INSTITUTE FOR SYSTEMS, INFORMATICS AND SAFETY

FURTHER DEVELOPMENT OF THE PHENOTYPE-GENOTYPE CLASSIFICATION SCHEME FOR THE ANALYSIS OF HUMAN ERRONEOUS ACTIONS

E. Hollnagel, P. Marsden

JOINT RESEARCH CENTRE
EUROPEAN COMMISSION

1996

EUR 16463 EN

181 g
CIRI 21277
CL4, 15 10
0 P 20

INSTITUTE FOR SYSTEMS, INFORMATICS AND SAFETY

FURTHER DEVELOPMENT OF THE PHENOTYPE-GENOTYPE CLASSIFICATION SCHEME FOR THE ANALYSIS OF HUMAN ERRONEOUS ACTIONS

E. Hollnagel, P. Marsden

Human Reliability Associates Ltd.
School House, Higher Lane, Dalton, Lancs. WN8 7RP, UK.

JOINT RESEARCH CENTRE
EUROPEAN COMMISSION

1996

EUR 16463 EN

Catalogue : CL-NA-16463-EN-C

Printed in Italy

PREFACE

Human errors have a clear impact on safety and economical aspects, for example it has been estimated that 30 % of the automatic scrams in nuclear power plants is due to human errors. This situation is also common to other domains: an analysis of the data for air carrier operations over a 24 years period shows that about two thirds of all accidents are attributable to the cockpit crew errors.

These statistics, although very important, must be properly interpreted since, often causes and effects are mixed-up. It is therefore mandatory to properly distinguish among causes and manifestations and to refer to the more correct expression of human erroneous actions.

At the beginning of the nineties E. Hollnagel developed the phenotype-genotype taxonomy to attempt to rationalise the task of classifying human erroneous actions and in order to fulfil the above mentioned requirements.

The taxonomy and its further developments have been at the bases of more applicative works and research programmes on Human Factors and Root Cause Analysis performed at the JRC-ISEI during the IV framework programme in the Reactor Safety, Working Environment and third party work areas. This resulted in a long and fruitful scientific collaboration among ISEI and the authors of this report.

The purpose of this study is twofold, first to improve and further develop a human error taxonomy for retrospective use of event analysis, second to explore the feasibility of a prospective use of the taxonomy, exploring thus the difficult area of errors of commission. This last task opens a breakthrough towards the establishment of the new generation of Human Reliability Analysis (HRA) methods of general applicability.

Giacomo Cojazzi, Pietro Carlo Cacciabue

European Commission
Joint Research Centre
Institute for Systems Informatics and Safety,
21020, Ispra, Varese, Italy

TABLE OF CONTENTS

FURTHER DEVELOPMENT OF THE PHENOTYPE-GENOTYPE CLASSIFICATION SCHEME FOR THE ANALYSIS OF HUMAN ERRONEOUS ACTIONS

The road to wisdom?—Well, it's plain
and simple to express:
Err
and err
and err again
but less
and less
and less.

Piet Hein

Abstract. This report presents a method for systematic error/event analysis and reliability prediction that is based on the principle of distinguishing clearly between phenotypes (manifestations or error modes) and genotypes (causes). It is argued that the essential parts of any analysis approach must be the specific method, the classification scheme, and the underlying model - which in the case of human performance must be a model of cognition. On this basis a number of representative error analysis approaches are analysed. It is found that traditional HRA approaches use a simple classification scheme, but have only weak links to a model of cognition. Information processing approaches can produce detailed explanations in terms of mental processes, but are weak in accounting for causes that have their origin in the working environment. While cognitive approaches may avoid both problems, they are still under development, and few have reached a level where they can be practically applied.

The report describes the basis of a specific approach that is based on a non-hierarchical, bi-directional classification scheme, and which therefore can be used for both performance analysis and performance prediction. The basis of classification scheme is described in detail, and the classification groups are documented in an appendix. The methods for event analysis and performance prediction are presented as step-by-step guidelines. The application of the phenotype-genotype approach is illustrated by two examples. It is argued, that the qualitative performance prediction provided by this method is a good basis for a second generation HRA method.

1. OBJECTIVES

The present report describes the outcome of a contract on **Further Development Of The Phenotype-Genotype Classification Scheme For Analysis Of Human Erroneous Actions,** carried out by Human Reliability Associates (HRA) for the Institute of Systems Engineering

and Informatics (ISEI) at the Joint Research Centre. The work had the following two purposes:

- To **revise** and **improve** the classification scheme for analysis of human erroneous actions and man-machine interaction. This involves critically assessing the structure of tables which are part of the classifications scheme (the phenotype-genotype classification groups) and the way in which the underlying cognitive model is used.

- To consider the **feasibility** of the classification scheme for retrospective event **analysis** and performance **prediction**. Analysis and prediction may differ in the method they use, but should not put different requirements to a classifications scheme. It is therefore important to assess the feasibility of the classifications scheme for either purpose.

The objectives of the first phase of the work (April - July 1994) were:

- to assess the of results from existing applications of the classifications scheme;

- to compare it with available domain specific classifications schemes (aviation, nuclear, etc.);

- to perform a critical assessment of the underlying cognitive model, SMoC, and of the impact of the model on the use of the classification scheme, and

- to propose a revised classification scheme and revised links between the classification scheme and the cognitive model.

The objectives of the second phase of the work (August - November 1994) were:

- to develop scenarios for retrospective and predictive applications of the taxonomy;

- to define requirements to use from the two types of analysis;

- to assess the applicability of the taxonomy for retrospective and predictive analyses;

- to compare it with existing classifications schemes.

The present report describes the overall outcome of the project, and does therefore not retain the chronological order of the work items. Rather, an attempt has been made to present at cogent argument for the improvements in error analysis developed by the project. Although the work has been carried out by HRA, the results have taken advantage of a number of discussions with ISEI staff, in particular Dr. Giacomo Cojazzi and Dr. Pietro Carlo Cacciabue.

1.1 Background and Motivation

The background for this report is a line of work which started in the late 1980s as a survey of existing theories and models for human erroneous actions. It was generally felt that the then available approaches were insufficient both on practical and theoretical side.[1] These issues had

[1] To some extent this is still the case in 1995.

been extensively discussed in a NATO sponsored workshop in 1983 (Senders & Moray, 1991), and it is perhaps warranted to see this workshop as the initiating event, at least in the sense of starting a more deliberate attempt to identify exactly what the problems were. The survey was conducted as part of the work in developing an expert system for plan and error recognition (Hollnagel, 1988). The result of the survey was a clear realisation of the need to distinguish between manifestations and causes of erroneous actions.

This view was presented by Hollnagel (1991) as a proposal to distinguish between error phenotypes and error genotypes, and was later discussed more extensively in Hollnagel (1993a). In parallel to these theoretical clarifications, a suggestion to use the phenotype / genotype distinction to construct a workable classification scheme was developed by Hollnagel & Cacciabue (1991). The same paper also developed the notion of the Simple Model of Cognition (SMoC) as the rationale for the structure of the classification scheme. The main contribution to the contents of the classification scheme came from a project to develop a human reliability assessment method for the European Space Agency (Hollnagel et al., 1990). This work produced a highly useful summary of the many descriptive terms that had been developed through years of practical accident and error analysis, but which had rarely been systematically evaluated.

The situation in the early 1990s was therefore that an alternative to the traditional approaches to error analysis had begun to emerge, and that the ideas were viewed with interest by a growing number of people. The first serious attempt at practically applying the classification scheme and describing the method by which it should be done was undertaken in 1992-93 by Mauro Pedrali as part of the work for a Ph.D. thesis. This work produced a detailed analysis of an aviation accident which demonstrated the value of the principles (Cacciabue et al., 1993). In parallel to that, ISEI carried out a project to survey and compare existing error taxonomies (Cojazzi et al., 1993). At the end of 1993 there was therefore a fairly well developed classification scheme and an associated method which had three characteristics. Firstly, it maintained the important distinction between phenotypes and genotypes, as a way of structuring the description of an accident or event. Secondly, it made explicit use of an underlying cognitive model, the Simple Model of Cognition, to structure the causal links, hence to provide consistent explanations. And thirdly, it had developed a relatively simple method of analysis which had proved valuable to unravel and understand even very complex accidents.

The motivation for the project reported here was to consider the phenotype-genotype classification in the light of this experience, and to developed an improved event analysis method. In addition, there was also a need to consider whether the principles behind the method could be used for performance prediction as well as event analysis. This need was partly caused by the vigorous debate in the PSA/HRA community about the possible shape of a second generation HRA method (Dougherty, 1990). Although the development of a viable alternative for a second generation HRA approach was not an objective of the current project, it was nevertheless felt that it was an important issue that should be addressed. As it turned out, the results are very encouraging and are being pursued separately. This report is confined to providing an account of the work done in the project, as well as the main conclusions and possible further developments.

2. INTRODUCTION

The history of error analysis - or, more properly, the analysis of failures caused by human actions - is long and varied. An excellent description of the background for the current state is given by Reason (1990). Since the present report is not an academic treatise, no attempt is made to explain the background for the current work in any detail (although some indication has been given in the preceding section on "Background and Motivation"). Instead, this introduction will present the basic approach to error analysis that is the basis for the current work.

The object of the study is human actions, and in particular incorrect or erroneous actions. It has always been important for analysts - be they psychologists or engineers - to understand the cause-effect relationships that can be used to explain erroneous actions. An erroneous action is, of course, not wrong in an absolute sense, but only because it leads to or contributes to the occurrence of an unwanted outcome; it is erroneous *post hoc* rather than *ante hoc*. There is a very practical need better to understand how erroneous actions can manifest themselves and how they can be explained. The former correspond to the **phenotypes** which describe the observable forms of erroneous actions. Other terms which commonly are used to denote this are **error modes**, **surface forms**, or **manifestations**. The latter correspond to the **genotypes** or the causes of the erroneous actions. Whereas the phenotypes or error modes can be observed, the genotypes or causes can only be inferred - leaving out the case of introspection. The phenotype-genotype categories are thus primarily applied to acknowledged erroneous actions. The categories are presented because they provide a more consistent way of describing and accounting for erroneous actions, both as part of an event analysis and in performance prediction.

2.1 Method, Classification, Model

The development of a system to support the analysis of accidents and events[2] must include a **method** by which the analysis can be performed and a **classification scheme**.[3] The purpose of the classification scheme is to provide a consistent basis for describing the details of the event and to identify the possible causes.

---

2. In the following we shall refer to the situation being analysed as the **event**. In most cases the event is of a specific kind, e.g. an accident or an incident, but since the method described here is intended to be of use in many different cases, the more neutral term is preferred.

3 We deliberately use the term classification scheme rather than the term taxonomy. Although the two terms are often used interchangeably, taxonomy should be reserved for the biological sciences - and perhaps also linguistics. There are two main reasons why the classification scheme proposed here is **not** a taxonomy. Firstly, it is not strictly hierarchical. Secondly, it does not describe *taxons*, i.e., members with a common evolutionary background. Even with considerable imagination it is not possible to describe the causes of human erroneous actions as having a common evolutionary background.

2.1.1 Classification Scheme

It is necessary to have a consistent basis to define the data that should be recorded and to describe the details of an event. Analyses very often start from existing event descriptions, e.g. reports from the field. Such event reports are, however, of a varying quality because the initial description depends on local practice, i.e., the guidelines or procedures that have been established for a field or an application. In cases where the job is heavily regulated, for instance in aviation or nuclear power, there are specific and well-defined reporting systems in place. In other cases, where there is less public concern for safety, reporting may be of a mixed quality. It is therefore necessary to bring event descriptions to a common form before an analysis is attempted. It is in particular necessary to ensure that the information provided is as complete as possible. This can best be achieved by referring to a systematic classification scheme.

A consistent classification scheme is also necessary in order to analyse the event and identify the possible or probable causes. Hollnagel (1993a) has argued at length for the importance of having a complete and consistent classification scheme, and in particular for maintaining a separation between manifestations and causes - the so-called phenotypes and genotypes. The argument is that a systematic study of erroneous actions must necessarily keep observation and interpretation apart. If the two are mixed, as when we use our intuitive understanding of human behaviour to classify an action in terms of its causes, then it is difficult to guard the consistency and reliability of the analysis. In addition, it becomes impossible to either undo or revise the analysis. The term "erroneous action" is itself an expression of this principle. An erroneous action is defined as "an action which fails to produce the expected result and which therefore leads to an unwanted consequence" (Hollnagel, 1993a). In contrast, the term "human error" can be used to mean either the action, the causes for it, or the outcome.

The phenotypes and genotypes represent two fundamentally different ways to consider erroneous actions. The **phenotype** is concerned with the manifestation of an erroneous action, i.e., how it appears in overt action, how it can be observed, hence the empirical basis for a classification. The **genotype** is concerned with the possible cause or, i.e., the functional characteristics of the human cognitive system that are assumed to contribute to an erroneous action - or in some cases even be the complete cause!). The phenotypes and genotypes require two different set of categories, one for observation and one for interpretation. In cases where a strong or well established theory exists there is little risk in using the categories for interpretation to define the categories for observation, hence to combine observation and interpretation to some extent.[4] But in the absence of a strong theory, this combination should be avoided as far as possible. The behavioural study of erroneous actions - whether it is called psychology, human factors , or cognitive engineering - is an example of this; in fact, there are no strong theories of human action in general. Observation and interpretation should therefore be kept separate in empirical investigations.

Analyses have often started by using a simple set of categories, the best known being "omission - commission" and "slip - mistake - lapse", only later to realise that the categories are in need of extension. But if the basis for the initial categories is not known or has not been explicitly described, it may be very difficult to extend them. This is clearly shown by the many

[4] This is not the case for psychology, although it can happen in other sciences such as nuclear physics, chemistry, etc.

problems the HRA community has in accommodating the concept of a "cognitive error", which refuses to fit neatly into any of the present schemes.

The present project has looked at a particular classification scheme, i.e., a specific set of categories organised in subsets or groups. The underlying argument for the need to have a classification scheme will, however, remain valid even if another set of categories is used. We believe that the classification scheme proposed here is a useful one, but will make no claims as to its general superiority.

2.1.2 Method

A classification scheme is an essential component but it must be used rigorously, i.e., the opportunities for subjective interpretations and variations must be limited. This means that the classification scheme must be accompanied by a **method**. The method is necessary to ensure that the classification scheme is used in a uniform way both for description and for analysis, and that the variability between analysts is reduced as much as possible. If the variability is reduced, the reliability of the method is increased.

A method is a regular or systematic way of accomplishing something. In this context a method is a detailed description of the way in which the analysis of actions should be performed in order to describe the erroneous actions, in particular of how the classification scheme should be applied.[5] The elements of the classification scheme are intended to describe specific relations between details of the observed event - although the details themselves can be the results of inference rather than observation.[6] Since the analysis cannot be done mechanically or automatically, the method is necessary to ensure that the classification scheme is used in the same way, as far as possible.

2.1.3 Model

In addition to having a method and a classification scheme, it is also necessary to have a **model**. A classification scheme must, by definition, refer to an underlying model or description of the domain. This is so whether we are talking about event analysis in the domain of human performance, or analysis of other phenomena; biological taxonomies are a prime example of that. In the present case, the underlying model must refer to the principles that govern human action, and in particular human cognition.

The model is a convenient way of referring to the set of assumptions, beliefs, and facts about human cognition that form the basis of how we view the world and the events that happen in it. It is also a useful reminder that we are not dealing with an objective reality. This is particularly important in the field of behavioural sciences - to which performance analysis clearly belongs - because differences in the basic view may easily be lost in theoretical elaboration. Thus, if discussions only take place on the level of the classification scheme, disputes about the

---

[5] Strictly speaking we should refer to the analysis of actions rather than the analysis of erroneous actions, since the latter term in a sense presupposes that the result of the analysis is already known.

[6] In the case of cognitive functions, the basis is inference rather than observation. Cognitive functions are covert processes and can therefore not be observed - not even by introspection.

meaning of terms and the proper way of applying them (i.e., the method) may be difficult to resolve. If, however, a clear reference can be made to the underlying model, it will be easier to determine what the causes of the differences are, and possibly also to resolve them - or at least to acknowledge them fully.

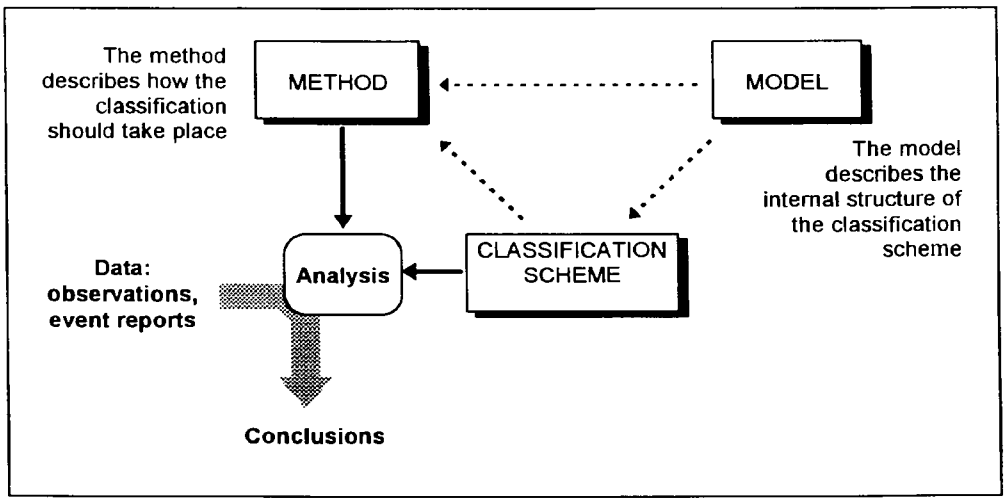

Figure 1: The relation between model, classification scheme, and method.

We therefore end up with a system that has three essential elements, which are related as shown in Figure 1. We shall refer to this as the MCM framework - for **M**ethod, **C**lassification scheme, and **M**odel - in analogy with the Root Cause Analysis (RCA) framework proposed by Cojazzi & Pinola (1994). The first element is a viable **model of human cognition.** This model makes it possible to link the description of specific system failures to the principles of the cognitive model - relative to the context in which the behaviour occurs. The second element is the **classification scheme** itself. Definitions of the categories of erroneous action embodied within the scheme should follow naturally from a consideration of the workings of the model of cognition, hence be a subset of the set of actions in general. This requirement is essential if assignment of causes for observed behaviour is to be justified on psychological grounds. Finally, the system must also incorporate a **method** which describes the links between the cognitive model and the classification of causes. The utility of analysis systems that lack a clear method is strictly limited. The absence of a method easily becomes a potential source of inconsistency when the classification scheme is used by different investigators, or applied by the same investigator working on different occasions.

2.2 The Role Of Data

In addition to method, classification scheme, and model a few words need be said about data. As indicated in Figure 1, data play a role as the input for the analysis. These data do not appear by themselves, but must be specified and collected. The data therefore depend on the assumptions inherent in the MCM framework, as discussed in the following section. But data also plays a role in a different sense. In cases where the purpose of the analysis is to make predictions of one kind or another - either quantitative or qualitative - data are important as the basis for the predictions. The analysis helps to describe the specific features of the situation or the action, but predictions can only be made if these are seen in relation to the general

characteristics, e.g. as frequencies, probabilities, levels, modes, etc. In this sense data constitute the basis for generating specific results from the analysis, as illustrated in Figure 2.

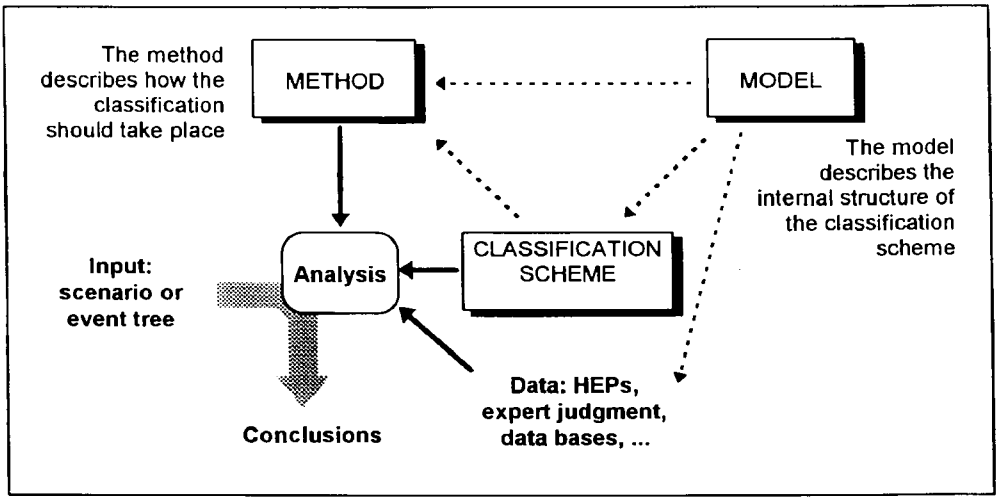

Figure 2: The MCM framework for performance prediction.

2.3 Data Analysis

As mentioned above, data types may vary from one source to another; the range may go from routine event reports and field studies to sessions from training simulators and protocols from controlled experiments. Similarly, the purposes of data collection and analysis may be quite different. The way in which the raw data are analysed depends upon their type as well as the purpose of the activity. Observation and analysis of human performance data may conveniently be described by a series of steps, derived from the work reported in Hollnagel et al. (1981):

- **Raw data** - constitute the basis from which an analysis is made. Raw data can be regarded as **performance fragments**, in the sense that they do not provide a coherent or complete description of the performance, but rather serve as the necessary building blocks or fragments for such a description. Raw data can be defined as the elementary level of data for a given set of conditions. The level of raw data may thus vary from system to system and from situation to situation.

- **Intermediate data format** - represents the first stage of processing of the raw data. In this stage the raw data are combined to provide a coherent account of what actually occurred. It is thus a description of the **actual performance** using terms and language from the raw data level rather than a refined, theoretically oriented language. The step from raw data to the intermediate data format is relatively simple since it basically involves a rearrangement rather than an interpretation of the raw data. A typical example of that is a time-line description.

- **Analysed event data** - here the intermediate data format is transformed into a description of the task or performance using formal terms and concepts. These concepts reflect the theoretical background of the analysis. The transformation changes the

description of the actual performance to a **formal description** of the performance during the observed event.

The step from the intermediate data format to the analysed event data involves the use of the classification scheme, since the analysed event data are expressed in terms of the defined categories. The transformation is one from task terms to formal terms. The emphasis is also changed from providing a description of **what** happened to providing an explanation of **why** it happened, i.e., to finding the causes.

- **Conceptual description** - aims at presenting the common features from a number of events. By combining a formal description of performances one may end up with a description of the **generic** or **prototypical performance**. The step from the formal to the prototypical performance is often quite elaborate and requires an analyst with considerable experience, in addition to various specialised translation aids. It also involves the use of the classification scheme.

- **Competence description** - is the final stage of the data analysis, and combines the conceptual description with the theoretical background. The description of **competence** is largely synonymous with the model of cognition, i.e., it is the description of the behavioural repertoire of a person independent of any particular situation - though, of course, still restricted to a certain class of situations. The step from the conceptual description to the competence description may be quite elaborate and require that the analyst has considerable knowledge of the relevant theoretical areas as well as a considerable experience in using that knowledge. The analyst must be able to provide a description in task independent terms of the generic strategies, models, and performance criteria which lie behind the observed performance.

