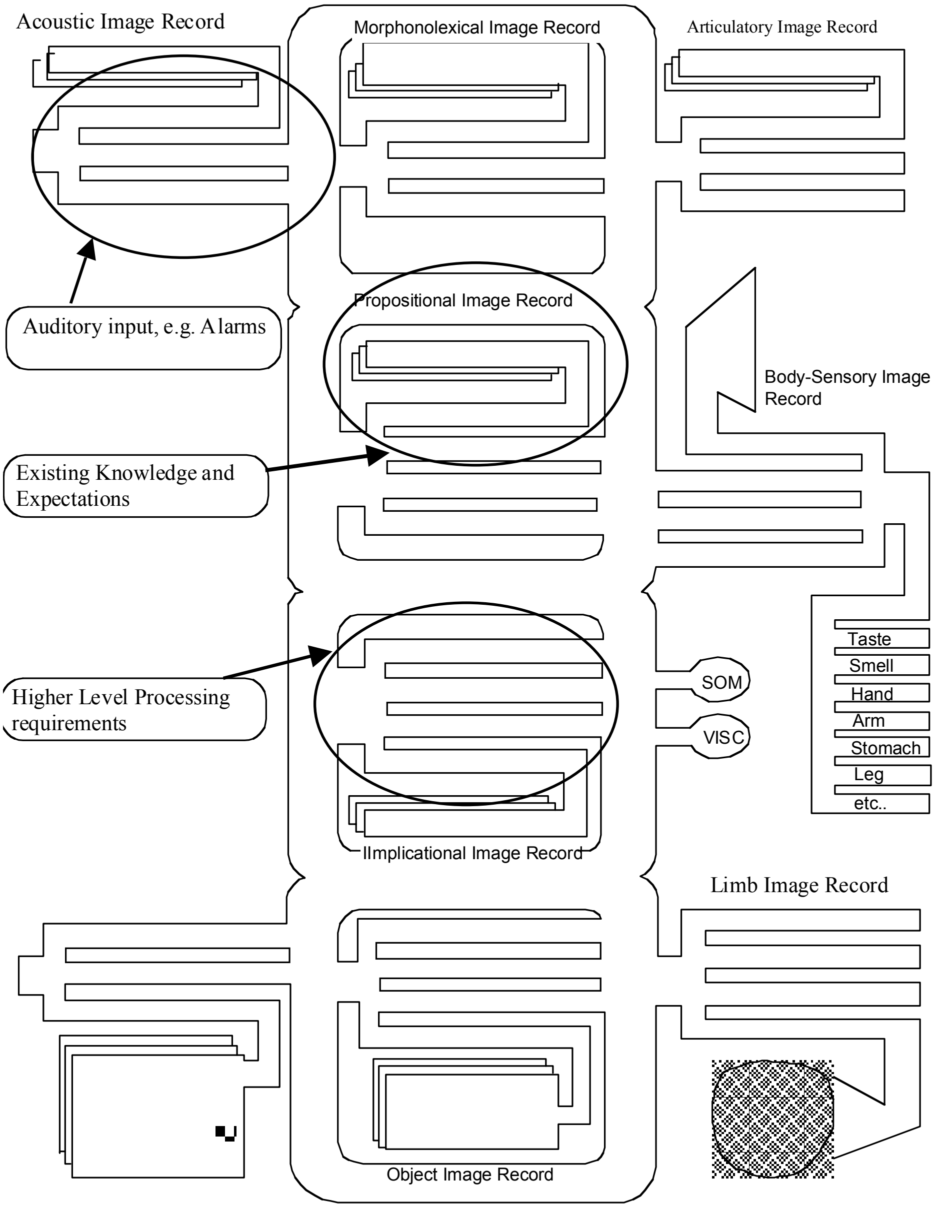


**Figure 6-7 -** ICS guidance on incident investigation

The incident report listed as “contributing factors” only that “each nurse had 2 IC babies in her care”. There were no minimizing factors listed. Suggestion for future avoidance was to have resources available to achieve a “one nurse : one patient” ratio. AIMS classifications arrived at in the analysis team were “High unit activity level”; “Insufficient staff”; “Inappropriate staff/patient allocation”; “Distraction/inattention”; and finally, “stress”.

Once more, it can be seen that the AIMS classifications of this incident are not conducive to either further understanding of the incident’s causation, nor to the identification (or even evaluation) of constructive safety recommendations. Meaningful quantitative analysis based on such classifications also seems hard to attain.

However, when the “human error expert” on the investigation team further probed the clinical and nursing staff as to possible causes and contexts of this incident, a different, more meaningful picture emerged. After lengthy discussion on the number of checks nursing staff are obliged to run on each patient, and the inherent high risk of missing one, and the importance of “vigilance” for staff to detect and prevent potential adverse events, questioning slowly led the investigation team to consider the technological component in this incident. It was noted that the intravenous infusion pump is built to react to pressure and would start sounding an alarm if the pressure limit is reached (which would be the case if the infusion became dislodged out of the vein onto the tissue). Thus to investigate this incident is was important to know whether an alarm had sounded or not, since if the pressure limit had been reached, and the pump properly set, but no alarm sounded, a technical failure could have been identified.

In reaction to this, and to the general literature on the importance of alarms in safety-critical environments (Stanton, 1994), a specific question regarding the presence of alarms was included in the incident reporting form (see earlier this chapter).

In the case that there was no technological failure implicated in this incident, a further causal option should be investigated: was the pump's pressure limit set appropriately? Was the limit set at all (or left at a previous setting)? Was the alarm switched off? Was the alarm sounding, but could not be heard? Was the agility state of the baby misjudged, which, when too agile, often precipitates these kinds of incidents?

Again, these questions probe further into the context of the incident's causation, and its interaction with the key players' mental state, knowledge, assumptions, and available processing capacities, i.e. in short, their cognitive processes. Thus, a cognitive framework is called for in order to further elucidate the events, internal or external, that led to this incident.

One further benefit of the suggested cognitive error analysis approach that is underlined by this incident case study is the importance and the nature of the questions that are brought up by the "human error expert" and that guide the discussion in the investigative team. Furthermore, rather than in lengthy verbal protocol transcripts, these discussion items could be documented in an appropriate modelling framework for future reference. The benefit of e.g. timelines and other diagrammatic modelling frameworks (such as fault trees, see the previous chapter) in elucidating accident causation (Johnson, 1998) or complex work processes (De Keyser and Nyssen, 1997) has been well demonstrated. A diagrammatic modelling framework that is based and grounded in cognitive theory, and which can also show and model an individual's cognitive processes *in interaction with* their environment, will thus support the elicitation, capture, and documentation of the reasoning underlying the identification of the cognitive precursors of incidents. Figure 6-8 demonstrates this by means of the above described incident case 2.

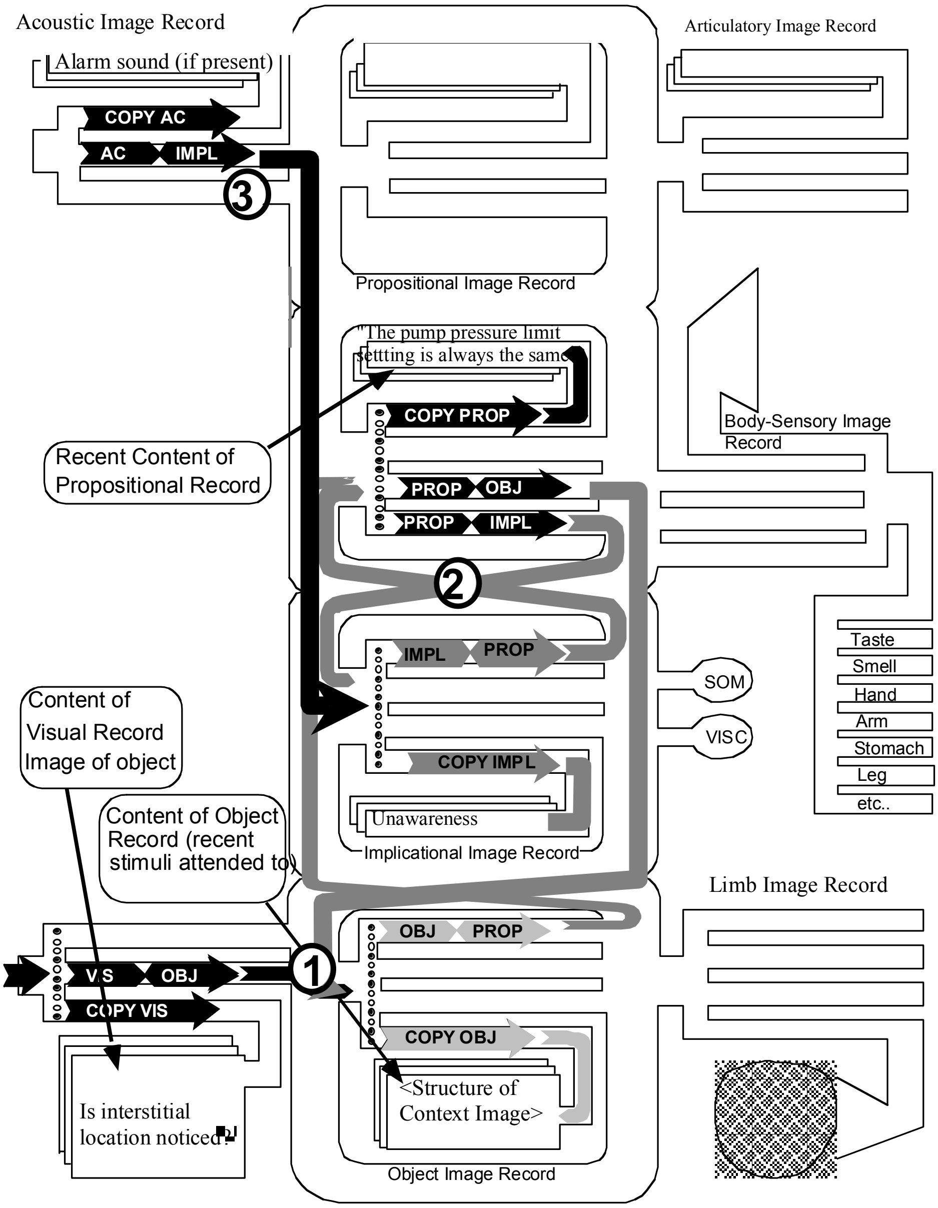


**Figure 6-8 -** Interstitial Intravenous Infusion

The figure illustrates the interplay of auditory input, higher-level mental processing, and data flow in the human cognitive system. The model can be used to play through "what if" scenarios, for instance concerning the nursing staff's assumptions on pump pressure limit settings, and how this would be affected by the presence of alarms. This can then be structured and captured in the model, and the reasoning process in the investigation can be documented. Furthermore, safety recommendations that arise from such modelling are unlikely to focus on staff vigilance as the remedy to such incidents. They are more likely to scrutinize the appropriateness of the technology and work context for the capacities, boundaries, and idiosyncrasies of the human cognitive processing system, including their dealing with alarms in an overly noisy and alarm-saturated environment, or with the detailed technical implementation of the pump pressure limit setting in terms of providing a user-friendly, safety-oriented system design.

## FORM VALIDATION

As part of the effort to take the local context of the incident reporting scheme into account, a questionnaire survey was carried out as an investigation of staff perception of the Critical Incident Reporting scheme as part of existing safety management at the unit. The study should also present an opportunity for staff to suggest improvements of the incident reporting scheme itself, as well as unit safety management in general.

22 questionnaires were administered to two 11 person day shifts on the ward. 21 were returned and 19 were included in the analysis. The two outliers that were excluded from the analysis were responses from temporary staff. Both medical and nursing staff contributed to the study.

Out of a total of 19 responses, 13 indicated awareness of the scheme, 4 said they had participated, and 11 could define "Critical Incident" (see Figure 6-9). Several valuable suggestions (see Table 6-1) on form and scheme improvement were given by the study participants. Most survey participants responded positively to the overall impact of the scheme as part of safety management at the unit. Most participants stressed the crucial importance of feedback of incident data and analysis results back to unit staff. Raised awareness of existing error potential can be inferred from the 'shock reaction' that was reported on seeing the (anonymized) incident data and analysis results. Understandably, this led to frequent calls for urgent safety intervention on the part of the survey respondents.

| **Survey Question** | **Staff Suggestions** |
|---|---|
| Form improvement | Include question on staff baby ratio<br>Include question on workload levels |
| Effectiveness of the scheme | Good for Awareness (N=4)<br>Need Action (N=3)<br>Need Feedback (N=2) |
| Safety management improvement | Need Extra Staff (N=6)<br>Need Review (N=5)<br>Need Action (N=3)<br>Need Training (N=2) |

**Table 6-1 -** Summary of Staff Suggestions for Form Improvement

The level of “safety culture” at the unit seems encouraging, with 2/3 of staff aware of the scheme and of the nature of critical incidents. Temporary staff, however, are still largely unaware. The scheme’s intended implications for safety management needs better publication. Further measures are needed to feed analysis and action results back to the staff. Reassurance of staff is needed to let them know that the identified safety deficiencies are urgently and thoroughly addressed through safety interventions where possible.

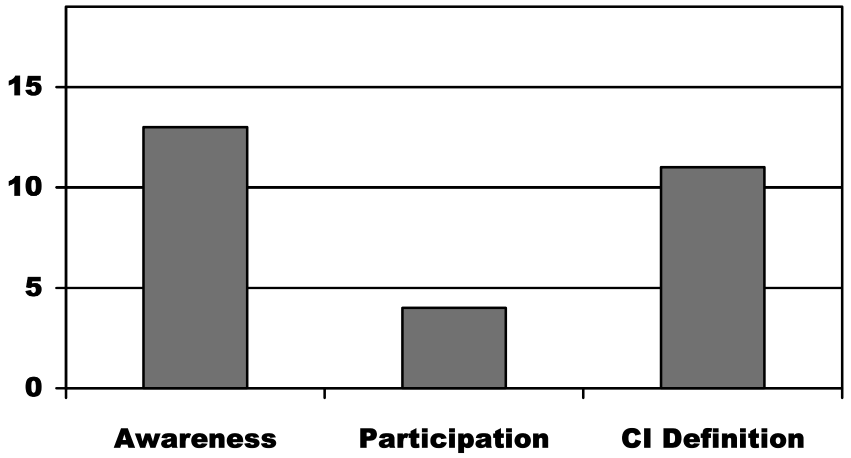


**Figure 6-9** - Staff Awareness, Participation, and Ability to define ‘Critical Incident’

# CONCLUSIONS

The NICU incident reporting scheme succeeded in achieving staff engagement and participatory design of the reporting form. Most importantly, it impacted the overall safety culture in the unit, by raising awareness of clinical incidents, and raising the belief that the occurrence of ‘mistakes’ can be dealt with in a constructive way by higher level management (rather than following a ‘punitive perfection model’ (Leape, 1994)). Equipment failures could be followed up by either contacting the manufacturer directly, or also, for instance, by being able to refer to the incidents as evidence of insufficient design of specific devices. This data provides the basis to pass on valuable lessons to other NICUs that e.g. use similar devices.

The UK Department of Health (2000) stated as their main conclusion to the Chief Medical Officer’s report: “We believe that, if the NHS is successfully to modernize its approach to learning from failure, there are four key areas that must be addressed. In summary, the NHS needs to develop:

- unified mechanisms for reporting and analysis when things go wrong;
- a more open culture, in which errors or service failures can be reported and discussed;
- mechanisms for ensuring that, where lessons are identified, the necessary changes are put into practice;
- a much wider appreciation of the value of the system approach in preventing, analyzing and learning from errors.”

Additional to this, however, it needs to be stressed that the incorporation of the specific local conditions and requirements are necessary for the successful maintenance of an incident reporting scheme. Furthermore, if a unified mechanism for analysis is determined, it needs to take into account that the classification scheme should not only aid ‘causal factor counting’, but also their meaningful analysis. This thesis proposes a cognitive error modelling approach that supports the meaningful analysis of incidents and their classification. This chapter worked through real-life examples applying this cognitive incident analysis approach.

# CHAPTER 7 CONCLUSIONS AND FURTHER RESEARCH

## THESIS CONTRIBUTION

It has been established that human involvement often plays a major role in the occurrence of accidents in safety critical systems such as in aviation, or medicine. Accident reports often resort to naming human error (“pilot error”, “operator error”) as ‘the reason’ why the accident happened.

However, identifying human involvement in the accident’s causation does not necessarily mean that ‘human error’ is the cause, and the only cause, of that accident. Human action (or inaction) never takes place in a vacuum, and is always stimulated by either the external context or an internal motivation (such as a set of expectations, or a mental goal). Therefore, the identification of human involvement in the events that led to an accident that is often misleadingly called ‘human error’ should be the starting point for further investigation, rather than provide a convenient category where blame, and thus the ‘cause’, have apparently been found.

This thesis scrutinized the current use of error cause taxonomies that is widespread in safety-critical domains such as aviation, process control, and clinical safety management. Several weaknesses were identified that related either to the taxonomies themselves (even if grounded in psychological theory), or to their use, which is often characterized by a “common sense” approach to categorization. The link to the generation and evaluation of safety recommendations was also highlighted in this thesis.

This thesis proposed the use of a cognitive error analysis approach that addresses critical weaknesses in current taxonomical approaches. The thesis' main contribution to the field, next to identifying and illustrating these weaknesses of current taxonomical approaches, is the elaboration of the potential benefits of the proposed cognitive error analysis approach in addressing these. The thesis presented these potential benefits through worked examples covering two safety-critical domains, in which accident case studies were analysed retrospectively, and in which the cognitive error analysis approach's feasibility was validated in a "live" clinical setting, through the implementation of integrated safety management measures.

It is put forward in this thesis that the proposed approach would stand a better chance to successfully address recurring system weaknesses than more traditional accident analysis approaches. Future work will also need to validate the effectiveness of the proposed approach in actually reducing the number of incidents and accidents (e.g. with a longitudinal study following up the implementation of the NICU incident reporting scheme), especially in direct comparison with alternative approaches.

It is reasoned that there are two core benefits of the proposed approach as compared to more traditional approaches: the method, involving an expert in "the human factor" (most likely a trained psychologist), and giving this expert a tool that will help them explore the context of the accident, and specifically the interaction of the human (who was involved in the accident's causation) with the system that suffered the accident. One of the core concepts in this thesis is that an accident's causation is viewed in this thesis as the result of a potential mismatch between the human system component and the rest of the system, rather than a system failure that was brought about by "human error" in isolation. The method put forward in this thesis (involving the expert and the tool) is proposed as an approach to address this viewpoint on accident investigation. This thesis presented an elaborate proof of concept for this approach.

The core points that were the focus of this thesis are:

- Human error analysis, if grounded in a cognitive theoretical framework, steers away from a common sense approach to analysis, by exposing implicit assumptions on the factors underlying human involvement in an accident's causation.

- The cognitive error analysis framework provides the tools and the vocabulary to reason about human error. Using the cognitive framework, alternative hypotheses about the underlying cognitive mechanisms of human error can be compared and substantiated.

- The modelling approach helps to document and communicate this reasoning process by explicating it in the (diagrammatical) cognitive framework

- Safety recommendations that are based on the analysis of Human Error need to be evaluated as to their theoretical validity and their compatibility with the findings of the accident/incident investigation. The cognitive error analysis approach provides the tools to support both these evaluations.

Specifically, in Chapter 4 (this thesis), an extensive table listed points of critique of existing error analysis approaches, and whether the proposed cognitive error analysis approach addresses the criticisms (table 4.1). The following section will summarize the key contributions of the proposed cognitive error analysis approach to the problem of effectively categorizing and analysing 'human error' data.

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Analysis of human behaviour only in terms of "error" (no room for non-erroneous behaviour) | The cognitive error analysis approach embodies erroneous as well as error-free behaviour and thought processes. Error is seen as the other side of the coin of an otherwise efficient thought, emotion, and action apparatus (as illustrated throughout the thesis examples, especially in chapter 4, where non-erroneous behaviour is explicitly modeled).. |
| Analysis of 'human error' only in terms of isolated behaviour | Human behaviour is analyzed in its interaction with external situation and events (this is shown, for instance, in modelling a pilot interacting with the co-pilot, situated in a moving plane, as done e.g. in chapter 3). |

**Table 7-1** - Benefits of the Proposed Approach (I)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| There is a gap between describing human behaviour as implicated in the accident's causal chain, to assigning a label to it through categorization | Cognitive error analysis models the behavioural description and its underlying cognition in a theoretical framework. The framework can be used to reason about competing categorizations (as illustrated e.g. in chapter 4 , 5, and 6). This justification process, the rationale, is thus traceable and documented. |
| Verbal description of human behaviour as implicated in an accident's causation. | The framework supports diagrammatic reasoning, which has been shown to to aid understanding and communication of the analysis results (Moran and Carroll, 1996).<br>The symbolic implementation of the model (ICSpert; May et al., 1993; Barnard et al., 1988) will aid precision, flexibility, standardization, and predictive power (see also Burns (2000) for the value of formal accident modelling). |
| Human error predictions based on implicit assumptions | The cognitive architecture allows 'what-if' scenarios to be executed, and presents a framework to reason about potential mental and behavioural consequences of cognitive precursors (this is an inherent property of modelling frameworks such as ICS, and is illustrated in competitive modelling (ie modelling of competing hypotheses) throughout the thesis). |

**Table 7-2** - Benefits of the Proposed Approach (II)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Taxonomical labels alone often lack explanatory power | Modelling the cognition underlying the error’s categorization contextualizes its causal processes, and thus adds explanatory power grounded in a theoretical framework (as illustrated in e.g. the modelling of the Gatwick incident. Explanatory power here refers not to supporting a layman’s understanding of the causal process, but an explanatory power that rises from a theoretically grounded, guided exploration of the incident’s context and causal processes). |
| Taxonomical labels do not support the generation of safety recommendations, and might even mislead the analyst (such as in the case of “reminder statements”) | The cognitive error model can help to identify intervention points for future safety measures in the causation processes. It can also be used to validate (or reject) suggested safety recommendations, by grounding both (the error analysis and the simulation of safety measures’ impacts) in a common theoretical framework (as illustrated e.g. in the TAP examples, especially in comparison with “reminder statements” as generated by the previously used approach to error modelling in the Intensive Care Unit). |
| Taxonomic labels do not provide any support for evaluating safety recommendations. | The cognitive error model can be used to compare competing recommendations, help predict their outcomes, and test the compatibility of recommendations with analysis evidence and results (see especially Chapter 4 for an illustration). |

**Table 7-3 -** Benefits of the Proposed Approach (III)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Error categorizations do not sufficiently take team factors into account. | It is possible to model interactions between several individuals in the proposed cognitive error framework, and also the interplay of team members' communication and their knowledge and mental processes (this was not the focus of this thesis, but a pilot-copilot communication scenario is included for instance in Chapter 4). |
| Taxonomies are finite, they cannot be exhaustive, and typically cover only a limited range of errors. | Since the proposed cognitive error analysis approach is grounded in a generic psychological theory, the range of behaviour and cognition that can be described is much wider than a restricted enumerations of categories (by definition). |
| Taxonomies often list mental and emotional states (such as stress, fatigue) as causal categories, but fail in contextualizing the human's mental state in terms of its impact on cognitive processes or action in a situated work activity. | The cognitive error framework can model emotions (Teasdale and Barnard, 1993) and can thus contextualise the human actor's mental state in terms of their interactions with their environment. |

**Table 7-4 -** Benefits of the Proposed Approach (IV)

| **Critique of existing error analysis approaches** | **How the proposed cognitive error analysis approach addresses these criticisms** |
|---|---|
| Seemingly 'conclusive' labelling of instances of human behaviour, rather than starting point for human error analysis. | Encourages in-depth analysis of the contextualized cognitive precursors defining the human involvement in the accident's causation (the framework provides the analyst in essence with a 'checklist' of subsystems and subsystem interactions (configurations) for consideration). |
| Common sense analysis, based on implicit assumption on theories of accident causation as well as of the human psyche. | Encourages use of appropriate expertise on human thought, decision-making, and action processes (the framework's vocabulary targets human factor experts). |
| Varying levels of "Goodness of Fit" to the accident investigation process and overall safety-management strategy. | The feasibility of the approach in live hospital settings was evaluated in this thesis, and the approach was put to use in collaboration of the human factors experts (using ICS modelling) and the hospital staff (the domain experts). |

**Table 7-5 -** Benefits of the Proposed Approach (V)

As detailed in Table 7.1, the proposed cognitive error analysis approach holds promise to address a list of weaknesses of current error analysis approaches that are utilized in accident and incident investigation. This thesis contributed to identifying and exploring these weaknesses (such as the lack of error modelling capabilities of existing cognitive models, the potential inadequacy of safety recommendations when derived with existing analysis approaches, without guidance by human error analysis expertise, and the ambivalence and lack of explanatory power of taxonomic approaches), as well as illustrating how a cognitive modelling framework could be used to support the analysis of human error in a way that would address these weaknesses (e.g. by including error modelling capabilities, by grounding the generation and evaluation of safety recommendations in a holistic theoretical model, and by adding explanatory power by utilizing the theoretical framework as a constraining (and thus structure-giving) vocabulary). This problem space was explored, and issues were identified and refined in the process. Future work will need to set about conducting a rigorous validation of the suggested benefits of the approach (such as using measures of total increase or decrease of numbers of incidents as an indicator of the method's merits).