The relations between the five steps described in the preceding can be shown as in Figure 3. The right side of Figure 3 describes the steps in going from raw data to competence description. This is characteristic of any data driven analysis, whether of experimental results or event reports. The basic trend is an aggregation of the various data types and a removal of the context - i.e., going from the specific to the generic. The left side of Figure 3 shows the complementary development from the level of competence to the level of performance fragments. This is typical of experimental design, planning of observations, etc. The basic trend is here an increasing level of detail and context - i.e., going from the generic to the specific.

In relation to the work in this project, the data analysis corresponding to the right side of Figure 3 can be seen as exemplifying the analysis of events to find the causes. The analysis moves from the level of the raw data to the level of a conceptual description - i.e., an explanation. The classifications scheme is used throughout - and in particular in the translation from the intermediate data format, via the analysed event data, to the conceptual description. The model, i.e., the competence description, serves as the point of reference for the analysis.

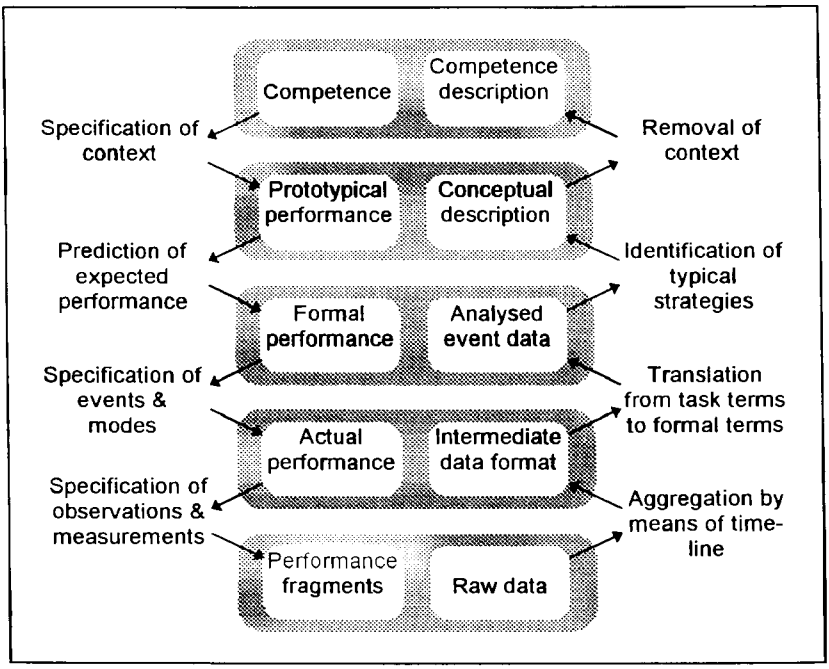

Figure 3: The dependence between data collection and data analysis.

3. MODELS OF COGNITION

As argued above, it is important that the classification scheme refers to a model for human cognition, but it is also important that this model is relatively simple. The model is necessary to define the relationship between components of the classification scheme, in particular the ways in which actions are typically produced, hence the ways in which erroneous actions may come about. The model must be rich enough to describe a set of cognitive functions that are adequate for analysing human erroneous actions. Yet it must not be so detailed that it introduces categories or concepts that do not have a practical value, i.e., which cannot be related to observations and which cannot be linked to remedial actions.

As an example, consider the case of short-term memory. Many models of cognition refer to short-term memory, and there is clearly little reason to doubt the existence of a short-term memory in human cognition.[7] But it would not be very useful if an analysis identified short-term memory *per se* as the root cause of an erroneous action. Firstly, because the status of short-term memory as a concrete entity is debatable (see previous footnote). Secondly, because there is little one can do about the functional characteristics of short-term memory. It would, indeed, be much more useful if the analysis ended up by identifying the set of conditions that **in combination** with the functional characteristics of a short-term memory could explain the erroneous action, hence the event. And finally, because short-term memory is a "passive" element in the sense that it does not contribute to actions, i.e., it is not a necessary part of explanations of how actions are produced.

3.1 A Simple Model Of Cognition

Earlier versions of the classification scheme made use of a simplified model of cognition called **SMoC** - meaning **S**imple **M**odel **o**f **C**ognition (Hollnagel & Cacciabue, 1991). Figure 4 presents an overview of the SMoC, as it was originally formulated. The SMoC described the basic features of human cognition, implying a typical path from observation over interpretation and planning to execution. The limited set of cognitive functions reflect a general consensus on the characteristics of human cognition, as it has developed since the 1950s. Each of the functions - or rather functional groups - would depend on memory, which therefore was shown as a general background.

The two fundamental features of the SMoC were (1) the distinction between observation and inference and (2) the cyclical nature of human cognition. The former emphasised the need to distinguish clearly between what can be **observed** and what can be **inferred** from the observations. Strictly speaking, and leaving out the thorny issue of introspection, what can be observed is overt behaviour, which match the two categories of **observation**[8] and **action execution**. The remaining cognitive functions can only be inferred from observations. The cyclical nature of human cognition means that cognitive functions unfold in a context of past events - as well as anticipated future events. Action execution, for instance, can be preceded

---

[7] It may, however, be debated whether short-term memory is an actual structure in the brain or rather a persistent functional characteristic, i.e., a functional structure.

[8] Perception is taken to be the cognitive process, while observation is the surface manifestation of it.

(or caused) by planning, by interpretation, or by observation. Observation in turn can follow as the consequence of an action, as well as of an external stimulus. The cyclical rendering of cognition serves to emphasise the multiple ways in which observable actions can depend on both the unobservable actions and the other events that may take place. A cyclical model, such as the SMoC, can therefore generate any sequential model, including the well-known step-ladder model (Rasmussen, 1986).

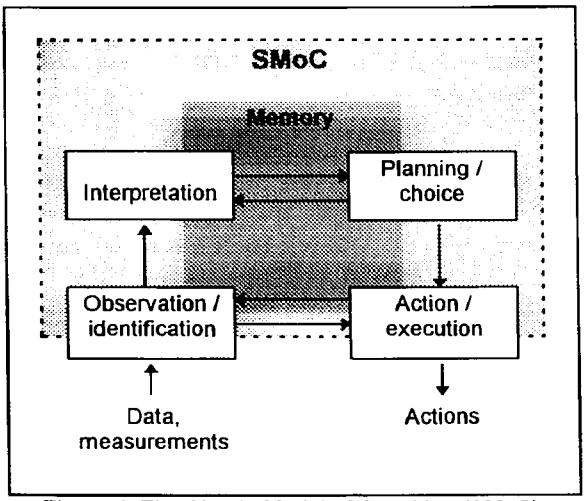

Figure 4: The Simple Model of Cognition (SMoC).

3.2 Competence And Control

Cognition is not only a question of processing input and constructing a reaction, but just as much a question of revision and review of goals and intentions (Bainbridge, 1993), i.e., a "loop" on the level of interpretation and planning. It is reasonable to assume that this occurs in parallel with whatever may happen on the other levels, while still in some way being determined by that. Cognition should therefore not be described as a sequence of steps, but rather as a controlled use of the available competence (skills, procedures, knowledge). This, of course, has significant implications for attempts to develop detailed models of cognition. In the present context, where the purpose is to develop a classification scheme and an associated method, the important implications are with regard to how the analysis is carried out. A strictly sequential model of cognition would correspond to a strictly hierarchical ordering of the concepts and causes, hence also to a well-defined path or set of paths through the classification scheme (which in this case even might be called a taxonomy). A non-sequential model of cognition means that the analysis is guided by the possible causal links between the various cognitive functions, as these unfold in a particular context. These links can not be defined *a priori*, but must reflect the prevailing conditions, i.e., the conditions that are assumed by the analysis. The basic assumption is that human performance is an outcome of the **controlled** use of competence, adapted to the requirements of the situation.[9] The analysis principle must reflect that assumption.

[9] This does not imply that the operator is in complete control of what he does. But the fundamental principle is that human actions are controlled, i.e., the result of deliberate intentions, rather than the

The non-sequential nature of cognition can be accounted for simply by weakening or removing the links between the cognitive functions in the SMoC. This would, however, lead to an "anarchic" or "anomic" type of model with no obvious links between the cognitive functions. An alternative, put forward by Hollnagel (1993b), is to use a modelling approach where **competence** and **control** are described in equal terms. Competence can be defined in terms of a relatively small range of cognitive functions which appear, to a greater or lesser extent, in most contemporary attempts to model the essential characteristics of human cognition. Control can be described by referring to a continuum, going from a situation where a person has no control over events to conditions where events are under complete control, and by emphasising characteristic modes of control along the continuum. Hollnagel (1993b) suggested, as a minimum, the following four control modes: (1) scrambled control, (2) opportunistic control, (3) tactical control, and (4) strategic control. Altogether this leads to a replacement of the SMoC with the Contextual Control Model (COCOM), which is shown in Figure 5.

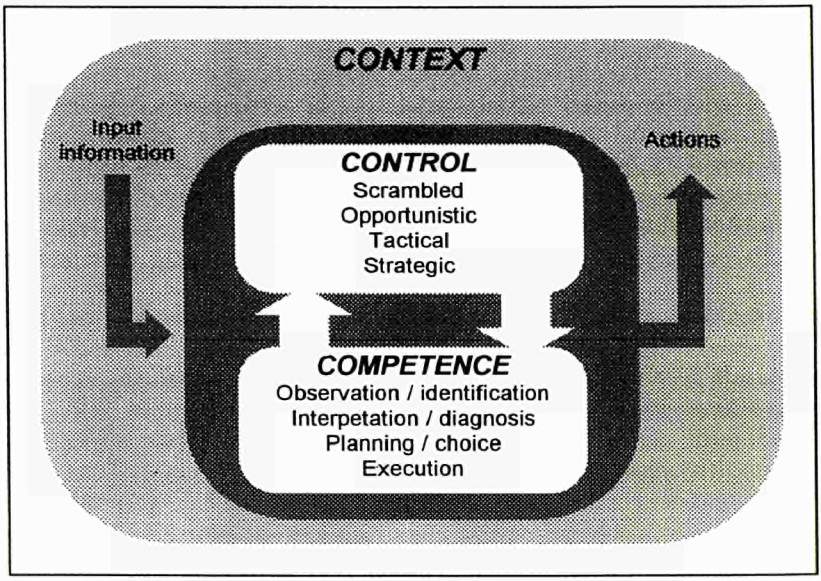

Figure 5: The Contextual Control Model of Cognition.

The basic difference between the COCOM and the SMoC is that the links between the cognitive functions have been relinquished and replaced with the control layer. This means that there are no pre-defined cause-effect relations defined in the model. In the first version of the analysis method, it was assumed that causes should be traced backwards through the chain of planning $\Leftarrow$ interpretation $\Leftarrow$ observation. This assumption imposed a constraint on the classification scheme. On further analysis it turned out that this constraint was superfluous, and that its removal would improve the range of possible cause-effect links. In the SMoC this corresponds to allowing links between all the functions; in Figure 4 this would mean introducing bi-directional links between observation and planning, as well as between interpretation and action execution. In that case there would be bi-directional links between all four functions; the links would cease to have a clear function, and it is therefore better to

outcome of pre-determined sequences of responses to events. This can also be emphasised by noting that human action is **intentional**.

remove them. The functions are still retained as part of the competence, but the structure has been relinquished. Instead, the structuring of actions is provided by the notion of control. While there is no *a priori* sequence in which the functions must be configured, there is in each particular case an **actual** configuration. At the present stage of development, the control model is not actually used for incident analysis; instead, the analysis is guided by the way in which groups of causes are associated, as described later in this report. The control model is, however, important for future developments. Firstly, it provides a way to include the influence of external conditions, which is different from the simple-minded notion of Performance Shaping Factors. Secondly, it opens the possibility of linking the classification scheme and the method with a dynamic model of cognition. This will in particular be of interest for attempts to base HRA on models of cognition.

4. TAXONOMIES AND CLASSIFICATION SCHEMES

In every system where human action is needed to ensure the proper functioning, actions may go wrong and unwanted consequences may occur. Although it is generally realised that the variability of human performance has positive as well as negative consequences, there has been an understandable tendency to focus on the cases where things can go wrong, since these may lead to a loss of material, money, and even human life. Unexpected positive outcomes are gratefully accepted and praised, but rarely give cause to any further deliberations - such as how one should reinforce them. Unexpected negative outcomes, on the other hand, are treated as situations that must be avoided and over the years significant efforts have been put into that.

The concern for the occurrence of unwanted outcomes is considerable in industries and processes where the cost is high and / or where public opinion is important, as for instance aviation and nuclear power production. In these cases one can therefore find many attempts to deal with the problem of human actions as a causal factor, usually referred to as "human error" or human erroneous actions. There have typically been two main concerns. The first is to develop classification systems or taxonomies that will enable the identification of the specific causes of unwanted consequences, in particular specific human actions or performance conditions. This is usually accompanied by the development of methods for event reporting and data analysis. The second concern is to develop methods to predict the possible occurrence of human erroneous actions, typically in the form of human reliability assessment (HRA). Both concerns have received much attention during last 30 years, and are still the focus of extensive interest and controversy (Dougherty, 1990).

The current project was neither the time nor the place to enter into this discussion as a whole. A number of books have dealt with this issue in recent years, e.g. Dougherty & Fragola (1988), Gertman & Blackman (1994), Hollnagel (1993b), Kirwan (1994), Reason (1990) and Woods et al. (1995). In relation to the project described here, a survey was made of a broad range of proposals for classification of erroneous actions, taken mainly from the aviation and nuclear industries. This survey was complemented by a more analytical overview of the principles for describing and explaining human erroneous actions, which is reported in a following chapter.

A classification scheme typically provides a way of describing the links between occurrences or events and causes or explanations. This conforms to the universal assumption of causality. Specifically, if something occurred - leading to an unwanted consequence - then it is assumed that there **must** have been a preceding cause. The proposed classification schemes have by and large been tailored to the specific domains, both in terms of the events that are described and in terms of the causes and conditions that are offered as explanations. The following is an attempt to characterise the typical classification schemes.

4.1.1 Descriptions Of Possible Events

The basis for any analysis must be the possible manifestation of erroneous actions and other contributing factors. The manifestations are usually referred to as the error modes or - in the terms of this report - the phenotypes.

Although there are a number of proposals for event descriptions, very few of these are pure phenotypes. As an example, consider the Critical Inflight Event Model (Rockwell & Giffen,

1987). This describes the errors that can occur during flight, but as Table 1 shows, the categories are a mixture of manifestations and causes.

Table 1: Critical inflight event model

1	Inadequate pre-flight checks.	2	Fails to recognise early warnings of problems
3	Fails to do sequence check	4	Decides to fly despite system discrepancies
5	Fails to recognise early warnings	6	Fails to monitor instrument readings
7	Fails to notice small discrepancies in flight sensations	8	Fails to notice lack of agreement of related instruments
9	Diagnostic error	10	Error is estimation of urgency
11	Improper corrective action	12	Poor emergency flying skills

Of the twelve events listed in Table 1, only numbers 2, 3, 5, 6, 7, and 8 are proper error modes; all of them are actually omissions. The remaining events are more in the nature of causes, in the sense that they are difficult to observe and in most cases must be inferred.

The lack of a sharp distinction between error modes and causes is characteristic of most classification schemes. This may lead to problems in defining unambiguous event reporting schemes. It is clearly important for the quality of event recording that there is a separation between observation and analysis, even though both steps may be performed by the same person. This separation, however, requires as a minimum that the categories used for event description have the smallest possible overlap with the categories used to describe the causes. If that is not the case, it becomes difficult to ensure that the chain of inferences that lead from manifestations to causes is distinct, hence to verify the conclusions. The imprecision that is a result of mixing error modes and causes may go some way towards explaining why event reporting schemes usually only have limited success - and why it is very difficult to combine different types of event reporting.

4.1.2 Descriptions Of Specific Psychological Causes

The suggestions for the description of causes fall into two major categories. One category contains the causes that are specific to a particular domain. In such cases it is of relatively little importance if the classification scheme refers to an underlying psychological model, or indeed if the causes are theoretically comprehensive. One could say that specificity is more important than generality, and that the main objective is to produce a classification scheme that is highly efficient within a given domain, both in terms of identifying frequent causes and in terms of being easy to use for domain experts.

Table 2 shows five examples of descriptions of psychological causes that have been used to account for observed errors. Four of these are from aviation, while the fifth is from the nuclear domain. Table 2 only provides the main categories; each example is described in more detailed in the source documents.

Table 2: Examples of specific psychological causes.

Human performance factors (Stoklosa, 1983).	Behaviour Medical Operational Task Equipment design Environmental	
Information transfer problems (Billings & Cheaney, 1981)	Instructions Errors involved in briefing or relief controllers Human errors associated with co-ordination failures	
Human failure (aviation) (Caeser, 1987)	Active failure (aware) Passive failure (unaware) Proficiency of failures Crew incapacitation	
Accident investigation checklist development (Feggetter, 1982)	Cognitive system	
	Social system	
	Situational system	Physical Environmental stress Ergonomic aspects
Managing human performance (SAE, 1987)	Behavioural aspects of sensing and mental processing Error evoked by sensing and mental process problems Verbal and written communications Defects in training contents and methods Work place environment	

The proposed causes listed in Table 2 vary both in their nature and in the level of detail. Some aim to address a broad range of events, while others are clearly more focused on a narrow set of occurrences. Table 3 shows a further, and more detailed, example also taken from the aviation domain. This describes not only the psychological (in the meaning of individual) causes, but also other factors that may contribute to the occurrence of the event.

Table 3: Causes of pilot related aircraft accidents.

Pilot error-related aircraft accidents (Kowalsky et al., 1974)		
Critical condition categories	Experience	Distraction
	Crew co-ordination	Neglect
	Air Traffic Control	"Decisions"
	Work/rest (fatigue)	Machine (plane)
	Airport	Weather
Critical decision categories	Decision resulting from out of tolerance conditions Decisions based on erroneous sensory input Decisions delayed Decision process biased by necessity to meet schedule Incorrect weighting of sensory inputs or responses to a contingency Incorrect choice of two alternatives based on available information Correct decisions Overloaded or rushed situation for making decisions Desperation or self-preservation decisions	

The causes given in Table 3 aim to be comprehensive, and do in fact represent a mixture between a specific and generic classification scheme. An even more detailed classification scheme is shown in Table 4. This is clearly an attempt to be complete in terms of the possible causes that must be considered. The number of variables that are contained in this list is very

large, and although the corresponding analysis may be complete, it will probably be very labour intensive and time consuming.

Table 4: Bureau of Air Safety Investigation accident analysis guide.

Bureau of Air Safety Investigation		
HF guide for the conduct of aircraft accident investigation	Anthropometrics /physical condition Physical strength Physical fatigue	Physical task saturation Physical co-ordination.
Physiological variables	Illusions Other factors related to illusions Vision, hearing and smell Visions Reaction time Nutritional factors Circadian rhythm Acute/transient fatigue	Cumulative/chronic fatigue Skill fatigue Hypoxia Hyperventilation Acceleration Decompression sickness Trapped gas effects Motion sickness
Psychological variables	General adaptation Information processing Attention level	Moods Personality
Psycho-social variables	Professional variables Familial factors Motivation	Financial factors Habit patterns
Pathological variables	Drugs Organic pathology	Functional pathology
Selection and training	24 specific factors relating to the training status of the pilot.	
Command and control	17 specific factors describing the adequacy of supervision and guidance provided to and by the pilot	
Operational requirements	Flight type Flight urgency Manoeuvre type Tactics employed	Time constraints Operating location Availability of resources
Support agencies	Other crew members Aircraft designers	Aircraft manufacturer
Morale considerations	Five specific factor	

On the whole, the specific classification schemes can achieve high efficiency and specificity in identifying possible causes. The main limitation is that there is only a loose association with established psychological theory; the classification scheme in Table 3 is probably the most explicit in that respect. This means that there may be problems in applying the classification scheme in a predictive fashion, and also that the relation between analysis and design may be weak. The lack of a general conceptual basis will make it difficult to recommend specific countermeasures (to the interface, the task definition, the training, etc.) based on the outcome of the analysis. The lack of generality in the concepts used by the classification schemes also means that transfer of results from one analysis to another may be difficult.

4.1.3 Descriptions Of General Psychological Causes

The other category of suggestions to describe causes contains classification schemes that are of a general nature. This specifically means that the terms are associated with a commonly accepted psychological theory, although the theory itself may not always be explicitly identified.

There are two main groups of classification schemes which have been generally accepted. One group puts emphasis on the **manifestation** of faulty actions, although it does not amount to a full description of error modes or phenotypes. Examples of that are shown in Table 5.

Table 5: General classification of manifestations.

Source	Proposed classification	
Slips and mistakes Reason (1985)	Skill-based slips Rule-based mistakes Knowledge-based mistakes	
Categorisation of action slips (Norman, 1981).	Action slips	
	Errors of omission	
	Errors of commission	
	Errors of substitution	Error in formation of intention Faulty activation of schemas Faulty triggering of active schemas

The other group puts emphasis on a description of the **prototypical information processing** that is assumed to be the substratum for human action. Here the ubiquitous model is the step-by-step description of decisions making which has been popularised by Rasmussen (1986). An example of that is shown in Table 6.

Table 6: General classification of causes.

Human error classification scheme		
General classification scheme (Rouse & Rouse, 1983)	Troubleshooting live aircraft power plants (Johnson & Rouse, 1982)	Supertanker engine control room (Van Eckhout & Rouse, 1981)
Observation of system state	Observation of system state	Observation of system state
Choice of hypothesis	Choice of hypothesis	Identification of fault
Testing of hypothesis		
Choice of goal		Choice of goal
Choice of procedure	Choice of procedure	Choice of procedure
Execution of procedure	Execution of procedure	Execution of procedure
	Consequence of previous error	Contributing factors

These two groups are only mentioned briefly here, since they will be dealt with in more detail later in the report. One common characteristic is that they are relatively independent of a specific application or domain. This is clearly demonstrated by the classification schemes shown in Table 6; although they have been proposed for three different applications, they are basically variations of a single underlying principle, which in this case is the traditional sequential information processing model.

4.1.4 Summary

On the whole, the classification schemes that can be found in the open literature are presented by themselves, i.e., without an accompanying method or indeed without any reference to their conceptual basis. In practically all cases the classification schemes imply an analysis in the form of a step-wise decomposition, going backwards in the event tree from the observed event until a likely cause has been found. This corresponds to the commonly known principles of root

cause analysis, and the advantages and limitations of this have been discussed by Cojazzi & Pinola (1994). In several cases the classification schemes are only partial, focusing either on phenomenological descriptions of the events or on the set of psychological factors that may cause the events. The importance of making and maintaining this distinction is, however, hardly ever recognised. Most of the schemes are also limited in the sense that they attempt to address a specific problem rather than to be generic. This state of affairs reflect the lack of a generally acknowledge "strong" theory or model of erroneous actions, which was typical of the field until the late 1980s.

4.2 Comparison Of Existing Taxonomies

It is by now commonly accepted that a theory or model for human erroneous actions must include three main sets of factors or influences, corresponding to the fact that erroneous actions take place in a context and that it is the combination of **individual**, **technological** and **organisational** factors.[10] These factors determine whether the action has any unwanted consequences, hence whether it is classified as an erroneous action.

- The first set of factors relates to the individual who carries out the action, and in particular to the characteristics of human cognition; they are here called **person related factors**.

- The second set of factors describes the technological characteristics of the system and in particular the various failure modes for the system, the sub-systems and the system components; they are here called **system related factors**.

- The third and final set of factors relates to the organisational context, e.g. established practices for communication and control, performance norms, and - in particular - the possibility of latent failures and system resident pathogens (Reason, 1991); they are here called **organisation related factors**.

The link between the three sets of factors can be shown in different ways, depending on how they are going to be applied. In the case where the purpose is the analysis of human erroneous actions or the prediction of human performance, it is natural to focus on the person- and system-related factors, and to see the organisational factors as providing the general background or context. We can illustrate the relation between the three sets of factors as in Figure 6, which shows human erroneous actions as influenced directly by person-related and system-related factors, and indirectly by organisational factors.

[10] This is often referred to as the Man, Technology, Organisation (MTO) triad.

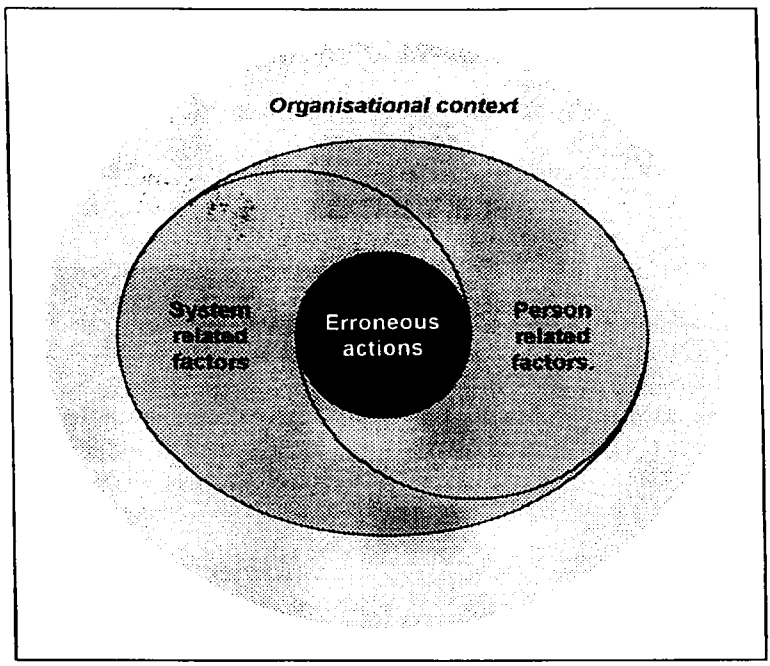

Figure 6: The relation between person-related, system-related, and organisation-related factors.

4.3 Factors Influencing Vulnerability To Error

Many empirical studies of human performance have treated human error in terms of a simple cause-consequence model of erroneous behaviour (e.g. Otway & Misenta, 1980; Canning, 1976). However, several investigators have taken issue with such a view and proposed that there are a number of factors in addition to stimulus-response considerations which are relevant to the classification of error events. Rasmussen et al. (1981), for example, has suggested a sevenfold classification of factors relevant to investigations of error, which is shown in summary form in Figure 7.

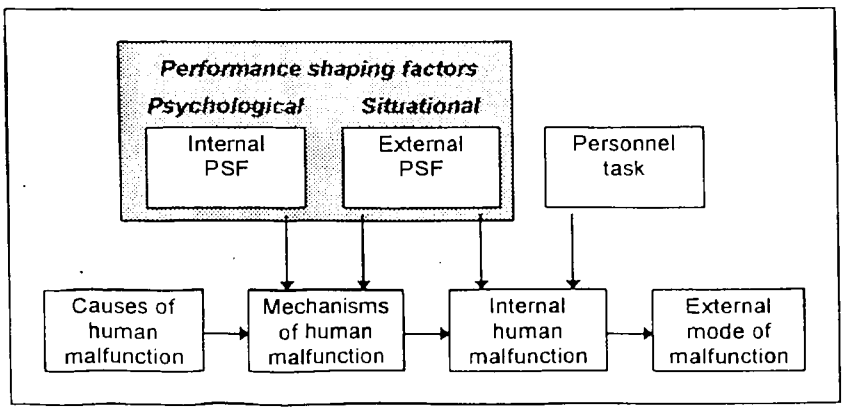

Figure 7: The CSNI taxonomy.

When Figure 7 is viewed **hierarchically** or top-down, the boxes in the lower tier indicate that a sequence typically begins with the **occurrence of an event** in the environment ("cause of human malfunction") which activates a psychological "failure mechanism". This in turn invokes a malfunction in human behaviour which may or may not manifest itself in the operating environment as an observable error ("external mode of malfunction"). In contrast to this the boxes in the top row of the figure shows the number of factors that can contribute to increase the likelihood that an error mechanism will release a malfunction. According to the CSNI scheme these are primarily factors known to influence the adequacy of human performance (e.g., Swain, 1967; Embrey, 1980) and sociological considerations which define the interrelationships between the relevant actor(s) and the current situation. They are commonly known as Performance Influencing Factors (PIFs).

Another way to read Figure 7 is with regard to how the different factors are presumed to **interact,** thus affording alternative explanations of error events as shown in Figure 8. For example, the two boxes shown at the bottom right hand side of the figure specify the factors that describe **what** error occurred - either in terms of external events (e.g. in an aviation context, failure to set flaps) or in terms of the human activity which went wrong (e.g. pilot forgot to implement check-list item or failed to use an established procedure). The box labelled "personnel task" immediately above these describes **who** committed the error (e.g. control room operator, maintenance / flight-deck crew, dispatcher, etc.).

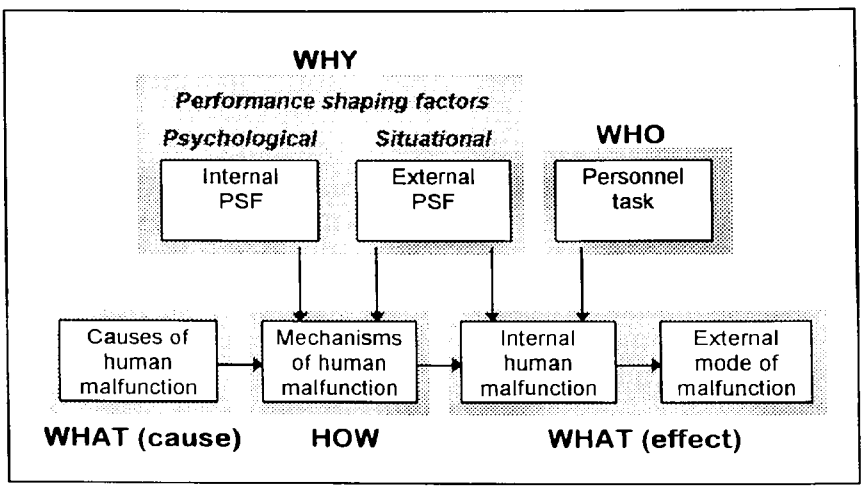

Figure 8: The CSNI taxonomy as describing interaction.

Conversely, the two boxes labelled "performance-shaping factors" essentially describe **why** a malfunction occurred. The "cause of human malfunction" category specifies the events that **caused** a process to deviate from the norm. Examples are situations where task demands exceed human performance capabilities or equipment malfunctions that produce misinformation and induce an erroneous belief among operators regarding actual system states. Also included here would be any off-normal event which acted to divert an individual's attention away from the task in hand.

Internal and external PIFs overlap considerably with the causes category in a conceptual sense. However, the main distinction being made here is with regard to the immediacy of effects. PIFs are not normally seen to be causal in a strict sense, but are generally thought of as factors that

contribute to error production. Thus, the presence of any single factor should not by itself lead to a failure in information processing. Rather, a combination of causal and contributory factors act cumulate to create a situation that is conducive to certain kinds of human malfunction. For example, in a transport environment a high level of personal stress combined with bad weather in busy conditions might cause task demands to exceed the individual's capability to perform appropriately, and this could be a catalyst for the occurrence of an error of some type; Rouse (1983) has provided a further discussion of the distinction between causes and catalysts of error.

Finally, the box labelled "mechanisms of human malfunction" denotes **how** an error occurs with reference to the underlying mental process which in a specific instance acted as the main error mechanism. For example, if a person forgets a critical action this could be interpreted either in terms of an error of omission (an error category at the behavioural level) or, alternatively, the error could be discussed in terms of a distraction brought about by competing memory demands (e.g. Baddeley, 1990).

The CSNI taxonomy provides a useful framework for investigating the causal pathways in the analysis of the erroneous actions taken by human operators. It is important to recognise, however, that the approach provides little guidance regarding assignment of operator errors to specific categories, nor to performance prediction. For these reasons the CSNI classification has not proved to be particularly useful for reliability analysis or accident investigation.

4.4 Taxonomies Based Upon First-Generation HRA Approaches

The earliest approach to error classification comes from the field of human reliability analysis. In this approach the objective has been to characterise how the actions of human operators lead systems states away from those which are expected or desirable. The primary tool used to describe performance deviations of this type has been the HRA event tree or one of its derivatives (see Figure 9, from Swain & Guttmann, 1983).

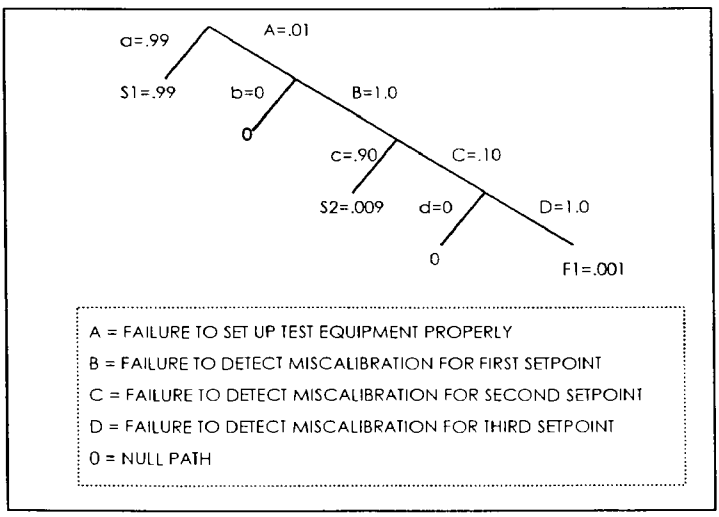

Figure 9: Example of an HRA event tree for a hypothetical calibration task.

Three important points need to be made in relation to an event tree such as the one shown in Figure 9. First, the event tree methodology represents work activities solely in terms of a linear and immutable sequence of operations that are enacted in a space-time continuum. Thus, an event tree conveys little or no information regarding possible covert cognitive activities which may be implicated in error causation (e.g. the collection and synthesis of evidence that a person may engage in when performing a problem-solving activity). Second, the event tree does not allow the occurrence of extraneous actions; it is basically a closed representation. Third, in an event tree each step of the linear sequence is treated as a binary choice node in which the only possible outcome is success or failure. This means that inefficient behaviour (i.e., behaviour which is not incorrect but still not desirable) cannot be represented in an event tree and this is an obvious limitation for the kind of classification schemes which may be developed for the purpose of assessing human reliability. This last observation leads naturally to a fourth point which relates to the kinds of error classification scheme that one finds in probabilistic risk assessments. There is a tendency for HRA analysts to confine their attentions to schemes based on the binary classification of errors of commission and omission such as the one shown in Table 7 below.

Table 7: Simplified example of error classes implied by HRA analysis

Error of omission		
Error of commission	Timing error	Too early Too late
	Sequencing error	
	Force error	Too little Too much
Extraneous error	Wrong act performed	

4.4.1 Evaluation Of HRA Classification Schemes

The most obvious shortcomings of error taxonomies devised for the purpose of risk assessment relates to the fact that such classifications do not refer to a viable model of human cognition; therefore such schemes cannot be validated on psychological grounds. In their favour, however, first-generation HRA taxonomies incorporate a rigorous application method which must be followed to produce the desired results. Error assignment represents a natural extension of the characterisation of operator actions in terms of an event tree and for this reason HRA techniques have proved popular with many investigators.

The extent to which the distinction between errors of commission and omission represents a valid classification of operator actions is a difficult question to answer. In a relatively simplistic task, for example, making a distinction between what the operator **does** and **does not** do, can often serve as a useful first approximation for discussing system failures. However, in relation to human performance in more complex environments, it is clear that the distinction has several weaknesses which limit the practicality of the scheme. The problems associated with classification such as the one proposed by Swain & Guttmann (1983) have been discussed at length by both Singleton (1973) and Reason (1986). Both authors point out that such schemes (Table 7) typically confound two important variables: (1) whether actions are taken or not taken, and (2) whether the outcome is correct or erroneous. Given that it is easy to envisage a situation where a specified action is omitted from a sequence but the system still behaves as desired, or alternatively, where a prescribed intervention is made by the operator but the

system still deviates from an expected course, then the taxonomy fails to establish whether the actions of the operator are to be considered erroneous. Criticisms such as these raise serious doubts about the general utility of schemes based upon the binary classification of error in terms of commission and omission errors, given the inherent philosophical problems that they appear to embody.

4.5 Taxonomies Based On Human Information Processing

The first-generation HRA approaches are deeply rooted in the stimulus-response paradigm which dominated psychology, particularly in the US, until the mid-1960s. Although the information processing metaphor served to enrich the understanding of the human mind, the basic dependence on a stimulus or event as the starting point for processing was retained. The information processing approach can be seen as a way of extending the "O" in the S-O-R paradigm, and although most of the details were developed for the information processes that were assumed to take place between stimulus and response, the fundamental principle of a sequence with a clearly defined beginning and end was not abandoned.[11] In relation to error classification, the transition from behavioural to information processing approaches nevertheless caused something of a revolution.

4.5.1 The Step-Ladder Model

The "step-ladder" model of dynamic decision-making (Rasmussen & Jensen, 1974; Rasmussen, 1976) is probably the information-processing model that has been most widely used as the basis for error classification. The step-ladder model proposes that there is a normal and expected sequence of information-processing stages which people engage in when performing a problem-solving or decision-making task, but that there are many situations where people do not perform according to the ideal case. To exemplify this an eight-stage model of information processing was developed, as shown in Figure 10.

A central theme of the step-ladder model is the idea that shunting (Gagné, 1962; shown by the dotted paths in Figure 10) between cognitive stages represents an efficient form of information-processing behaviour because it can reduce the amount of cognitive effort that needs to be invested in the performance of a task. In highly familiar circumstances, operators were believed unlikely to undertake each stage of processing and several types of behavioural short-cuts have been identified (Rasmussen & Jensen, 1974). However, these strategies can increase a person's vulnerability to making errors because they depend on the appropriateness of past experience.

[11] Although some people, like Miller at al. (1960) and Neisser (1976), early on realised the limitation of the sequential information processing paradigm, the mainstream of cognitive psychology failed to do so for many years. In Europe the different traditions in many cases diluted the influence from mainstream US psychology.

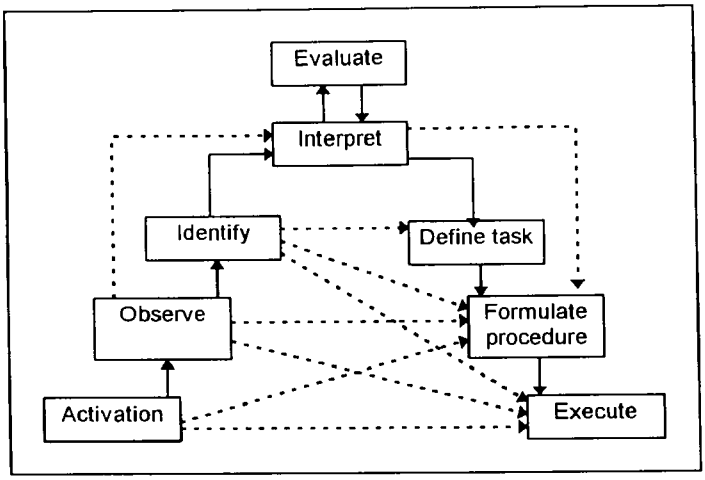

Figure 10: The step-ladder model for decision making.

4.5.2 Pedersen's Classification Of Error In Accident Causation

Pedersen (1985) made an interesting attempt to develop the error classification component of the step-ladder model by using it as the dynamic component of an error taxonomy designed for incident investigation. The resultant classification of erroneous actions can best be illustrated in the form of a question-answer list such as the one shown in Figure 11. This figure has been slightly modified from the original to be consistent with the representation of the step-ladder model shown in Figure 10.

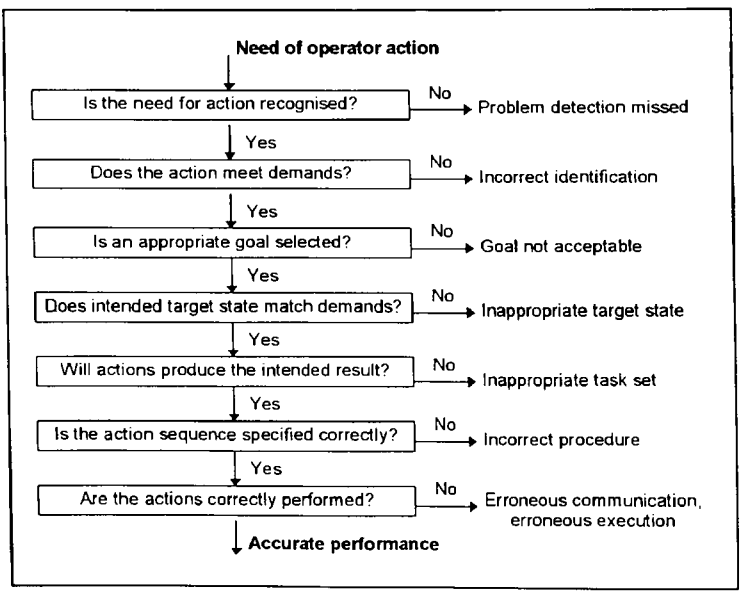

Figure 11: Pedersen's (1985) guide for error identification.

4.5.3 Generic Error Modelling System (GEMS)

Reason (1990) used a cognitive model, similar in form to the step-ladder model, as the technical basis for the Generic Error Modelling System or GEMS, in which the objective was to develop a context-free model of human error. Reason suggested that the emphasis on cognitive factors, as opposed to environmental or context related factors would permit the error classification embodied within GEMS to be applied to the analysis of error in a variety of industrial situations (e.g., nuclear power, process-control and aviation, etc.). In essence, GEMS extended a second important feature of the step-ladder model, namely the assumption that three distinct levels of cognitive functioning can be distinguished relative to a person's familiarity with a task or situation. Skill-based behaviour was assumed to be characteristic of highly familiar situations where the control of individual actions was delegated to stored patterns of "pre-programmed" motor sequences operating with little or no attention resources. The performance of routine tasks in familiar situations, on the other hand, was seen as rule-based. In this case the attainment of goals is presumed to require the development and maintenance of an action plan (see also Reason, 1976; 1979) whereby conscious control of action is required at critical choice points. Finally, the knowledge-based level covers situations where the individual must rely upon resource limited forms of information processing, such as reasoning. Hannaman & Spurgin (1984) made two important points in relation to the quality of performance at the knowledge-based level. First, they suggested that, in general, knowledge-based reasoning is expected to be more prone to misjudgements and mistakes, due to inherent limitations in the capacity of the human information processing system. Second, they argued that the identification of problem solutions will take longer because of the need to rely on more resource intensive cognitive activities such as deductive reasoning, etc. A representation of the SRK framework is shown in Figure 12.

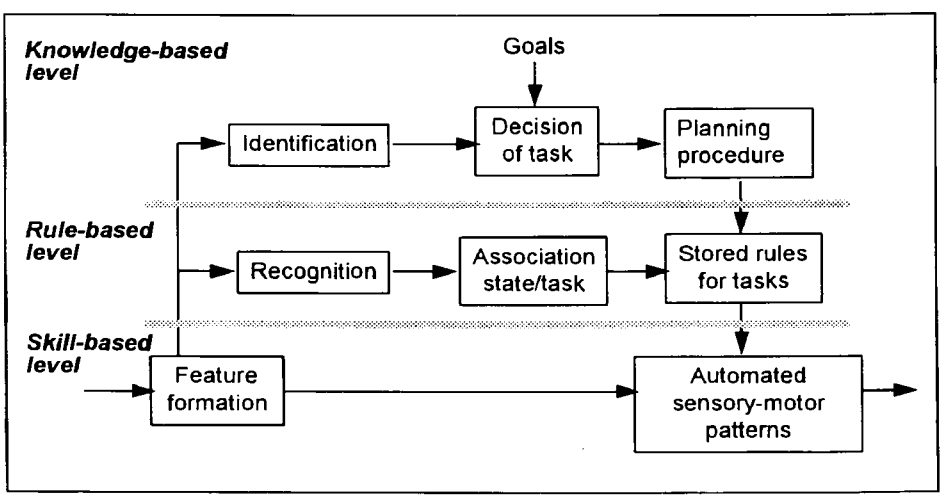

Figure 12: The SRK framework.

Reason's major contribution to error classification was to develop further the error types that were thought to be associated with the SRK framework; the resultant classification is shown in Table 8. In this scheme, the SRK framework has been blended with the distinction commonly made between slips and mistakes. The various types of slips identified by Reason are shown opposite the skill-based and rule-based columns, while categories of cognitive malfunction implicated in operator mistakes are shown in the column opposite knowledge-based mistakes.

Table 8: Major error types proposed within GEMS (Reason, 1990)

Cognitive Control Mode	Error Type	
Skill-based	Recency of prior use Frequency of prior use	Environmental signals Shared "schema" properties Concurrent plans
Rule-based	Mind-set Knowledge availability	Matching bias Oversimplification Over-confidence
Knowledge-based	Selectivity errors Short term memory limitations Bounded rationality Thematic vagabonding	Encystment Reasoning by analogy Errors of deductive logic Incomplete mental model Inaccurate mental model

4.5.4 Rouse's Operator Error Classification Scheme

A third taxonomy developed on the basis of the step-ladder model was proposed by Rouse and his colleagues at the Centre for Man-Machine Systems Research in Atlanta, Georgia. This scheme made an explicit attempt to blend together the approach to error classification supported by the step-ladder model with more traditional methods of error classification used by reliability analysts. The result was an error classification based on information processing, designed to be used to identify the probable causes of human error.

The cognitive modelling component of Rouse's scheme is shown in outline form in Figure 13. As can be seen from this figure, the basic model used to hypothesise probable causes of human error is similar in both form and content to the step-ladder model, although some of the major stages of information processing have been modified. Described in overview, the model proposes that during normal operations the operator of a human-machine system cycles through a sequence of activities that involves at least three stages: (a) observation of a system state, (b) the selection of a procedure, and (c) the subsequent execution of that procedure. In this interpretation, it should be noted that the term "procedure" is used by Rouse in a generic sense to include forms of script-based reasoning where the operators follow a pre-established pattern of actions from memory. Conversely, Rouse proposed that when a system state is such that one or more state variables have values outside the normal range, the situation may be considered abnormal and under these conditions the operator will usually need to engage in problem-solving (i.e., involving the formulation and testing of hypotheses). In this mode of operation the stages of information-processing are presumed to be particularly vulnerable to failure due to the high task demands placed on human cognition by problem-solving and decision-making activities.

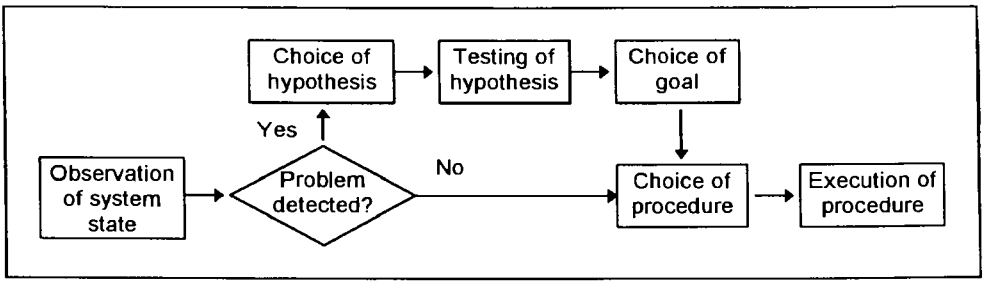

Figure 13: Rouse's conceptual model of the human operator.