Thus, this thesis concentrated on demonstrating the broad range of weaknesses of current error analysis approaches (summarized in tables 4.1 and 7.1), and the corresponding broad value of a systematic error analysis approach that is grounded in a theoretical framework that can account for error-free as well as erroneous human-machine interaction (or human-human interaction). It also examined the proposed analysis approach in a series of exploratory case studies. In doing this, an analysis of the process of the cognitive error analysis approach's application there was not the explicit focus of this thesis. However, in the course of the theoretical and practical exploration of the issues involved in error analysis in accident and incident investigation, several interesting points regarding an analysis process emerged (such as the advisory role of the human error analysis expert in conjunction with the domain experts' analysis process, and the role of a modelling tools like ICS as an expert's tool that is used by the human error analyst to systematically ground their reasoning process and their contextualization of the error analysis process in a cognitive theoretical framework). Future work will need to explore and formalize considerations on the process of applying the technique, including the role of the expert in the analysis process, and a proceduralized way ICS can be used to support the analysis process.

## CRITIQUE OF THE SUGGESTED COGNITIVE ERROR MODELLING APPROACH

ICS, compared to other Cognitive Task Analysis methods, is a rich and expressive modelling approach. The cognitive primitives are well defined, and no assumptions regarding a certain level of performance (such as novice/expert) are made. It also presents the possibility to model temporal information (such as concurrency and sequentiality), the lack of which Hollnagel (1991) pointed out among typical human error modelling techniques.

However, there are a number of shortcomings of the approach presented above. If the in-depth error modelling approach is used (Barnard and May, 1998; Barnard et al., (in press)), then the specification of the cognitive processing modelled is required at a very low level. It requires, next to modelling time and effort, an amount of equally well-defined error data, which is typically hard to come by. However, the design of the incident reporting scheme could counter-act this by targeting the data elicitation accordingly. Also, the incident reporting scheme could be voluntarily confidential (rather than anonymous) and include follow-up interviews with the incident reporters to clarify their responses.

Furthermore, cognitive modelling in ICS requires a degree of expertise and modelling craft skill (May and Barnard, 1994), which leads to training requirements for the analyst, and a background in cognitive psychology will be highly desirable. Ideally, a human error expert with training in relevant psychological theories will be involved in the investigation process, as illustrated in the case studies in this thesis.

Even if a high degree of expertise and skill is employed in the modelling process, a conclusive validation of the obtained modelling results is difficult, since it is not based on data derived from an empirical, laboratory-based, experiment, but from contextual field data. Thus, inter-judge reliability of the approach is estimated to be low. The complexity of human behaviour and cognition outperforms the fine-tuned and fine-grained ICS specification (as is typically the case with low-level, predictive processing models (Dix, 1997)). However, if the framework is utilized to model possible error processes and to reason about alternative hypothesis of accident causation, it is not the conclusive validity of each individual model that is required, but the explanatory power of the models in comparison.

This presumed low inter-rater reliability is shared with the human error model approaches presented in Chapter 2, where more complex causal analyses are concerned. As is noted above, the techniques only give guidance on which error shaping factors are likely to apply to the identified error forms, but no specific guidelines. A methodology that enables the modelling product to be validated is not provided either. It is very much left up to the analyst's insight and experience to identify human error forms. It was argued in this thesis that complementary cognitive modelling, even if difficult to validate on an individual model basis, can act as a tool to inform the error analysis process.

Another weakness which both cognitive as well as error modelling approaches share is the lack of methodology for translating the modelling product into design guidelines, or post-accident recommendations (e.g. Sanderson and Harwood, 1988). May et al (1993) attempted to embody this translation process in an expert system. This, however, never became fully operational. The generation of safety recommendation is, again, left up to the analyst's expertise. However, the cognitive error modelling approach as proposed in this thesis supports the analyst by enabling the identification of intervention points through the explicit detailing of the error processes. Furthermore, the impact of safety recommendations can be evaluated against the original findings and analysis model in the cognitive error analysis approach.

# FUTURE RESEARCH DIRECTIONS

## Detection Factors and Safety Recommendations

It is important to note that human involvement in incidents or near-miss accidents is not limited to 'human error', but humans also play a major role in detecting an 'accident waiting to happen', and in its avoidance and recovery. This insight mirrors Reason's (1990) metaphor of 'human error' being only one side of the coin of a well-functioning and efficient cognitive system. These aspects are typically neglected in the human error literature, and in human error taxonomies.

Seeing human error as one side of the 'cognitive coin' also resets the focus on the inevitability of erroneous behaviour, especially when interacting with complex systems. This leads to the conclusion that not only error detection and recovery mechanisms need to be further investigated, but also that systems need the capability to 'absorb' error (Rasmussen and Vicente, 1989) and to degrade 'gracefully' in the case of an incident or accident. Thus, system design needs to be based on a sophisticated understanding of error production as well as human error handling mechanisms. Future work could scrutinize the relationship between detection factors and safety recommendations systematically, thus exploring one further potential source for the generation and validation of safety recommendation.

## Beyond Individual Cognition

There has been increasing research into the analysis of team errors in safety-critical environments. Taxonomies such as these described in this thesis usually do not focus on team factors that play into the causation of accidents or incidents. One team factor that is usually included in taxonomies is “communication” (though this category typically runs the risk of being used as a ‘catch all’ for a very generic class of error forms – these might range from ‘protocol’ errors e.g. during shift hand-over, to patient records not being updated in time, to describing a ‘cranky’ surgeon that does not communicate appropriately with the anaesthetist). The generic nature of this category leads to decreased differentiation and explanatory power when it is used to classify human involvement in an accident’s causation. However, recent research has recognized the need for a whole vocabulary that specifically describes the facets found in team interactions, and especially the breakdown of these interactions (Harris et al., 1999; Kostopoulou and Shephard, 1999; Sasou and Reason, 1999).

This report has focused on the cognitive aspects of an individual’s involvement in accidents and incidents. The impact of teamwork aspects on the tasks in intensive care units (as well as in the cockpit – as illustrated in chapter 4) cannot be overemphasised, however. In ICUs, medical staff and nursing staff are continuously interacting in the process of carrying out their tasks (Kostopoulou and Shephard, 1999), and especially when dealing with emergency situations. The impact of this team interaction on an individual’s performance and on ‘human error’ thus needs to be investigated further. As a prerequisite, it thus needs to be examined whether ‘team error’, pragmatically, forms a separate entity from individuals’ errors, as current error taxonomies suggest. Initial research in this direction has already been pursued by the thought leaders in this field (Sasou and Reason, 1999). More work is needed to fully understand the team processes in these safety-critical environments, and to incorporate them successfully into the analysis of incidents and accidents.

## Towards a ‘Total System’ Approach

Reason maintains that active failure is usually associated with the performance of ‘front-line’ operators (such as pilots, nurses). Latent failure, however, is most often generated by those at the ‘blunt end’ of the system (e.g. designers, high-level decision-makers, and managers) and may lie dormant for a long time.

So far in this work we have concentrated on ‘human errors’ leading to active failure. The analysis of these is most likely to benefit from cognitive modelling techniques. We suggest that latent failures are best analysed in ‘total system’ approaches that go beyond an individual’s cognition and also take organisational aspects into account. This information is only implicitly embodied in cognitive models.

Recent research has suggested that it is valuable to investigate whether knowledge of the regulation, organisation, and management of complex systems can be incorporated in the analysis process and in safety management (Jeffcott and Johnson, 2001). The ‘external’ influences on human cognition and task performance need to be clearly demarcated by the analysis method. The scope of current modelling techniques can be delineated, and potential extensions or integration with system-oriented models could be investigated. This means that, ideally, there would be an integration of the individual error model and a model of the total system, and their interactions. Models of overall work systems usually do not lend themselves to modelling error or accidents, and to integrating models of finer grain, such as models of individuals’ cognition.

However, two approaches to modelling overall work systems could be identified that promise to be appropriate for modelling not only ideal, error-free work environments. One, Cognitive Work Analysis (Vicente, 1999), is based on Rasmussen’s cognitive engineering approach, which itself holds as a premise the inevitability of occurrences of human-machine mismatch.

Another promising work modelling approach might be based on activity theory (Vygotsky, 1978). Activity theory is one psycho-social theory of human action that confronts the concept of “conflict” as one of its core concerns. This concept of conflict often accompanies descriptions of mismatches, errors and accidents in activity theoretical case studies (Nardi, 1995), and it warrants further investigation in as much a work model based on activity theory might add value to the process of analysing the human involvement in accidents and incidents.

# CONCLUSION

The case studies detailed in this thesis have shown how human error taxonomies on their own, even if rooted in cognitive theory, may not sufficiently contribute to human error analysis in accident or incident investigation. They provide limited explanatory power, and they also do not provide constructive value in that they do little to support the generation of safety recommendations. Although error categorisation does provide a basis for labelling analysis results, the resulting quantitative categorisation data itself does not aid the understanding of the error or the substantiation of safety recommendations. Labelling of instances of human behaviour does provide the basis for larger scale statistical analysis of error data, and aids the prioritization of further in-depth analysis and intervention. However, the categories can be misleading, and the value of the resulting statistical analysis might thus be doubtful. We demonstrated in this thesis how a cognitive error modelling approach can be used to address those weaknesses, when applied to human error analysis in accident and incident investigation, across safety-critical domains. The cognitive error analysis approach is not a panacea, however, and ideally, analysis of team factors and organizational work context factors also need to be taken into account in accident and incident analysis.

# REFERENCES

AAIB, A. A. I. B. (1989). “Aircraft Accident Report Number 2/89. Report on the Incident involving BAC-1-11 G-AYWB, and Boeing 737 EI-BTZ on 12 April, 1988 at Gatwick Airport.”

AAIB, A. A. I. B. (1997). Aircraft Accident Report Number: - 4/97 - (EW/C96/10/8). Report on the accident to Aerospatiale AS 355F1 Twin Squirrel, G-CFLT near Middlewich, Cheshire on 22 October 1996.

Allison, R. (1997). Inquiry into the fire on heavy goods vehicle shuttle 7539 on 18 November 1996. London, Her Majesty’s Stationary Office.

Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, N.J., Lawrence Erlbaum.

Annett, J., K. D. Duncan, R. B. Stammers and M. J. Gray (1971). Task Analysis. London, HMSO.

Apostolakis, G. (1991). *Probabilistic Safety Assessment and Management.* New York, Elsevier.

Ashcraft, M. H. (1994). *Human Memory and Cognition*, Harper Collins College Publishers.

Bainbridge, L. (1987). Ironies of Automation. *New Technology and Human Error*. J. Rasmussen, K. Duncan and J. Leplat, John Wiley & Son Ltd.

Bainbridge, L. (1993). “Types of Hierarchy Imply Types of Model.” *Ergonomics, Special Issue on Cognitive Processes in Complex Tasks* **36**(11): 1399-1412.

Barnard, P.J., Wilson, M., and MacLean, A. (1988). “Approximate modelling of cognitive activity with an expert system: A theory based strategy for developing an interactive design tool.”*The Computer Journal, pp. 445-45.*

Barnard, P. and J. May (1998). “Representing Cognitive Activity in Complex Tasks.” *Special Issue of Human-Computer Interaction*.

Barnard, P., J. May, D. Duke and D. Duce (in press). “Systems, Interactions, and Macrotheory.” *Transactions on Computer-Human Interaction*.

Barnard, P. J. (1987). Cognitive Resources and the Learning of Human-Computer Dialogs. *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. J. M. Carroll, MIT Press, Cambridge, MA.

Barnard, P. J., May, J. (1993). Cognitive Modelling for User Requirements. *Computers, Communication and Usability*. P. F. Byerley, P. J. Barnard and J. May. Amsterdam, Elsevier.

Barnhart, W., C. Billings, G. Cooper and e. al. (1975). A method for the study of human factors in aircraft operations. Washington, D.C., National Aeronautics and Space Administration (NASA).

BASI, D. o. T. B. o. A. S. I. (1994). Investigation Report 9301743 Piper PA31-350 Chieftain VH-NDU Young, NSW, 11 June 1993. Australia.

Battles, J. B., H. S. Kaplan, T. W. van der Schaaf and C. E. Shea (1998). “The Attributes of Medical Event-Reporting Systems.” *Arch Pathol Lab Med* **122**.

BBCNews (2001). Medical accidents - unstoppable? *<http://news.bbc.co.uk/hi /english/health/newsid_1171000/ 11717 12.stm>*.

Benner, L. (1975). *Accident Theory and Accident Investigation*. Proceedings of the Society of Air Safety Investigators Annual Seminar, Ottawa, Canada.

Benner, L. (1981). *Methodological Biases Which Undermine Accident Investigation*. Proceedings of the International Society of Air Safety Investigators, 18th Annual Seminar, Washington, D.C.

Berger, J. T. (2001). “Book Review of Margin of Error: The Ethics of Mistakes in the Practice of Medicine.” *New England Journal of Medicine* **344**(5): 389-390.

Berwick, D. M. (1998). *Taking Action to Improve Safety: How to Increase the Odds of Success*. MedErrors, California.

Bes, M. O. and C. W. Johnson (1998). *Towards a Framework for the Use of Cognitive Models to Analyse Accidents*. Second Workshop on Human Error, Safety and Systems Development, Seattle, USA,, Safeware Engineering Corporation.

Bibbings, R. (2001). Learning from Safety Failure, Royal Society for the Prevention of Accidents, (http://www.rospa.co.uk) (Jan 2002).

Bignell, V. and J. Fortune (1984). *Understanding Systems Failures*, Manchester University Press.

Billings, C. (1998). *What can and What can't be Learned from Self-Reported Errors?* Fourth Annual Symposium on Human Interaction with Complex Systems - HICS'98, Fairborn, OH, USA.

Blalock, H. M. (1972). *Social Statistics*. New York, McGraw-Hill.

Bogner, M. S., Ed. (1994). *Human Error in Medicine*, Lawrence Erlbaum Associates.

Booth, P. A. (1991). Modelling the User: User-System Errors and Predictive Grammars. *Human-Computer Interaction and Complex Systems*. J. L. Weir G.R.S. and Alty, Academic Press Ltd.

Buckingham-Shum, S., Blandford, A.E., Duke, D., Good, J., May, J. Paterno', F., Young, R.M. (1996). "Multidisciplinary Modelling for User-Centred System Design: An Air-Traffic Control Case Study". Sasse, A., Cunningham, J. and Winder, R. (eds.) *People and Computers XI.*

Burns, C. P. (2000). Analysing Accident Reports using Structured and Formal Methods. *Computing Science*. Glasgow, Glasgow University.

Busse, D. (1998). PhD First year Report. *Computing Science*. Glasgow, Glasgow University.

Busse, D. and C. Johnson (1998). *Modelling Human Error within a Cognitive Theoretical Framework*. ECCM '98, Nottingham, UK, Nottingham University Press.

Busse, D. K. (2000). "Staff Perception of Critical Incident Reporting in an Intensive Care Unit." *Proceedings of the British Psychological Society (BPS).*

Busse, D. K. and C. W. Johnson (1999). *Identification and Analysis of Incidents in Complex, Medical Environments*. First Workshop on Human Error and Clinical Systems, Glasgow, UK.

Busse, D. K. and D. J. Wright (2000). "Classification and Analysis of Incidents in Complex, Medical Environments." *Topics in Health Information Management Special Issue on Human Error in Clinical Information Systems* **20**(4): 1-11.

Cacciabue, P. C. (1998). *Modelling and Simulation of Human Behaviour in System Control –Advances in Industrial Control*, Springer Verlag London Ltd.

Card, S. K., T. P. Moran and A. Newell (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, N.J., Lawrence Erlbaum.

Carroll, J. M. and R. L. Campbell (1989). "Artifacts as Psychological Theories: the Case of Human-Computer Interaction,." *Behaviour and Information Technology* **8**: 247-256.

Chappell, S. L. (1996). *Using Voluntary Incident Reports for Human Factors Evaluations*. The International Society of Air Safety Investigators.

Cullen, H. L. (1990). The Public Inquiry into the Piper Alpha Disaster. London, HMSO.

Davies, L. M. (1979). The Three Mile Island Incident. Harwell, Atomic Energy Technical Unit.

Dawson, S., P. Garside and F. Moss, Eds. (1998). *Organisational Change - the Key to Quality Improvement*. Quality in Health Care, BMJ Publishing Group, RCN Publishing Company.

De Keyser, V. (1989). “Human Error.” *Recherche* **20**(216): 1444.

De Keyser, V. and A.-S. Nyssen (1997). *The Management of Temporal Constraints in Naturalistic Decision Making. The Case of Anesthesia*. Proceedings of the Fourth Conference on Naturalistic Decision Making, Warrenton, Virginia.

Dix, A. (1997). *Closing the Loop: Modelling Action, Perception, and Information*. Proceedings of AVI'96, ACM Press.

DoE (1992). Root Cause Analysis Guidance Document. Washington, D.C., US Department of Energy.

DoE (1997). Conducting Accident Investigations. Washington, D.C., U.S. Department of Energy.

DoH Department of Health, C. M. O. (2000). An Organisation with a Memory. UK, Crown Copyright.

Duke, D. J., P. J. Barnard, D. A. Duce and J. May (1995). Syndetic Modelling, Amodeus-2.

Eagle, C. J., J. M. Davies and J. Reason (1992). “Accident Analysis of large-scale technological disasters applied to an anaesthetic complication.” *Canadian Journal of Anaesthesia* **39**(2): 118-122.

Flanagan, J. C. (1954). “The Critical Incident Technique.” *Psychological Bulletin* **51**: 327-358.

Freitag, M. (1997). Structure of Event Analysis. *After the Event - From Accident To Organisational Learning*. A. Hale, B. Wilpert and M. Freitag, Pergamon Books**:** pp 11-21.

Fullwood, R. R. and R. E. Hall (1988). *Probabilistic Risk Assessment in the Nuclear Power Industry: Fundamentals and Applications*, Pergamon Press.

Gaba, D., M. Maxwell and A. DeAnda (1987). “Anesthetic mishaps: Breaking the chain of accident evolution.” *Anesthesiology* **66**: 670-676.

Grant, A. S. and J. T. Mayes (1991). Cognitive Task Analysis? *Human-Computer Interaction and Complex Systems*. W. G.R.S. and J. L. Alty, Academic Press Ltd.

Grant, S. (1996). *Developing a Cognitive Architecture for Modelling and Simulation of Cognition and Error in Complex Tasks,*. RoHMI project meeting, Valenciennes.

Green, R. G. (1985). “Stress and Accidents.” *Aviation Space and Environmental Medicine* **56**(7): 638 - 641.

Greenwood, M. and H. M. Woods (1919). A Report on the Incidence of Industrial Accidents Upon Individuals With Special Reference to Multiple Accidents, British Industrial Fatigue Research Board.

Hale, A., B. Wilpert and M. Freitag (1997). *After The Event - From Accidents to Organisational Learning*, Pergamon Books.

Harris, M., A. P. Jagodzinski and K. R. Greene (1999). *Human Error in the Context of Work Activity Systems: A Case Study of Knowledge Based Computer Systems in Routine Use*. Human Error and Clinical Systems, Glasgow, UK.

Harvard University (1990). *Medical Practice Study: Patients, Doctors, and Lawyers: Studies of Medical Injury in New York.* Boston, Harvard University Press.

Heinrich, H. W. (1936). *Industrial Accident Prevention*. New York, NY, McGraw Hill.

Helmreich, R., J. Wilhelm, S. Gregorich and T. Chidester (1990). “Preliminary results from the evaluation of cockpit resource management training: performance ratings of flight crews.” *Aviat Space Environ Med* **61**: 576-9.

Helmreich, R. L. (2000). “On error management: Lessons from aviation.” *British Medical Journal* **320**: 781-785.

Hollnagel, E. (1983). *Why Human Error is a Meaningless Term*. NATO conference on Human Error, Bellgio, Italy, Published on http://www.iav.ikp.liu.se/staff/hollnagel/HE%20discussion.htm (current at Jan 2002).

Hollnagel, E. (1991). The Phenotype of Erroneous Actions: Implications for HCI Design. *Human-Computer Interaction and Complex Systems*. G. R. S. Weir and J. L. Alty. London, Academic Press Ltd.

ICAO, I.C.A.O. (1993). *Investigation of Human Factors in Accidents and Incidents*. Montreal, ICAO.

ICAO, I.C.A.O. (1994). *International Standards and Recommended Practices: Aircraft Accident and Incident Investigation - Annex 13*, ICAO.

Jeffcott, M. and C. W. Johnson (2001). *The Role of User Participation in the Failure of Public-Sector IT Projects*. IHM-HCI, Lille, France.

Johnson, C. W. (1995). The Application of Petri Nets to Represent and Reason about Human Factors Problems During Accident Analyses, Editors. 1995, Springer Verlag: Berlin. p. 93-112. *The Design, Specification And Verification Of Interactive Systems*. P. Palanque and R. Bastide. Berlin, Springer Verlag**:** p. 93-112.

Johnson, C. W. (1998). "Representing the Impact of Time on Human Error and Systems Failure." *Interacting with Computers* **11**: 53-86.

Johnson, C. W. (1999). *The Application of User Modelling Techniques to Reason about the Human Contribution to Major Accidents*. User Modelling'99.

Johnson, C. W. (1999). *Visualizing the Relationship between Human Error and Organizational Failure*. Proceedings of the 17th International Systems Safety Conference, Unionville, Virginia, United States of America, The Systems Safety Society.

Johnson, C. W. (2000). *Viewpoints and Bias in Accident and Incident Reporting*. Foresight and Precaution: Proceedings of European Safety and Reliability Conference ESREL 2000, Balkema.

Johnson, C. W. (2002). *Reasons for the Failure of Incident Reporting in the Healthcare and Rail Industries*. Components of System Safety: Proceedings of the 10th Safety-Critical Systems Symposium, Springer Verlag.

Jordan, B. (1993). *Ethnographic Workplace Studies and Computer Supported Cooperative Work*. Interdisciplinary Workshop on Informatics and Psychology, Schaerding, Austria.