Rouse used the model of operator behaviour shown in Figure 13, to guide the definition of possible failure modes in human performance. In an interesting departure from the work of both Pedersen and Reason, Rouse introduced a two-fold classification of error causation that identified a general category (error causation relative to the operator model) as well as a specific category (relative to the behavioural episode). Table 9 reproduces the general and specific categories of error causation proposed by Rouse, that are related to the cognitive model.

Table 9: Rouse's proposed classification scheme

General Category	Specific Category	
Observation of system states	a. excessive b. misinterpreted c. incorrect	d. incomplete e. inappropriate f. lack
Choice of hypotheses	a. inconsistent with observations b. consistent, but unlikely	c. consistent, but costly d. functionally irrelevant
Testing of hypotheses	a. incomplete b. False acceptance of wrong hypothesis	c. false rejection of correct hypothesis d. lack
Choice of goal	a. incomplete b. incorrect	c. unnecessary d. lack
Choice of procedure	a. incomplete b. incorrect	c. unnecessary d. lack
Execution of procedure	a. Step omitted b. Step repeated c. Step added d. Steps out of sequence	e. inappropriate timing f. incorrect discrete position g. incorrect continuous range h. incomplete i. unrelated inappropriate action

4.5.5 HEAT

Bagnara et al. (1989) have attempted to strengthen the methodological component of taxonomies based upon an information-processing standpoint. In this work Bagnara and his research team have applied the techniques of a knowledge-based systems approach to make inferences about the nature of causes of human error in an industrial situation. As in the case of the schemes considered above, the model of human malfunction utilised in the classification scheme was a variant of the step-ladder model and this has been strengthened by the inclusion of three further classes of error causation which Bagnara labelled: (a) decision-making, (b) socio-organisational condition and (c) external situation. These additional classes correspond

approximately to the range of factors identified by Rasmussen as being implemented in error causation, as shown in Figure 7.

A general flavour of the type of error classification that falls out of the Bagnara et al. approach can be exemplified as shown in Table 10. This table reproduces the types of error identified as arising from a consideration of the general category "problems with human performance".

Table 10: Example of the HEAT Taxonomy (Bagnara et al. 1989)

Human Performance Category of Failures		
Phenomenological Appearance	*Time*	Faulty activity Proper Activity
	Task not performed due to	Task omission Act omission Inaccurate performance Inappropriate timing Actions in wrong sequence Other Not applicable
	Erroneous act on system	Wrong act, right equipment Wrong equipment Wrong time Other Not applicable
Cognitive Function	*Detection*	
	Identification of system state	Faulty or incomplete monitoring of system state Faulty or incomplete assessment of system state
	Decision	Selection of goals Selection of system target state Selection of task
	Action	Specifying the procedure Carrying out the action
	Evaluation	Outcome inappropriate to goal Outcome inappropriately related to action
Cognitive Control Mechanism		Knowledge-based Skill-based Rule-based

4.5.6 Evaluation of Information Processing Taxonomies

It should be clear that each of the taxonomies outlined above incorporate, to a greater or lesser extent, aspects of the three elements thought to be essential to error taxonomies, i.e., a model, a method, and a classification scheme. In essence, the model(s) in the information processing approach have been variants of the step-ladder model, which effectively represents the state-of-the-art around 1990. While the step-ladder model has proved to be an invaluable tool for understanding the ways that human information processing can go wrong, it has been less successful when it has been used to specify categories of human error. The discussion above illustrates that the same model has been used to guide the specification of error categories on at least three separate occasions (e.g., Pedersen, Reason and Rouse) and that on each occasion a different classification scheme has resulted. **This finding suggests that categories of error do not necessarily follow logically from a consideration of the underlying model.** If they did, there would be better agreement among investigators regarding what constitutes the basic categories of operator error.

At least part of the problem is the failure among the investigators to agree upon a method which specifies how the error researcher should move from the model to error assignment. Of the three schemes discussed above, for example, only the scheme proposed by Pedersen provides any guidance about error assignment. The process of error assignment involves evaluating success or failure during each stage of information-processing such that if the response to a question is negative, then an error is assigned to that particular stage of processing. Clearly, such descriptions can be valuable for providing general explanations of error causation, although the results fall far short of an exhaustive classification of error causation in human cognition.

A second major problem with the taxonomies derived from the step-ladder model relates to the fact that this model essentially provides an explanation of errors made by experts (i.e., errors which are based in the habit-strength of the operators). The major concern is providing a description of errors made by skilled operators in familiar circumstances and these errors represent only a small proportion of the total error corpus. It could reasonably be argued that an analysis of the factors which underpin operator mistakes (actions as planned, but where the plan is inadequate) are at least equally important. These errors lie outside the scope of either of the schemes outlined above.

The challenge to the adequacy of a classification scheme based upon an information processing analysis relates to the more fundamental problem of whether it is appropriate to base explanations of error tendencies on what amounts to "design-defects" in human cognition. For example, investigators working from the standpoint of the cognitive systems perspective would argue that the same process that underpin error on one occasions are the same as those which underpin accurate performance on another. This suggests that the search for the psychological causes of operator error requires a much broader range of analysis than is implicit in the information processing taxonomies considered above.

4.6 Summary

The survey of error taxonomies shows that a number of classification schemes currently available within the scientific literature can be used to guide empirical investigations of operator error. To a large degree the abundance of error taxonomies is indicative of the amount of effort that has been put into researching this important question in recent years. The extent to which the dearth of error classification schemes represents the strength of error research, however, is much less clear. The fact that so many schemes have been produced by investigators could be interpreted as evidence of significant gaps in our understanding of the cognitive factors that are implicated in error causation and it is this theme which has been given prominence in the present work.

A key factor in error classification is the extent to which categories of human malfunction can be related to a viable model of human cognition. Traditional HRA methods are particularly weak in this area because their analysis of human performance is based entirely upon the observable behaviours that "cause" a system to deviate away from a desired state. Such observations are clearly important in developing an understanding of why systems fail. Yet confining the focus of the analysis to the consideration of external manifestations of error events tells us nothing about how and/or why a particular error event occurred. For this kind of explanation it is necessary to refer to a model of the human operator that will aid the

investigator in the process of making inferences about the possible psychological (or indeed sociological) causes that underpin the occurrence of the observed error.

Error taxonomies based upon the view of the operator as an information processing system provide a more rounded explanation of error events. This is because they are based upon a concrete model of human cognition in which information is generally perceived, a decision made and an action executed.[12] While no-one would take issue with the idea that information-processing must involve cognitive activity in each of these three broad areas, it can be questioned whether the information processing viewpoint provides the analyst with the most appropriate metaphor for discussing human thought processes. Since the mid 1980s there has been growing feeling that the analogy may have been over-utilised to the detriment of alternative developments. A major problem for information processing models that aim to specify error modes in human performance is that explanations of error events are couched in terms of breakdown in the natural course of information processing. Thus, such models are not particularly useful for dealing with errors that have their origins in the environment, such as those which arise due to false or misleading signals. Put differently, information processing models tell us little about the **contextual** causes of error, nor do they allow us to interpret why a particular course of action was selected for use by the human operator.

In contrast, classifications developed from a cognitive systems perspective are predicated on the assumption that erroneous actions require explanations which refer to a contextual model of human cognition, in which a model of competent human performance is related to factors present in the environment. The event is therefore interpreted relative to factors that define the operating environment at the time the error occurred. In this respect, the cognitive systems classifications view errors as a type of **mismatch** that occurs between cognition and context. Thus, cognitive systems taxonomies can be seen as representing a break with both traditional and information processing approaches to the problem of error classification insofar as they aim to provide explanations of error causation at several cause and effect levels. The major disadvantage of the cognitive systems approach to error classification relates to the fact that it is still in its infancy and that its concepts are continually evolving.

The second important issue is the specification of a method to guide practical error assignments. On this issue, information processing models were found to be quite weak because the different schemes incorporate methods that involve a **high degree of subjectivity**. Thus, confusion and disagreement regarding the treatment of specific categories of error are commonplace in information processing investigations and this tends to weaken the overall utility of the general approach. The methods associated with HRA analyses are generally sound, although the assignment of numerical estimates of error probabilities has been vigorously debated in the scientific literature in recent years. Techniques for assisting estimates of probability have been devised by several investigators, although none appear to have caught the imagination of all workers in the field. In contrast to this, the application method described in the following, which is also a part of **CREAM**, occupies an intermediate position in relation to the error assignment issue insofar as it permits limited interpretations of error causation / effects to be made within a particular analysis. The analysis method also allows consideration

[12] Strictly speaking, however, cognition is treated as an epiphenomenon of the information processing. In a proper cognitive model this will not be the case.

of "probable" causes / effects when sufficient data is unavailable to discriminate between alternative explanations of the error event.

5. CLASSIFICATION SCHEME

In the most basic form, the classification scheme expresses the relations between the phenotypes / manifestations and the genotypes / causes. It is, however, important from the start to make clear that the phenotypes are the result of an **interaction** between the genotypes and the context or environment, rather than a result of the genotypes alone.[13] We thus have the fundamental relation shown in Figure 14. The analysis of an event, the search for possible causes, naturally goes in the direction of the large arrow, i.e., from phenotypes to genotypes.

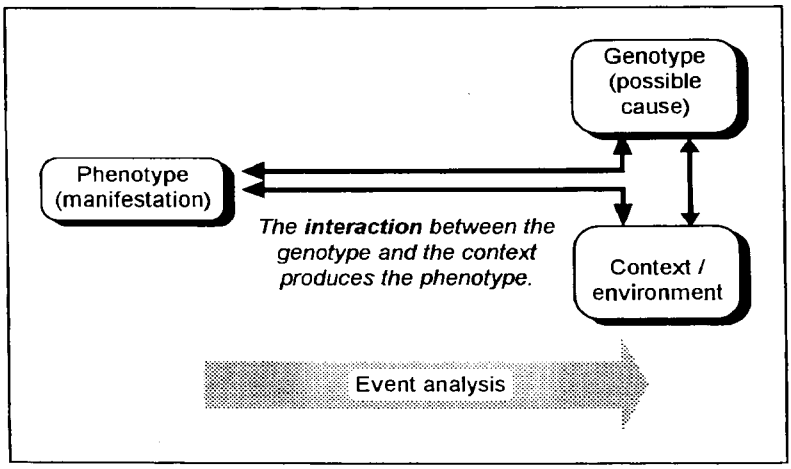

Figure 14: Phenotypes result from the interaction between genotypes and environment.

5.1 Basic Principles Of The Classification Scheme

In order to be useful, a classification scheme must obviously contain a larger number of details. While still remaining at the overall level, it is possible to distinguish between three major categories of causes, cf. the previous reference to the combination of individual, technological and organisational factors: (1) causes that have to do with the person, i.e., individual factors; (2) causes that have to do with the technological system; and (3) causes that have to do with the environment.

The first category contains the genotypes that are linked to an individual, for instance relating to cognition, to psycho-physiological variables, to the emotional state, to personality traits, etc. Depending on the psychological approach adopted, these genotypes can either be limited to those causes that have an immediate and clear link to behaviour in a situation (such as most of the cognitive factors) or be extended to include factors that are more remote, e.g. personality traits. It may be useful to refer to the classical distinction between **proximal** and **distal** variables (Brunswik, 1956). Proximal variables or proximal genotypes are those that have a direct influence on the person's behaviour, while distal variables or distal genotypes are those

[13] This view is in good agreement with the biological analogy on which the phenotype / genotype distinction is based.

that have an indirect influence.[14] In practice we are only interested in the proximal genotypes, i.e., the causes for which a direct link can be established to the event characteristics.

The second category consists of the genotypes that are linked to the technological system, in particular to the state of the system and to state changes. This category includes everything that has to do with the state of components, failure of components and subsystems, observable changes, etc. It also includes everything that has to do with the man-machine interaction, the man-machine interface (information presentation and control), etc. A further distinction can be made between causes that have to do with technological hardware, and causes that have to do with software. This might define a more detailed set of classification groups, but will not be pursued further here.

The third category contains the genotypes that characterise the environment and the interaction between people. Examples could be permanent features of the system (whereas temporary features would be included in the second category), aspects of the organisation (the local or the global organisation), and environmental conditions such as noise, temperature, etc. The third category might be seen as a garbage can for genotypes that do not belong in either the first or the second categories. In reality it is much more than that. Environmental causes are important in their own right, and human erroneous actions can only be explained fully by referring to the combination of personal, technological and environmental causes.

The role of the three main sub-categories is shown in Figure 15.

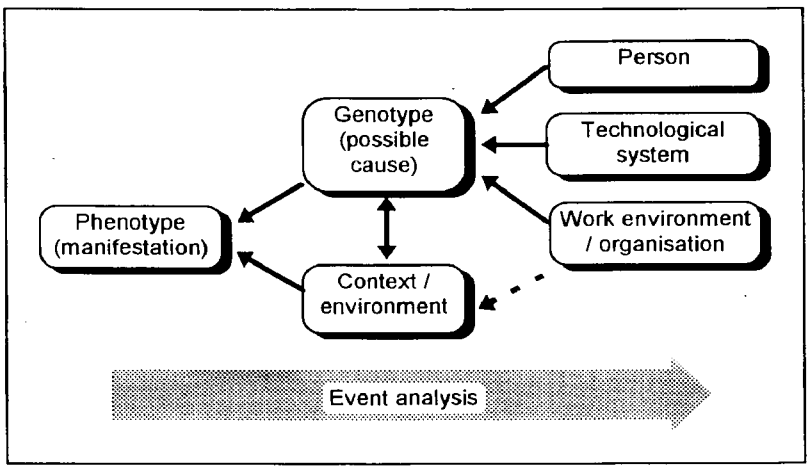

Figure 15: Differentiation of Genotypes.

The three main sub-categories can be further expanded to support a more detailed analysis. As shown in Figure 16, the genotypes make a distinction between person related causes, system related causes and environment related causes. In the person related causes a further differentiation is made between specific functions - which in turn refer to the underlying cognitive model - and general functions which can be either temporary or permanent. In the

[14] The use of proximal and distal variables does not completely solve the problem, since the terms require their own definition of what is meant by direct.

system related causes, a distinction is made between components, procedures, and interfaces - the latter further being divided into temporary and permanent causes. Finally, in the environment related causes major subgroups are communication, organisation, and ambient conditions. The details of these classification groups will be described in the following.

In addition to the finer differentiation among the genotypes, a small change has also been made to the phenotypes. This is simply to add the category of general error consequences, i.e., the effects of the event in the system being analysed. The error modes denote characteristics of actions that the operator has made, while the error consequence describes what the effects are in the system. The error analysis will not lead to a specification of the consequences, but the consequences will in many cases be the starting point for the analysis. A typical analysis would begin either by a consideration of the error consequences (as would be appropriate for an incident analysis) or from the perspective of probable error modes (as when sources of potential risk are assessed).

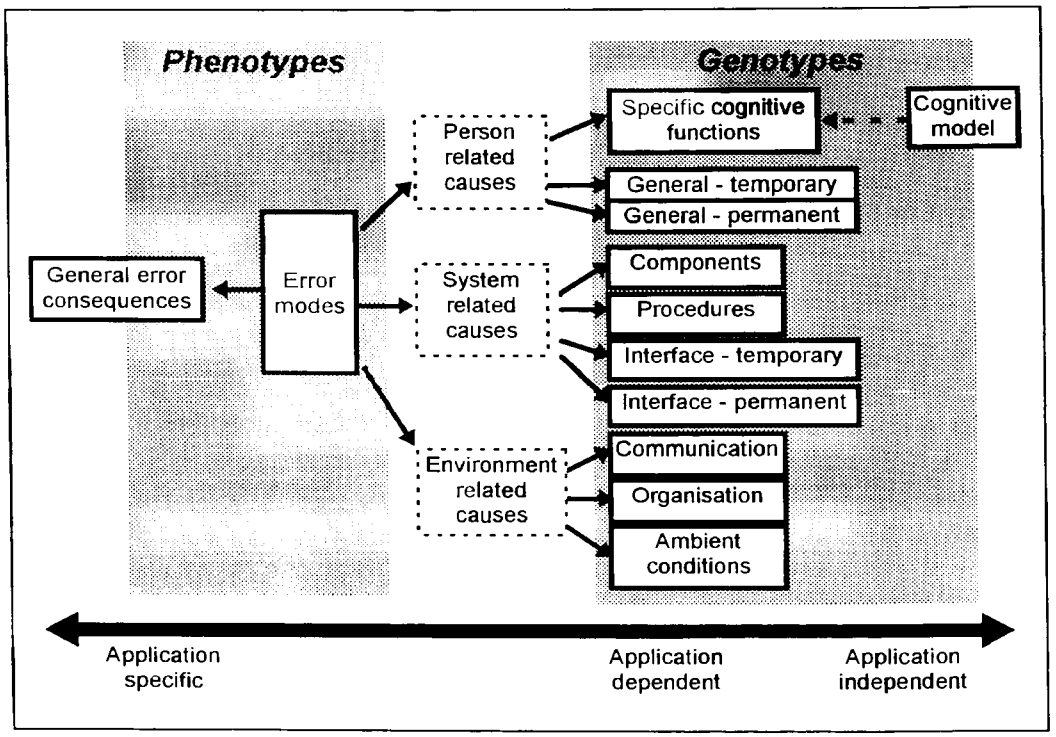

Figure 16: Overall Grouping of Phenotypes and Genotypes.

Figure 16 also indicates how each category in the classification scheme depends on the context. This is important for a predictive analysis, and also for appreciating the importance of various phenotypes and genotypes. Clearly, the error consequences are completely application specific. Thus, the consequence of forgetting an action will depend on which system or which task is being considered, for instance whether it is a blast furnace or an aircraft. The main phenotypes and genotypes are application dependent, i.e., the details of the classification scheme may vary according to the application (this will be discussed in greater detail in the following). It specifically means that it is not feasible to develop a completely general

classification scheme. The categories must always be specific either to a particular application or to a type of applications - for instance aviation or nuclear.

Finally, a subset of the genotypes will be application independent. This is, for instance, the case for the genotypes that are part of the cognitive model, such as the basic cognitive functions. The cognitive functions are typical of the human operator, and will therefore be potentially present in all situations and for all applications. The cognitive functions may express themselves differently depending on the application and the context, but they will in principle always be there. They can therefore be described in more general terms, and do in fact provide the link between the classifications scheme and the generally accepted psychological facts.

As shown in Figure 16, the categories "person related causes", "system related causes", and "environment related causes" do not belong to the genotypes. Indeed, in the classification scheme proposed here these three categories are empty and only serve as a convenient labelling for a group of more specific categories. Similarly, neither the cognitive model nor the general error consequences belong to the classification scheme. The cognitive model is the basis for the "specific cognitive functions" category, while the general error consequences describe the phenomenology of the application. The genotypes are consequently not hierarchically structured. This has important consequences for the associated method.

The classification scheme shown in Figure 16 is based on the outcome of several efforts and projects reported in e.g. Hollnagel et al. (1990), Hollnagel & Cacciabue (1991), Cojazzi et al. (1993). It is also being used as part of the development of **CREAM**.

5.2 Classification Groups

The classification scheme does not take the form of a strict hierarchy of classes and subclasses. There are two reasons for this. Firstly, there is not enough knowledge about the causes of human actions to produce a consistent hierarchical classification. This goes for human actions seen in isolation, and even more so if human actions are seen as part of a context. Secondly, a hierarchical classification system forces the analysis to become strictly sequential, i.e., to go from one end of the hierarchy to the other - from top to bottom or from bottom to top. This means that the depth of the analysis is pre-defined, so to speak, and also that the transitions between categories is given in advance. It follows that a hierarchical classification system either must be correct for all applications or domains, or be limited to a specific range of applications - which in turn means that there must be more than one classification system. But since a hierarchical classification system must reflect an underlying ordering principle, it is difficult to see how there could be several different hierarchical classification systems.

Instead of having the classification scheme as a strict hierarchy, it will be composed of a number of **classification groups**.[15] The groups shall reflect the principles of differentiation or specialisation described above (cf. Figure 16). The number of groups shall be so large that all reasonable sets or clusters of causes can be recognised. This is important to maintain sufficient overlap between the classification scheme proposed here and already existing classification

[15] In previous work these groups have been referred to as **tables**. We will, however, make a distinction between the classification groups as conceptual entities, and classification tables as the practical implementation.

schemes. The number of groups shall, however, not be so large that it creates a problem for their application in an event analysis. Performing an analysis must represent an acceptable trade-off between the level of detail and the amount of work required. If there are too many classification groups, the task of managing them and maintaining an overview of the analysis may become prohibitively large. If there are too few classification groups - say, only two - it becomes comparatively easy to do the analysis, but the value of the outcome may be limited.

Based on the summary provided by Figure 16, we can define the following classification groups.

- **Error modes.** The error modes describe the manifestations on the level of observable behaviour. Error modes can refer either to the phenomenal feature of the time-space continuum or to the systematic phenotypes (Hollnagel, 1993a).

The group "person related causes" is a label for all causes that clearly relate to the person or the user. The group itself is empty; the detailed causes can be found in the following three groups:

- **Specific cognitive functions.** The specific cognitive functions must reflect the principles of the underlying model of cognition. In the present work the basis will be taken in the Contextual Control Model (COCOM), as described above.

- **General person related functions (temporary).** Temporary person related functions are typically psycho-physical states - or emotional states - that are characteristic of the person at a given time. A classical example from this classification group is circadian rhythm or time pressure.

- **General person related functions (permanent).** Permanent person related functions are constant person characteristics, as for instance colour blindness.

The group "system related causes" is a label for all causes that clearly relate to the technological system. The group itself is empty; the system related causes are described by four detailed categories:

- **Procedures.** This classification group refers to the existing procedures or prescriptions for how a task shall be performed. It may conceivable overlap with the "organisation" group.

- **Components.** This group refers to the purely technological elements, such as mechanical or electronic components (including software), sub-systems, control systems, etc. There may be a potential overlap with temporary interface causes.

- **Interface (temporary).** This group describes causes that come from temporary conditions relating to the man-machine interaction, such as failure of information presentation, limited access to controls, etc.

- **Interface (permanent).** This group describes causes that come from permanent features of the man-machine interface, typically design flaws or oversights.

The group "environment related causes" is a label for all causes that clearly relate to the environment, such as the working conditions in general. The group itself is empty; the environment related causes are described by three detailed categories:

- **Communication.** This classification group refers to everything that has to do with the communication between operators, or between an operator and the technological system. There is a potential overlap with the temporary interface group, and with the ambient conditions group.

- **Organisation.** This classification group refers to causes that have to do with the organisation in a large sense, such as safety climate, social climate, reporting procedures, lines of command and responsibility, etc.

- **Ambient conditions.** This final classification group refers to causes that characterise the working conditions, such as temperature, time of day (or night), noise, etc. These are all factors that have an impact on the well-being of the operator, hence on his efficiency.

5.3 Details Of Classification Groups

The classification groups describe the possible error modes and the probable causes. In principle, each group could just list the error modes or causes that were relevant, i.e., as an unstructured list. There are, however, two important considerations to keep in mind.

Firstly, since the classification scheme is not organised as a strict hierarchy, it is necessary to provide a different principle to link the classification groups. This can be achieved by noting that an **effect** of one classification group may also appear as a **cause** of another group, and vice versa.

Secondly, when an analysis is made there may be varying amounts of information available. In some cases it will be possible to describe details of the event and be very specific about possible causes. In other cases it will only be possible to describe the event in broad terms, and hence not possible to be precise in the analysis or in the identification of causes. In order to account for this, the concepts or terms in the classification groups should exist on different levels of specificity. In practice, this requires at least two levels, generic and specific. In combination with the causes and effects categories this produce four categories called **general causes**, **specific cause**, **general effect**, and **specific effect**.

In the following the details of the classification groups are described, using the classification groups listed above. A description of the links between the classification groups follow later in the report.

5.3.1 Error Modes (Basic Phenotypes)

The error modes are the categories which describe the ways in which an incorrect action can manifest itself, i.e., the possible phenotype. Since actions must take place in a four-dimensional time-space continuum, it is possible to define an exhaustive set of error modes. (The same is not the case for the causes.) The possible error modes are illustrated in Figure 17.

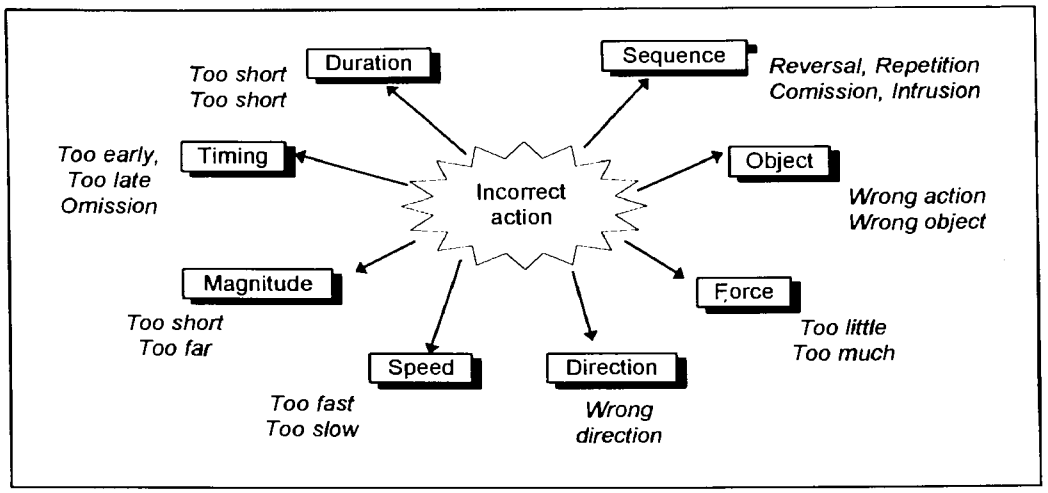

Figure 17: Dimensions of Error Modes.

Rather than define eight different classification sub-groups for the error modes, it is practical to divide them into the following four sub-groups.

- **Action at the wrong time**, which includes the error modes of timing and duration. The contents of this sub-group is shown in Appendix A, Table 2.

- **Action of wrong type,** which includes the physical characteristics of force, magnitude, speed, and direction. The contents of this sub-group is shown in Appendix A, Table 3.

- **Action at wrong object**, which only includes the error mode of object. The contents of this sub-group is shown in Appendix A, Table 4. This is, in principle, a question of wrong direction and wrong magnitude, since that can be used to account for a position in three-dimensional space. However, it makes sense to use the simpler description in terms of the wrong object.

- **Action in wrong place**, which only includes the error mode of sequence. The contents of this sub-group is shown in Appendix A, Table 5.

Finally, an action can of course also go as planned, corresponding to the category of **no erroneous action**. This can either be included in each of the classification groups, or kept as a separate group - depending on how the analysis is performed. In this report the latter option has been chosen (Appendix A, Table 6).

5.3.2 Person Related Causes

First among the person related causes are the specific cognitive functions, which must refer to the underlying model of cognition. The detailed model of cognition used in this classification scheme was described earlier in this report.

It is fundamental trait of human cognition that it is **covert**, i.e., that it cannot be observed. While there is general agreement about the main characteristics of human cognition, both in terms of types of functions and in terms of functional characteristics - and in particular

limitations - there is less agreement about the details. It seems that the more detailed a model of cognition is, the less likely it is to be correct. The reason for that must be found in our limited knowledge of human cognition, despite more than a century of psychological research and experimentation.

For the purpose of an error analysis it is, of course, highly desirable to be able to identify the cognitive functions that may be the causes of observed actions. The analysis is made to determine **why** an event occurred and to find out **what** can be done about it. Consequently, the causes should be described on a level where they can be used as a basis for recommending changes. This means that they should not refer to the specific details of a theory of cognition - or a model of cognition - unless this model has a strong link to actual performance. Causes that refer to hypothetical cognitive mechanisms should therefore be avoided, since hypothetical cognitive mechanisms have limited practical applications. The cognitive functions that are used as a basis for the classification should also be arguably correct, i.e., not too speculative. For this reason, the division into specific cognitive functions is kept as simple as possible.

The cognitive functions that are the basis for thinking and decision making can be described in many different ways. One of the simplest is to differentiate between **analysis** and **synthesis**. Analysis refers to the functions that are used when a person tries to determine what the situation is, typically including observation, identification, recognition, diagnosis, etc. Synthesis refers to the functions that are used when a person tries to decide what to do and how to do it; this typically includes choice, planning, scheduling, etc.

In the present classification scheme the category of analysis includes observation and identification. It thus describes all aspects of receiving data and information from the process, either reacting or responding to signals or events or actively looking for information. The details are provided in Appendix A, Table 7.

Interpretation is used as a common term for understanding, diagnosis, and evaluation. It thus refers to a group of cognitive processes that have to do with the **analysis** of the observed information. This group may possible be divided into further detail (sub-groups). Current classification schemes have usually been quite rich in referring to the various facets of analysis. The difficulty with a very detailed classification is, however, to specify the links and dependencies between the sub-groups. The details of the Interpretation group are provided in Appendix A, Table 8.

Corresponding to the analysis there is a phase of **synthesis** in which the results from the analysis are used to develop a specific line of action (or just an intention to act). This group includes all functions that have to do with setting out the detailed course of action, i.e., choosing and scheduling. It may be a matter of belief - or preference - whether the actual decision or choice is put together with analysis or with synthesis. In the present classification scheme the choice has been included in the analysis group. The reason is that it may be possible to analyse and interpret a situation without actually doing anything. A process of synthesis is needed before the choice is turned into actual actions. The details of the Planning group are provided in Appendix A, Table 9.

General person related functions are not directly linked to a specific cognitive function. It is common to make a distinction between **temporary** (transient, sporadic) and **permanent** person related functions. The temporary functions only exist for a short period of time, hence

do not exert a constant influence on performance. The details of the temporary person related functions are provided in Appendix A, Table 10.

The permanent person related functions are present in all situations, hence exert a constant influence. In some cases a specific function may belong to either group - but never to both at the same time. Memory problems, for instance, can be either temporary or permanent. In the current version of the classification scheme memory problems are considered as a temporary cause. Were they permanent they might express themselves as a cognitive style, since presumably the operator would learn to cope with them by adopting a suitable strategy. The details of the permanent person related functions are provided in Appendix A, Table 11.

5.3.3 System Related Causes

The system related causes include everything that can be traced directly to technological aspects or parts of the system. This includes in particular the technological malfunctions that may occur, inadequacies of the operational support systems - and in particular of procedures - and general issues of the interface.

A conspicuous feature of the technological system - the process, the interface - is that it may fail due to problems with the hardware or the software, together referred to as equipment failures. The failure of a component or subsystem is usually one of the initiating causes for an event, but is rarely a contributing cause for a human erroneous action. Failures of the interface may, on the other hand, directly affect human performance. The details of the equipment failures are provided in Appendix A, Table 12.

Much of human performance in an industrial setting is guided by procedures. Experience has shown that deficiencies in procedures or discrepancies between procedures and the working environment can be an important cause for human erroneous actions. This is therefore included as a separate classification group, as shown in Appendix A, Table 13.

A properly functioning man-machine (human-computer) interface is an important prerequisite for the operator's ability to perform the tasks in an adequate fashion. The interface can temporarily malfunction in several ways. Whereas explicit malfunctions of physical equipment are included in Appendix A, Table 12, the contents of the temporary interface failure are shown in Appendix A, Table 14. Similarly, interface problems of a more permanent nature - often due to deficiencies in the design, or in modifications to the system which have made the design assumptions invalid - are shown in Appendix A, Table 15.

5.3.4 Environment Related Causes

The third main part of the classification includes all those factors that can be attributed to the environment, rather than to the operator or the technological system.

The first group contains the factors that have to do with communication, i.e., with the exchange of information among the operators or between operators and sources outside of the control room. These are shown in Appendix A, Table 16.

The second group contains the factors that relate to the organisation as such. This group of factors can easily become exceedingly large. It is, however, important to keep in mind that the

purpose of this classification scheme is to identify causes that are proximal genotypes, i.e., which have a direct impact on the operator's performance. This may serve to limit the number of organisation factors. A representative set is shown in Appendix A, Table 17.

The final group of factors refer to the general working conditions, here called the ambient conditions. These could conceivable be divided into the social and the physical environment. For practical reasons the former is included in the Organisation group. The details of the Ambient Conditions group are shown in Appendix A, Table 18.

5.4 Summary

The classification scheme described in the preceding sections is composed of phenotypes and genotypes. Both phenotypes and genotypes are further divided into more detailed classification groups. Each classification group makes a distinction between: (1) **general manifestations**, (2) **specific manifestations**, (3) **general causes**, and (4) **specific causes**. There are no pre-defined specific relations between the classification groups. This means that the classification scheme is not hierarchically organised. Instead of the hierarchical organisation, explicit links between the categories in the groups are defined, as described below.

The phenotypes are the description of the error modes. The error modes are divided into four classification groups called: (1) **action at wrong time**, (2) **action of wrong type**, (3) **action at wrong object**, and (4) **action in wrong place**.

The genotypes are the description of the error causes. The error causes are divided into ten different classification groups, which in turn can be assigned to three main groups. One main group is **person related causes** which is further divided into: (1) **observation**, (2) **planning**, (3) **interpretation**, (4) **temporary person related causes**, and (5) **permanent person related causes**. The first three groups refer to the underlying model of cognition. The second main group is the **system related causes**, which is further divided into: (1) **components**, (2) **procedures**, (3) **temporary interface problems**, and (4) **permanent interface problems**. Finally, the third main group is the **environment related causes**, which is divided into three more detailed groups: (1) **communication**, (2) **organisation**, and (3) **ambient conditions**. The overall structure of the classification scheme is shown in Figure 16.

6. LINKS BETWEEN CLASSIFICATION GROUPS

The separation of control and competence recognises that cognitive functions evolve in a **context** consisting of past events as well as anticipated future events. This contrasts with a strictly sequential modelling of cognition, where one action follows the next in a pre-defined pattern. The principles of COCOM, shown in Figure 5, make it is possible to explain how the execution of a particular action, for instance, can be preceded (or caused) by planning, by interpretation, or by observation, depending on the context and the mode of control. It is these causal connections that must be unravelled by the analysis. There is no *a priori* defined causal chain which links the different cognitive functions. In order to perform an analysis it is therefore necessary to begin by establishing an understanding of what the likely context is. From this it should be possible to infer the likely mode of cognitive control.

> There is a clear difference between using the COCOM as an approach to modelling of cognition, and using it as the basis for an analysis of human erroneous actions. In the former case it is necessary to account explicitly for how control is implemented and which specific mode of control one should assume for a given set of conditions. This is obviously a task that requires substantive effort. In the latter case, for the analysis, the impact of using COCOM is less demanding in terms of effort. It simply means that the analysis is released from the need to follow a predetermined route through the classification scheme. In the lack of any specific information the analysis will proceed on the basis of a standard or default context. But it can obviously be made more precise if the conditions of the event can be provided.

Using the principles described above, the retrospect analysis of error events, or alternatively, the prediction of human reliability, is based on a distinction between causes, effects, and consequences. **Consequences** are the systemic results of operator actions that are being performed. In a reliability analysis, the emphasis is naturally placed on potential unwanted consequences that result from erroneous actions. Such outcomes are typically described in terms of common error modes (i.e., the erroneous actions are usually not sufficient themselves to cause an accident, but are important as an intermediate stage in the analysis). In a retrospective event analysis, the consequences usually form the starting point. Considerations of the **causes** of erroneous behaviour, on the other hand, involve the determination of the occurrence of a certain "inappropriate" effects. In the classification scheme - and in **CREAM** - each observable action can have one or more causes. For example, the cause can be an external event or, more likely, an intervening cognitive function. An action is, however, rarely the result of a single cognitive function but rather relates to a complex of functions, each of which also has one or more causes. Thus, a misinterpretation may result in an incorrect diagnosis, which may lead to an inappropriate plan, which may lead to an erroneous action. Causes and effects are thus used to account for how cognitive functions depend on each other; in the sense that the effects of one function are the cause of the next - and *vice versa*.

It is useful to make a distinction between general and specific causes, and general and specific effects (see Figure 18). The use of both general and specific causes / effects serves to simplify the analysis. If a specific cause or effect is known or if it can be predicted, then it should clearly be used. If, however, there is insufficient information available the general causes / effects should be used as defaults. The specific causes / effects will, of course, be more precise than

the general ones, but will also require supplementary knowledge and effort. It may therefore be reasonable to begin a predictive analysis with the general causes / effects and only proceed to the specific ones when it is needed. This will, among other things, prevent the analysis from becoming unnecessarily complex. In the case of a retrospective analysis, the starting point is a specific event or outcome. It can normally also be assumed that sufficient details will be available to allow the identification of specific causes.

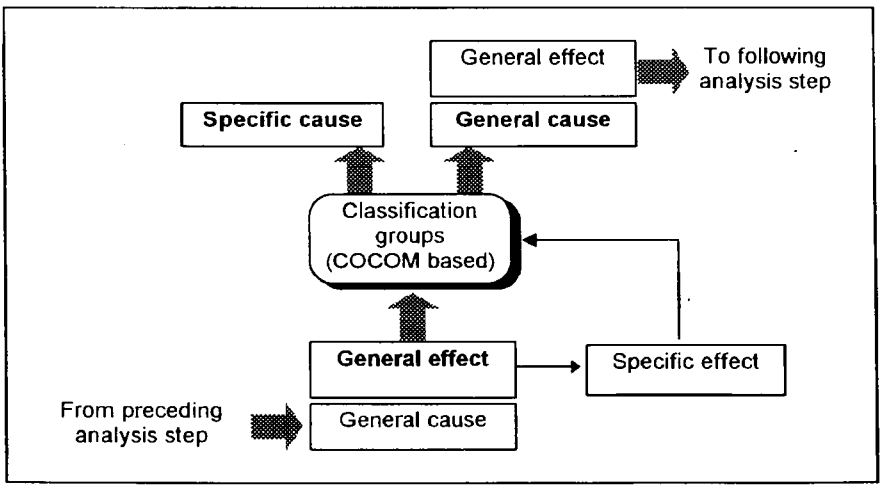

Figure 18: Links Between Effects and Causes for a Retrospective Analysis.

The analysis method assumes that there is a relation between the effects of a cognitive function at any stage in the development of an intention / action and the outcome of the cognitive functions that follow it. This is due to the fact (in the model) that cognitive functions do not have pre-defined links. The specific relations are based on the classification scheme of error modes, causes, and effects as described in separate classification groups. The groups describe, for instance, how the effect of an erroneous interpretation can become the cause of inappropriate planning; the inappropriate planning may in turn become the cause of an inappropriate action (consequence). The causal relations between the cognitive functions are very important for how an analysis should be carried out for a real context, as it will be seen below. With this framework in mind it follows that erroneous actions (error modes) can have many different causes and that the actual instantiation of the causes is important to understand how inappropriate behaviour can come about. The causal relations create the conditions for capturing the dynamic aspects of the MMI which are crucial to carry out a comprehensive qualitative analysis as part of the error and risk management efforts.

As shown in Figure 18, the link is between general causes and general effects. In a retrospective analysis, the starting point is a specific effect - i.e., the observed consequence. The analysis moves backwards step by step, until a probable general cause has been found or until a specific cause has been found. In the classification scheme there are no links **from** specific causes; once a specific cause has been found, the analysis has come to a conclusion. General and specific effects may, however, co-exist, i.e., it may be possible to identify a specific effect in addition to a general effect. Both will, however, point to the same general cause.

7. CONTEXT DEPENDENCE OF CLASSIFICATION GROUPS

The context dependence of the various groups has already been touched upon in the discussion relating to Figure 16. All the genotypes are context dependent, although the degree of dependency may vary. In the two cases where a distinction was made between temporary and permanent groups (general person related functions and interface functions), it is reasonable to assume that the groups of temporary factor are more context dependent than the group of permanent factors. For the other groups the variation within groups is probably larger than the variation between groups.

The degree of context dependence of the groups and the factors is important for both analysis and prediction. An analysis will always be easier to make and yield more precise results if the categories and concepts used have a specific relevance for the current conditions. It is **easier** because there is no need to include causes that clearly are irrelevant or impossible. It is **more precise** for almost the same reason, i.e., that the causes used to explain the event are appropriate for the event and the working conditions.

In a retrospective analysis the context of the event is known in advance, and this may be used to select a subset of classification groups and / or causes that is particularly relevant. In the case of a prediction no such prior selection is possible, because the nature of the situation is unknown. There are, however, ways in which the context dependence can be ascertained.

7.1 Possible Manifestations And Probable Causes

As mentioned several times in the preceding, the classification scheme and the classification groups are not intended to be used rigidly as they are described. Instead, groups of causes and effects should be adjusted to match the current conditions of the event that is being analysed. The event analysis must therefore start by defining or characterising the context, go on to describe the likely error modes, and only then describe the possible causes.

In a retrospective analysis, the context is defined as the conditions that existed for the event that is being analysed. The first step should therefore be relatively easy to accomplish. It may, however, be useful to try to formalise the description of the context, i.e., to bring it on a common form. This will ensure both that all necessary information is available, and also that possible comparisons between different analyses are made easier. A suggestion for the common form is the set of Common Performance Conditions (CPCs), as described by Hollnagel (1993b). Briefly stated, the CPCs include the categories shown in Appendix A, Table 1. The table also shows a possible set of descriptors for each CPC, which at least will provide a minimal level of discrimination. Although the CPCs have been limited to the basic characteristics of man-machine systems, they may still require some work to be filled out. The information that is provided by an event report, for instance, may not be sufficient to describe all the CPCs. An event report concentrates on the actual event, but the CPCs also include information about the general conditions and the general quality of the system / organisation. Filling out the CPCs may therefore provide a useful complement to the event report, which in turn may facilitate the event analysis and the identification of the causes.

When the Common Performance Conditions have been determined, the next step is to identify the possible error modes. As far as the initiating event goes this, of course, describes itself. If

the initiating event is the general error consequence (e.g. controlled flight into terrain), it must be related to the possible error modes. If the initiating event is an observed or inferred action, it is still important to limit the number of possible error modes to avoid possible incorrect classifications.

In most cases the analyst can make a distinction between error modes that are **possible**, i.e., which can actually occur, and error modes that are **impossible**, i.e., which cannot occur. An error mode may be impossible because of the working context, the design of the interface, the nature of the task, etc. For example, if the input to the system is provided via touch panels, it is impossible to do something with too much or too little force; the action is either registered (i.e., done) or not registered and does not depend on the pressure exerted. Similarly, the functionality of the control system may make it impossible to do something with the wrong speed, in the wrong direction, etc. In addition to excluding certain error modes as impossible, this preparatory step can also be used to define more precisely the conditions for certain error modes. In the case of incorrect timing it is, for instance, important to be able to set the limits for what is too early and what is too late. Depending on the type of application, the limit may be in terms of seconds or minutes.

Another way of refining the possible manifestations is to increase the precision of the descriptors used for the error modes. For instance, in the case of timing it is important to say precisely **when** an action is too early or too late, whether the critical limit is five seconds or ten minutes. The same goes for the other error modes. Wherever these refer to physical dimensions or to time, the analyst should replace the general descriptors with specific ones.

When it comes to error causes the same clear distinction between what is possible and impossible cannot be made. The permanent person related causes do, of course, not depend on the context. The very fact that they are permanent means that they will be the same across different conditions. The temporary person related causes, as well as the specific cognitive functions may, however, be influenced by the context. In the case of specific cognitive functions it is not possible to exclude any of them as potential causes. Any cognitive function can play a role in the explanation of how an event occurred. There will, however, be some causes that are more **probable** than others. The same goes for the temporary person related causes. If, for instance, the specification of the Common Performance Conditions have shown that the adequacy of the man-machine interface is only tolerable, then the causes that involve the MMI are more likely. The difference may not be quantifiable, but is perhaps rather a difference in priority or possibility. In any event it means that the probable causes are the ones that should be examined first, although all causes should remain included in the classification groups.

The specification of possible modes and probably causes is a way of making the analysis more efficient and more precise. Without this sharpening of the categories, the analysis will always have to go through the full set. This will not only require effort that is unnecessary, but also fail to use the information that is available in terms of a description of the context.

8. ANALYSIS METHOD

As described in the beginning, the three main elements of a system to support the analysis of accidents and events are a model, a classification scheme, and a method. This chapter will provide an overview of the method as it can be applied to retrospective analysis. Further details about a method for prediction will be developed in the second half of this project.

8.1 General Analysis Method

The basic principles of the analysis is to start from the description of the initiating event and go backwards step by step until a reasonable cause - or set of causes - has been found. The term "reasonable" is used deliberately, since even a root cause on closer inspection will reveal itself as being relative rather than absolute (Cojazzi et al., 1993b).

An important part of the analysis method is the **stop rule**, i.e., the criteria which are used to determine when the analysis has gone far enough. In a hierarchical classification system or a taxonomy the stop rule is implicitly given: when the analysis reaches the level of terminal nodes (leaves), then it has by definition been completed. Conversely, if the terminal nodes have not been reached, then the analysis must continue.

Since the classification scheme proposed here is not hierarchical, it is necessary to provide an explicit stop rule. This can be done by pointing out the difference between terminal and non-terminal causes. All classification groups are described in terms of general and specific causes. Of these, the specific causes are **terminal causes**. This means that a specific cause is seen as being sufficient in itself, and that it therefore does not refer to a preceding cause. In the present classification system "mismatch to actual equipment" is a specific cause of "error in procedure". Thus, when the analysis reaches "mismatch to actual equipment" it has reached a natural end point.

> It may, of course, be possible to continue the analysis. This will require that the classification scheme is extended. In that case "mismatch to actual equipment" should be a general cause rather than a specific cause, and it should be matched by one or more entries under general effects in other classification groups. But even for the extended classification scheme the main principle is that a specific cause terminates the analysis.

In contrast to the specific causes, the general causes are considered to be **non-terminal**. This means that it is possible to continue the analysis from a general cause. The principle is that a general cause matches a general effect in one or more of the other classification groups. (This is actually one of the defining characteristics of the classification scheme and of the separation between general / specific effects and causes.) If, for instance, the general cause of the error mode "wrong object" has been determined as "access difficulty", then the analysis can proceed by looking for classification groups where "access difficulty" is among the general effects. This would in the current version lead to the group for temporary interface characteristics, where "access difficulty" is found. Since "access difficulty" can have either two general or five specific causes, the analysis can either stop at this level (if one of the specific causes is chosen), or continue one level further if the general cause is chosen.