Kanse, L. and T. van der Schaaf (2000). Recovery from failures - understanding the positive role of human operators during incidents. *Human System Interaction: Education, Research and Application in the 21st century*. D. de Waard, C. Weikert, J. Hoonhout and J. Ramaekers. Netherlands, Shaker Publishing.

Kennedy, I., R. Howard, B. Jarman and M. Maclean (2000). The Inquiry into the management of care of children receiving complex heart surgery at The Bristol Royal

Infirmary. Interim Report: Removal and retention of human material. UK, Crown Copyright.

Kieras, D. and D. E. Meyer (1995). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Ann Arbor, University of Michigan, Electrical Engineering and Computer Science Department.

Kieras, D. E. and P. G. Polson (1985). "An Approach to the Formal Analysis of User Complexity." *International Journal of Man-Machine Studies* **22**: 365-394.

Kieras, D. E., S. D. Wood and D. E. Meyer (1997). "Predictive Engineering Models Based on the EPIC Architecture for a Multi-Modal high-Performance Human-Computer Interaction Task." *ACM Transactions on Computer-Human Interaction* **4**(3): 230-275.

Kirwan, B. (1992). "Human Error Identification in Human Reliability Assessment. Part 1 (Overview of Approaches) and Part 2 (Detailed Comparison of Techniques),." *Applied Ergonomics* **23** (5 and 6): 299 – 318 and 371 – 381.

Kirwan, B. (1998). "Human Error Identification Techniques for Risk Assessment of High Risk Systems Part 1 and 2" *Applied Ergonomics* **29** (3 and 5).

Kirwan, B. and L. K. Ainsworth (1992). *A Guide to Task Analysis*. London, Taylor and Francis.

Kleinfield, N. R. (2001). "The Crash of Flight 587: The overview; 260 on jet die in Queens crash; 6 to 9 missing as 12 homes burn; U.S. doubts link to terrorism." *New York Times* **Late Edition**(Final, Section A): Page 1, Column 6.

Knowles, C. J. (1997).
A Qualitative Approach to the Assessment of the Cognitive Complexity of an Interface. *Computing Science*, University of London.

Knox, N.W., and R.W. Eicher. MORT User's Manual (SSDC-04R3). Rev 3. US Department of Energy, System Safety Development Center: EG&G Idaho, Inc., Indaho Falls. (1980)

Kostopoulou, O. and A. Shephard (1999). *Fragmentation of Care and the Potential for Human Error in Neonatal Intensive Care*. Human Error and Clinical Systems, Glasgow, UK.

Laird, J., P. Rosenbloom and A. Newell (1986). *Universal Subgoaling and Chunking*. Boston, mass., Kluwer Academic Publishers.

Leape, L. L. (1994). “Error in Medicine.” *JAMA: The Journal of the American Medical Association* **272**(23): 1851-57.

Leape, L. L., T. A. Brennan, N. Laird, A. G. Lawthers and H. Hiatt (1991). “Adverse Events and Negligence in Hospitalised Patients.” *Iatrogenics* **1**: 17-21.

Leape, L. L., D. D. Woods, M. J. Hatlie, K. W. Kizer, S. A. Schroeder and G. D. Lundberg (1998). “Promoting Patient Safety by Preventing Medical Error.” *JAMA: The Journal of the American Medical Association* **280**(16): 1444-47.

Lekberg, A. K. (1997). “Different Approaches to Incident Investigation - How the Analyst Makes a Difference.” *Hazard Prevention* **33**(4).

Leveson, N. G. (1995). *Safeware: System Safety and Computers, A Guide to Preventing Accidents and Losses Caused By Technology*, Addison Wesley Publishing Company.

Leveson, N. G. and C. Turner (1993). “An investigation of the Therac-25 accidents.” *IEEE Computer*: 18-41.

Logan, G. D. a. C., W.B. (1984). “On the Ability the Inhibit Thought and Action - A Theory of an Act of Control.” *Psychological Review,* **91**(3): 295-327.

Luxhoj, J. T., D. N. Arendt, T. P. Williams and T. G. Horton (1997). *An Application of Advanced Information Technology for Assessing Aircraft Accident Causation*. The International Society of Air Safety Investigators.

Macias-Chapula, C. A. (1997). An Approach to Identifying the Role of 'Information' in a Health Care System: Implications for the Quality of Health. *Social Science, Technical Systems, and Cooperative Work - Beyond the Great Divide*. G. C. Bowker, S. L. Star, W. Turner and L. Gasser, Lawrence Erlbaum Associates.

Maddox, M. E. and J. Reason (1996). *Human Factors Guide for Aviation Maintenance*. Washington, DC, Federal Aviation Administration/Office of Aviation Medicine.

May, J. and P. Barnard (1994). *Supportive Evaluation of Interface Design*. The Proceedings of the First Interdisciplinary Workshop on Cognitive Modelling and User Interface Development, Vienna.

May, J. and P. Barnard (1995a). Cinematography and Interface Design. *Human-Computer Interaction: Interact ‘95*. K. Nordby, P. H. Helmersen, D. J. Gilmore and S. A. Arnesen. London, Chapman and Hall**:** 26-31.

May, J., & P. Barnard (1995b) Towards supportive evaluation during design. *Interacting with Computers*, 7 , 115-143.

May, J., P. J. Barnard and A. Blandford (1993). “Using Structural Descriptions of Interfaces to Automate the Modelling of User Cognition.” *User Modelling and User Adaptive Interfaces* **3**: pp. 27-64.

Merry, A. F. and D. J. Peck (1995). “Anaesthetists, errors in drug administration, and the law.” *The New Zealand Medical Journal* **108**(1000): 185-7.

Merry, A. F. and C. S. Webster (1996). “Labelling and drug administration error.” *Anaesthesia* **51**: 987-988.

Merry, A. F., C. S. Webster and D. J. Mathew (2001). “A New, Safety-Oriented, Integrated Drug Administration and Automated Anesthesia Record System.” *Anaesth Analg* **93**: 385-90.

Mills, C. W. (1970). *The Sociological Imagination*, Pelican Books.

Moran, T. P. and J. M. Carroll, Eds. (1996). *Design Rationale: Concepts, Techniques, and Use*. Computers, Cognition, and Work. New Jersey, Lawrence Erlbaum Associates.

Morone, J. G. and E. J. Woodhouse (1986). *Averting Catastrophe - Strategies for Regulating Risky Technologies*. Berkeley/Los Angeles/London, University of California Press.

Nardi, B. A., Ed. (1995). *Context and Consciousness: Activity Theory and Human Computer Interaction*. Cambridge, MIT Press.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA, Harvard University Press.

Newman, W. and M. Lamming (1995). *Interactive Systems Design*, Addison Wesley.

Nielsen, J. (1990). “A Meta-Model for Interacting with Computers.” *Interacting with Computers* **2**: 147-160.

Norman, D. A. (1981). “Categorisation of Action Slips.” *Psychological Review*(88/1): pp. 1-15.

Norman, D. A. (1983). “Design Rules based on Analyses of Human Error.” *Communications of the ACM* **26**(4).

Norman, D. A. (1986). Cognitive Engineering. *User-Centered System Design: New Perspectives in Human-Computer Interaction*. N. D.A. and D. S.W. Hillsdale, NJ, LEA.

Norman, D. A. (1988). *The Psychology of Everyday Things,*. N.Y., Basic Books.

Norman, D. A. (1990). "Commentary: Human Error and the Design of Computer Systems." *Communications of the ACM* **33**(1).

Norman, D. A. (1993). *Things That Make Us Smart – Defending Human Attributes In The Age Of The Machine,* Addison-Wesley.

NTSB (1998). *National Transportation Safety Board: Revised Accident/Incident Database Coding Manual*. Washington, DC.

O'Hare, D. and B. Lawrence (2000). "The Shape of Aviation Psychology: A Review of Articles Published in the First 5 years of the International Journal of Aviation Psychology." *International Journal of Aviation Psychology* **10**(1): 1-11.

O'Hare, D., M. Wiggins, R. Batt and D. Morrison (1994). "Cognitive Failure Analysis for Aircraft Accident Investigation." *Ergonomics* **37**(11): 1855-1869.

Perrow, C. (1984) "*Normal Accidents: Living with High Risk Technologies*". Basic Books.

Popper, K. R. (1934). *Logik der Forschung (The Logic of Scientific Discovery)*, Springer.

Porter, D., Ed. (1981). *Society and the Social Sciences*, Open University Press.

QuIC, Q. I. C. T. F. (2000). Doing What Counts for Patient Safety: Federal Actions to Reduce Medical Errors and Their Impact. Report of the Quality Interagency Coordination Task Force (QuIC) to the President, February 2000. Washington, DC, Quality Interagency Coordination Task Force.

Rasmussen, J., Pedersen, O. M., Mancini, G., Carnino, A., Griffon, M. & Gagnolet, P. (1981). *Classification system for reporting events involving human malfunctions (Risø-M-2240), SINDOC(81) 14*. Risø National Laboratory: Roskilde, Denmark.

Rasmussen, J. (1982). "Human Error: a Taxonomy for Describing Human Malfunction in Industrial Installations." *Journal of Occupational Accidents* **4**: 311-333.

Rasmussen, J. (1983). "Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models." *IEEE Transactions: Systems, Man and Cybernetics* **SMC-13**: 257-267.

Rasmussen, J. (1985). "Trends in Human Reliability-Analysis." *Ergonomics* **28**(8): 1185-1195.

Rasmussen, J. (1986). *Information processing and human machine interaction: An approach to cognitive engineering*. New York, NY, North-Holland.

Rasmussen, J. (1987). The Definition of Human Error and a Taxonomy for Technical System Design. *New Technology and Human Error*. J. Rasmussen, K. Duncan and J. Leplat, John Wiley & Sons Ltd.

Rasmussen, J. (1990). "Human Error and the Problem of Causality in Analysis of Accidents,." *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences* **327**(1241): 449-462.

Rasmussen, J. (1990). "Human error in organising behaviour." *Ergonomics* **33**.

Rasmussen, J. (1990). Mental Models and the Control of Action in Complex Environments. *Mental Models and Human-Computer Interaction*. D. Ackermann and M. J. Tauber. North Holland, Elsevier Science Publishers B.V.

Rasmussen, J., K. Duncan and J. Leplat, Eds. (1987). *New Technology and Human Error*. Wiley Series: New Tecnology and Work. Berlin, John Wiley & Son Ltd.

Rasmussen, J. and K. Vicente (1989). "Coping with Human Errors through System Design: Implications for Ecological Interface Design." *International Journal of Man-Machine Studies* **31**: 517-534.

Reason, J. (1988). Framework Models of Human Performance and Error: A Consumer Guide. *Tasks, Errors, and Mental Models*. L. P. Goodstein, H. B. Andersen and S. E. Olsen, Taylor and Francis Ltd.

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

Reason, J. (1997). *Managing the Risks of Organizational Accidents*, Ashgate Publishing Limited.

Reason, J. and M. E. Maddox (1996). Human Factors Guide for Aviation Maintenance. Washington, DC, Federal Aviation Administration/Office of Aviation Medicine**:** 14/1-14/45.

Reece, W. J. and S. G. Hill (1995). *Human Performance Analysis of Industrial Radiography Radiation Exposure Events*. Human Factors and Ergonomics Society 39th Annual Meeting.

Rizzo, A., O. Parlangeli, E. Marchigiani and S. Bagnara (1996). "HCI in Italy: The Management of Human Errors in User-Centered Design." *SIGCHI Bulletin* **28**(3).

Roth, E. M., D. D. Woods and H. E. Pople (1992). "Cognitive Simulation as a Tool for Cognitive Task Analysis." *Ergonomics* **35**: 1163-1198.

Rouse, W. B. and W. J. Cody (1989). “A Theory-Based Approach to Supporting Design Decision-Making and Problem-Solving.” *Information and Decision Technologies* **15**(4).

Rouse, W. B. and N. M. Morris (1987). “Conceptual Design of a Human Error Tolerant Interface For Complex Engineering Systems.” *Automatica* **23**(2): 231-235.

Runciman, W. B., A. Sellen, R. K. Webb, J. A. Williamson, M. Currie, C. Morgan and W. J. Russell (1993). “Errors, Incidents, and Accidents in Anaesthetic Practice.” *Anaesthesia and Intensive Care* **21**(5): 506-519.

Sanderson, P. M. and K. Harwood (1988). The Skills, Rules, and Knowledge Classification: A Discussion of its Emergence and Nature. *Tasks, Errors, and Mental Models*. L. P. Goodstein, H. B. Andersen and S. E. Olsen, Taylor and Francis Ltd.

Sasou, K. and J. Reason (1999). “Team Errors: Definition and Taxonomy.” *Reliability Engineering and Safety* **65**: 1-9.

Sasse, M. A. (1992). Users' models of computer systems. *Models in the Mind*. Y. Rogers and A. Rutherford. London, Academic Press.

Sasse, M. A. (1997). Eliciting and Describing Users’ Models of Computer Systems, Faculty of Science, University of Birmingham.

Senders, J. W. and N. P. Moray, Eds. (1991). *Human Errors. Their Causes, Prediction, and Reduction.* Hillsdale, NJ, LEA Inc.

Simon, T. (1988). Analysing the Scope of Cognitive Models in HCI: A Trade-Off Approach. *People and Computers IV, Proceedings of the Fourth Conference of the British Computer Society*. D. M. Jones and R. Winder, Cambridge University Press.

Snowdon, P. (2002). “*An Investigation Into Using Informal Models Of Argumentation To Support The Analysis Of Accident Reports*.” PhD Thesis, Computing Science Dept, Glasgow University.

Snowdon, P. and C. W. Johnson (1998). “The Impact of Rhetoric on Accounts of Human 'Error' in Accident Reports.” *ISASI Forum online* **4**(31).

Stanhope, N., C. Vincent, S. E. Taylor-Adams, A. M. O'Connor and R. W. Beard (1997). “Applying Human Factors Methods to Clinical Risk Management.” *British Journal of Obstetrics and Gynaecology* **104**: 1125-1232.

Stanton, N. A. (1994). *Human Factors in Alarm Design*. London, Taylor & Francis.

Svenson, O. and P. Sjöström (1997). *Two Methods for Accident Analysis: Comparing a Human Information Processing with an Accident Evolution Approach*. Proceeding of CSAPC'97, Sixth European Conference for Cognitive Science Approaches to Process Control, Baveno, Italy.

Teasdale, J. and P. Barnard (1993). *Affect Cognition and Change: re-modelling depressive thought.* Hove, Erlbaum Associates.

Tognazzini, B. (1991). *TOG on Interface,*. Reading, MA, Addison-Wesley.

Van der Schaaf, T. W. (1996). *PRISMA: A Risk Management Tool Based on Incident Analysis*. International Conference and Workshop on process Safety Management and Inherently Safer Processes, Orlando, Florida, American Institute of Chemical Engineers.

Van der Schaaf, T. W., D. A. Lucas and A. R. Hale, Eds. (1991). *Near Miss Reporting as a Safety Tool*, Butterworth-Heinemann Ltd.

Vicente, K. (1999). *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*, LEA.

Vygotsky, L. S. (1978). *Mind and Society*. Cambridge, Harvard University Press.

Webster, C. S., A. F. Merry, L. Larsson, K. A. McGraths and J. Weller (2001). “The Frequency and Nature of Drug Administration Error During Anaesthesia.” *Anaesthesia and Intensive Care* **29**(5).

Wedderburn, D. (1998). “Where Does the Clinical Audit fit in?” *Quality in Health Care* **7**: 1-4.

Westrum, R. (2001). *11th Meeting of the Advisory Committee on Blood Safety and Availability*. Department of Health and Human Services, Advisory Committee on Blood Safety and Availability Meeting, Wash., D.C.

Woods, D. D. (1987). “Cognitive Engineering in Complex and Dynamic Worlds.” *International Journal of Man-Machine Studies* **27**: 571-585.

Woods, D. D. (1994). *The Critical Incident Technique - 40 Years Later*. The Human Factors and Ergonomics Society 38th Annual Meeting.

Woods, D. D., L. J. Johannesen, R. I. Cook and N. Sarter (1994). *Behind Human Error: Cognitive Systems, Computers, and Hindsight*, CSERIAC Crew Systems Ergonomics Information Analysis Center.

Woods, D. D. and E. M. Roth (1988). “Modelling Human Intention Formation For Human Reliability Assessment.” *Reliability Engineering and System Safety* **22**(1-4): 169-200.

Wright, D. (1999). *Ten Years of a Medical Incident Reporting System*. First Workshop on Human Error and Clinical Systems, Glasgow, UK.

Wright, D., S. J. MacKenzie, I. Buchan, C. Cairns and L. E. Price (1991). “Critical Incidents in the Intensive Therapy Unit.” *The Lancet* **338**: 676-78.

Young, R. M. and G. D. Abowd (1994). *Multi-Perspective Modelling of Interface Design Issues: Undo in a Collaborative Editor*. People and Computers IX, Proceedings of HCI '94, Cambridge University Press.

Zotov, D. (1996). *Reporting Human Factors Accident*. ISASI Proceedings Part II, International Society of Air Safety Investigators.

# **APPENDIX A** – PROOF OF CONCEPT (PUBLICATION)

Busse, D.K, and Johnson, C.W. (1998)
**“Modelling Human Error within a Cognitive Theoretical Framework”**
Published in the Proceedings of the 2nd European Conference of Cognitive Modelling (ECCM-98), Ritter, F.E., and Young, R.M. (eds.), Nottingham University Press, 1998, pp 90-97.

# Modelling Human Error within a Cognitive Theoretical Framework

## 1 INTRODUCTION

Current cognitive user models enable interface designers to describe, analyse and predict aspects of user cognition. However, none of the major cognitive user models such as ICS, or GOMS tackle the human error aspect of cognition explicitly. The represented operator performance is constrained to be error-free, expert performance. This report argues that human-machine system analysis as well as current error-modelling techniques will greatly benefit from representing a cognition-based error model within a cognitive architecture, such as ICS. The Netscape Internet browser acts as a case study throughout. The resulting approach is shown to aid the analysis of human error. Reasoning about potential error causes as well as the generation of design recommendations can thus be grounded in cognitive theory.

### 1.1. Integrating Error Models and Cognitive Architectures

Cognitive architectures seek to represent the building blocks of human cognition. They provide the basis for cognitive user models, which strive to represent some aspects of the user's understanding, knowledge, or cognitive processing. These models can then contribute to our understanding of the cognitive limitations of an operator performing a task, for example the effects of cognitive load on user performance (Barnard and May, 1993; Ashcraft, 1994).

Erroneous task performance highlights precisely these limitations of human cognition. It is surprising, therefore, that the major cognitive user models do not explicitly tackle issues associated with erroneous performance based on cognition. They strive to represent error-free performance, assuming expert performance in some perfect context (see for instance Simon, 1988; Grant and Mayes, 1991; Booth, 1991). This idealises real-life conditions of task performance.

User error can point to problems in human-system interaction that need to be resolved in order to enhance the system's usability. Human error taxonomies aid the prediction and detection of error classes. They can thus be exploited for error prevention and recovery mechanisms (Reason, 1990; Taylor, 1988). Those can then be incorporated into the interface design.

On the other hand, stand-alone human error theories highlight possible sources of erroneous performance without providing a language in which to express these error tendencies when applied to human cognitive task performance. Section 4 and 5 use a cognitive architecture as a vehicle for expressing not only expert task performance but also the more realistic error-prone thought and action sequences processed by the human operator. By doing this, the error modelling capability implicit in the comprehensive ICS cognitive architecture is made the focus of inquiry into the underlying cognition of user performance. Such explicit modelling of erroneous performance can thus help to communicate user cognition analyses, and to ground design decisions in a cognitive theoretical framework.

As a running example, error modelling will be applied to tasks concerning the use of Netscape Navigator™. This example is appropriate because it represents a mass-market application where errors frequently lead to high levels of frustration during common tasks (Johnson, C., 1997).

## 1.2. Interacting Cognitive Subsystems (ICS) and GEMS

We will use Interacting Cognitive Subsystems (ICS) (Barnard and May, 1993) to illustrate the modelling of human error within a cognitive architecture. ICS provides a comprehensive account of human cognition. It has proved powerful in explaining cognitive phenomena such as the stability of users' mental models during dual task interference effects (Duke, et al. 1995). It has been applied to real-life systems and tasks, such as cinematography (May and Barnard, 1995a).

Alternative cognitive user models, such as Task Analysis for Knowledge based Descriptions (TAKD) (Johnson, P. et al., 1994), User Action Notation (UAN) (Hartson et al., 1990), or Soar (Newell, 1990) might have been used. However, they lack the level of detail in ICS's representation of cognitive processes, or, in the case of Soar, the inherent constraints these have to satisfy (Wilson et al., 1988; Kjaer-Hansen, 1995). ICS was designed to provide a theoretical framework within which to place user cognition. It attempts to "satisfy the need for applicable theory" (Barnard, 1987). ICS, therefore, bridges the gap between theory-oriented cognitive architectures and task-oriented cognitive user models (Grant and Mayes, 1991; Simon, 1988).

As noted in Section 2, Reason's taxonomy of human error (Reason, 1990) represents a conceptual classification of error, as opposed to a contextual or a behavioural one. The latter, exemplified for instance by Hollnagel's (1991) classification of error phenotypes, does not lend itself to the in-depth analysis of the underlying cognitive sources of error. For instance, a behavioural error category might include errors that exhibit the same surface characteristics without sharing the same cognitive basis.

## 1.3. The Netscape Navigator Case Study

According to user population estimates, the Internet is gaining roughly 150,000 new users per month, joining 20 million existing Internet users (Pitkow and Recker, 1994). Internet browsers facilitate global communication by providing supporting hypertext navigation. Familiarity with such browsers, and therefore their usability constitutes a prerequisite for taking part in this novel information exchange. Maximising this usability therefore represents a continuous concern for designers of successively modified versions of Internet browsers. The Netscape Interface (see Figure 1) will be used for illustration throughout this report.

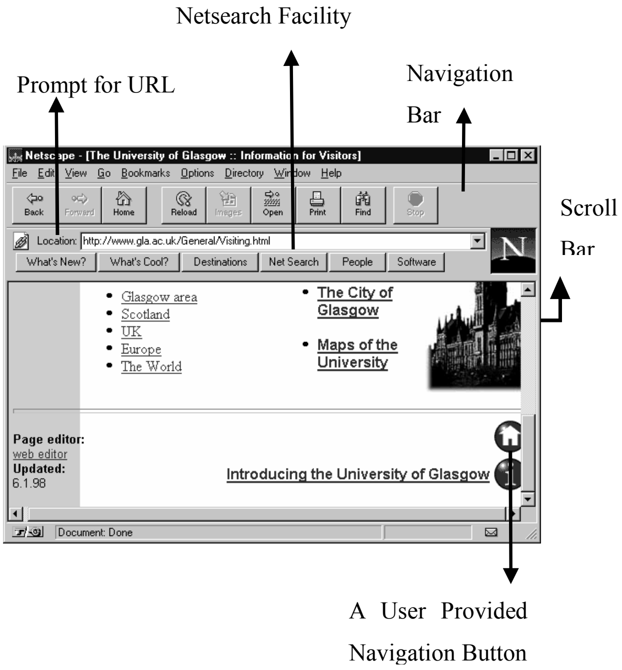


**Figure 1** - The Netscape Internet Browser

## 2 A COGNITIVE ARCHITECTURE AND A HUMAN ERROR MODEL

This section reminds the reader Barnard's ICS model and Reason's human error taxonomy introduced earlier. ICS will provide the framework in which the representation of erroneous operator interaction can be placed.

| | | |
|---|---|---|
| **Sensory subsystems:** | | |
| **VIS** | visual: | hue, contour etc. from the eyes |
| **AC** | acoustic: | pitch, rhythm etc. from the ears |
| **BS** | body-state: | proprioceptive feedback |
| **Effector subsystems:** | | |
| **ART** | articulatory: | subvocal rehearsal & speech |
| **LIM** | limb: | motion of limbs, eyes etc. |
| **Structural subsystems:** | | |
| **OBJ** | object: | mental imagery, shapes etc. |
| **MPL** | morphonolexical: | words, lexical forms |
| **Meaning subsystems:** | | |
| **PROP** | propositional: | semantic relationships |
| **IMPLIC** | implicational: | holistic meaning |

**Figure 2** - The Cognitive Subsystems

## 2.1. Interactive Cognitive Subsystems (ICS)

Cognition is represented in ICS as the flow of information between a number of different subsystems, and the processing performed on this data. Each of the subsystems has associated with it a unique mental code in which it represents the information it receives and processes. It will transform its data output into the corresponding mental code of the subsequently receiving subsystems. Each subsystem can receive several input streams and achieve a blending of these data streams under certain circumstances as described below (May and Barnard, 1995a). Each subsystem also has at its disposal a local image store. This serves as an episodic memory buffer of infinite size. A copy of any input the subsystem receives will automatically be copied to the local image store, before being further processed.

The nine subsystems can be grouped into four categories. Figure 2 above presents an overview.

### 2.1.1. Modelling a Netscape Task in ICS

Figure 3 illustrates how the error-free performance of a task of locating an object (an up-arrow, such as shown in the visual subsystem) is modelled in ICS in terms of information flow between the subsystems, and thus the different resources that are employed. Visual information concerning the target arrives at the visual subsystem and is copied into the local store. It is then transformed into object code (1). The propositional subsystem has generated a representation of the target of the location task (by conferring with its local buffer) and transforms this into object code (2). This is sent to the object subsystem, and can there be blended with the incoming structurally encoded visual information (3). The matching representation can be sent back to the propositional subsystem – the target has been located.

Thus, Figure 3 illustrates how human mental processing underlying error-free performance can be represented within ICS. In the case of erroneous performance, however, usability designers might resort to an error classification scheme in order to analyse this particular instance of user behaviour. The following section will introduce the relevant aspects of Reason's taxonomy. We will then go on to show how a more detailed, cognitive analysis can be based on initial error classification, and thus provide a further perspective on user behaviour.

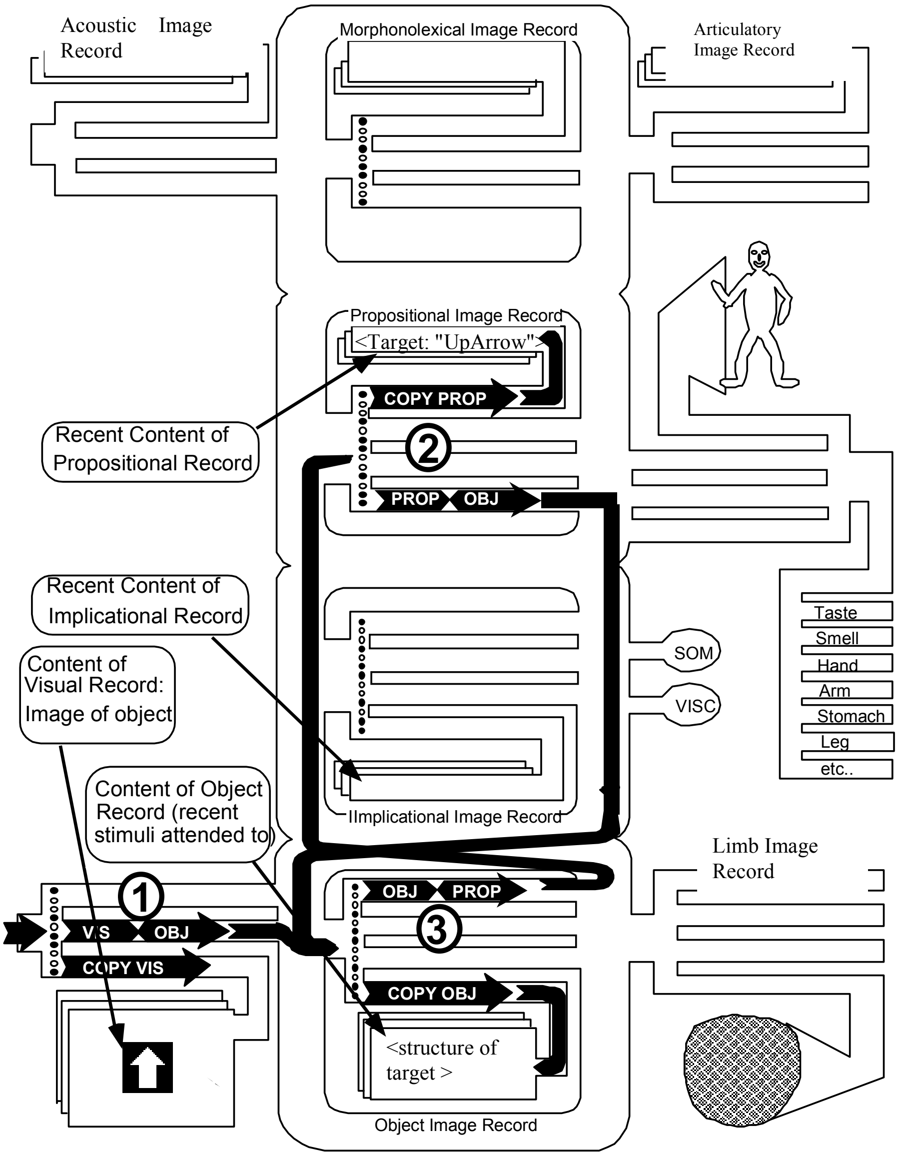


**Figure 3 -** Processing associated with the task of locating an icon on Netscape

## 2.2. Reason's Error Taxonomy

Reason (1990) investigated the more general underlying error production mechanisms within human cognition and produced a conceptual classification of error types. He bases his error classification skill-based slips and lapses on the one hand, and rule- and knowledge- based mistakes on the other (see also Norman, 1981, and Rasmussen, 1983).

Reason furthermore asserts that instances of his three basic error types are indirect results of what he calls the ‘underspecification’ of cognitive operations. In case of an ambiguity of the situational requirements, the cognitive system defaults to contextually appropriate, high frequency responses. This idea of default assignments features in most other cognitive theories, such as Bartlett’s (1932) theory of schemata, and is well backed up by empirical evidence.

ICS provides for this cognitive principle by referring to the depository role of image records attached to the individual subsystems. Thus ambiguous external input is complemented by internal input. In this way, ICS can be used to examine Reason’s elementary concept of cognitive underspecification.

### 2.2.1. Skill-based Slips and Lapses

Slips and lapses are error types that these manifest themselves as actions or states that deviate from the current intention due to execution failures (slips) and/or storage failures (lapses). Slips and lapses are observed at the skill-based level of performance, and originate from either the omission of attentional checks (inattention) during the routine action sequence or making an attentional check at an inappropriate moment (overattention).

A *slip* caused by inattention occurs in particular when current intention is to deviate from common practice. For instance, entering a well-known URL of a website constitutes a routine task. If the URL is changed and the user, although aware of that change, still happens to enter the old URL, then this is a typical example of an action slip.

A *lapse* might arise from what Reason calls 'Reduced Intentionality'. For instance, if selecting a link on the current site results in a considerable delay for this site to be loaded, the users might become distracted, and then experience disorientation upon facing the loading site. This can be seen as one of Reason's described reduced intentionality states, such as a 'what-am-I-doing-here' experience (see below).

Skill based errors such as these contribute to the sources of user frustration when accessing the World Wide Web (as described in more detail in Johnson, C., 1997). These errors need to be taken into account in future design decisions. Applying Reason's categorisation of error helps to identify error classes and presents a step towards dealing with the underlying usability problems of the system.

However, error taxonomies such as Reason's typically confine themselves to broad error categories such as slips and lapses. A more detailed, lower level description of such classes might aid the further investigation of its instances. Thus, the design process might be tuned more finely to the usability needs pointed to by the user error.

Cognitive modelling techniques such as ICS can provide a more precise vocabulary to augment the general descriptions of error taxonomies. Examples of this lower level modelling of classes of human error are given below.

### 2.2.2. Rule-based Mistakes

Mistakes are apparent in actions that may run according to plan, but where the plan is inadequate to achieve its desired outcome. For any task, rules must be selected by the cognitive system which describes methods to reach a given (sub)goal. The selection occurs according to certain criteria. These include best match, specificity, and rule strength. Rule strength is defined to be the number of times a rule has performed successfully in the past. Occasionally, rule strength might override the other factors resulting in misapplications of otherwise 'good' rules to inappropriate situations (what Rasmussen calls a ‘procedural trap’ – in “The Human as a System Component”, 1980).

As an example, an animated icon at the bottom of a page, near the contact information is quite often the mail-me icon (commonly found are self-folding envelopes, self-writing letters, or moving mailboxes). A corresponding rule will be formed and strengthened over several successful applications. In the case of a home-page icon being animated and located at a similar position in the screen layout, this rule might be applied and could lead to non-intended actions such as clicking on the icon when intending to mail the author of the page.

Such error classes can be predicted as increasingly adding to usability deficiencies as the use of animated icons accelerates in web page design (Nielsen, 1997). By being able to predict these errors, preventative measures can be taken and further user frustration (Johnson, C., 1997; Ramsay et al., 1998) can be curbed.

## 3 USING ICS TO EXPRESS REASON’S ERROR TYPES

In this section, we will examine more closely the modelling of errors as identified by Reason's taxonomy within the ICS architecture.

Commonly occurring errors and usability problems when interacting with Internet browsers’ interfaces gave rise to numerous design guidelines and principles[21]. Interface design issues such as the use of counter-intuitive icons and download delays are all well known to aggravate usability problems (see for instance Nielsen, 1996; Johnson, C., 1997; Ramsay et al., 1998). Rarely, however, are the errors resulting from those usability problems described in detail, or even analysed in terms of

---

[21] See for instance Yale C/AIM WWW Style Manual (URL: “http://info.med.yale.edu/caim/-manual/index.html“ current at 08.12.1997) or The Ten Commandments of HTML (URL:”http://www.visdesigns.com/design/commandments.html“ current at 08.12.1997

underlying psychological factors (Johnson, C., 1998). Expressing such errors within a cognitive model will allow us to investigate and reason about their underlying psychological causes. The model is thus used as a tool for reasoning about user error on a further, more detailed level.

## 3.1. Analysis of Errors and their Underlying Cognition

High download latency of web pages was identified as major source of frustration and decreased satisfaction with the downloading site and also as attenuating user performance (Ramsay, Barabesi and Preece, 1998; Johnson, C., 1997). For instance, as introduced above, if selecting a link on the current site results in a considerable delay for this site to be loaded, the users might become distracted, and then experience disorientation upon facing the loading site.

This disorientation can be classed as the effect of a phenomenon which Reason termed ‘Reduced Intentionality’. If a delay occurs between the formulation of an intention to do something and the time for this activity to be executed, the intention needs to be periodically refreshed. Other cognitive processes such as secondary intentions will otherwise claim the workspace resources. This mechanism can lead to lapses in the form of reduced intentionality states, the above described surprise and disorientation.

The cognitive processes underlying this scenario can be represented in ICS as shown in Figure 4.

After processing the goal hierarchy for selecting a link, the cognitive system shifts its focus back onto the current page (3 and 4). If novel external (1) and the current internal input are not coherent, and thus cannot be blended (2), a decision must be made as to which of those to accept as valid input. The longer the delay, the stronger the influence of the novel input grows, with it eventually replacing the internal propositionally influenced representation (3). The recognition of this mismatch will lead to a lapse as described above.

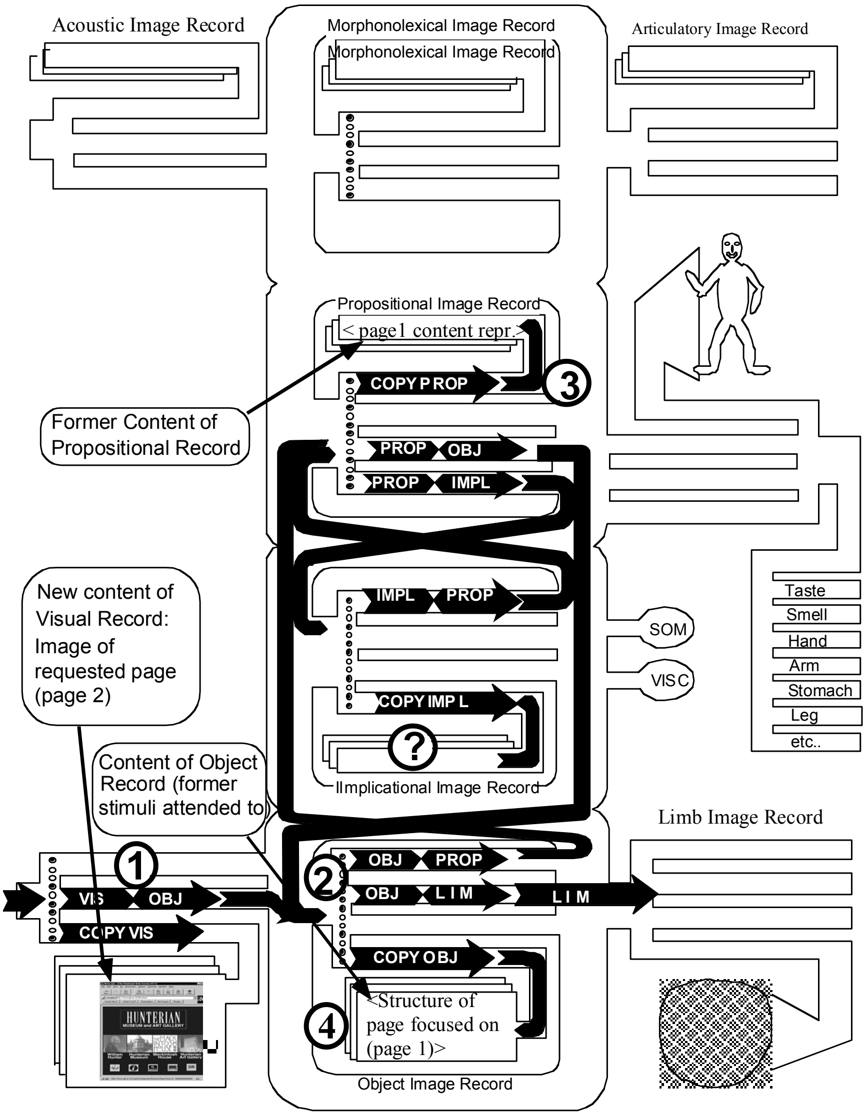


**Figure 4 -** Reduced Intentionality: A Lapse

By modelling the underlying mechanisms of manifestations of attenuated performance, such as user error, and the causes of decreased satisfaction within ICS we can shed some light on the processes fundamental to the production of the user error as mediated by the described usability problems.

## Reasoning about Alternative Analyses of Error Causes

Misinterpreting user interface icons is a common source for user error in interactive systems (Norman, 1988, 1993). However, the mistake might be grounded in varying cognitive processes, and not stem from one kind of cognitive mechanism alone.

Typically, user interface design manuals and textbooks stress the importance of intuitiveness of the icons chosen (Preece, 1994) and thus identify 'counter-intuitiveness' as a source of faulty identification of icons. However, further insight into the source of such user error can be obtained by investigating it in greater detail. As will be shown below, mistaking for instance a mail-me button with a homepage icon can be modelled in respect to two differing underlying cognitive mechanisms.

Unless these two different causes are considered these designs might misdiagnose an important problem in user utilisation of icons. Using a cognitive architecture to reason about the potential underlying cognitive error production processes allows designers to investigate the detected usability problem in a systematic way.

The above described user error could according to Reason's scheme be classified as a slip termed 'Perceptual Confusion'. In perceptual confusion, something that looks like the proper object, is in the expected location, or does a similar job is accepted as a match for the proper object. These slips could arise because, in a routine set of actions, it is unnecessary to invest the same amount of attention in the matching process. Thus acceptance criteria concerning the expected input might degrade, and result in rough and ready matches.

The processing carried out can be modelled in ICS as shown in Figure 5.

The visual data is received at the visual subsystem (1), sent to the object subsystem for the recovery of a structural description (2), and finally interpreted by the propositional subsystem (3). A loop is entered in order to maintain a stable cognition. The resulting interpretation on the propositional level influences the further view of the object. If, however, the object subsystem receives ambiguous visual information, it will make use of its local image record and fill in the assumed missing information. This principle of ICS resembles closely what Reason describes as the cognitive system's reaction to underspecification of a mental operation as described above.

The data thus acquired from the image record of the object subsystem might also fit in with the propositional interpretation of what is perceived, and thus stabilise in the cognitive system. If the assumption underlying the choice of what data is used to eliminate the underspecification is wrong, however, the representation of what is thought to be perceived will also be incorrect. The wrong icon will be chosen, and the information necessary for a mouse click sent to limb subsystem (4).

This represents one possible underlying cause of the described error. However, the same manifestation of user behaviour might also point towards a second, different underlying cognitive mechanism. Employing Reason's taxonomy, the mistaking of an icon can be classed as a perceptual slip as modelled above. On the other hand, it could also be classed as a rule based mistake. Using ICS to model the underlying cognition of the error provides a means to further investigate the behaviour trace and its associated usability problem.

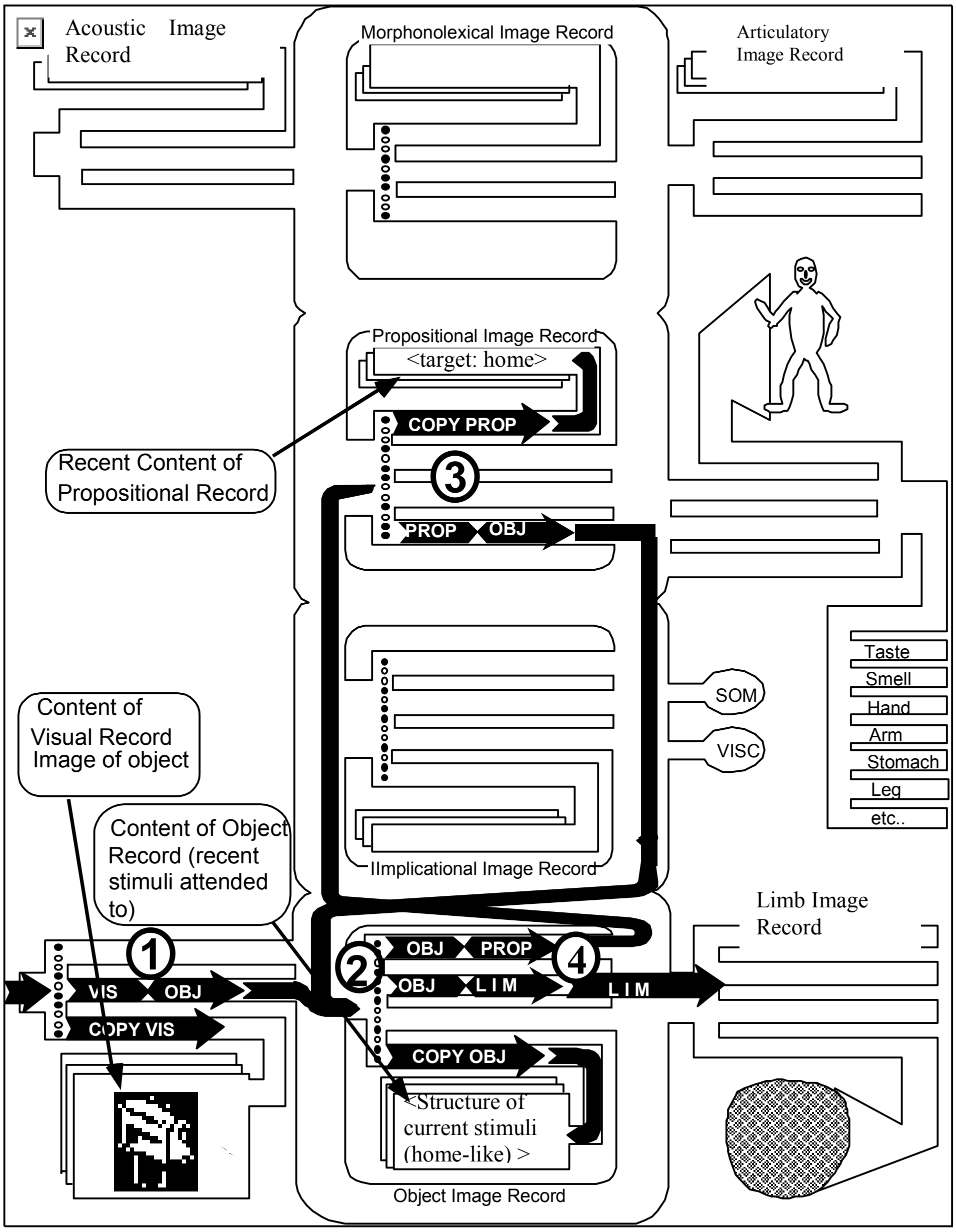


**Figure 5-** Perceptual Confusion: A Slip

Thus, the error described above could be classed as a rule-based mistake as opposed to a slip. Identifying the home-icon might well be based on rules that are utilised by the cognitive system in order to discriminate different sets of icons. Features which positively discriminate icons fulfilling one function from those fulfilling another might be listed in the set of conditions which when matched cause to fire the rule. Indiscriminative features in icons might thus lead to a rule wrongly being fired.

This can be modelled in ICS (see Figure 6) similar to the modelling approach applied to the perceptual confusion approach, but this time with the implicational subsystem playing the major role in accepting information augmented wrongly by the propositional subsystem and its local image store. Thus for the goal ‘press home button’, a subgoal hierarchy can be formulated as ‘if locate home button, move cursor to click on it’, and ‘if object has X features, it is the home button’. By mistaking the icons on a propositional level, the mail-me button might be clicked instead.

The examples elaborated above show clearly how one overt form of user error can stem from several different ‘errors’ within the cognitive processing taking place. This M:N relationship between cause and error might have gone undetected if systematic error modelling within a cognitive architecture had not taken place, this helps analysts to explicitly consider the detailed causes of usability problems.

## Generating Design Recommendations

Since underspecification proved the major source of error in the above example, once for perceptually and then for semantically discriminative features of the icon, this should be targeted by designers to remedy misidentification of icons. Thus, two functionally dissociated sets of icons should not share the same superficial perceptual features.