In general, the analysis comes to a halt when it is not possible to continue further, i.e., when there are no general causes for a given effect. If there are specific causes, then one of these should be chosen - if sufficient information is available to warrant that choice. If no specific causes are provided, then the analysis stops with the last general cause. The reason why the analysis did not simply stop at the preceding level, i.e., by changing the general cause to a specific cause, is that the general cause, seen as a general effect, may include some specific effects. Noting what these are may be important to provide a full description of the event analysis.

8.2 Specific Analysis Method

In the following a description is given of the detailed analysis method for retrospective analyses. The description is aimed to be sufficient for a manual analysis, although it will clearly be more efficient if the analysis can be supported by a computerised version of the classification scheme. Such as computerised tool has been developed as part of **CREAM**.

In accordance with the basic principles of context dependent analysis, the first steps are to describe the context and refine the classification groups. This is achieved through the following steps.

1. **Determine or describe the context.** This is done using the CPCs. It may require a detailed analysis of aspects of the application that are not usually taken into account or which are not contained in the event report.

2. **Describe the possible error modes**. This description is to be given for all possible actions, i.e., without considering a specific action. The description uses the knowledge of the application and the context to produce a limited set of error modes, and also to spell out the criteria for certain error modes (e.g. when is an action too late).

3. **Describe the possible error causes.** The description of error causes can be used to identify causes that are more probable in the given context. In the case where error causes are cognitive functions, it is not possible to rule out any of them. For any given context there will, however, be some that are more likely than others. Thus the context may enforce compliance with rules, encourage deviations, support learning of skills, have a bad interface, hence promote misunderstandings or execution errors, etc.

4. **Perform the more detailed analysis of main task steps**. This step will try to trace the possible causes for the noted error modes. The analysis looks both at the error modes and the causes, trying to find the most reasonable links. The causes are both the internal causes and external causes (system events, system conditions) which have been described as a part of the CPC development.

The principal steps of this method are illustrated in Figure 19 below. It is, however, necessary to expand the description of the last step, i.e., the detailed analysis of main task steps.

The detailed analysis of a main task step begins by looking for the most likely error mode. This is based on a description of the initiating event, including a description of the Common Performance Conditions. The error mode is taken from the four classification groups that describe error modes: (1) action at wrong time, (2) action of wrong type, (3) action at wrong object, and (4) action in wrong place. If there is sufficient information available, the general

error mode (general effect) may be supplemented by a description of the specific error mode (specific effect).

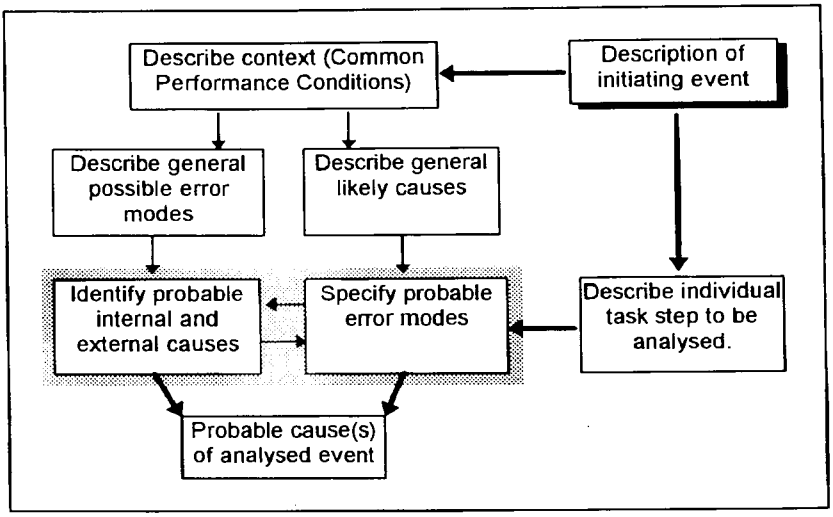

Figure 19: Overall Method for Retrospective Analysis.

Once the general / specific error mode for the initiating event has been described, the analysis can proceed to find the causes. The next step is to select one of the causes linked to the error mode. This can either be a general cause or a specific cause.

- If the outcome is the identification of a specific cause, then the analysis has been completed. It is quite conceivable that the analysis only needs to go one step to find the case. If, for instance, the general effect was "wrong object" and the specific cause was found to be "incorrect label", then the analysis has been completed in a single step.

- If the outcome is the identification of a general cause, then the analysis must proceed. The next step is to check the classification groups to see if there are any general effects that match the general cause.[16] In other words, the general cause on one level of the analysis must match a general effect on the next level, as illustrated in Figure 18. When a relevant general effect has been found, the analysis continues from there.

The analysis can continue by supplementing the general effect with the selection of a specific effect, provided that the necessary information is available. When that has been done, the relevant general cause(s) or specific cause(s) are identified. As before, if the outcome is the identification of a specific cause, then the analysis has reached its end. Similarly, if there are no general causes - in which case there most probably also are no specific causes either - then the analysis must stop. In all other cases the outcome will be a general cause, and that is then matched with the general effects of the classification groups, as described before. In this way the analysis continues by applying the same principle recursively, until a stop criterion is reached. The principles of the detailed analysis are shown in Figure 20.

[16] It is guaranteed that there will always be a match, since this is one of the construction criteria for the classification groups and the classification scheme as a whole.

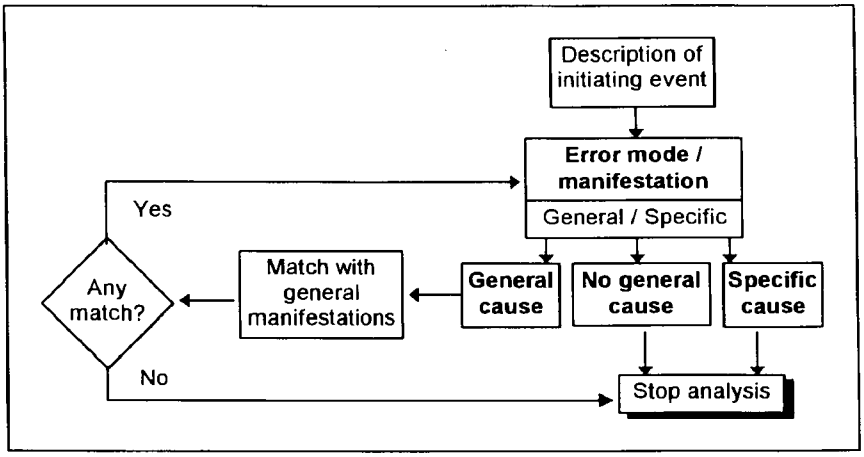

Figure 20: Detailed Method for Retrospective Analysis.

The use of the retrospective analysis is illustrated in Appendix B.

9. PRINCIPLES OF PERFORMANCE PREDICTION

Predictions of the future are never anything but projections of present automatic processes and procedures, that is, of occurrences that are likely to come to pass if men do not act and if nothing unexpected happens; every action, for better or worse, and every accident necessarily destroys the whole pattern in whose frame the prediction moves and where it finds its evidence.

Hannah Arendt

The quotation above refers to political issues and affairs of the state, but is in many ways also pertinent for performance predictions and HRAs. In HRA, and particularly in relation to PSA, the whole purpose is to predict what will the outcome will be if nothing unexpected happens, which is another way of saying that there must be no discrepancy between the models and reality. Fortunately, the whole foundation is not destroyed when an accident happens, although some accidents do give rise to comprehensive reformulations of the conceptual foundations.

The art of performance prediction is to describe what is likely to happen **if** a specific initiating event occurs, **and if** the model of the world bears an acceptable correspondence to the real world. In the situation addressed in this study, the interest is focused on human action, hence on the underlying model of the person (the operator, the user, etc.), rather than on the model of the process or the interaction.

9.1 The Role Of Context

The context refers to the circumstances in which an event occurs. In order to understand how an event **has** developed or how an event **will** develop, it is necessary to know the context. In the case of event analyses or accident analyses, the first step is always to describe the situation or the circumstances. The analysis is an attempt to find the most probable causal chain, going backwards from the observed event. The main reason why this is possible is that there is information available about the conditions in which the events took place. In the investigation of a plane crash, for instance, a lot of effort is put into describing the context, almost on a second-by-second basis. Only then will it be possible to identify the most likely cause (Cacciabue et al., 1993).

In the case of performance prediction, context description must also be the first step. It stands to reason that we can only discuss what is likely to happen if we know what the circumstances are likely to be. Human action, after all, is not spontaneous or stochastic, but (usually) intentional and directed towards a specific goal. The goals that people assume and the ways in which they try to reach them will all depend on the context.

Figure 21 shows the principle of retrospective and predictive analyses relative to the context. An **event** or **accident analysis** is concerned with events (accidents) that have occurred and tries to find the most probable causes (root causes). The purpose is to understand better something that has already happened. Because the context is known, it is possible to follow the event step-by-step in the direction **from** the focal event (the observed consequence) **to** the probable cause (Cojazzi & Pinola, 1994). If the information is incomplete, inferences about

missing details can be made with great certainty. Note, however, that the links between the focal event and the probable cause(s) are not in the simple form of an event tree, but rather in the form of a network. The event tree rather can be seen as the instantiation of a limited set of paths through the causal network.

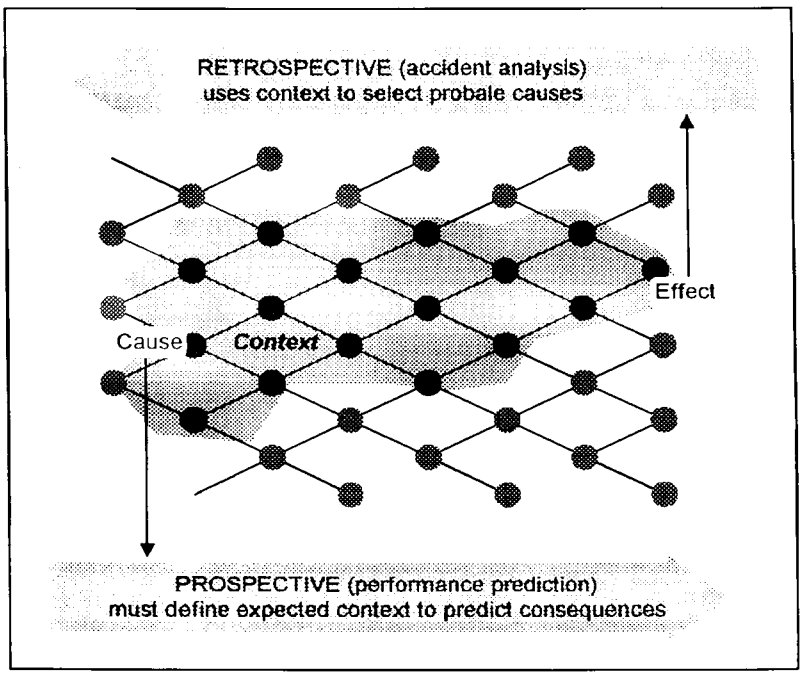

Figure 21: The central role of context in performance analysis and prediction.

Performance prediction is concerned with describing how an event may possibly develop, in particular how the variability of human performance may influence the propagation of events. The purpose is to find out what **may** possibly happen under given conditions, e.g. if a component fails, if there is insufficient time to act, or if a person misunderstands a procedure. On the level of individual behaviour, each action will generate so many possibilities that the total quickly becomes unmanageable. This is because a mechanical combination of taxonomic categories inevitably leads to a combinatorial explosion. The focus can be improved only if the context can be defined. Performance prediction must therefore specify the context that is most likely to exist, before it specifies the actions that will occur.

In principle the task is "simply" to find a path between causes and consequences. This path, however, only exists for a given context or set of conditions. If the conditions are even slightly changed, the path may look completely different. **The basic prerequisite for performance prediction is therefore that a probable context has been described.** The essence of a predictive analysis, such as a predictive HRA, should therefore be to estimate the probable performance conditions rather than to predict specific events!

9.2 Performance Prediction In First-Generation HRA

The traditional approach to performance prediction in HRA makes use of the event tree representation. As an example, consider the event tree shown in Figure 22. (An equivalent

form is the THERP tree, which usually is drawn vertically with the initiating event at the top, cf. Figure 9).

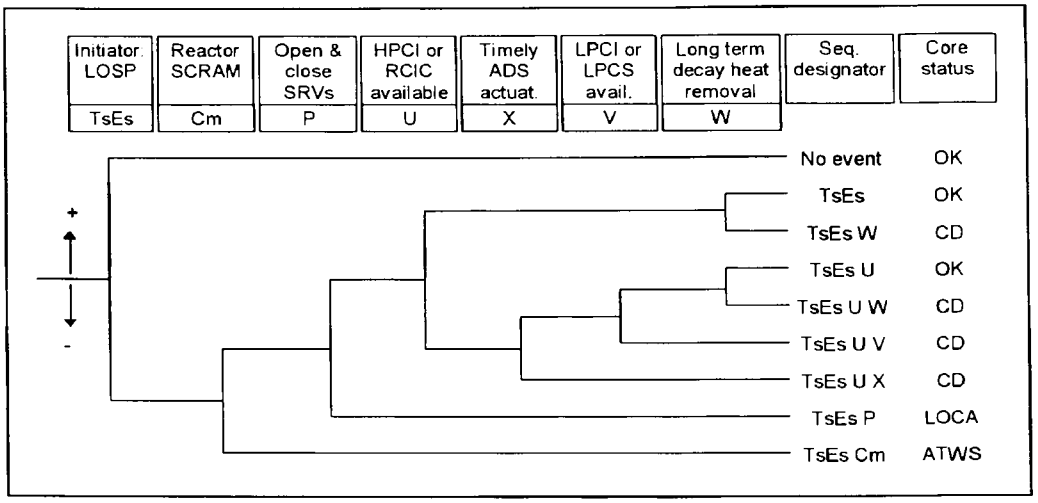

Figure 22: Event tree representation.

As Figure 22 shows, the event tree representation actually contains all possible predictions, i.e., all the possible combinations of events and conditions that are being considered. The events are described as a simple linear sequence, and the seeming complexity of an operator action tree, or an event tree, arises because the representation include the branch-points and the possible alternative developments. The analysis actually considers only **one** sequence of events. The prediction addresses how likely a specific outcome is, rather than to what may actually happen in the sense of the events in the sequence (Heslinga & Arnold, 1993). The HRA is therefore concerned with assigning probabilities to the set of specified events, but not concerned with defining the events as such (cf. below). This is usually done in the PSA, or in a task analysis associated with a PSA node. Even assigning the probabilities, however, requires detailed knowledge of the context.[17]

This approach means that first-generation HRA does not really address the problem of performance prediction in the sense of trying to predict how the situation may develop after the initiating event and how the consequences may propagate through the system. This may partly be because the possibilities for doing so are quite limited, due to the simplicity of the classification schemes. As noted above, the traditional approach to classification of incorrect action is based on the sequential binary decisions associated with the omission-commission scheme. It is hardly surprising that this scheme can deliver no more than an event tree, since it is functionally isomorphic to it. In order to predict performance, it is necessary to have a classification scheme which contains categories of both causes and effects (manifestations) and which can account for the links between them going in the direction from cause to effect.[18]

---

[17] In first-generation HRA this important condition is usually neglected. While the qualitative effects of the context are captured in the initial task analysis, the quantitative effects are brought to bear only after the probabilities have been assigned, by means of adjustments from the PSFs / PIFs.

[18] The other direction, from effect to cause, can only be used for event analysis.

9.3 The Separateness Between Analysis And Prediction

In first-generation HRA, the emphasis is on the prediction of the likelihood of errors, and the classification schemes and models have been developed to support that. For historical reasons, first-generation HRA adapted the approach used in reliability analysis for technical systems. Thus, Miller & Swain (1987) noted that "the procedures of THERP are similar to those employed in conventional reliability analysis, *except that human task activities are substituted for equipment outputs*" (italics added). This means that the development of operator models has been almost incidental to the HRA approach, and it is therefore not surprising that the models only contain enough detail to satisfy the demands from PSA/HRA applications. Consequently, models associated with first-generation HRA are ill-suited to performance prediction, except in the narrow sense of probability estimates for simple error manifestations.

In the information processing approaches the emphasis is on the analysis of events and the explanation of the psychological causes for erroneous actions. The specific paradigm for explanation, i.e., the information processing system, was taken as a starting point, and the main effort was put into reconciling this with detailed introspective reports. The results, as described above, was a number of very detailed theoretical accounts, although in most cases with limited validity. On the whole, there was very little concern for performance prediction. Even in the case of the more detailed accounts, such as the step-ladder model, the descriptions referred to how decision making **should** take place, but could not easily be used to predict exactly how it **would** happen. Strangely enough, most of the information processing models are of limited use even for event analysis, the most notable exception being Pedersen's guide (Figure 11). The reason for this seeming discrepancy is that the models try to explain the causes for action from the point of view of the information processing "mechanism", i.e., as a process. Although this can serve as the basis for an analysis method, it requires a reformulation which few of the approaches have bothered to make.

In cognitive systems approaches, and particularly in the case of the phenotype-genotype approach which is the basis for this project, the emphasis is on a principled way of analysing and predicting human erroneous actions. The cognitive systems approach is based on the MCM framework (cf. above), and therefore provides the best basis for supporting both retrospective analysis and performance prediction. In particular, the phenotype-genotype approach is based on non-directional links between classification groups or tables. This means that the same system can be used for retrospective and predictive purposes.

The first parts of this report described in detail how the approach could be used to support event analysis, and in particular gave an account of the method that should be used. With respect to performance prediction, we have so far emphasised the importance of describing the context before looking at the details of how an event may develop. In this respect the role of the Common Performance Conditions (CPCs) is crucial. In the retrospective method the CPCs are used to delimit the possible effects and the probable causes. In the method for performance prediction, the principle must be exactly the same, i.e., the CPCs must be used as a means of constraining the propagation of events, by effectively eliminating some of the links between causes and effects.

10. PREDICTIVE USE OF THE CLASSIFICATION SCHEME

As argued in the preceding, the important aspect of performance prediction is to be able to develop the likely event sequences that may occur, rather than to calculate or assign probabilities to individual events. (Or at least the two things should be done in that order.) We can distinguish between the two by naming them **qualitative performance prediction** and **quantitative performance prediction** respectively (Figure 23). The purpose of qualitative performance prediction is to find out which events are likely to occur, in particular which are the possible outcomes. The purpose of quantitative performance prediction is to find out how probable, in a probabilistic sense, it is that a specific event will occur.

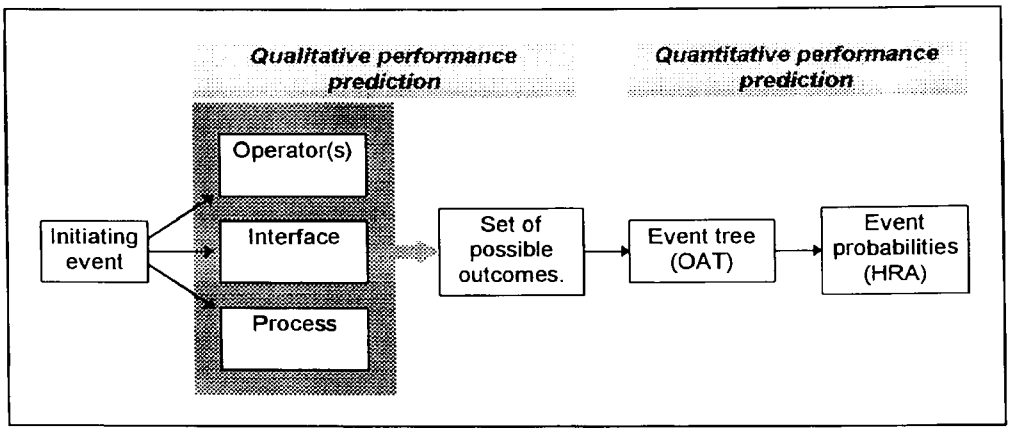

Figure 23: Qualitative and quantitative performance prediction

The qualitative performance prediction will generate the set of outcomes that may be the result of various event developments. The **validity** of the set depends on the assumptions on which the analysis is based, in particular the details of the process description, of the operator description, and of the interaction description. If the assumptions are accepted as reasonable, the set of outcomes will in itself provide a good indication of the reliability of the system, and whether unwanted outcomes can occur at all. That may, in the first instance, be sufficient. It may only be necessary to proceed to a quantification of specific probabilities if a number of unwanted consequences are part of the set. This latter step will, however, not be addressed in this report.

10.1 Combinatorial Performance Prediction

In the case of a predictive analysis, the basic problem is that if a classification is used in a simple, mechanical fashion it will produce far too many alternatives. If we consider the case of a *distraction* this can, according to the currently used version of the classification scheme, have effects on the *execution* (speed), *observation* (wrong identification), *interpretation* (decision error), *planning* (inadequate plan + inappropriate scheduling), and *communication* (communication failure). Each of these can have further effects through one or several iterations. This kind of combinatorial prediction will clearly generate too many alternatives to be useful. Furthermore, there will be no indication of whether one alternative is more likely or reasonable than another. Yet for any specific condition, some alternative developments will

clearly be more likely than others (or, in other words, not all alternatives will be equally likely). The purpose of the predictive analysis is to be able to say which they are. It follows, that this can only be done if the context is sufficiently well known.

To illustrate the perils of combinatorial performance prediction, we can consider the above example of a distraction a little more closely. First of all, the *distraction* itself can have several causes, namely functional impairment, equipment failure, and communication failure. For each of these further causes can be sought for, but in the example we assume that distraction is the initiating event. If we move forward in the classification system, using *distraction* as the starting point (and remaining on the level of general manifestations/general functions), we find that distraction can lead to five different effects in the first iteration. Of these only *execution* is terminal, i.e., it does not lead any further. The others are "internal", i.e., they can be found as general causes in the classification scheme. If we consider each of these "internal" effects in the second iteration, the outcome for all of them is either an execution error or an "internal" effect. Most of these lead back to a previous "internal" effect; this is due to the cyclical nature of cognition, as described by both the SMoC and COCOM. One of the effects requires a third iteration, but after that the already identified cause-effect links can be repeated. The complete example is shown in Table 11. Note, however, that the categories used in Table 11 refer to the general causes / effects. If the specific effects had been added, the table would have been considerably larger.

Table 11: Example of combinatorial performance prediction.

Initiating event	First iteration	Second iteration	Third iteration
Distraction ⇒	Execution		
	Observation ⇒	Execution Planning	
	Interpretation ⇒	Execution Observation Planning	
	Planning ⇒	Execution Observation	
	Communication ⇒	Execution Observation	
		Person (temporary) ⇒	Execution Observation Interpretation Planning Communication

10.2 Context Dependent Performance Prediction

The example shows clearly that a combinatorial approach to performance prediction can be a futile exercise. The basis for a predictive analysis must clearly be a description of the likely context or the likely working conditions, and this in turn must be based on a valid description of the tasks. The context description must, however, be in a form which matches the performance description. In first-generation HRA, performance has been described by means of the events in the PSA event tree, and the context has been described as the factors that could influence performance. The approach has furthermore been predicated on the PSA requirement to express the results in a quantitative fashion. The performance shaping or performance influencing factors have usually been a conglomerate of factors which empirically

have been recognised as important but with few attempts to structure them systematically - the best known exception being STAHR (Phillips et al., 1983) - or to consider dependencies and overlaps. The use of the performance shaping factors also seems to have been confined to the behavioural approaches. Although the information processing approaches recognise the importance of performance conditions, there have been few attempts of integrating them into the models. Instead, there has been a further specialisation of the field of organisational risk and reliability (Reason, 1992).