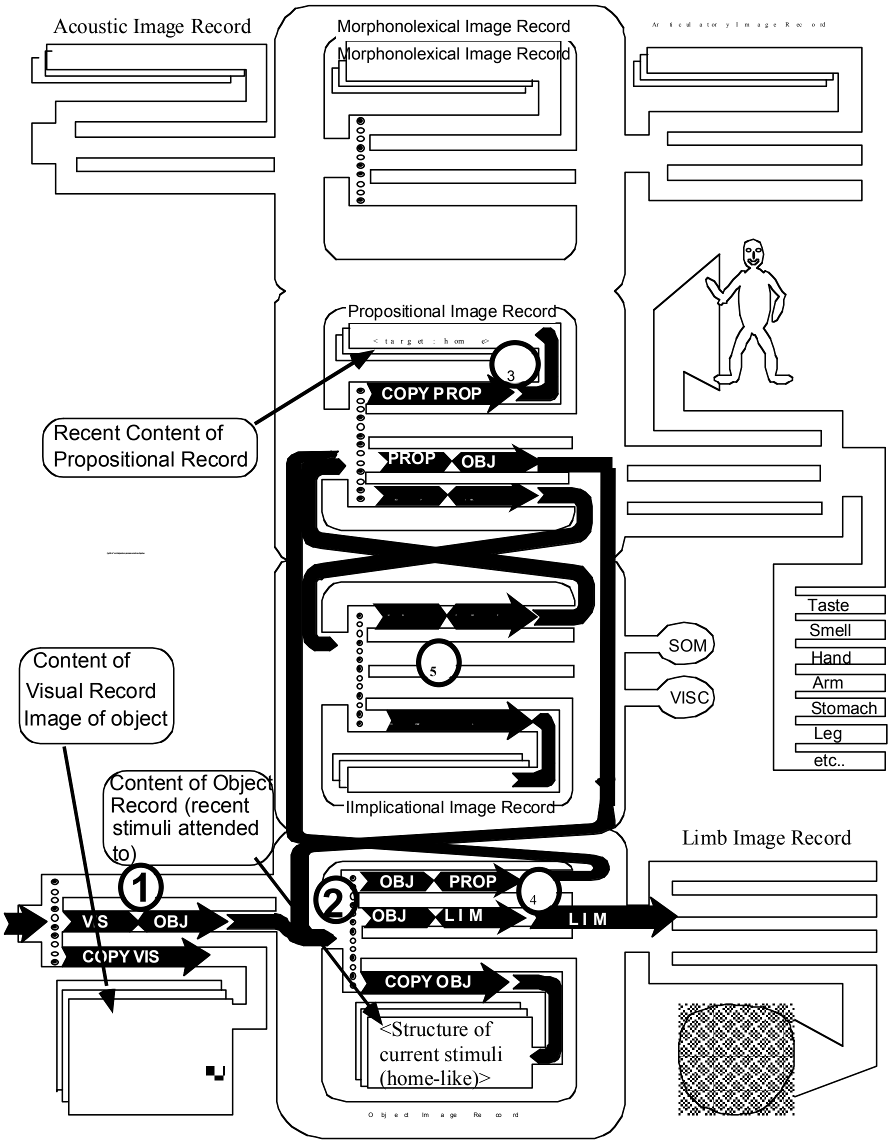


**Figure 6 -** Rule Strength: A Mistake

Features commonly used to discriminate one set of icons from another should be taken into account when designing future sets (Moyes, 1995). These feature considerations should not limit themselves to ambiguity concerning structural characteristics of icons, but also to features such as those mentioned in the examples earlier. This included as discriminative features of mail-me buttons not only their shape and internal composition, but also for instance the location of the icon on the screen, and characteristics commonly unique to mail-me buttons such as animation as present in self-folding envelopes, self-writing letters, or moving mailboxes.

The important point to highlight here is that the modelling approach described does present a method for providing a grounded rationale for design decisions, and can guide the designer in making informed choices when faced with design alternatives.

Another example of how this modelling technique can aid the generation of design decisions is introduced as this section progresses.

Johnson (C., 1997) describes how download latency of web pages affects the usability of the World Wide Web.  The effects range from user dissatisfaction with time investment to the psychological devaluation of the anticipated page (Ramsay et al., 1998). Consider the following scenario of user error resulting from download latency: After having selecting a link on the current site, a delay in downloading might lead to attention being focused on reading the current page. An intention to scroll down the page just before the new page is downloaded might lead to the scrolling action being carried out on the new page instead.

This scenario fits Reason's description of 'behavioural spoonerisms', namely slips based on interference errors. As defined above, a slip is an action that deviates from intention due to failure in the execution stage of processing operations. An interference error occurs, when two concurrent actions compete for control over cognitive processing and a transposition of actions within the same sequence takes place. For instance, intending to speak and perform an action at the same time can lead to inappropriate blends of speech and action. In our example, waiting for the new page to load, and scrolling the old page can be seen as two concurrent actions interfering and leading to an execution failure, the scrolling of the new page.

This can be modelled in ICS very similarly to the skill-based example of reduced intentionality. Only this time the focus is not on the delay but on the shift of focus back to the current page. A ‘mental model’ of the current page will be constructed (or reactivated). The unexpected appearance of the new page might lead to a blending of representation and the action included in one cognitive configuration carried out as part of a secondary one.

As a consequence, future browser designers should beware of the error-inducing character of non-interrupted browser functionality when downloading a site. Alternatively, browser functionality should only be available to the current site accessed. A clear distinction should be made when transferring functionality to the downloading site to alert users to the new context. This design flaw in Internet Browsers has not received much attention. We hypothesise that it may become increasingly important as the interweaving of the user population of the Internet grows and the World Wide Web becomes an increasingly common tool for communication and information exchange. Detailed, error-oriented cognitive analysis of such design problems can help to predict future generations of interface problems.

## CONCLUSION

Cognitive user modelling enables engineers to gain a deeper understanding of the complexities of human task performance. Current techniques typically constrain this performance to be idealised, error-free and often at an expert level. However, human error during performance represents a major source of insights into the workings and limitations of operator cognition, and therefore into usability problems.

By being based on cognitive models, the possibility of representing erroneous performance is inherent in these techniques. Few modelling techniques to date explicitly represent human error precisely, as embedded in cognitive theory.

This section showed the adoption of Reason's error taxonomy and Barnard's ICS for the systematic representation of user error within a theoretical cognitive framework. The utilisation of such a combined approach was illustrated to benefit several areas of application. User error can be described more precisely by linking it to its underlying cognition. Analysis can reach beyond surface categorisation, and it is made possible to reason about the actual causes of error. As a consequence, an informed choice concerning competing design options is facilitated. This paves the way for usability design that takes full advantage of the insights expressed in cognitive theory.

## References

Ashcraft, M.H., *Human Memory and Cognition*, 2nd ed., Harper Collins College Publishers, 1994

Barnard, P.J., Cognitive Resources and the Learning of Human-Computer Dialogs. In: Carroll, J.M., (ed.) *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, ch.6. MIT Press, Cambridge, MA, 1987

Barnard, P.J., May, J., Cognitive Modelling for User Requirements. In Byerley, P.F., Barnard, P.J., & May, J. (eds) *Computers, Communication and Usability* (North Holland Series in Telecommunication) Elsevier: Amsterdam (1993).

Bartlett, F.C., *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press, 1932

Booth, P.A., Modelling the User: User-System Errors and Predictive Grammars. In Weir G.R.S. and Alty, J.L. (eds.) *Human-Computer Interaction and Complex Systems*, Academic Press Ltd., 1991

De Keyser, V., Human Error, *Recherche*, 1989, Vol. 20, No. 216, p.1444

Duke, D.J., Barnard, P.J., Duce, D.A., and May, J., *Syndetic Modelling,* Amodeus-2 Technical Report ID/WP49, 1995

Grant, A.S. and Mayes,J.T., Cognitive Task Analysis? In Weir G.R.S. and Alty, J.L. (eds.) *Human-Computer Interaction and Complex Systems*, Academic Press Ltd., 1991

Green, R.G., Stress and Accidents, *Aviation Space and Environmental Medicine* 1985, Vol. 56, No. 7, pp.638 - 641

Hartson, H.R., Siochi, A.C., Hix, D. The UAN: A User-Oriented Representation for Direct Manipulation. *ACM Transactions on Information Systems*, 8(3), pp. 181-203, July 1990

Hollnagel, E., The Phenotype of Erroneous Actions: Implications for HCI Design. In: Weir, G.R.S. and Alty, J.L., (Eds.), *Human-Computer Interaction and Complex Systems*, Academic Press, 1991

Johnson, P., Diaper, D., and Long, J.B. Tasks, Skills, and Knowledge: Task Analysis for Knowledge based Descriptions. In: B. Shackel (ed.) *Human-Computer Interaction - INTERACT 1994*. Amsterdam: North Holland.

Johnson, C.W. Electronic Gridlock, Information Saturation and the Unpredictability of Information Retrieval Over the World Wide Web. In: Palanque, P. and Paterno, F. *Formal Methods in Human Computer Interaction: Comparison, Benefits, Open Questions.* Springer Verlag, Berlin, 1997

Johnson, C.W. *Why CHI (Computer-Human Interaction) Has Failed to Improve the Web*, at URL: "http://www.dcs.gla.ac.uk/~johnson/papers/web98.html" current at 02.02.1998

Kjaer-Hansen, J., Unitary Theories of Cognitive Architectures. In: Hoc, J., Cacciabue, P., Hollnagel, E., (Eds.), *Expertise and Technology: Cognition and Human-Computer Interaction*. LEA, Inc, 1995

May, J. and Barnard, P. (1995) Cinematography and Interface Design. In Nordby, K., Helmersen, P.H., Gilmore, D.J. and Arnesen, S.A. (eds.) *Human-Computer Interaction: Interact '95*. Chapman and Hall: London pp.26-31

Moyes, J., *Putting Icons in Context: The Influence of Contextual Information on The Usability of Icons*, PhD Thesis, Glasgow University, Computing Science Dept, 1995

Newell, A., *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press, 1990

Newell, A. and Simon, H.A., *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972

Nielsen, J. Alert Box for May 1996: *Top Ten Mistakes in Web Design* at URL: "http://www.useit.com/alertbox /9605.html" current 08.12.1997

Norman, D.A. Categorization of action slips. *Psychological Review*, 1981, 88, 1-15.

Norman, D.A., *The Psychology of Everyday Things*. New York: Basic Books. 1988

Norman, D.A., *Things that make us smart*, Reading, MA: Addison-Wesley, 1993

Pitkow, J.E. and Recker, M.M., *Results from the first World-Wide Web User Survey*, 1994, at URL: "http://www.gvu.gatech.e du/user_surveys/survey-01-1994/survey-paper.html" current 02.12.1997

Preece, J., *Human-Computer Interaction*, Addison-Wesley, 1994
Ramsay, J., Barabesi, A., and Preece, J., A Psychological Investigation of Long Retrieval Times on the World Wide Web. In: *Interacting with Computers*. Special edition on CHI and Information Retrieval, 1998, in press

Rasmussen, J. Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions: Systems, Man and Cybernetics*, 1983, SMC-13, 257-267.

Reason, J., *Human Error*, Cambridge University Press, 1990

Rouse, W.B., Morris, N.M., Conceptual Design of a Human Error Tolerant Interface For Complex Engineering Systems, *Automatica*, 1987, Vol. 23, No. 2, pp. 231-235

Simon, T., Analysing the Scope of Cognitive Models in HCI: A Trade-Off Approach. In: Jones, D.M. and Winder, R. (eds.) *People and Computers IV, Proceedings of the Fourth Conference of the British Computer Society*. Cambridge University Press, 1988

Taylor, J.R., Using Cognitive Models to make plants Safer: Experimental and Practical Studies. In: Goodstein, L.P., Andersen, H.B., and Olsen, S.E. (eds.) *Tasks, Errors and Mental Models,* Taylor and Francis, London, 1988

Wilson, M.D., Barnard, P.J., Green, T.R, and Maclean, A. Knowledge-Based Task Analysis for Human-Computer Systems. In: Van Der Veer, G., Green, T.R., Hoc, J., and Murray, D.M. (eds.) *Working with Computers: Theory versus Outcome* Academic Press, 1988

# APPENDIX B – THE EDINBURGH SCHEME (PUBLICATION)

Busse, D.K. and Wright, D.J. (2000)

**“Classification and Analysis of Incidents in Complex, Medical Environments”**

Published in the Special Edition of Topics in Health Information Management on Human Error and Clinical Systems (THIM) 20:4, May 2000.

# Classification and Analysis of Incidents in Complex, Medical Environments

## ABSTRACT

Risk management in medicine is often seen as lagging behind other safety-critical domains, where there has been considerable research into incident causation models. In this paper, incident analysis theory and methodology from outside medicine is applied to an incident reporting scheme in an Edinburgh Intensive Care Unit. The incident analysis model used emphasises the importance of latent, organisational factors and complex, multi-layered incident causation. It also takes the role of cognitive performance shaping factors into account. This provides an analysis framework that supports the identification of distal causal factors and reasoning about alternative causal hypotheses.

## INTRODUCTION

In 1990, the Harvard Medical Practice Study [1] investigated the occurrence of patient injury caused by treatment, so-called adverse events. This study found that nearly 4% of patients suffered an injury that prolonged their hospital stay or resulted in measurable disability. Leape [2] pointed out that, if these rates are typical of the US, then 180.000 people die each year partly as a result of iatrogenic ('treatment-caused') injury. Since most of the precursors to iatrogenic injuries are perceived to be 'Human Error' [3], there is an increasing interest in human involvement in adverse events [4]. The cost of adverse events is high, not only in human suffering, but also in compensation claims and the need for prolonged treatment of afflicted patients.

In order to address this problem, a number of clinical incident reporting schemes have been set up in the last decade. These look at 'critical incidents', which are typically defined as near miss adverse events, where iatrogenic injury could have occurred but did not. In these schemes, however, in-depth analysis and search for root causes of adverse events typically does not take place [2].

In safety-critical domains other than health care, research into accident causes and prevention has made considerable advances. Accident causes formerly described as ‘Human Error’ have come under close scrutiny. Notably Rasmussen [5], Reason [6], Hollnagel [7], and Hale [8] have systematically analysed the cognitive mechanisms underlying human error as well as organisational influences on the evolution of accidents. This has led to a shift from ‘blaming the human’ to the understanding that error invariably occurs in complex systems. This has encouraged the creation of error-tolerant systems, which absorb errors through ‘system defences’, and that can allow for error recovery. The design of such systems can be guided and informed by in-depth analysis of ‘human error’ data. This data can be obtained by implementing an anonymous and non-punitive incident reporting scheme.

In aviation, the US Federal Aviation Administration (FAA) established a confidential reporting system (the Aviation Safety Reporting System - ASRS) for safety infractions as early as 1975. Self-reports of incidents were collected and a closed feedback loop was ensured by disseminating the results of data analysis and suggested remedial measures. The system’s success was found to be based on its confidential non-punitive approach. In medicine, there has been a belief in a punitive ‘perfectibility model’, where there will be no mistakes if staff are properly trained and motivated [2]. Outside medicine, however, it is now widely accepted that punishment of individuals as a means to prevent future adverse events is the wrong approach to error management [9]. Thus, there is a need in medicine to recognise the inevitability of error and adverse events, and to adapt a more constructive error management approach.

Safety culture that takes this into account in clinical systems is still lacking [10]. There have been some notable exceptions in the recent past where incident reporting schemes were implemented and the identified incidents analysed. In this paper, we will look at one such example, an incident reporting scheme at the Western General Hospital (WGH) in Edinburgh.

The Edinburgh incident reporting scheme was set up in an adult intensive care unit (ICU) in 1989 [11]. The unit has 8 beds, and there are usually 1-3 junior medical staff, 1 consultant, and 8 nurses per shift. Equipment includes monitors that display information such as heart rate and blood pressure, life support equipment such as lung ventilators, and drug delivery systems such as pumps and drips. Complex tasks include recording observations and measurements, altering the settings of equipment according to changing situations, and carrying out such procedures as inserting intravascular lines, endotracheal tubes, and chest drains.

This paper reviews and analyses the incident classification and analysis process in the Edinburgh scheme as compared to the approaches in other safety-critical domains. The paper firstly describes the Edinburgh system. It then reviews the relatively simple process of incident classification and analysis that has been used in running it. Finally, the data is re-examined using theory and methodology from other safety-critical disciplines, such as aviation.

## THE EDINBURGH WGH ICU INCIDENT REPORTING SCHEME

By the 1970s, critical incident studies had been used to investigate anaesthetic mishaps [12]. Modelled on this work, ICU critical incidents have been investigated at the WGH since 1989. For data collection, a questionnaire was used that was based on one that had been developed by Williamson and colleagues [13] for use in anaesthetic practice.

A critical incident was defined to be an occurrence that might have led (if not discovered in time), or did lead, to an undesirable outcome. The incident must have been caused by an error made by a member of staff or by a failure of equipment. It had to be described in detail by a person involved in or who had discovered the incident. It had to have occurred while a patient was under the care of the ICU staff (though not necessarily in ICU). It had to be clearly preventable.

Such incident reporting schemes need a form for describing the incidents; a system for collection of the forms; classification and analysis of the incidents; a review of the incidents by senior staff that includes proposals for strategies to reduce the likelihood of future incidents; and feedback to staff that includes summaries of the incidents and the proposed preventative strategies.

The essence of such studies is anonymity and absence of criticism and these points should be constantly emphasised.

The scheme was set up in 1989 and its first year's experience was published in 1991 [11]. By early 1999, over 700 incident forms had been completed. This paper presents an analysis based on 710 reports from the period January 1989 to February 1999.

Earlier studies confirm how important human error is in the generation of critical incidents [12; 13]. Similarly, in the Edinburgh study most incidents were associated with errors by staff, only a minority being associated with equipment failure. These human errors fell into four groups (see table 1): those related to ventilation, those related to vascular lines, those related to drug administration and a miscellaneous group.

| | |
|---|---|
| Human Error | |
| Drug related | 204 |
| Ventilation related | 202 |
| Vascular line related | 94 |
| Miscellaneous | 108 |
| Equipment Related | |
| Non disposable equipment | 44 |
| Disposable equipment | 39 |
| Miscellaneous | 7 |
| Not a Critical Incident | 12 |
| Total | 710 |

**Table 1 –** Edinburgh Classification System

When the incidents were analysed and classified, a number of contributing factors could be coded from the data (table 2 lists the commonest). Factors such as inattention/thoughtlessness, inexperience, and failure to check equipment were identified. In addition, a number of factors could be noted as contributing to the detection of incidents (table 3 lists the commonest). Such factors were the presence of experienced staff, repeated regular checking, protocols, and the presence of alarms.

| | |
|---|---|
| Inattention/Thoughtlessness | 197 |
| Inexperience | 175 |
| Failure to Check Equipment | 109 |
| Poor Equipment Design | 86 |
| Poor Communication | 86 |

**Table 2 –** Commonest Contributing Factors

| | |
|---|---|
| Presence of Experienced Staff | 212 |
| Repeated Regular Checks/Protocols | 165 |
| Presence of Alarms | 86 |
| Handover Checks | 79 |
| Hearing an Unusual Noise | 11 |

**Table 3 –** Incident Detection Factors

Nurses completed just under 90% of the forms (table 4) and the vast majority of incidents recorded had no serious sequelae for patients (table 5).

| | |
|---|---|
| Nursing | 621 |
| Medical | 77 |

**Table 4 –** Person Reporting the Incident

| | |
|---|---|
| Nothing/Not Serious | 688 |
| Serious/Fatal | 10 |

**Table 5 –** Outcome of Incident

The incidents, together with suggestions for preventative strategies, were periodically summarised in draft reports. These were then reviewed by a core group of senior staff before being issued as formal reports to all staff. This feedback allowed the opportunity to change practice and attitudes.

Preventative strategies which aim at reducing the likelihood of future events have included writing new and redrafting existing protocols, emphasising the importance of handover checks, changing equipment suppliers and providing photographs of correctly assembled equipment as visual aids.

Notably, while the frequency of the equipment related incidents has reduced markedly over the 10 years, the type of incident related to human error has changed little over this time (the frequency of the two commonest fields of errors, ventilation related and drug related incidents being relatively constant).

## REVIEW OF THE EDINBURGH INCIDENT CLASSIFICATION

In accident causation research, the categorisation and analysis of ‘human error’ has received considerable attention. Several taxonomies exist, ranging from behavioural classifications [7] to classification of error according to underlying cognitive mechanisms [6]. Behavioural analysis is based on the consideration of human error in terms of their external manifestation (such as omission, commission, and inappropriate timing). It does not provide a link to the cognitive origin of error types, but adopts a purely descriptive approach. Thus, behavioural classifications do not aid the understanding of the underlying mechanism of erroneous actions. The distinction between varieties of human error according to their cognitive origin, however, plays a significant role in accident analysis. Different types of error require different methods of error management and remediation [14; 15].

Rasmussen's error modelling framework [16] has been widely applied in safety-critical domains. Rasmussen sees the behavioural manifestation of human error as 'External Mode of Malfunction', which is brought about by a combination of task and work environment factors, such as time pressures, mental load, and equipment design. These Performance Shaping Factors (PSF) [17] impact on task performance and produce erroneous behaviour that, in turn, might lead to an accident.

The Edinburgh classification in its current form represents a hybrid system, combining PSFs and behavioural categories. The initial classification system [11] had covered 12 categories (see table 6). Throughout the evolution of the reporting scheme, the collected incident data was coded into novel categories to fit the current requirements. Thus, 11 further categories were created through informal coding of the narrative incident data.

In the initial taxonomy, the 'contributing factors' represented mainly Performance Shaping Factors (such as Fatigue, Unit Busy, and Night Time). Poor Communication could also be considered a PSF. The one factor notably not a PSF was Thoughtlessness. The only factor that clearly denoted a latent failure was Poor Equipment Design.

The categories added since the initial scheme (see table 6) are of a notably different type. The ongoing coding approach led to a predominantly domain-specific, behavioural classification scheme. The revised categories do not reflect potential underlying cognitive mechanisms, nor organisational influences. Mostly, they denote where in the task sequence a step had been omitted, or had been executed wrongly (e.g. Turning the Patient).

| Initial Categories | Added Categories |
|---|---|
| • Inexperience with equipment<br>• Shortage of trained staff<br>• Night time<br>• Fatigue<br>• Poor Equipment Design<br>• Unit Busy<br>• Agency nurse<br>• Lack of Suitable equipment<br>• Failure to check equipment<br>• Failure to perform hourly check<br>• Poor Communication<br>• Thoughtlessness | • Presence of students/teaching<br>• Too many people present<br>• Poor visibility/position of equipment<br>• Grossly obese patient<br>• Turning the patient<br>• Patient inadequately sedated<br>• Lines not properly sutured<br>• Intracranial Pressure Monitor not properly secured<br>• Endotracheal tube not properly secured<br>• Chest drain tube not properly secured<br>• Nasogastric tube not properly secured |

**Table 6 –** Edinburgh Classification of Contributing Factors

These categories can provide the basis for descriptive statistics on reported incidents. They do not help in the analysis of the underlying causes of incidents, since only a superficial, descriptive account of the incident is given. In the next section we will demonstrate an approach to a more in-depth analysis of ‘human error’ and incidents.

# INCIDENT ANALYSIS

Incident analysis ought to uncover *what* happened in an incident occurrence (e.g. wrong drug administered), *how* it happened (e.g. drug confusion), and, most importantly, *why* it happened (e.g. illegible handwriting on drug container; unusual storage location; similarity of appearance). Current incident analysis approaches often only record the ‘what’ and ‘how’ of an incident occurrence [18; 19; 20]. In the Edinburgh system, this is exemplified by the classification of incidents focussing on ‘proximal causes’ rather than on possible ‘root causes’, and by employing a behavioural classification.

Furthermore, research in other disciplines has shown that incidents are typically not caused by a single, unique factor, but by a concatenation of conditions and events [21].

## Multiple Levels of Causation

According to Rasmussen, incidents and accidents are occurrences of a human-system mismatch, which can only be characterised by a multi-faceted description. Thus, faults and errors cannot be sufficiently defined by considering only a single ‘cause’. Any incident occurrence is precipitated by a number of factors that can be organised in a ‘causal tree’ [17; 9; 22].

In such a causal tree, the incident is preceded by several levels of causal factors organised in a hierarchy. The factors immediately preceding the incident are seen as ‘proximal’, whereas those further removed from the actual incident occurrence are seen as ‘distal causal factors’.

Reason [6] introduced the related concepts of ‘active failure’ and ‘latent failure’ to express the multi-level nature of incident causation. Reason maintains that active failure is usually associated with the performance of ‘front-line’ operators (such as pilots, or nurses) and has an immediate impact upon the system. Latent failure is most often generated by those at the ‘blunt end’ of the system (equipment designers, decision-makers, managers, etc.) and may lie dormant for a long time.

In the Edinburgh study, the classification of events involves a non-formalised classification and analysis process. Classification takes place according to what was identified as the main contributing factor to the incident occurrence. However, as noted above, the categorised data often seems to present behavioural descriptions of proximal ‘causes’.

Following Reason’s accident causation and analysis model (see figure 1), we classified the ‘contributing factors’ into latent failure types and work conditions failure types (intermediate/distal causal factors), and active failures (proximal causal factors), as shown below in table 7.

## Multiple Contributing Factors

It is argued that the categorisation of ‘causes’, or precursors, of the incident needs to reflect the hierarchical nature of the presumed causal chain. This chain involves latent system failures that influence work conditions, which in turn provide the context for the proximal incident cause to occur.

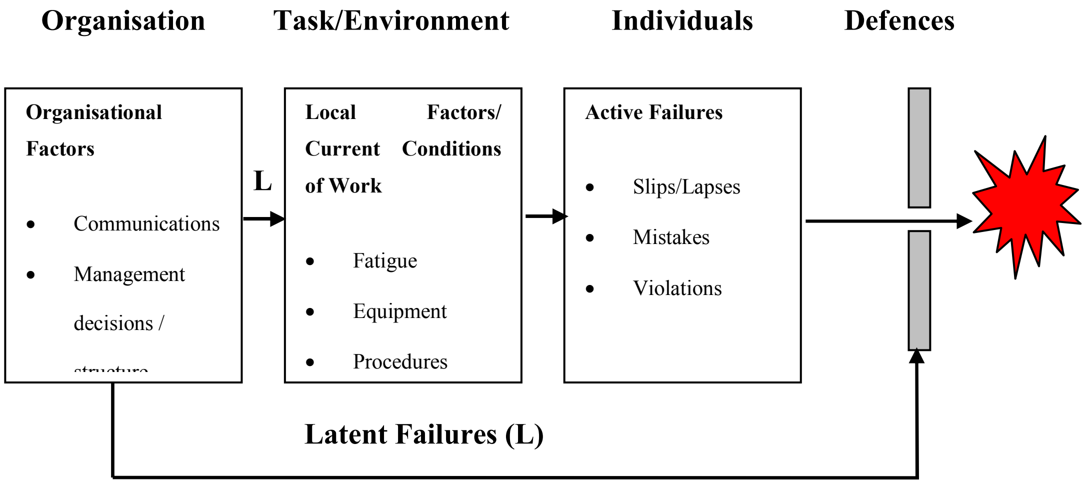


**Figure 1 -** Reason's Organisational Accident Model

(adapted with kind permission from Reason, J., Managing the Risks of Organizational Accidents, 1997, © Ashgate Publishing Ltd.)

Rasmussen's model also reflects this hierarchy of incident causation. Most of his PSF can be seen in relationship to Reason's latent failures. This perspective on accident causation is mirrored in most accident causation and analysis models.

For instance, the National Transport and Safety Board (NTSB) accident/incident database uses a multiple-cause classification system, and allows for a causal tree, or hierarchy, being constructed. In this classification system, ‘causal factors’, ‘conditions facilitating the occurrence’, and ‘suggested root causes’ can all be distinguished. Still, the classification still often stops short of an assessment of *why* the error was made. Even the probable cause statements are largely descriptive in nature, without reference to latent failures, or how the ‘error’ is related to human information processing constraints. Without such information, it is difficult to develop preventive strategies [23].

| Proximal Causal Factors | Intermediate and Distal Causal Factors |
|---|---|
| 1. Failure to check equipment | 1. Inexperience with equipment |
| 2. Failure to perform hourly check | 2. Shortage of trained staff |
| 3. Thoughtlessness | 3. Night time |
| 4. Turning the patient | 4. Fatigue |
| 5. Patient inadequately sedated | 5. Poor Equipment Design |
| 6. Lines not properly sutured | 6. Unit Busy – refined by #9, 10 |
| 7. ICP monitor not properly secured | 7. Agency nurse |
| 8. En. tube not properly secured | 8. Lack of Suitable equipment |
| 9. Chest drain tube not properly secured | 9. Presence of students/teaching |
| 10. Nasogastric tube not properly secured | 10. Too many people present |
| | 11. Poor visibility of equipment |
| | 12. Poor Communication |
| 0. Grossly Obese Patient | |

**Table 7** - Failure Type Categorisation

In the Edinburgh system, multiple factors may be combined in the classification of the contributing factors, and combinations of factors frequently occur in the data sample. The data analysis might thus also be based on noting the frequency or likelihood of certain factors being correlated with one another, rather than on one factor's occurrence only. In that way, conclusions can be drawn that are justified by a richer data set. To date, however, this has not been the focus of the Edinburgh categorisation, and such benefits of multi-causal categorisation go as yet unheeded.

Associating multiple contributing factors with one incident occurrence embeds behavioural factors within latent system and work condition factors. Thus, factors that warrant further explanation (such as Thoughtlessness) can be placed in context by considering the more diverse, multiple categorisation. A single factor classification neglects this facet of analysis and therefore sheds a misleading light on the collected incident data.

For instance, a closer look at the data reveals that a significant proportion of Thoughtlessness incidents are not single factor categorisations, but are placed in combinations with Fatigue, Inexperience With Equipment, Poor Communication, and Unit Busy.

To illustrate multi-causal categorisation, we will consider some commonly occurring ‘contributing causes’ in more detail in the following section.

## Causal Tree Analysis

The US Department of Energy developed a diagrammatical root cause analysis method [24] for the investigation of industrial incidents. ‘Root causes’ are defined as the causal factors that, if corrected, would prevent recurrence of the incident. Root causes are derived from and generally encompass several ‘contributing causes’. These are higher-order, fundamental ‘causal factors’ that address classes of deficiencies, rather than single problems or faults. Root causes can include system deficiencies, management failures, performance errors, and inadequate organisational communication. The incident causation factors are then listed in a tree-shaped diagram, which guides the data collection, interpretation, and analysis. Thus, root cause analysis provides a ‘tool for thought’ and a documentation means for causal chain analysis. The final aim is to identify management weaknesses, or in Reason's terminology, latent organisational failures.

Currently, the Edinburgh study does not proceed much beyond the “what and how” stage of incident analysis [25]. Given the contributing factors classification above, root cause analysis can be used to reflect the variety of levels in the incident causation tree. Multi-causal categorisation can then be used to reconstruct a root cause analysis. For instance, the common ‘contributing factor’ Inexperience With Equipment is often combined with several other factors in the data classification. This can be illustrated in a causal tree as shown in figure 2.

The incident (e.g. ventilator related) was ‘caused’ by staff Inexperience With Equipment. This was then hypothesised to be mediated by Lack Of Suitable Equipment, which in turn pointed to Poor Equipment Design. Thus, a causal chain is established and documented. During root cause analysis, it could also be argued that Lack Of Suitable Equipment contributed directly to the occurrence of the incident. This hypothesis, again, has different implication for potential system redesign. Rather than focusing solely on targeting staff inexperience (by training, or worse, by blaming the individual’s motivation), the provision of suitable equipment ought to be considered from an organisational and managerial point of view.

Thus, this kind of analysis takes several levels of causation into account and aids reasoning about the ‘root causes’ of incident occurrence. Other factors that could be considered in lower levels of causation include work conditions, latent system failures, and alternative ‘causes’ and contributing factors (see figure 2). Rather than being a causal category, ‘human error’ should be seen as representing a symptom, and a starting point for investigation [26].

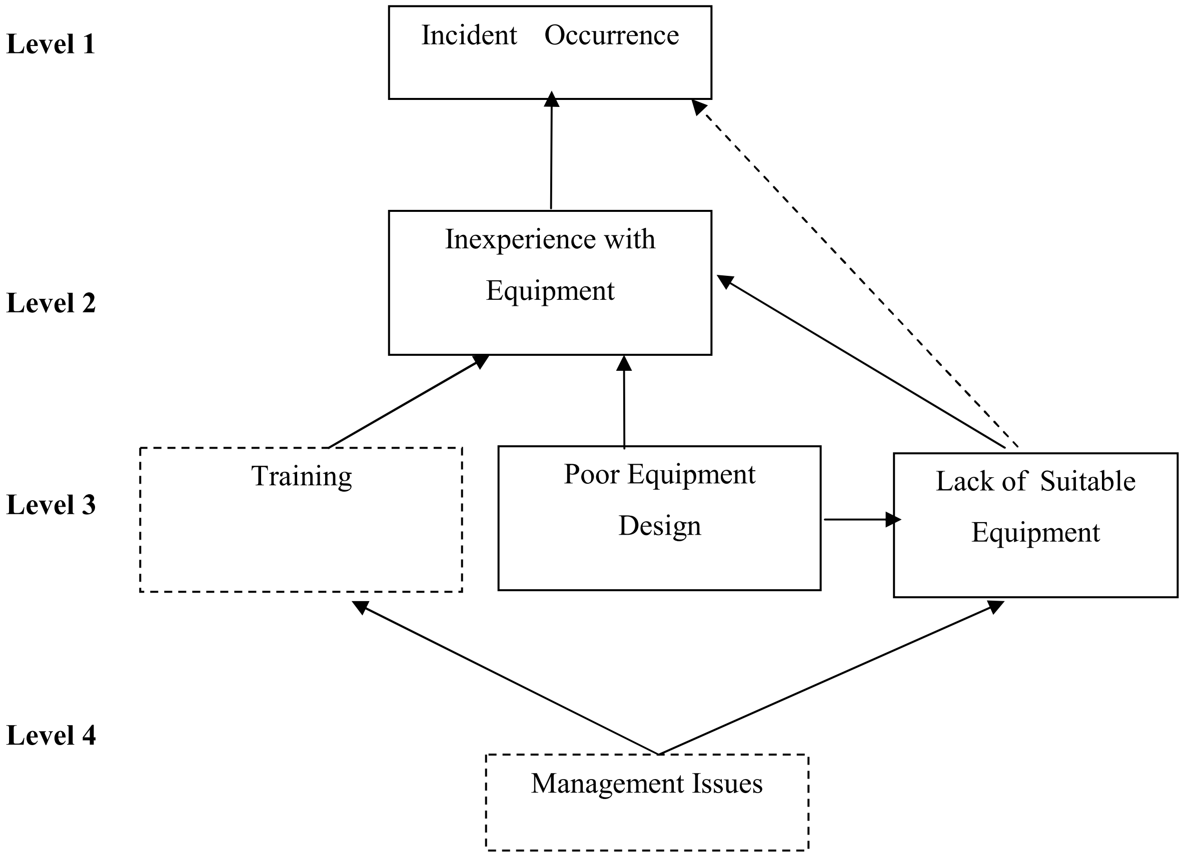


**Figure 2** - Incident Root Cause Analysis I

Furthermore, the classification of incidents might be influenced by human thought biases such as the Fundamental Attribution Error [27]. This bias leads us to favour a situational perspective when explaining our own mistakes, but to overestimate personal factors, such as inattention, clumsiness, or thoughtlessness, when explaining others' behaviour. Clearly, this cognitive bias can exert an undesirable influence on error analysis, warranting a structured, formalised analysis approach. For instance, Inexperience represents a personal factor that was listed as one of the commonest ‘contributing factors’ in the incident data of 10 years (see table 2). Above we showed how this factor can be put into perspective by employing a formalised analysis approach that embeds personal factors in system factors such as the design or availability of equipment.

The most common ‘contributing factor’ overall in the Edinburgh scheme is Thoughtlessness (see table 2). This represents a personal factor that on its own does not present much of a starting point for the analysis of latent factors potentially underlying the incident. The consideration of the other factors that were typically associated with Thoughtlessness, however, emphasises the benefits of multi-causal classification, as shown in figure 3.

Figure 3 illustrates one common constellation of factors preceding an incident that involved Thoughtlessness. A Failure to Check Equipment led to these types of incident. Other contributing factors listed were Unit Busy, Night Time, and Thoughtlessness. The Edinburgh classification does not indicate the main contributing factor, or which factors are considered proximal or distal causal factors. There are several different interpretations of such a list of multiple contributing factors: Thoughtlessness could be considered the main causal factor; alternatively, Thoughtlessness could be considered a negligible causal factor, with PSFs (such as Night Time) being of prime importance in the incident’s causation. In the Edinburgh classification system, there is no inherent way of expressing relative impact of different causal factors, or of reasoning about these alternative causal hypotheses. A causal tree analysis can show the factors’ relative importance, whether they are considered proximal or distal, and thus it can be used to evaluate contrasting causal hypotheses.

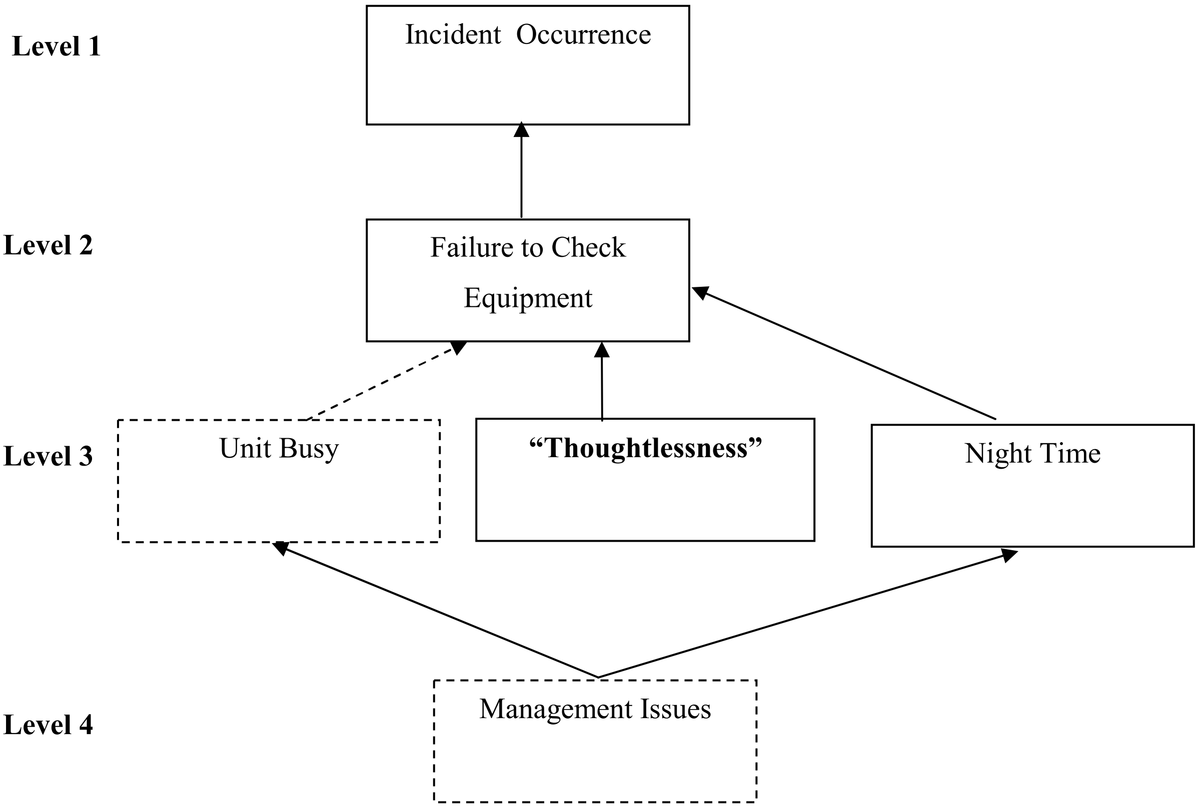


**Figure 3** - Incident Root Cause Analysis II

Figure 3 also demonstrates how descriptive statistics that are based on single factor occurrences can be misleading in that they misrepresents a single factor’s contribution to the incident. More importantly, a causal tree analysis such as shown in figure 3 can also detail intervention points for preventative strategies. Since Thoughtlessness as causal factor on its own can seemingly only be targeted by reminding staff to be more thoughtful, the tree in figure 3 points towards alternative intervention points, such as staff fatigue at night time, or a busy unit. These can then specifically be addressed by strategies that target lower levels of causation, such as management issues.

## CONCLUSIONS

We have illustrated in this paper how research into incident causation from safety-critical domains other than medicine and its resulting understanding and methodology can be applied in a clinical setting.

In many medical incident reporting schemes, in-depth analysis and a search for the root causes of adverse events does not take place [2; 28]. Thus, although active failures may be noted, the potentially more important latent failures may not be identified. In this paper, we applied a theory-driven analysis framework that supported the identification of latent factors. It also supported the reasoning about alternative hypotheses on incidents' causation.

Incident investigation schemes often neglect formalised, in-depth analysis of single incidents in favour of a quantitative surface analysis. Stanhope et al. [29] suggested that a more systematic approach dealing with a smaller number of cases in more depth is likely to yield greater dividends in understanding incident causation and generating action recommendation than the 'many' cases currently analysed quite briefly and hence less effectively. We have shown in this paper how in-depth analysis of a few types of incidents can lead to valuable insights.

## Acknowledgements

This work was supported by UK EPSRC Grant No GR/L27800.

## REFERENCES

1. Harvard University. Medical Practice Study: Patients, Doctors, and Lawyers: Studies of Medical Injury in New York. (Harvard University Press, Boston, 1990).

2. Leape, L. L. Error in Medicine. JAMA: The Journal of the American Medical Association 272, 1851-57 (1994).

3. Runciman, W. B. et al. Errors, Incidents, and Accidents in Anaesthetic Practice. Anaesthesia and Intensive Care 21, 506-519 (1993).

4. Bogner, M. S. (ed.) Human Error in Medicine (Lawrence Erlbaum Associates, 1994).

5. Rasmussen, J., Duncan, K. & Leplat, J. (eds.) New Technology and Human Error (John Wiley & Son Ltd., Berlin, 1987).

6. Reason, J. Human Error (Cambridge University Press, 1990).

7. Hollnagel, E. The Phenotype of Erroneous Actions: Implications for HCI Design. In Human-Computer Interaction and Complex Systems (eds. Weir, G. R. S. & Alty, J. L.) (Academic Press Ltd., London, 1991).

8. Hale, A., Wilpert, B. & Freitag, M. After The Event - From Accidents to Organisational Learning (Pergamon Books, 1997).

9. Reason, J. Managing the Risks of Organizational Accidents (Ashgate Publishing Limited, 1997).

10. Leape, L. L. et al. Promoting Patient Safety by Preventing Medical Error. JAMA: The Journal of the American Medical Association 280, 1444-47 (1998).

11. Wright, D., MacKenzie, S. J., Buchan, I., Cairns, C. & Price, L. E. Critical Incidents in the Intensive Therapy Unit. The Lancet 338, 676-78 (1991).

12. Cooper, J. B., Newbower, R. S., Long, C. D. & McPeek, B. Preventable Anaesthesia Mishaps: A Study of Human Factors. Anaesthesiology 49, 399-406 (1978).

13. Williamson, J. A., Webb, R. K. & Pryor, G. L. Anaesthesia Safety and the “Critical Incident” Technique. Aust Clin Rev, 57-61 (1985).

14. Maddox, M. E. & Reason, J. Human Factors Guide for Aviation Maintenance (Federal Aviation Administration/Office of Aviation Medicine, Washington, DC, 1996).

15. Busse, D. K. & Johnson, C. W. Human Error in an Intensive Care Unit - A Cognitive Analysis of Critical Incidents. In 17th International System Safety Conference (ed. Dixon, J.) (Florida Chapter of the System Safety Society, Orlando, FL, USA, 1999).

16. Rasmussen, J. The Definition of Human Error and a Taxonomy for Technical System Design. In New Technology and Human Error (eds. Rasmussen, J., Duncan, K. & Leplat, J.) (John Wiley & Sons Ltd, 1987).

17. Rasmussen, J. Human Error: a Taxonomy for Describing Human Malfunction in Industrial Installations. Journal of Occupational Accidents 4, 311-333 (1982).

18. Eagle, C. J., Davies, J. M. & Reason, J. Accident Analysis of large-scale technological disasters applied to an anaesthetic complication. Canadian Journal of Anaesthesia 39, 118-122 (1992).

19. Bogner, M. S. The “Why” of Error in Clinical Systems. In Human Error and Clinical Systems (ed. Johnson, C. W.) (Glasgow, UK, 1999).

20. O'Hare, D., Wiggins, M., Batt, R. & Morrison, D. Cognitive Failure Analysis for Aircraft Accident Investigation. Ergonomics 37, 1855-1869 (1994).

21. Wagenaar, W. A. & Groeneweg, J. Accidents at Sea: Multiple Causes and Impossible Consequences. International Journal of Man-Machine Studies 27, 587-598 (1987).

22. Johnson, C. W. Visualizing the Relationship between Human Error and Organizational Failure. In International Systems Safety Conference (Orlando, FL, USA, 1999).

23. Luxhoj, J. T., Arendt, D. N., Williams, T. P. & Horton, T. G. An Application of Advanced Information Technology for Assessing Aircraft Accident Causation. In The International Society of Air Safety Investigators (1997).

24. Department of Energy Root Cause Analysis Guidance Document. (US Department of Energy, Washington, D.C., 1992).

25. National Transportation Safety Board Revised Accident/Incident Database Coding Manual (Washington, DC, 1998).

26. Woods, D. D., Johannesen, L. J., Cook, R. I. & Sarter, N. Behind Human Error: Cognitive Systems, Computers, and Hindsight (ed. Schopper, A. W.) (CSERIAC Crew Systems Ergonomics Information Analysis Center, 1994).

27. Jones, E. E. & Ellsworth, E. Attribution: perceiving the causes of behaviour (General Learning Press, Morristown, N.J, 1972).

28. Chappell, S. L. Using Voluntary Incident Reports for Human Factors Evaluations. In The International Society of Air Safety Investigators (1996).

29. Stanhope, N., Vincent, C., Taylor-Adams, S. E., O'Connor, A. M. & Beard, R. W. Applying Human Factors Methods to Clinical Risk Management. British Journal of Obstetrics and Gynaecology 104, 1125-1232 (1997).

# APPENDIX C – THE GLASGOW SCHEME (PUBLICATION)

Busse, D.K. and Holland, B. (in press)
**“Implementation of Critical Incident Reporting in a Neonatal Intensive Care Unit”**
To be published in the Special Edition of the Journal of Cognition, Technology, and Work (CTW), to appear 2002.

# Implementation of Critical Incident Reporting in a Neonatal Intensive Care Unit

## ABSTRACT

The occurrence of medical adverse events is a growing cause for concern worldwide. Critical incident reporting schemes have recently been suggested as an effective means to tackle the problem of medical adverse events. There are few comprehensive frameworks that accommodate the specific requirements of local settings as well as generic issues in incident reporting. The local setting radically influences a scheme's successful implementation and maintenance. Major issues that impact the overall success of incident reporting schemes concern the format of data collection, and especially, a meaningful data analysis. This paper reports on the introduction of a critical incident reporting scheme to a neonatal intensive care unit (NICU). Issues concerning the implementation and maintenance of the reporting scheme are discussed. Incident analysis is described in terms of the process and the results of incident categorization. The implications of such a viewpoint are considered.

## 1. INTRODUCTION

Adverse events in medicine have become a major public concern in recent years, after hospital tragedies such as the Bristol Baby Case in the UK (Kennedy et al. 2000). This spurned scientific studies into the nature of medical adverse events, their incidence, and whether their characteristics contain any clue as to how they can be remedied (Leape, 1994). Both the UK and US national governments have acknowledged the issue and have called into being organizations such as the US National Patient Safety Foundation, and the UK National Patient Safety Agency. The need to reduce ‘human error’ in medicine has thus been publicly recognized, and there is an urgent demand to implement safety measure that can tackle the problem effectively.