For analytical purposes it is clearly necessary to use some kind of simplification or abstraction. It also seems quite reasonable to describe the context with reference to a limited number of factors or dimensions, as long as the properties of these dimensions are explicitly defined. In the description of the retrospective analysis method use was made of the set of Common Performance Conditions (CPCs) which have been proposed by Hollnagel (1993b). The retrospective analysis demonstrated how the CPCs could be related to the classification scheme, and how this could be used to focus the analysis - and, incidentally, also to support the interpretation of the conclusions.

In relation to performance prediction it is therefore necessary to consider whether the CPCs can be used in a similar way and to provide a detailed description of how this should be done. Based on the discussions in the preceding, and in particular the distinction between qualitative and quantitative performance prediction, it is possible to describe an overall approach as shown in Figure 24. The numbers in the figure refer to the following steps of the procedure.

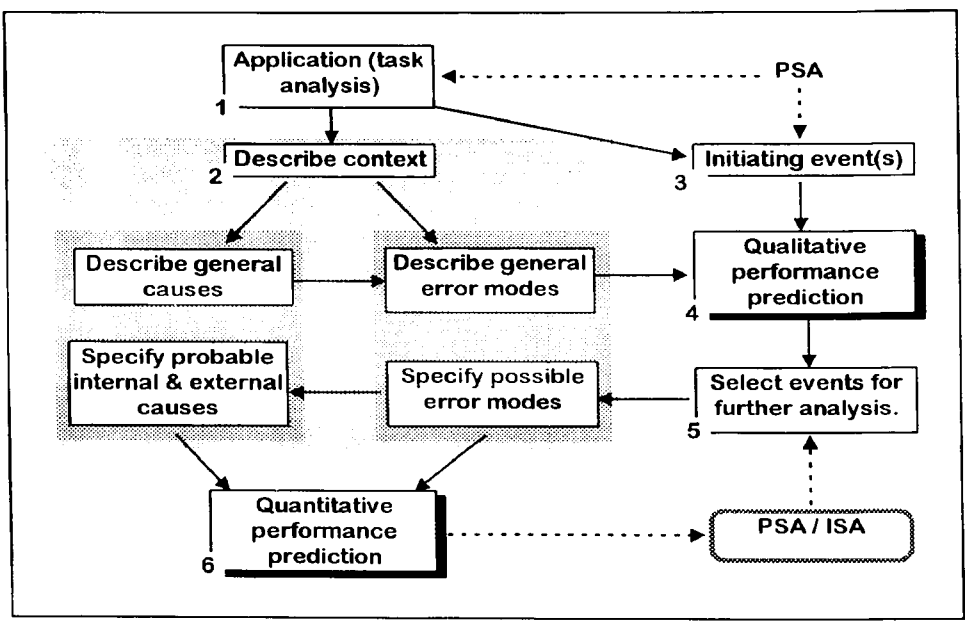

Figure 24: General method for performance prediction.

1. **Application analysis**. It is first necessary to analyse the application and the situation. This may in particular involve a task analysis, where the tasks to be considered can be derived from e.g. the PSA. The analysis must, however, also consider the organisation and the technical system, rather than only the operator and the control tasks. Depending on whether the analysis is made for an existing system or for one which is being

designed, sources of information may vary. If the system in question is yet non-existent, it is important to use information from similar systems and from operating experiences in general.

2. **Context description**. The context is described by using the CPCs (cf. above). The principle for the context description is exactly the same as for the retrospective analysis, the difference being that the level of detailed information may vary. In some cases it may be necessary to make assumptions about aspects of the design or the process which are not precisely known, or particularly about aspects of the organisation.

   The general context description can be used to prime the classification groups, just as for the retrospective analysis, by specifying more precisely the probable external and internal causes and to specify the possible error modes. The specification of the internal causes is, in particular, important and considerable care should be taken to ensure that a balanced result is achieved. It is a question of striking a proper equilibrium between, on the one hand, constraining the analysis to avoid unnecessary propagation paths and, on the other, to ensure that potentially important paths are not neglected or eliminated. This is, of course, the dilemma that is faced in any kind of HRA. But in this case it is more pertinent because the event tree is to be produced by the method rather than simply taken over from the PSA.

3. **Specification of initiating events.** The initiating events for the human actions / performance can be specified from several points of view. An obvious one is the PSA, since the PSA event trees will define the minimum set of initiating events that must be considered. Another is the outcome of the application and task analysis. A task analysis will, in particular, go into more detail than the PSA event tree, and may thereby suggest events or conditions that should be analysed further. The outcome of this step is the set of initiating events for which a performance prediction should be made.

4. **Qualitative performance prediction**. The qualitative performance prediction uses the classification scheme, as modified by the context, to describe the ways in which an initiating event can be expected to develop. Initially this may be done manually using a paper representation of the events and the classification scheme. Eventually this is a step that should be supported by a software tool, since it might otherwise become too laborious to be accomplished in practice.

   In a manual version, the performance prediction can be done using the matrix shown in Annex C. In this matrix the rows show the possible effects or manifestations, while the columns show the possible causes. In both cases the categories are grouped according to the major components of the classification scheme. The matrix obviously shows the complete set of categories; the priming or filtering will have to be done in each case, for instance by marking the rows / columns that are improbable. The analysis would start by finding the initiating event in the column headings (e.g. "person (temporary) / distraction"). The next step would be to find all the rows that have been marked for this column. With the exception of the error mode, each row will point to an effect which in turn may be found among the possible causes. The effects should be pursued both from the specific effect (now cause) and from the general group of causes. In this way the prediction continues in a mechanical fashion until there are no further paths. As the preceding illustration showed, this is easy to do but not necessarily very useful in terms

of practical results, unless knowledge of the context is applied to constrain the propagation.

5. **Selection of task steps for analysis**. The qualitative performance prediction, properly moderated by the context, may in itself provide useful results, for instance by showing whether there will be many or few unwanted outcomes. If a quantitative performance prediction is going to be made, it is necessary to select the cases that require further study. This can be done from the set of outcomes from the qualitative performance prediction, or from the PSA/ISA[19] input.

6. **Quantitative performance prediction.** The last step is the quantitative performance prediction. This aspect was not covered by the present project, and also falls outside the scope of the phenotype-genotype classification scheme. The issue of quantification is, of course, the philosophers' stone of HRA.[20] The lesson to be learned from the previous discussion is that one should not attempt the quantification without having first established a solid qualitative basis or description. If the performance prediction identifies potentially critical tasks or actions, and if the failure modes can be identified, then it is perhaps not necessary to quantify beyond a conservative estimate. In other words, the search for specific HEPs for specific actions may be somewhat unnecessary. To the extent that a quantification is required, the qualitative analysis may at lest be useful in identifying possible dependencies between actions. The description of the context in terms of the CPCs may also serve as a basis for defining ways of preventing or reducing specific types of erroneous actions through barriers or recovery.

These six steps provide a high-level description of how the phenotype-genotype classification scheme can be used for performance prediction. The description demonstrates that the principle of the classification scheme is equally well suited to retrospective and predictive applications, and thus confirm the arguments presented previously. The detailed categories will most certainly have to be modified, since the present version has been developed with the retrospective application in mind. However, the main principles of the classification scheme - the non-directedness and the dynamic development of the links between classification groups - should provide the necessary basis for further refinement.

Appendix C illustrates the principles of a predictive use of the phenotype-genotype classification. This is presently being developed into a proper second-generation HRA.

---

[19] ISA stands for Integrated Safety Analysis.

[20] In a very practical sense, it will have the power to turn base material into gold!

11. COGNITIVE MODELLING AND THE PHENOTYPE / GENOTYPE CLASSIFICATION SCHEME

Throughout this project the cognitive model has been used as the basis for the classification scheme. This was by itself a distinct advantage over previous approaches. It is, however, possible to use the cognitive model more explicitly, particularly in the case of qualitative performance prediction.

The model in question must be on the level of cognitive control, i.e., COCOM rather than SMoC. The essence of cognitive control is that it can co-determine the way in which events propagate. Thus, if cognitive control is on a low level, it is likely that most cognitive functions may go wrong. Indeed, it is almost certain that more complex activities - such as diagnosis or planing - will fail. Conversely, if cognitive control is on a high level, fewer failures should be expected and actions will be better adapted to the tasks.

In the case of qualitative performance prediction an additional outcome of the context description could be the specification of the probable initial control mode. This could be used to improve the precision of the performance prediction, provided that the links between the classification groups were extended to take the control mode into account. Such an undertaking would indeed be feasible, although the practical implementation would require the development of an appropriate software tool. Otherwise there would simply be too much housekeeping to take care of during the analysis. The state of the cognitive model, in particular the control mode, would have to be adjusted as the analysis went forward, in good agreement with the principle of context dependence. Eventually, a system of this type could develop into a fully-fledged JOSSI - a Joint System Simulation.

12. POSTSCRIPT

In the beginning of this report, the background and motivation for the project were described. It is clear from that, as well as from the contents of the report as a whole, that the work reported here is part of a larger context. In parallel to the current project work has been done on the development of a more comprehensive theoretical and practical framework for event analysis and performance prediction.

This framework has been presented on several occasions under the name of **CREAM**, for **C**ognitive **R**eliability and **E**rror **A**nalysis **M**ethod. The purpose of **CREAM** is to provide explicitly a useful and efficient candidate for a cognitive reliability analysis (CORA) method, in response to the challenge to develop a second generation HRA approach. Although much of the work reported here has contributed to the development of **CREAM**, the objectives of **CREAM** are more ambitious and it includes additional features. One of these is a software tool to support the practical implementation of the method, for both retrospective and predictive analyses. This development is in parallel to, but independent of, the continued work of Mauro Pedrali to develop a software implementation of the phenotype-genotype analysis method. There are probably other strains of activity which either are parallel to or overlap with what has been reported here. This should only be taken as encouragement, since it shows that there is a common acceptance of the ideas and that there is a real need for additional developments. This development can only be furthered by continued discussions and exchange of ideas. The authors therefore welcome all reactions to this report, be they positive or negative.

13. REFERENCES

Baddeley, A. (1990). *Human memory: Theory and practice*. London: Lawrence Erlbaum Associates.

Bagnara, S., Di Martino, C., Lisanti, B., Mancini, G. & Rizzo, A. (1989). *A human taxonomy based on cognitive engineering and on social and occupational psychology*. Ispra, Italy: Commission of the European Communities Joint Research Centre.

Bainbridge, L. (1993). *Building up behavioural complexity from a cognitive processing element*. London: University College, Department of Psychology

Billings, C. E. & Cheany, E. S. (1981). *Information transfer problems in the aviation system* (NASA Technical Paper 1875). Moffett Field, CA: NASA.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Cacciabue, P. C., Pedrali, M. & Hollnagel, E. (1993). *Taxonomy and models for human factors analysis of interactive systems: An application to flight safety* (ISEI/IE 2437 / 93). Paper presented at the 2nd ICAO Flight Safety and Human Factors Symposium, Washington D. C., April 12-15.

Caeser, C. (1987). *Safety statistics and their operational consequences*. Proceedings of the 40th International Air Safety Seminar, Tokyo, Japan.

Canning, J. (1976). *Great disasters*. London: Octopus Books.

Cojazzi, G. (1993). *Root cause analysis methodologies. Selection criteria and preliminary evaluation* (ISEI/IE/2442/93). JRC Ispra, Italy: Institute for Systems Engineering and Informatics.

Cojazzi, G., Pedrali, M. & Cacciabue, P. C. (1993). *Human performance study: Paradigms of human behaviour and error taxonomies* (ISEI/IE/2443/93). JRC Ispra, Italy: Institute for Systems Engineering and Informatics.

Cojazzi, G. & Pinola, L. (1994). *Root cause analysis methodologies: Trends and needs*. In G. E. Apostolakis & J. S. Wu (Eds.), Proceedings of PSAM-II, San Diego, CA, March 20-25, 1994.

Dougherty, E. M. Jr. (1990). Human reliability analysis - Where shouldst thou turn? *Reliability Engineering and System Safety*, *29*(3).

Dougherty, E. M. Jr., & Fragola, J. R. (1988). *Human reliability analysis. A systems engineering approach with nuclear power plant applications*. New York: John Wiley & Sons.

Embrey, D. (1980). *Human error. Theory and practice*. Conference on Human Error and its Industrial Consequences. Aston University, Birmingham, UK.

Feggetter, A. J. (1982). A method for investigating human factors aspects of aircraft accidents and incidents. *Ergonomics, 25,* 1065-1075.

Gagne, R. M. (1965). *The conditions of learning.* New York, Rinehart and Winston.

Gertman, D. I. & Blackman, H. S. (1994). *Human reliability & safety analysis data handbook.* New York: John Wiley & Sons, Inc.

Hannaman, G. W. & Spurgin, A. J. (1984). *Systematic human action reliability procedure.* EPRI NP-3583. Palo Alto, CA. Electric Power Research Institute.

Heslinga, G. & Arnold, H. (1993). *Human reliability: To what extent can we consider humans as system components.* International ENS Topical Meeting Towards the next Generation of Light Water Reactors, April 25-28, The Hague, Netherlands.

Hollnagel, E. (1988). Plan recognition in modelling of users. In G. E. Apostolakis, P. Kafka, & G. Mancini (Eds.), *Accident sequence modelling: Human actions, system response, intelligent decision support.* London: Elsevier Applied Science.

Hollnagel, E. (1993a). The phenotype of erroneous actions. *International Journal of Man-Machine Studies, 39,* 1-32.

Hollnagel, E. (1993b). *Human reliability analysis: Context and control.* London: Academic Press.

Hollnagel, E., Cacciabue P. C. (1991). *Cognitive Modelling in System Simulation.* Proceedings of Third European Conference on Cognitive Science Approaches to Process Control, Cardiff, September 2-6, 1991

Hollnagel, E., Pedersen, O. M., & Rasmussen, J. (1981). *Notes on human performance analysis* (Risø-M-2285). Roskilde, Denmark: Risø National Laboratory, Electronics Department.

Hollnagel, E., Rosness, R. & Taylor, J. R. (1990). *Human Reliability And The Reliability of Cognition.* Proceedings of 3rd International Conference on "Human Machine Interaction And Artificial Intelligence In Aeronautics And Space" Toulouse-Blagnac, 26-28 September.

Johnson, W. E. & Rouse, W. B. (1982). Analysis and classification of human errors in troubleshooting live aircraft power plants. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-12,* 389-393.

Kirwan, B. (1994). *A practical guide to human reliability assessment.* London: Taylor & Francis.

Kowalsky, N. B., Masters, R. L., Stone, R. B., Babcock, G. L. & Rypka, E. W. (1974). *An analysis of pilot error related to aircraft accidents* (NASA CR-2444). Washington, DC: NASA.

Miller, D. P. & Swain, A. D. (1987). *Human error and human reliability.* In G. Salvendy (Ed.) Handbook of Human factors. New York: Wiley.

Miller, G. A., Galanter, E. & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

Neisser, U. (1976). *Cognition and reality.* San Francisco: W. H. Freeman.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88,* 1-15.

Otway, H. J. & Misenta, R. (1980). Some human performance paradoxes of nuclear operations. *Futures.* October. pp. 340-357.

Pedersen, O. M. (1985). *Human risk contributions in process industry. Guides for their pre identification in well-structured activities and for post-incident analysis* (Risø-M-2513). Roskilde, Denmark: Risø National Laboratories.

Phillips, L. D., Humphreys, P. C., & Embrey D. E. (1983*). A socio-technical approach to assessing human reliability* (83-4). Oak Ridge, TN: Oak Ridge National Laboratory.

Rasmussen, J. & Jensen, A. (1974). Mental procedures in real-life tasks. A case study in electronic troubleshooting. *Ergonomics, 17,* 193-207.

Rasmussen, J. (1976). Outlines of a hybrid model of the process plant operator. in T. B. Sheridan & G. Johannsen. (Eds). *Monitoring Behaviour and Supervisory Control.* New York. Plenum Press.

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

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.

Reason, J. T. (1976). Absent minds. *New Society,* 4. pp. 244-245.

Reason, J. T. (1979). Actions not as planned. The price of automatization. In G. Underwood & R. Stevens (Eds). *Aspects of consciousness* (Vol. 1). *Psychological issues.* London: Wiley.

Reason, J. T. (1986). *The classification of human error,* unpublished manuscript. University of Manchester.

Reason, J. T. (1990). *Human error.* Cambridge: Cambridge University Press.

Reason, J. T. (1991). Reducing the risks of organisational accidents in complex systems. *Paper presented to the Colloquium on Human Reliability in Complex Systems,* Nancy 17th-18th April, 1991.

Reason, J. T. (1992). The identification of latent organisational failures in complex systems. In J. A. Wise, V. D. Hopkin & P. Stager (Eds.), *Verification and validation of complex systems: Human factors issues.* Berlin: Springer Verlag.

Reason, J. T. (1985). Recurrent errors in process control environments: Some implications for the design of Intelligent Decision Support Systems. In E. Hollnagel, G. Mancini & D. D.

Woods (Eds.), *Intelligent decision support in process environments*. Heidelberg, F. R. Germany: Springer Verlag.

Rockwell, T. H. & Giffin, W. C. (1987). *General aviation pilot error modelling - again?* Proceedings of the Fourth International Symposium on Aviation Psychology. Columbus: Ohio.

Rouse, W. B. & Rouse, S. H. (1983). Analysis and classification of human error. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*, 539-549.

Rouse. W. B. (1983). *Elements of human error*. NATO Conference on Human Error. Bellagio, Italy.

Singleton, W. T. (1973). Theoretical approaches to human error. *Ergonomics*, 16. pp. 727-737.

SAE - Society of Automotive Engineers. (1987). *Human error avoidance techniques*. Conference Proceedings P-204. Washington, DC.

Stoklosa, J. H. (1983) *Accident investigation of human performance factors*. Proceedings of the Second Symposium on Aviation Psychology. Columbus, Ohio.

Swain, (1967). Some limitations in using the simple multiplicative model in behavioural quantification. in W. B. Askren (Ed). *Symposium on Reliability of Human Performance in Work*. AMRL-TR-67-88. Wright-Patterson Air Force Base. OH.

Swain, A. D. & Guttman, H. E. (1983). *Handbook of human reliability analysis with emphasis on nuclear power plant applications* (NUREG CR-1278). Washington, DC: NRC.

Van Eckhout, J. M. & Rouse, W. B. (1981). Human errors in detection, diagnosis, and compensation for failures in the engine control room of a supertanker. *IEEE Transactions on Systems, Man and Cybernetics. SMC-11*, 813-816.

Woods, D. D., Johannesen, L. J., Cook, R. I. & Sarter, N. B. (1995). *Behind human error: Cognitive systems, computers and hindsight*. Columbus, Ohio: CSERIAC.

14. GLOSSARY

COCOM	**C**ontextual **C**ontrol **M**odel
CORA	**C**ognitive **R**eliability **A**nalysis.
CPC	**C**ommon **P**erformance **C**onditions
CSNI	**C**ommittee for the **S**afety of **N**uclear **I**nstallations
HEP	**H**uman **E**rror **P**robability
HF	**H**uman **F**actors (or Human Factors Engineering)
HRA	**H**uman **R**eliability **A**nalysis (or **H**uman **R**eliability **A**ssessment)
ISA	**I**ntegrated **S**equence **A**nalysis
JOSSI	**J**oint **S**ystem **S**imulation
MCM	**M**ethod, **C**lassification scheme, and **M**odel
PIF	**P**erformance **I**nfluencing **F**actor (also called Performance Shaping Factor)
PSA	**P**robabilistic **S**afety **A**nalysis (or **P**robabilistic **S**afety **A**ssessment)
SMoC	**S**imple **M**odel **o**f **C**ognition
SRK	**S**kill-based, **R**ule-based, **K**nowledge-based
STAHR	**S**ocio-**T**echnical **A**pproach to assessing **H**uman **R**eliability
THERP	**T**echnique for **H**uman **E**rror **R**ate **P**rediction

APPENDIX A: PHENOTYPE-GENOTYPE CLASSIFICATION GROUPS

Table 1: Common Performance Conditions			
CPC name	**Level**	☑	**Comments**
Adequacy of organisation	Very efficient		
	Efficient		
	Inefficient		
	Deficient		
Working conditions	Advantageous		
	Compatible		
	Incompatible		
Adequacy of MMI and operational support	Supportive		
	Adequate		
	Tolerable		
	Inappropriate		
Availability of procedures / plans	Appropriate		
	Acceptable		
	Inappropriate		
Number of simultaneous goals	Fewer than capacity		
	Matching current capacity		
	More than capacity		
Available time	Adequate		
	Temporarily inadequate		
	Continuously inadequate		
Execution mode	Explicit, attention required		
	Skilled or automatic		

Table 2: Basic Error Modes - Action at wrong time			
General manifestation	***Specific manifestation***	***General cause***	***Specific cause***
Timing	Too early	Communication failure	Earlier omission
	Too late	Error in procedure	Loss of control
		Faulty interpretation	Temporal pressure
		Faulty planning	Trapping error
		Synchronisation	
Duration	Too long	Communication failure	Loss of control
	Too short	Error in procedure	Temporal pressure
		Faulty diagnosis	Trapping error
		Synchronisation	

Table 3: Basic Error Modes - Action of wrong type			
General manifestation	**Specific manifestation**	**General cause**	**Specific cause**
Force	Too little	Communication failure	
	Too much	Equipment failure	
		Error in procedure	
		Faulty interpretation	
		Faulty planning	
		Motor variability	
Magnitude	Too little	Communication failure	
	Too much	Equipment failure	
		Error in procedure	
		Faulty interpretation	
		Faulty planning	
		Motor variability	
Speed	Too fast	Communication failure	
	Too slow	Distraction	
		Equipment failure	
		Error in procedure	
		Faulty interpretation	
		Motor variability	
		Synchronisation	
Direction	Too far	Communication failure	Ambiguous label
	Too short	Error in procedure	Convention conflict
	Wrong movement type	Faulty interpretation	Incorrect label
		Faulty planning	

Table 4: Basic Error Modes - Action at wrong object			
General manifestation	**Specific manifestation**	**General cause**	**Specific cause**
Wrong object	Neighbour	Access difficulty	Ambiguous label
	Similar object	Communication failure	Incorrect label
	Unrelated object	Error in procedure	
		Faulty interpretation	
		Faulty planning	
		Motor variability	

Table 5: Basic Error Modes - Action in wrong place			
General manifestation	**Specific manifestation**	**General cause**	**Specific cause**
Sequence	Action overshoot	Communication failure	Branching
	Jump backwards	Error in procedure	Capture
	Jump forward	Faulty interpretation	Incomplete information
	Omission	Faulty planning	Trapping error
	Repetition	Memory failure	Underspecification
	Reversal	Priority error	
	Side-tracking		
	Task not completed		
	Wrong action		

Table 6: Basic Error Modes - Right action in right place at right time			
General manifestation	**Specific manifestation**	**General cause**	**Specific cause**
No error	Expected performance		

Table 7: Person Related Causes - Observation			
General effect	**Specific effect**	**General cause**	**Specific cause**
Observation error	Overlook cue / signal Overlook measurement Overlook change False reaction False recognition	Equipment failure Faulty interpretation Faulty planning Distraction Functional impairment Fatigue Error in procedure	information overload Multiple signals Parallax Noise
Wrong identification	Mistaken cue Partial identification	Insufficient training Missing information Faulty interpretation Conflicting indications	Ambiguous signals Erroneous information Information overload Habit, expectancy