Critical incident reporting schemes have been cited as a major safety tool to combat human error and adverse events in medicine. These kinds of schemes are established in safety-critical domains such as aviation and process control as complementary safety management tools. Incident reporting schemes facilitate the collection and archiving of data concerning critical incidents, or near-misses, that have occurred in those industries. Typically, a distinction can be drawn between standardized, national schemes of broad scope but less depth, and smaller, local schemes. Local schemes permit close scrutiny of in situ adverse events. Their implementation can also be fine-tuned to the local culture and conditions, a prerequisite for successful delivery of the scheme. This is at the expense of the benefits of statistical evaluation of the collected data that a more extensive, standardized data set offers. Both types of scheme operate by presenting employees with a data collection form, which prompts for a number of characteristics of the problem description. The nature of the form varies, but there are some fundamentals common to most schemes: general circumstances of the adverse events need first to be established (e.g. at what time did the incident occur?), as well as general facts about the reporters themselves (e.g. how experienced they are). Typically, this is followed by asking for a narrative description of the incident, and includes questions about the presumed ‘causes’ and contributing factors, and how the incident was detected. This is question is not typically included, but is clearly valuable for future incident avoidance. This data can then be used to instigate further, more in-depth investigation of the incident. It can also be categorized and archived for statistical purposes, if the data set permits. Major national safety schemes are NASA’s Aviation Safety Reporting System (ASRS), and its UK equivalent CHIRP (Confidential Human Factors Incident Reporting Programme). An example for a local incident reporting scheme that has been implemented in healthcare and maintained for over 10 years is described in (Busse and Wright, 2000).

This paper offers a report on the prerequisites to successfully running an incident reporting scheme. The implementation that is described here attempted to localize an incident reporting scheme by taking contextual factors such as safety culture into account. It had staff actively participate in the conception of the scheme and the design of the reporting form. This is argued to be crucial in achieving long-term staff participation, and is argued to optimize the scheme's efficiency. The scheme that is described here also attempted to incorporate standardization issues that are posed by schemes implemented on a grander, e.g. national, scale. The experiences with this standardization approach will be discussed. This paper thus presents a matrix of issues that still need to be addressed in any safety-critical domain that employs critical incident reporting schemes as part of ongoing safety management.

## 2. INCIDENT REPORTING

The cost of adverse events is high; not only in human suffering, but also in compensation claims and the need for prolonged treatment of afflicted patients. In 1990, the Harvard Medical Practice Study (Leape, 1994) investigated the occurrence of patient injury caused by treatment - so-called adverse events. It found that nearly 4% of patients suffered an injury that prolonged their hospital stay or resulted in measurable disability. Leape (1994) pointed out that, if these rates are typical of the US, then 180000 people die each year partly as a result of iatrogenic ('treatment-caused') injury. In the UK, a Department of Health report (2000) revealed that as many as 850,000 adverse incidents are happening in UK hospitals each year. This, in terms of litigation and the extra care needed by victims, added up to a £2bn bill. (BBC News 02/15/2001). There is an urgent need to make patient safety one of the highest priorities. The BBC also cites fears that the medical community is "complacent" about the toll of accidents, and notes that to date the National Health Service "did not even collect figures on the number of medical accidents".

Until recently, evidence of medical incidents (or near-misses) was mostly anecdotal. In the case their existence was acknowledged, the data typically did not leave a hospital’s boundaries. This lack of distribution of incident data can lead to the replication of similar, preventable incidents across hospitals. This not only concerns e.g. faulty or badly designed equipment which might lead to deadly consequences. It also concerns badly designed work procedures that might be in place in hospitals across the country, the safety threat of which might only be recognized locally and sporadically. Similarly, it concerns drugs that might have similar sounding names, but that have very different effects on a patient’s condition. In order to prevent incidents needlessly repeating themselves, incident data must be recorded, analyzed, and then made available for distribution. In industries such as process control for chemical plants and power stations, incident reporting schemes have often been used as ‘early warning schemes’. This has yet to translate fully to the medical domain.

Obstacles to the implementation of safety measure in medicine as established in aviation also lie in the differences between the two work domains. Not only are accidents in aviation comparatively infrequent but very visible, receiving high media attention, and often involve massive loss of life (Helmreich, 2000). In contrast, accidents in medicine typically only involve not more than one patient (or member of staff), with less or no media coverage, with news about adverse events often not leaving a hospital’s boundaries. Also, the type of standardized, unified safety management measures implemented in aviation often cannot translate to the less standardized, less regulated, and thus less clear-cut work environment that medicine presents. Doctors and clinical staff often learn ‘on the job’ to a large extent, in contrast to aviation or nuclear power plant operation (a domain that also has a long-term history of use of safety management measures such as incident reporting). Medicine, described recently again as a “humbling art and a complex team activity” (Berger, 2001), deals with humans, whose conditions and responses is typically less predictable than an aircrafts’ (Helmreich, 2000). Errors may be particularly difficult to recognize in health care because variations in an individual’s response to treatment is expected. In addition, medical professionals may not recognize that a particular product or procedure may have contributed to or caused the problem because the patient is already ill, the product is not expected to work perfectly at all times, or the event appears unrelated to the product or procedure (QuIC, 2000).

| Areas of Impact | Local Incident Reporting Schemes | 'Global' Incident Reporting Schemes |
|---|---|---|
| **Growth and Training** | ▪ Learning opportunity<br>▪ Reflect on local work practice<br>▪ Discuss with superiors and peers<br>▪ Co-analysis of incident data by staff representatives | ▪ Learning opportunity is somewhat limited<br>▪ Analysis removed from local setting |
| **Safety Intervention** | ▪ Identification of possible safety intervention in local context<br>▪ 'Trial and error', iterative solution development is appropriate and possible | ▪ Identification of possible industry-wide safety interventions<br>▪ Identification of appropriate, context specific safety solution is more difficult<br>▪ Frequent testing and revision of local solutions is not feasible |
| **Threat Awareness** | ▪ Keep staff 'in the loop' about error potential in their local work environment<br>▪ Inevitability of incidents is highlighted, as opposed to 'human error' | ▪ Staff awareness only of industry-wide safety threats<br>▪ Building safety culture<br>▪ Often focuses on 'human error' (guilt seems more easily assigned, since the involved parties seem 'anonymous' and distant) |
| **Safety Culture** | ▪ Building a safety community | ▪ Building safety culture, e.g. |

| | | |
|---|---|---|
| | through staff participation in scheme design and data co-analysis<br>▪ Raising safety culture through active participation and heightened awareness | through political weight given to the project<br>▪ public acknowledgement of importance of safety measures and awareness |
| **Current safety level** | Statistical significance of frequency data is limited | Can gather more valid data on industry-wide error frequency |
| **Safety level evolution** | Statistical significance of trend data is limited | Can gather more valid data on industry-wide error trends |
| **Business Case** | ▪ Incident data and generated analysis and recommendations can be used as evidence of safety threat and safety effort vis-à-vis management<br>▪ Incident data can be used to back-up business decision-making processes | ▪ Incident data and analysis can be used as business case e.g. for equipment procurement<br>▪ Incident data can be used for backing-up safety strategy decisions |

**Table 1 -** Local versus 'Global' Incident Reporting

Studies suggest, for instance, that uncertainty about the most effective diagnostic and therapeutic approaches is pervasive (Macias-Chapula, 1997). One area in medicine that resembles more closely the more proceduralized and well-defined and thus more predictable task space of aviation is often cited to be anaesthesia. Correspondingly, safety measures such as in-depth error analysis (e.g. Gaba et al., 1987) and incident reporting (e.g. Runciman et al., 1993) have been applied to anaesthesia prior to a more wide-spread adoption in other areas in medicine. The work of Runciman and his colleagues forms the basis of the incident analysis method described in this paper as applied to neonatal intensive care.

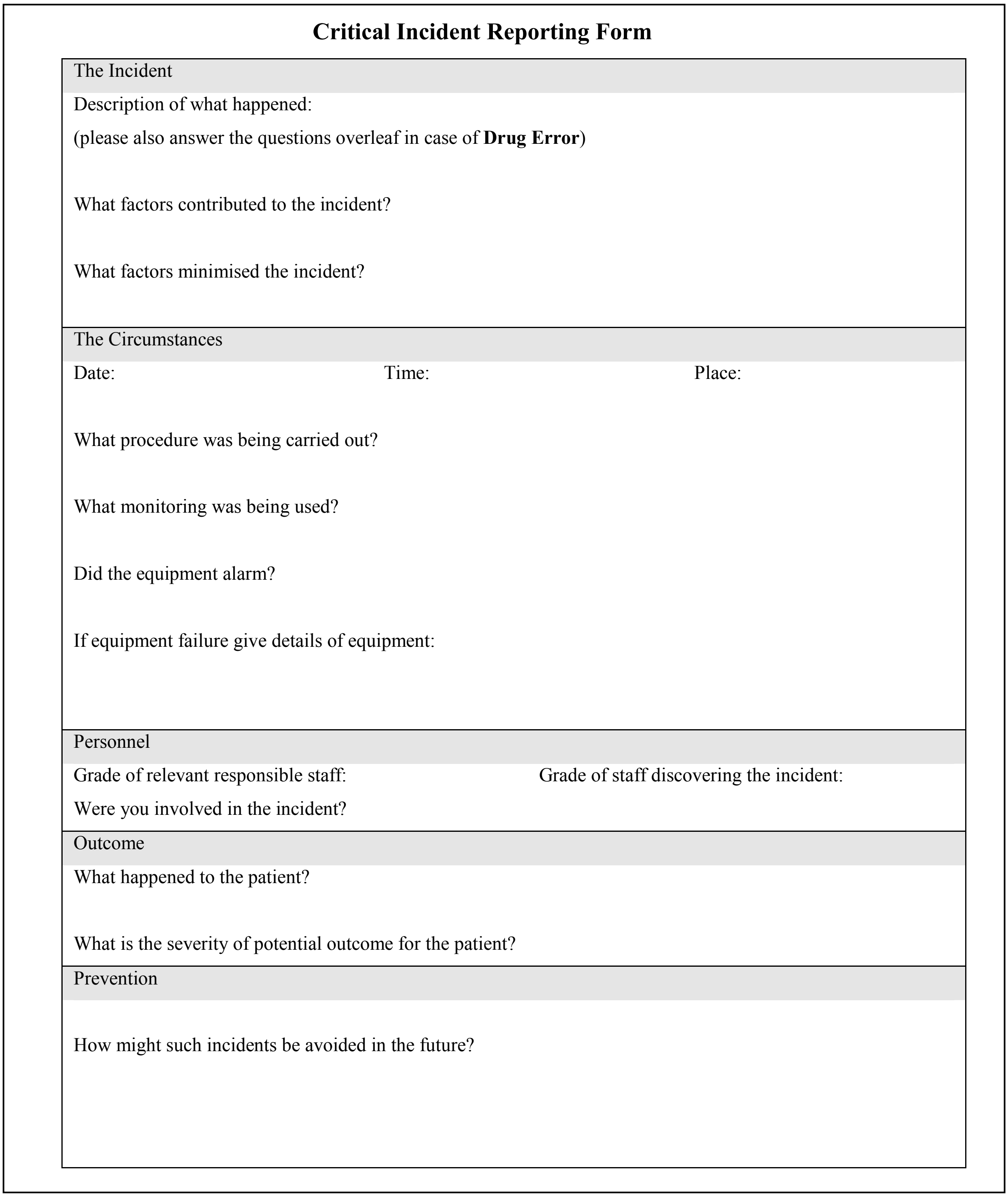
The tendency to lay blame on staff involved in the incident rather than e.g. error-prone equipment design (Busse and Johnson, 1999) further prohibits the use of incident reporting as a constructive safety measure in medicine. Thus, incidents might not necessarily be perceived by staff to be ‘accidents waiting to happen’ (Reason, 1990). Incidents might be seen as mere task characteristics, with mistakes seen as human fallibility, and with incident detection and recovery taken for granted. Incident reporters might also not be aware of ‘upstream precursors’ to the incident, such as underlying system faults (‘upstream’ since in systemic incident analysis, multiple layers of incident causation are assumed, with systems factors being the lowest layer). Staff might not acknowledge the significance of local workplace factors. For instance, if staff have been accustomed to working with substandard equipment, they may not report this as a contributing factor since they see it as the ‘normal’ work context; if they habitually perform a task that should have been supervised but was not, they may not recognize the lack of supervision as a problem (Reason, 1997). This tendency, and the associated ‘work-around’ culture in medicine, emphasizes the need for explicit scrutiny of potential upstream precursors (i.e. system factors) in incident reporting and analysis.

System factors might be organizational in nature, such as the notorious under-staffing in healthcare, with its associated stress on hospital staff, and the known increase in error-prone behaviour of individuals under stress. There might also be a lack of end-user consideration when choosing and procuring equipment, and also, for instance, neglect of training requirements on part of the organization purchasing the equipment (Jeffcott and Johnson, 2001). Accidents can often be traced back to equipment design that induces ‘human error’. It is rarely the case that accidents are caused by one single point of failure, and failure histories can usually be traced back through several layers of causation. While it is important for hospitals to prioritize error-tolerant and usable design and functionality in equipment procurement, there is often no empirical or analytical data available evaluating the system at hand in those terms. Thus, given difficult to use equipment design that seems obvious on scrutiny, accidents often seem preventable in retrospect. This seems particularly unforgivable in safety-critical work environments such as surgery or intensive care units.

Hospitals and healthcare trusts are increasingly confronting this issue of iatrogenic (treatment caused) illness as is shown by the recent debates in public media as well as in the increasing interest in safety management measures such as incident reporting. Incident reporting itself can help raise awareness of the risks of ‘human error’ in the day to day clinical work environment. It can also, of course, help gather data on near-miss incidents, events that could have led to a full-blown adverse event, but didn’t. These near-miss events were captured in time by observant staff, often enough coincidence playing a large role in their detection and recovery. Analyzing this data, and distributing it within the hospital unit and external to it, can help identifying system aspects that are prone to error. It can also help distinguishing random, coincidence based mishaps from *systematic*, design-induced error tendencies. This data can then provide valuable feedback to both the organization that targets a work environment with reduced risks associated with working its machines, and also to equipment designers and manufacturers that might not have been aware of the devices pitfalls, and that now are able to improve on their future design and also alert its current customer base. Equipment can be designed and developed in such a way as to minimize error and maximize safe productivity. Thus, device design can accommodate error prevention and tolerance to a certain degree. Medical equipment manufacturers, however, will need the appropriate feedback from the ‘real-life’, front-end users, in order to iteratively improve their systems. Current incident reporting schemes still often adhere to the ‘old-school’ punitive model, attempting to identify guilty individuals rather than constructively capturing improvement information. Still, even more notable attempts than punitive reporting often tend to be geared towards identifying error potential with the aim at finding work-arounds, with staff finding arrangements to ‘make substandard systems work’, rather than gathering information that would enable the generation of safety recommendations towards system re-design.

Introducing incident reporting into hospital wards is the first step towards recording information on incidents' nature and frequency. Summary data can then be used for trend analysis to identify systematic sources of error, and to prompt more in-depth analysis of potential causes. However, in order to base valid and relevant conclusions on this frequency counts, the classification of incidents clearly needs to be meaningful.

For instance, in current incident studies, most of the incidents' precursors are perceived to be 'human error' (e.g. Runciman et al., 1993). There are doubts, however, how meaningful this category, and its implications, really are. Often, the fact that an incident does not fit a category such as 'equipment failure' alone is seen as justifying labelling the incident as 'human error'. Such categorization might provide an initial filtering of immediately attributable equipment faults, but does not tell us much about how to prevent future instances of such 'human error'.



**Critical Incident Reporting Form**

The Incident

Description of what happened:

(please also answer the questions overleaf in case of **Drug Error**)

What factors contributed to the incident?

What factors minimised the incident?

The Circumstances

Date: Time: Place:

What procedure was being carried out?

What monitoring was being used?

Did the equipment alarm?

If equipment failure give details of equipment:

Personnel

Grade of relevant responsible staff: Grade of staff discovering the incident:

Were you involved in the incident?

Outcome

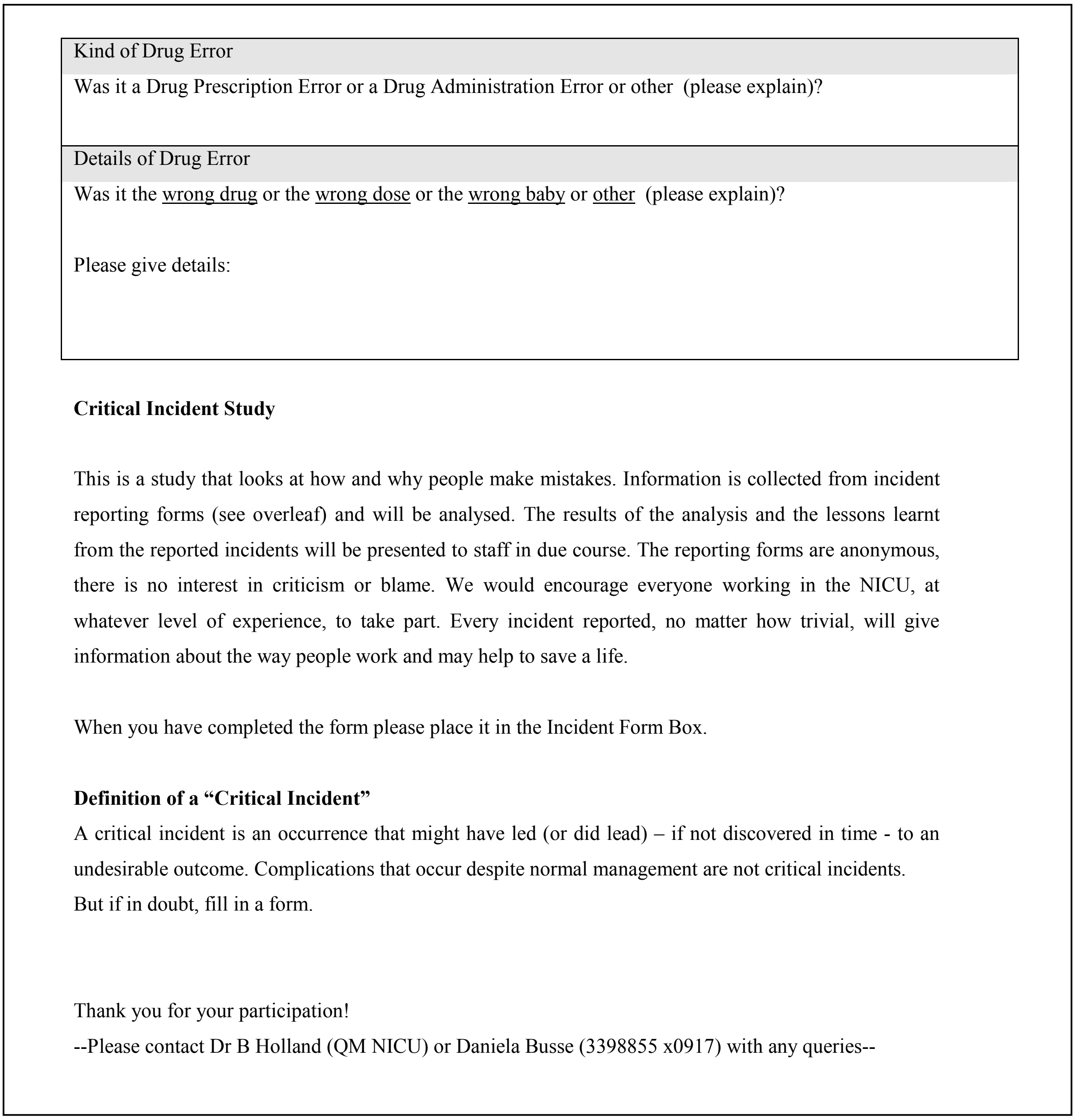
What happened to the patient?

What is the severity of potential outcome for the patient?

Prevention

How might such incidents be avoided in the future?

**Figure 1 (I) -** NICU Incident Reporting Form



| |
|---|
| Kind of Drug Error |
| Was it a Drug Prescription Error or a Drug Administration Error or other (please explain)? |
| Details of Drug Error |
| Was it the wrong drug or the wrong dose or the wrong baby or other (please explain)?<br>Please give details: |

**Critical Incident Study**

This is a study that looks at how and why people make mistakes. Information is collected from incident reporting forms (see overleaf) and will be analysed. The results of the analysis and the lessons learnt from the reported incidents will be presented to staff in due course. The reporting forms are anonymous, there is no interest in criticism or blame. We would encourage everyone working in the NICU, at whatever level of experience, to take part. Every incident reported, no matter how trivial, will give information about the way people work and may help to save a life.

When you have completed the form please place it in the Incident Form Box.

**Definition of a "Critical Incident"**
A critical incident is an occurrence that might have led (or did lead) – if not discovered in time - to an undesirable outcome. Complications that occur despite normal management are not critical incidents.
But if in doubt, fill in a form.

Thank you for your participation!
--Please contact Dr B Holland (QM NICU) or Daniela Busse (3398855 x0917) with any queries--

**Figure 1 (II) -** NICU Incident Reporting Form

Furthermore, as soon as poor equipment design is considered as an instance of equipment failure, there is no telling as to what constitutes error-inducing design (such as similarly named drug containers) and what constitutes human error (mistaking the drug containers). Thus, the artificial distinction between equipment failure and human error (more meaningfully described by Rasmussen as 'Human-Machine Mismatch') is cemented and perpetuated by such a classification. This clearly poses a very real problem for the validity of incident data and its analysis (a problem sometimes denoted as "GIGO" ("garbage in- garbage out)). However, this subjectivity in the incident classification process, and the resulting spurious precision of trend analyses based on the data, is not sufficiently recognized as would seem necessary for any wider distribution of incident data beyond the local setting.

There have been some notable attempts at creating grounded and relevant categorizations schemes in the recent past. An influential example is reported in Runciman et al. (1993) who studied anaesthesia incidents in Australian hospitals as part of the Australian Incident Monitoring Study (AIMS). This study was subsequently extended to also investigate intensive care unit (ICU) incidents. The AIMS categorization scheme presents an integrated summary of previous categorization schemes, and has had substantial impact on future ones. For the study reported in this paper, the AIMS-ICU categorization scheme was utilized to analyze incident data that had been collected in a neonatal intensive care unit. In the following sections, the outcome of this process is reported

# 3 THE NICU INCIDENT REPORTING SCHEME

In the Neonatal Intensive Care Unit (NICU) in which the study took place, current safety management included informal checks, communication and consultation with fellow members of staff, morbidity and mortality meetings, and an adverse events reporting scheme, which addresses incidents that in fact resulted in harm to the patient or to staff and that legally require investigation. Near-miss adverse events that do not require legal investigation, but that could also lead to harm to the patient or staff, were dealt with on a local and immediate basis. They were not documented or kept track of, and distribution of known sources of error in the system was at best infrequent. There was demand to complement the existing safety management measures with a critical incident reporting scheme.

## 3.1. Critical Incident Definition

A 'critical incident' was here defined as follows: "A critical incident is an occurrence that might have led (or did lead) – if not discovered in time - to an undesirable outcome. Complications that occur despite normal management are not critical incidents." Staff that participated in the study were also asked to fill in an incident reporting form "if in doubt". This reflected the intention to collect rich, qualitative data, rather than data that would be fit for exact statistical analysis.

## 3.2. Set Up

A 'critical incident' thus includes near-misses as well as actual adverse events. Reporting schemes involve staff reporting critical incidents using the provided reporting forms on a voluntary and anonymous basis. The incident reports are regularly analyzed and categorized. The main aim of the analysis is to identify factors contributing to the causation of incidents that may be rectified. Accordingly, similar incidents are hoped to be avoided in future.

## 3.3. Incident Reporting Form

The incident form was developed iteratively, and evaluated by means of a questionnaire survey of the unit staff (Busse, 2000). The current form covers the following questions: the first section asked for a "description of what happened"; 'Drug Confusion Error' is treated as a category distinct from other types of critical incident on the form, due to its known frequency (Bogner, 1994). This separate treatment allowed for more specific data to be gathered on drug errors. Other questions related to what factors contributed to the incident, and which factors minimized it. The next section covers details on the circumstances: which procedure was being carried out, which monitoring was being used, and which equipment failed (if any). A question on the presence of alarms was added to the form, since it was felt that incidents discovered through alarm sounding fell in a sufficiently distinct category of incident circumstances. This was then validated by the data that was collected. One section of questions touched on the personnel that were involved in the incident and its detection. The data was collected on an anonymous basis, so the reporting staff was not asked to provide contact details. However, experience levels and job titles were covered in the personnel section. Another section noted the estimated and actual outcome of the incident to the patient (or in terms of other costs). The final section provided an open-ended question regarding suggestions for improvements by the reporting staff - future prevention of similar incident being the primary goal of the incident reporting scheme.

## 3.4. Incident Analysis

The AIMS-ICU analysis scheme was used. AIMS used a reporting form that consisted primarily of given categories which were to be ticked off by the reporting staff (see Figure 2). This way of recording staff reports could potentially lead to decreased analysis time (since the reporters essentially did the categorization themselves). It could also be argued that this decreased the degree of indirection in the analysis process – categorization based on fairly subjective interpretation of gathered data could be replaced by the reporter’s own interpretation of the actual events. In the NICU study, the form that was used was specifically developed to address the local needs. However, the collected data was subsequently analyzed by assigning it to categories as listed in the AIMS. The breakdown of the results is shown in the following section.

SYSTEM-BASED FACTORS

| **Physical environment / infrastructure** | # |
|---|---|
| Lack of space / room | |
| Lack of facility | |
| Excessive noise | |
| High unit activity level | 4 |
| Staff mealtime | |
| Handover / ward round | |
| Lack of support staff | |

| **Equipment (including monitors)** | |
|---|---|
| Unavailable equipment | |
| Inadequate equipment | |
| Poor design | 4 |
| Poor maintenance | |
| Equipment failure | 3 |
| Inadequate inservice | |

| **Work Practices / Policies / Protocols** | |
|---|---|
| Communication problem | 1 |
| Inadequate assistance | |
| Lack of supervision | |
| Inadequate training | |
| Inadequate protocol | |
| Insufficient staff | 1 |
| Unable to contact staff | |
| Inapprop. staff / patient allocation | 1 |

HUMAN FACTORS

| **Knowledge-based error** | # |
|---|---|
| Lack or faulty knowledge | 1 |
| Error of: | |
| Judgement | 1 |
| Problem recognition / anticipation | |
| Diagnosis | |
| Treatment decision | |
| Use of investigation procedures | |
| Timing of investigation procedures | |
| Omitting intended treatment | 1 |
| Incorrect charting | 1 |
| Incorrect prescription | |
| Incorrect interpretation of information | |
| Information not sought | |
| Information not available | |

| **Rule-based error** | |
|---|---|
| Patient assessment inadequate | |
| Patient preparation inadequate | |
| Failure to check equipment | |
| Misuse of equipment | |
| Unfamiliar equipment | |
| Unfamiliar environment | |
| Unfamiliar patient | |
| Failure to follow protocol | 2 |
| Labelling error | |
| Calculation error | |

| **Skill-based error** | |
|---|---|
| Distraction / inattention | 1 |
| Fatigue | |
| Haste | |

**Figure 2** Incident Categorization and Analysis

## 3.2. RESULTS

As can be seen in Figure 2, 14 causes of incidents were classified as system factors, whereas only 11 where classified has human factors. This is in contrast to findings in comparable studies, where up to 80% of causal factors were classified as human error. The most commonly attributed subgroup of causal factors was ‘equipment’ with 7 occurrences, followed by ‘physical environment/infrastructure’ (4) and ‘knowledge-based errors’ (4). Furthermore, the categories ‘rule-based errors’, ‘skill-based errors’, ‘technical errors’, and ‘work practices’ were also all represented in the results. All causal factors could be categorized, and ‘other system factors’ and ‘other human factors’ were not assigned in this study.

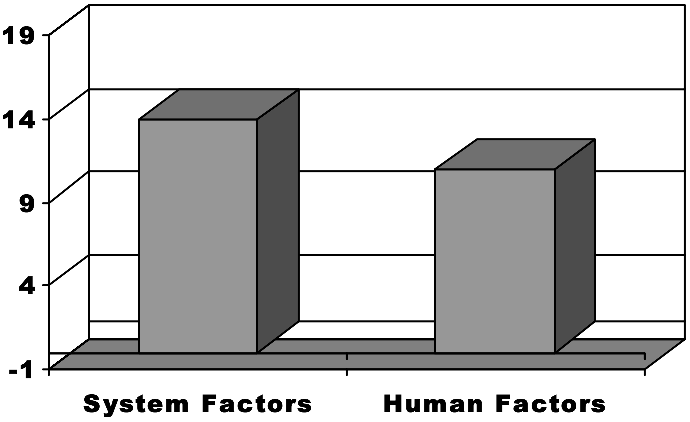


**Figure 3** – Distribution of System vs. Human Factors contributing to NICU incidents

However, it was found that for most incidents, multiple categorizations were necessary. There were mostly several causal factors per incident, for instance the categorizations listed above show at least two categorizations of causal factors per incident. The combination of factors proved to provide a more meaningful picture of the incident's causation and its potential future prevention, than single categorizations. This confirms previous findings (Busse and Wright, 2000). The most striking finding in this study was arguably the difficulty of arriving at a *meaningful* classification of incidents. The nature of a *meaningful* classification has not yet been discussed, let alone been operationalized, in the current discourse on incident reporting. Early work in the process control domain has covered substantial ground in delineating a meaningful analysis of Human Error (Rasmussen, 1982), but such work has still be addressed in incident categorization and analysis.

## 4. FORM VALIDATION

A questionnaire survey was carried out as an investigation of staff perception of the Critical Incident Reporting scheme as part of existing safety management at the unit. The study should also present an opportunity for staff to suggest improvements of the incident reporting scheme itself, as well as unit safety management in general.

22 questionnaires were administered to two 11 person day shifts on the ward. 21 were returned and 19 were included in the analysis. The two outliers that were excluded from the analysis were responses from temporary staff. Both medical and nursing staff contributed to the study.

Out of a total of 19 responses, 13 indicated awareness of the scheme, 4 said they had participated, and 11 could define "Critical Incident" (see graph below). Several valuable suggestions (see below) on form and scheme improvement were given by the study participants. Most survey participants responded positively to the overall impact of the scheme as part of safety management at the unit. Most participants stressed the crucial importance of feedback of incident data and analysis results back to unit staff. Raised awareness of existing error potential can be inferred from the 'shock reaction' that was reported on seeing the (anonymized) incident data and analysis results. Understandably, this led to frequent calls for urgent safety intervention on part of the survey respondents.

| **Survey Question** | **Staff Suggestions** |
|---|---|
| **Form improvement** | ▪ Include question on staff baby ratio<br>▪ Include question on workload levels |
| **Effectiveness of the scheme** | ▪ Good for Awareness (N=4)<br>▪ Need Action (N=3)<br>▪ Need Feedback (N=2) |
| **Safety management improvement** | ▪ Need Extra Staff (N=6)<br>▪ Need Review (N=5)<br>▪ Need Action (N=3)<br>▪ Need Training (N=2) |

**Table 2 -** Summary of Staff Suggestions for Form Improvement

The level of "safety culture" at the unit seems encouraging, with 2/3 of staff aware of the scheme and of the nature of critical incidents. Temporary staff, however, is still largely unaware. The scheme's intended implications for safety management needs better publication. Further measures are needed to feed analysis and action results back to the staff. Reassurance of staff is needed to let them know that the identified safety deficiencies are urgently and thoroughly addressed through safety interventions where possible.

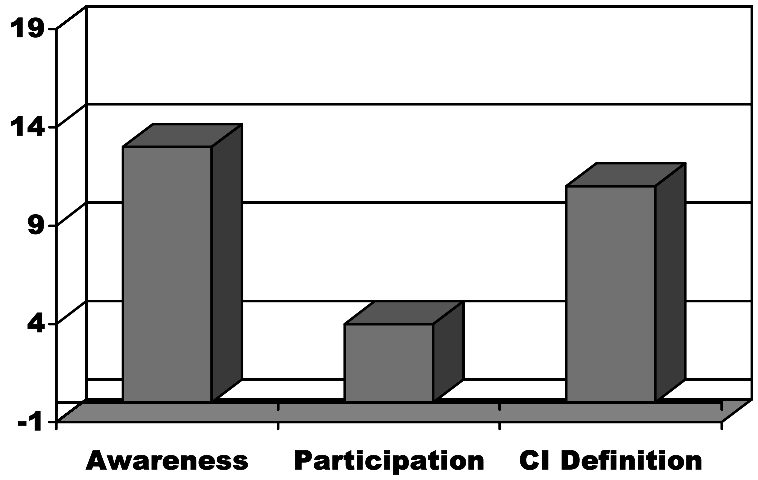


**Figure 4 –** Staff Awareness, Participation, and Ability to define 'Critical Incident'

## 6. CONCLUSIONS

The NICU incident reporting scheme succeeded in achieving staff engagement and participatory design of the reporting form. Most importantly, it impacted the overall safety culture in the unit, by raising awareness of clinical incidents, and raising the belief that the occurrence of 'mistakes' can be dealt with in a constructive way by higher level management (rather than following a 'punitive perfection model' (Leape, 1994)). Equipment failures could be followed up by either contacting the manufacturer directly, or also, for instance, by being able to refer to the incidents as evidence of insufficient design of specific devices. This data provides the basis to pass on valuable lessons to other NICUs that e.g. use similar devices.

The UK Department of Health (2000) stated as their main conclusion to the Chief Medical Officer's report: "We believe that, if the NHS is successfully to modernize its approach to learning from failure, there are four key areas that must be addressed. In summary, the NHS needs to develop:

- unified mechanisms for reporting and analysis when things go wrong;
- a more open culture, in which errors or service failures can be reported and discussed;
- mechanisms for ensuring that, where lessons are identified, the necessary changes are put into practice;
- a much wider appreciation of the value of the system approach in preventing, analyzing and learning from errors."

Additional to this, however, it needs to be stressed that the incorporation of the specific local conditions and requirements are necessary for the successful maintenance of an incident reporting scheme. Furthermore, if a unified mechanism for analysis is determined, it needs to take into account that the classification scheme should not only aid 'causal factor counting', but also their meaningful analysis.

## REFERENCES

BBC News, "Medical accidents - unstoppable?" http://news.bbc.co.uk/hi/english/health/newsid_1171000/ 11717 12.stm, Feb 15 2001.

Berger, J.T., "Book Review of Margin of Error: The Ethics of Mistakes in the Practice of Medicine" in New England Journal of Medicine, Vol 344, No. 5, pp. 389-390, 2001.

Bogner, M. S., (ed.) "Human Error in Medicine", Lawrence Erlbaum Associates, 1994.

Busse, D.K. “Staff Perception of Critical Incident Reporting as Part of Clinical Safety Management”, in Proceedings of the PsyPag Annual Conference, Glasgow University, 2000.

Busse D.K. and Johnson C.W., “Human Error in an Intensive Care Unit: A Cognitive Analysis of Critical Incidents” in J. Dixon (ed.) 17th International Systems Safety Conference, Systems Safety Society, Unionville, Virginia, USA, pp.138-147, 1999.

Busse, D.K. and Wright, D. “Classification and Analysis of Incidents in Complex, Medical Environments”, in Johnson, C.W. (ed.) Special Edition of Topics in Health Information Management on Human Error and Clinical Systems; THIM Vol 20 Issue 4.

Department of Health, UK, Chief Medical Officer. “An Organisation with a Memory”. Crown Copyright 2000.

Gaba DM, Maxwell MS, DeAnda A: “Anesthetic mishaps: Breaking the chain of accident evolution” Anesthesiology, 66:670-676, 1987.

Helmreich, R.L. “On error management: Lessons from aviation” British Medical Journal, 320, pp. 781-785, 2000.

Jeffcott, M. and Johnson, C.W. “The Role of User Participation in the Failure of Public-Sector IT Projects”, to be presented at HCM-HCI Lille, France, 2001

Kennedy, I., Howard, R., Jarman, B., and Maclean, M. “The Inquiry into the management of care of children receiving complex heart surgery at The Bristol Royal Infirmary. Interim Report: Removal and retention of human material”, Crown Copyright, UK, 2000.

Leape, L. L. “Error in Medicine.” JAMA: The Journal of the American Medical Association 272(23): 1851-57, 1994.

Macias-Chapula, C.A. “An Approach to Identifying the Role of ‘Information’ in a Health Care System: Implications for the Quality of Health”, in Bowker, G.C., Star, S.L., Turner, W., and Gasser, L. (eds.) “Social Science, Technical Systems, and Cooperative Work – Beyond the Great Divide”, Lawrence Erlbaum Ass., 1997.

Quality Interagency Coordination Task Force (QuIC) “Doing What Counts for Patient Safety: Federal Actions to Reduce Medical Errors and Their Impact”, 2000.

Rasmussen, J. “Human Error: A Taxonomy for Describing Human Malfunction in Industrial Installations.” Journal of Occupational Accidents 4: 311-333, 1982.

Reason, J. “Managing the Risks of Organizational Accidents”, Ashgate Publishing Limited, 1997.

Runciman, W. B., A. Sellen, Webb RK, Williamson JA, Currie M, Morgan C., et al. “Errors, Incidents, and Accidents in Anaesthetic Practice.” Anaesthesia and Intensive Care 21(5): 506-519, 1993.

# **APPENDIX D** – IMPACTING SYSTEM REDESIGN (PUBLICATION)

Busse, D.K. (2001)
**“Medical Incident Reporting and the Longitudinal Design Evaluation of Clinical Systems”**
Published in the Proc. of the 9th Int. Conf. on Human-Computer Interaction (HCI International 2001), New Orleans, Louisiana, USA, August.2001. Lawrence Erlbaum Ass., New Jersey, 2001. Poster Sessions: Abridged Proceedings.

# Medical Incident Reporting and the Longitudinal Design Evaluation of Clinical Systems

## ABSTRACT

Poor design of medical systems is a major factor in the causation of medical accidents. This paper argues for the use of incident reporting data in the longitudinal design evaluation of medical systems. Incident reporting schemes are first described, and then scrutinized for their failure of leading to safety recommendations that target system redesign. Similarly, the medical systems and human factors profession is called on to take advantage of the 'live' system use data offered by incident reporting by guiding form design and incident analysis.

## 1. INTRODUCTION

The costs of medical accidents are estimated to exceed billions of dollars, and thousands of preventable deaths per year are brought about by 'technical failure' or 'human error' in medicine (Leape, 1994). Hospitals now increasingly use incident reporting schemes that gather data on error potential in their day-to-day work and in their use of medical equipment (QuIC, 2000). Incident data reflects actual system use, and can offer a longitudinal evaluation of medical systems design. This paper argues that this data needs to feed back into the design and development of medical systems. Currently, there is no such direct link. This means that there is a wealth of error-oriented, 'live' human-system interaction data in hospital archives that is not put to use by medical system developers.

Healthcare work environments such as hospitals are often highly computerized, with information displays and devices fulfilling vital roles, e.g. in intensive care units, emergency medicine, radiology, and operating theatres. The **usability** of medical systems and devices has crucial impact on the success of the healthcare provided, and on the safety of the patient. Poor design of equipment (as well as its plain malfunction) still plays a major part in medical incidents and accidents, often with serious consequences.

Despite this, the implications of such ‘adverse events’ for system and device design often go unnoticed. Incidents relating to the same design faults keep re-occurring, threatening patient safety and the effective healthcare delivery.

Blame is often apportioned to the human operating the system rather than to the system design. Common responses to incidents thus include cautioning or reprimanding the staff involved, i.e. the users of the medical system. Evidently, such approaches are doomed to failure in preventing future incidents, since the actual ‘cause’ for the incident in such cases typically is not human malice or negligence, but inherent usability or design deficiencies (Busse and Wright, 2000).

## 2. MEDICAL SYSTEM DESIGN

There is typically no usability feedback to designers and manufacturers in the case of medical incidents. This is so especially where ‘human error’ is concerned, and where work-arounds can be found by the clinical staff to provide a ‘quick fix’ to the perceived problem. The ‘free lessons’ offered by system deficiencies as revealed in real-life, day-to-day system use, often go un-used. The lack of proactive error data collection on the side of medical system providers might also seem surprising since most cater for a highly specialized market, and the difficulties that do occur often have a high probability of re-occurring (Busse and Wright, 2000). Even the early THERAC-25 accidents (Jacky, 1991) are a case in point. They showed that not only can design flaws in medical systems lead to loss of life, but they also show that a lack of communication between system manufacturers and hospitals can lead to preventable accidents re-occurring in other hospitals. Incident reporting is one method to combat this lack of communication.

## 3. MEDICAL INCIDENT REPORTING

Incident Reporting is used in safety critical domains as part of systematic safety management that tries to ensure communication, and widespread dissemination, of potential sources of accident and error. An incident is typically defined as an occurrence that might have led (or did lead) to an undesirable outcome if not discovered in time. Thus, this definition includes near-misses as well as actual ‘adverse events’. Reporting schemes involve staff reporting incidents on a voluntary and confidential basis, using the provided reporting forms. Incident reports are regularly analyzed and categorized. The main aim of analysis is to identify factors contributing to the causation of incidents that may be rectified. Thus, similar incidents are hoped to be avoided in future.

An incident reporting form, can collect data on, for example: a description of what happened, factors that contributed to the incident, and factors that minimized it; which tasks were being carried out, and which equipment was used; whether there were any alarms, and what grade of personnel was involved in the incident and its detection; and also on the estimated and actual consequence of the incident to the patient. Finally, it can include an open-ended question regarding suggestions for future prevention of similar incidents. This example is drawn from a form that was developed iteratively and in interaction with clinical staff (Busse, 2000). Form design needs to reflect human factors concerns in order to yield information that benefits ergonomic system evaluation and redesign. Thus, form designs might also need to be adapted, for instance, to address questions on the context of collaborative work among clinical staff. Lack of communication is a typical cause for accidents arising out of collaborative work, and is often included as a category in incident taxonomies. However, incident forms rarely focus on information that would shed light on the particulars of the communication and its context, or how this source of accidents could be addressed.

## 4. DESIGN AND SAFETY RECOMMENDATIONS

If the collected incident data does come to use to inform system redesign, this typically occurs in the form of creating local ‘work-arounds’ to deficient equipment and work procedures. However, these work-arounds tend to increase the learning burden for new staff on training. They can often not be considered stable solutions that robustly tackle the accident source. And importantly, they are not communicated to other hospitals, or to the system manufacturer.

Furthermore, the safety recommendations generated from incident analysis frequently focus on human input to error avoidance and recovery, instead of providing design recommendations for more fail-safe, error-tolerant, and usable systems. Incidents are initially categorized as either caused by technical or human factors. The latter category frequently draws up to 80% of the total incidents. The routine response to this category of incidents can be described as a ‘blame and train’ approach, since, apparently, no technical failure was existent that could be addressed by redesign measures. However, so-called ‘human-error’ is often induced by ‘unkind work environments’ (Rasmussen, 1980) and system design flaws (Busse and Wright, 2000). The major distinction in error analysis schemes between ‘technical failure’ and ‘human error’ often creates the misleading impression that technical failures can be remedied by technical measures, whereas the human error ought to be remedied by ‘changing the human’. Given this distinction, it is routinely overlooked that the prime contributing factors to ‘human error’ still constitute technical deficiencies such as ‘poor design’, lack of consistency, misleading error messages, and confusing displays (Busse and Wright, 2000).

## 5. THE NEED FOR EFFECTIVE INCIDENT ANALYSIS

Therefore, an **effective** analysis method is needed in order to fully realize the positive potential of incident reporting. Data analysis should offer insights into an incident's existence and causation. But it also needs to lead to an understanding of the role of the human-system interaction that led to the incident, and how this can be addressed to avoid future re-occurrences. System redesign, rather than 'blame and train', should be envisaged as potential remedy. Lekberg (1997) observed that the analyst's background has crucial impact on the analysis result, and on the recommendations given. It might thus be argued that the focus on training as panacea is contingent on the professional background of the experts that are consulted for incident analysis. This suggests that if effective design recommendations are to be drawn from medical incidents, the system engineering perspective needs to play a major role in their analysis.

Unfortunately, there is currently not one theoretical analysis framework that details how analysis results might feed into design recommendations and improve the usability and safety of medical systems. This clearly exacerbates the lack of design recommendation drawn from incident analysis. It seems ironic that the initial conceptions which paved the way for research into the design implications of 'human error' (Rasmussen, 1980; Norman, 1983) have led to a flurry of error models that are now widely applied, but that do not include pointers towards potential system re-design. Current incident analysis models do not provide any support for the analyst in identifying potential systemic remedies. There is an urgent need to address this missing link between data collection and the generation of safety recommendations that target the design and usability of medical systems

## 6. CONCLUSIONS AND FURTHER WORK

Medical error is widespread, and many preventable deaths are its consequence. Usable and error-tolerant medical systems could prevent accidents from re-occurring. The healthcare community has recently embraced the use of incident reporting schemes to capture potential error sources, to the point of introducing national laws and regulations. Therefore, medical system designers and manufacturers should respond to this opportunity and take advantage of the information gathered in hospitals that put their systems to everyday use. Incident reporting offers a wealth of data on problematic human-system interaction. It also presents an opportunity to actively support the data gathering process by fine-tuning reporting forms, and to make the most of data analysis by directly focusing on recommendations for improved system design. Both of these areas are still largely unexplored by research. Incident diaries are not unknown in human factors research, just as error reports inform iterative system development in practice. However, incident reporting data is largely gathered on an ad-hoc basis, and data analysis techniques as well as safety recommendations need improving. Taking advantage of rich and relevant data sources like incident reporting is a prerequisite for the reduction of design-related medical accidents.

## REFERENCES

Busse, D.K. “Staff Perception of Critical Incident Reporting as Part of Clinical Safety Management”, in Proceedings of the PsyPag Annual Conference, Glasgow University, 2000.

Busse, D.K. and Wright, D.J. Classification and Analysis of Incidents in Complex, Medical Environments. In C.W. Johnson (ed.) Special Edition of Topics in Health Information Management on Human Error and Clinical Systems (THIM) Vol. 20 Issue 4; 2000.

Jacky, J. Safety-critical Computing: Hazards, Practices, Standards and Regulation. Computerization and Controversy, Dunlop and Kling (eds.), Academic Press, 1991.

Leape, L. L. Error in Medicine. Journal of the American Medical Association (JAMA) 272(23): 1851-57, 1994.

Lekberg, A.K. Different Approaches to Incident Inves-tigation - How the Analyst Makes a Difference Hazard Prevention, 33:4, 1997.

Norman, D. A. Design Rules Based on Analyses of Human Error. CACM, 4, 254-258, 1983.

Rasmussen, J., What Can Be Learned From Human Error Reports? In Duncan, K., Gruenberg, M., and Wallis, D (Eds.) Changes in Working Life, John Wiley & Sons Ltd., 1980.

Quality Interagency Coordination Task Force (QuIC), Doing What Counts for Patient Safety: Federal Actions to Reduce Medical Errors and Their Impact. Report of the Quality Interagency Coordination Task Force (QuIC) to the President, February 2000.