Table 8: Person Related Causes - Interpretation			
General effect	**Specific effect**	**General cause**	**Specific cause**
Faulty diagnosis	Incomplete diagnosis Incorrect interpretation Oversimplifcation	Insufficient training Wrong identification Cognitive biases	Confusing symptoms Error in mental model Misleading symptoms Multiple disturbances New situation Erroneous analogy
Wrong reasoning	Induction error Deduction error Wrong priorities Similarity matching Frequency gambling Overconfidence	Cognitive style Cognitive biases	Legal higher priority Prediction error Too short planning horizon False analogy Overgeneralisation Hindsight bias
Decision error	Decision paralysis Wrong decision Partial decision	Fear Social pressure Distraction	Lack of knowledge Shock Stimulus overload Workload
Delayed interpretation	No identification Increased time pressure	Error in procedure Equipment failure Fatigue	No indications Response slow-down

Table 9: Person Related Causes - Planning			
General effect	**Specific effect**	**General cause**	**Specific cause**
Faulty planning	Incomplete plan Wrong plan	Distraction Memory failure Faulty interpretation Insufficient training Error in procedure	Error in goal Model error Overlook precondition Overlook side effect Overlook subgoal Too short planning horizon Workload Time pressure
Priority error	Wrong goal selected Wrong task selected	Faulty interpretation	Legal higher priority Ambiguous criteria
Synchronisation	Incorrect prediction Incorrect adaptation	Observation error Delayed interpretation	Change not observed
Inappropriate scheduling		Distraction Faulty interpretation	Consequences misjudged

Table 10: General person related functions (temporary)			
General effect	**Specific effect**	**General cause**	**Specific cause**
Memory failure	Forgotten Incorrect recall Incomplete recall	None	Confusion of possibilities Daydreaming Long interval since learning Other priority Temporary incapacitation
Fear	Random actions Action freeze	None	Earlier error Possible consequences Uncertainty
Distraction	Task suspended Task not completed Goal forgotten Loss of orientation	Equipment failure Communication failure Functional impairment	Colleague / manager Comfort call Commotion Competing task Forced delay Telephone
Fatigue	Delayed response	Ambient condition	Exhaustion
Motor variability	Lack of precision Increasing misses	Equipment failure	Change of system character Illness Lack of training Haste Tiredness Temporary incapacitation
Inattention	Signal missed	Ambient condition	

Table 11: General person related functions (permanent)			
General effect	**Specific effect**	**General cause**	**Specific cause**
Functional impairment	Deafness Bad eyesight Colour blindness Dyslexia Speech problems Aphasia (motor) Aphasia (sensory)	None	
Cognitive style	Simultaneous scanning Successive scanning Conservative focusing Focus gambling	None	
Cognitive biases	Miscalibration Insensitivity to sample size Incorrect revision of probabilities Ignoring base rates Hindsight bias Attribution error Overconfidence Illusion of control Confirmation bias Hypothesis fixation	None	
Conformity	Self-censorship Group think	Social pressure	

Table 12: System Related Causes - Components / Technology			
General effect	**Specific effect**	**General cause**	**Specific cause**
Equipment failure	Actuator stick/slip Blocking Breakage Jamming Release Freeze Slow down No indications	Maintenance failure Software fault External event	Excess power Fire Flooding Loss of power Spray

Table 13: System Related Causes - Procedures			
General effect	**Specific effect**	**General cause**	**Specific cause**
Error in procedure	Ambiguous text Conflicting criteria Incomplete text Incorrect text Mismatch to actual equipment	Inadequate quality control	
Procedure not available	Misplaced procedure Wrong classification	None	
Inappropriate procedure format	Presentation style *Layout* No standards	None	

Table 14: System Related Causes - Interface (temporary)			
General effect	**Specific effect**	**General cause**	**Specific cause**
Access difficulty	Item cannot be reached Item cannot be found	Equipment failure Mislabelling	Design Distance Localisation problem Obstruction Temporary incapacitation

Table 15: System Related Causes - Interface (permanent)			
General effect	**Specific effect**	**General cause**	**Specific cause**
Access problems	Item cannot be reached Item cannot be found	Design failure	
Mislabelling	Incorrect information Incomplete information Ambiguous information Language error	Design failure Maintenance failure	
Layout problems	*Incorrect labelling* Inappropriate grouping Inappropriate demarcation Inappropriate location Mode error	*Design failure*	Non-standard design
Conflicting indications		Design failure Error in procedure	

Table 16: Environment Related Causes - Communication			
General effect	**Specific effect**	**General cause**	**Specific cause**
Communication failure	Message not received Message not understood	Distraction Functional impairment Inattention	Noise Presentation failure Temporary incapacitation
Missing information	No information Incomplete information Incorrect information Ambiguous information	Management error Design failure	Hidden information Presentation failure Language error Mislabelling Noise
Information not shared	*False consensus* Pluralistic ignorance	*Social pressure* Priority error	

Table 17: Environment Related Causes - Organisation			
General effect	**Specific effect**	**General cause**	**Specific cause**
Maintenance failure	Equipment not operational Indicators not working	None	
Software fault	Performance slow-down Information delays Command queues Information not available	None	
Design failure	Inadequate MMI	None	
Inadequate quality control	Inadequate procedures	None	
Management error	Unclear roles Dilution of responsibility Unclear line of command	None	
Social pressure	Group think Collective rationalisation Stereotyped perception	None	
Inadequate job management	Excessive work demands Excessive duration Inadequate scheduling of tasks Inappropriate staff allocation	None	
Inadequate job support	Inadequate task briefing Lack of training Adequate tools not available	None	
Information dissemination	No management response Inadequate reporting schemes	None	
Insufficient training	Overly domain specific Inadequate transfer Inadequate basic training Inadequate retraining	None	

Table 18: Environment Related Causes - Ambient conditions			
General effect	**Specific effect**	**General cause**	**Specific cause**
Temperature	Too hot Too cold	None	
Sound	Too loud Too quiet	None	
Humidity	Too dry Too humid	None	
Illumination	Too bright Too dark	None	
Other	Vibration	None	
External event	Tremor Fire Projectile	None	
Protective clothing & equipment	Too heavy Awkward to use	Design failure	

Table 19: Links between general causes and general effects			
General cause	**Table where general effect can be found:**	**General cause**	**Table where general effect can be found**
Access difficulty	Interface (temporary)	Access problems	Interface (permanent)
Ambient condition	Ambient conditions	Cognitive biases	Person related cause (permanent)
Cognitive style	Person related cause (permanent)	Communication failure	Communication
Conflicting indications	Interface (permanent)	Delayed interpretation	Interpretation
Design failure	Organisation	Distraction	Person related cause (temporary)
Equipment failure	Components / technology	Error in procedure	Procedures
External event	Ambient conditions	False observation	Observation
Fatigue	Person related cause (temporary)	Faulty interpretation	Interpretation
Faulty planning	Planning	Fear	Person related cause (temporary)
Functional impairment	Person related cause (permanent)	Inadequate quality control	Organisation
Inattention	Observation	Insufficient training	Organisation
Interpretation	Interpretation	Maintenance failure	Organisation
Management error	Organisation	Memory failure	Person related cause (temporary)
Mislabelling	Interface (permanent)	Missing information	Communication
Motor variability	Person related cause (temporary)	Observation error	Observation
Priority error	Planning	Social pressure	Organisation
Software fault	Organisation	Synchronisation	Planning
Wrong identification	Observation		

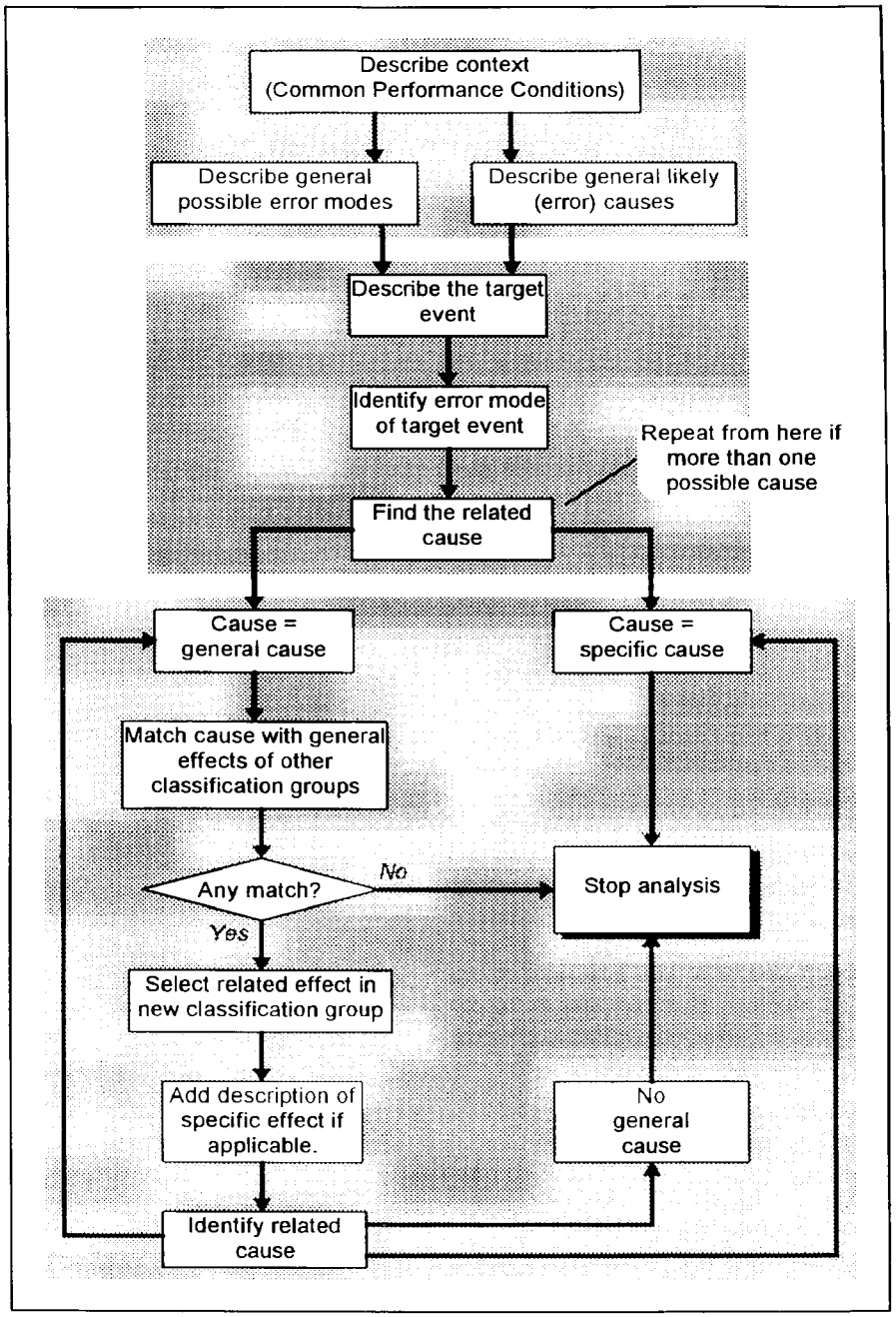

Figure 25: Steps in CREAM retrospective analysis.

APPENDIX B: ILLUSTRATION OF THE GUIDELINES IN USE

1. DESCRIPTION OF THE EXAMPLE

In order to demonstrate the use of the guidelines, and analysis has been made of an incident, taken from the description of the steam generator tube rupture at the Ginna nuclear power plant. This event has been documented in detail in a report from INPO in 1982, as well as in a NUREG. The purpose is not to reanalyse the incident as such, but to illustrate how the guidelines function in use.

1.1 Ginna Steam Generator Tube Rupture: Summary Of Event

As basis for the analysis, selected parts of the description of the incident are reproduced below. Since the analysis only concerned the first of several events, the description has been abridged accordingly. The emphasis has been added. All descriptions have been taken from the INPO report, except the one that is attributed to the NUREG.

About 09:25 on January 25, 1982, a single steam generator tube ruptured in the "B" steam generator at the R. E. Ginna Nuclear Plant. The plant, which had been operating normally at full power conditions, began a sudden primary coolant system depressurization as coolant rushed through the ruptured tube into the "B" steam generator. Alarms occurred indicating a possible steam generator tube rupture, and the operators began reducing power. A reactor trip on low pressure, followed by initiation of safety injection flow and containment isolation, occurred within three minutes. The reactor coolant pumps were stopped as required by procedures. During the first five minutes of the event, reactor coolant system (RCS) pressure decreased from about 2200 psig to 1200 psig.

The "B" steam generator was identified as the location of the rupture, and its isolation was completed 15 minutes after the rupture by termination of feedwater flow and closure of the "B" main steam isolation valve. (The analysis was made for the isolation of the steam generator only, and the continued development is therefore not included in the present description.)

1.1.1 Isolation Of Rupture

In order to isolate the spread of primary system radioactive nuclides and to reduce the primary-to-secondary leak rate through the ruptured tube, the operators next identified the damaged steam generator, stopped its feedwater and steam flow, and prepared to depressurize the reactor coolant system to the pressure of the damaged steam generator.

At 09:32, the motor-driven auxiliary feedwater pump to the "B" steam generator was stopped, and the steam supply valve from the "B" steam generator to the turbine-driven auxiliary feedwater pump (TDAFWP) was closed. Either at 09:32 or a short time later, the auxiliary feedwater flow from the TDAFWP to the "B" steam generator was isolated. The "B" steam generator water level continued to rise. At 09:40, the operators, satisfied that they had identified the steam generator with the ruptured tube, closed the main steam isolation valve

(MSIV) for the "B" steam generator. Fifteen minutes had elapsed since the rupture. This action isolated the steam generator and halted further spread of primary radioactive nuclides downstream of the "B" MSIV. Auxiliary feedwater continued to the "A" steam generator until level was established between 75-80 % on the narrow range instrumentation and then stopped at 09:48.

1.1.2 Operational Problems

Operational problems and dilemmas occurred during the Ginna event that complicated the recovery process and demonstrated that improvements in operating procedures, training, and design for tube rupture events are necessary. The problems encountered during the event led both directly and indirectly to the opening of a steam generator safety valve and to small radioactive releases to the environment. **The occurrence of operational problems for which the operating procedures and the operator's experience were incomplete or vague led the operators to take conservative actions to protect the reactor core**. These actions while conservative for maintaining core cooling, were not necessarily conservative for minimising releases from the steam generator safety valves. However, given even perceived threats to the reactor core and the limitations of the operating procedures and training for the problems encountered, the operators made the prudent choices. With the advantage of hindsight and post-event analysis, it can be seen that the reactor core was in no real danger, and actions could have been taken to avoid the steam generator safety valve opening.

The crew had to cope with a **novel event**. The **high cost of errors of commission**, the **inadequate design of the procedures**, the **various distractions** by the arrival of extra personnel to the control room, and the fact that the shift had been on duty for an hour and a half only, placed additional **stresses** on their performance. (NUREG.)

It should be emphasised that the Ginna operating staff performed well **under stressful and sometimes novel circumstances**. To better understand the operational problems that occurred at Ginna, let us begin by considering, from an operational perspective, the tension and the rush of events that occurred during the first five minutes after the tube rupture.

1.1.3 Isolation Of Ruptured Steam Generator - How Soon?

The first operational problem encountered by the Ginna staff was how soon to isolate the suspected SG. **Operating procedures required early isolation of the ruptured SG as soon it was positively identified** to minimise the spread of primary coolant contamination to secondary side systems and possible releases of radioactivity to the environment. **However, isolation of the wrong SG would require a delay to open the MSIV bypass valve in order to repressurize the downstream piping and unisolate the SG**. On a two-loop plant such as Ginna, a failure of the bypass valve to open would remove the capability for normal cooldown via the condenser dump valves and would necessitate using the unaffected SG power-operated atmospheric steam dump valve (ASDV). Use of the SG power-operated ASDV for cooldown would require a longer cooldown time since its flow capacity is considerably less than that of the condenser steam dump valves. Also, use of the ASDV increases the possibility of radiological releases in the event that the auxiliary feedwater supply should become contaminated or both steam generators should rupture. Therefore, cooldown via the condenser is preferred during tube rupture events.

Although the Control Operator was convinced, based on preliminary indications, that the rupture had occurred in the "B" SG, **the Shift Supervisor wanted more confirming information before isolating the steam generator**. **The Ginna procedure required positive identification prior to isolating a steam generator**. After the reactor trip, with auxiliary feedwater flow to both SGs, the operators noted that the "B" SG water level was increasing more rapidly than the A SG water level.

After the auxiliary feedwater pump supplying the "B" SG was stopped at 09:32 (0:07:00), the water level continued to rise at approximately 4 percent per minute. It was this continuing increase in the "B" level that convinced the Shift Supervisor that isolation was prudent. After observing the "B" SG water level increase steadily for eight minutes with feedwater secured, the Shift Supervisor ordered the "B" MSIV closed. A short time later the Health Physics Technician entered the control room and reported that radiation readings on the "B" SG blowdown line were 9 mrem/hr, compared to less than 1 mrem/hr on the "A" SG blowdown line. The SG blowdown lines were isolated by the containment isolation signal that occurred 3-4 minutes after the tube rupture. The readings by the Health Physics Technician were made several minutes after isolation of the blowdown lines. An auxiliary operator and a health physics technician were dispatched to obtain radiation readings on the "B" SG steam piping. They reported a reading of 30 mrem/hr upstream of the "B" MSIV and a reading of 2 mrem/hr on the steam header 3 feet downstream of the "B" MSIV.

The operators had correctly isolated the ruptured SG in 15 minutes. The use of technicians to monitor radiation conditions for diagnosis of the ruptured SG was in part due to the inoperability of portions of the radiation monitoring system designed to monitor radiation from the steam line.

1.1.4 Step 1: Determine Or Describe Context

The context is described using the Common Performance Conditions, as presented in Table 12. Considering the situation as described in the reports, and supplementing with general knowledge about the conditions in nuclear power plant control rooms, the following characterisation results.

Table 12: Common Performance Conditions

CPC name	Level	☑	Comments
Adequacy of organisation	Very efficient		
	Efficient		
	Inefficient	☑	Apparently there were many distractions; this is not efficient for an emergency situation
	Deficient		
Working conditions	Advantageous		
	Compatible		
	Incompatible	☑	Too many people in the control room; too many disturbances.
Adequacy of MMI and operational support	Supportive		
	Adequate		
	Tolerable	☑	No details given; MMI considered adequate for normal conditions but possibly less so for an emergency
	Inappropriate		
Availability of procedures / plans	Appropriate		
	Acceptable		
	Inappropriate	☑	Inadequate design of procedures; conflicting goals and priorities.
Number of simultaneous goals	Fewer than capacity		
	Matching current capacity		
	More than capacity	☑	This is assumed to be the case in an emergency in general.
Available time	Adequate		
	Temporarily inadequate	☑	Incident occurred early in shift.
	Continuously inadequate		
Execution mode	Explicit, attention required	☑	The operations were non-routine activity, hence required (and got) attention.
	Skilled or automatic		

1.1.5 Step 2: Describe The Possible Error Modes

Based on knowledge about the MMI in a typical nuclear power plant control room, it was possible to rule out only "force / magnitude" among the general manifestations. All the other error modes would be possible for the events that occurred.

1.1.6 Step 3: Describe The Probable Error Causes

Based on the description of the context, as summarised in Table 12 above, it was possible to distinguish between more and less likely error causes. The main aspects were the novelty of the situation, the ambiguity of the procedures, the concerns of the operators for not making the wrong decision, and the general working conditions. The probable error causes can either be marked on the separate tables, or be summarised as shown below (Table 13).

Table 13: Likelihood of general causes.

Likelihood			General causes
Hi	Med	Low	
		☑	Observation
☑			Interpretation
	☑		Planning
	☑		General person related functions (temporary)
☑			General person related functions (permanent)
☑			Procedures
☑			Components / technology
		☑	Interface (temporary)
		☑	Interface (permanent)
☑			Communication
☑			Organisation
		☑	Ambient conditions

1.1.7 Step 4: Perform the more detailed analysis of main task steps

This can be accomplished through a number of steps, as described below (cf. also Appendix A, Figure 1).

- **Describe Target Event**: The target event for the analysis was the delay in closing the Main Steam Isolation Valve in loop "B".

- **Identify Error Mode Of Target Event**: According to the description of the target event, the most likely error mode was easily determined as an error of timing, specifically an action that occurred too late. No other error modes were applicable for the target event.

- **Find The Related Cause(s):** A closer look at the possible general causes of the error mode for the target event led to the selection of two candidates. The first was "error in procedure" and the second "interpretation".

The delay was due to the ambiguity of the procedure, which both demanded fast action but also a high level of certainty. That in turn led to problems in identifying the steam generator to be closed. There were no problems in diagnosis of the event as SGTR, nor in communication or observation. Planning was also appropriate, in the sense that it was the supervisor's deliberate wait for further evidence that caused the delay. This wait was explicitly planned.

Since the outcome in either case was a general rather than a specific cause, the analysis was continued. First, the possibility of an "error in procedure" was investigated.

- **Match Cause With Other Classification Groups:** A search for a match to "error in procedure" led to the "procedures" group, using Appendix A, Table 19.

Going to the "procedures" group it was possible to find a relevant specific effect, being "conflicting criteria". In the "procedures" group there was only one possible general cause, which was "inadequate quality control".

Returning to Appendix A, Table 19, "inadequate quality control" pointed to the "organisation" group of causes. This confirmed that "inadequate procedures" was, indeed, a specific effect of "inadequate quality control". It also showed that there were no general causes, nor any specific causes, listed for "inadequate quality control". This meant that the analysis of "error in procedure" had come to an end. The findings of this analysis can be shown as in Figure 26. In each box, the boldface term denotes the **error mode** or **classification group**. In the case of a classification group, the category written in normal text denotes the general effect, while the category written in italics denotes the *specific effect*.

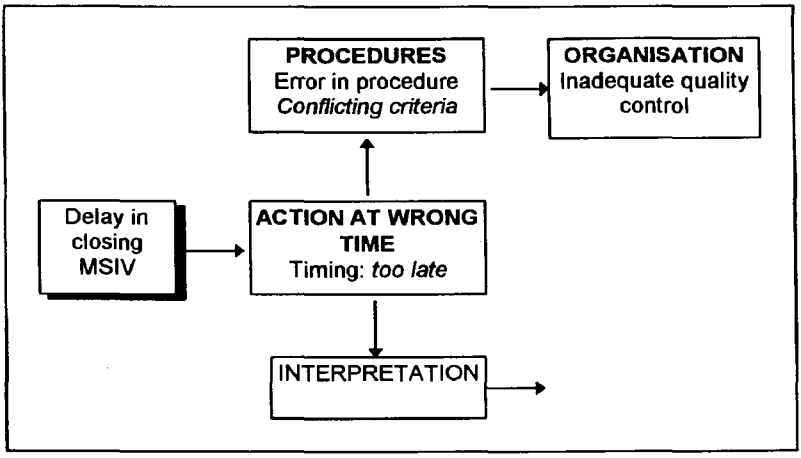

Figure 26: Analysis of Ginna example - first cause.

- **Repeat Search For Causes:** After having completed the analysis of "error in procedure", the guidelines were applied to the second general cause "interpretation". This led in the first hand to the group for "interpretation" where the general effect was "delayed interpretation" and the specific effects were "no identification" and "increased time pressure". The latter was simply due to the fact that waiting to confirm the interpretation meant that other actions (according to the procedures) had to be postponed while the event continued to unfold.

The general cause matching "delayed interpretation" was in this case determined to be "error in procedure" rather than any of the others. There were no known equipment failures and no sign of fatigue. This meant that the analysis of "interpretation" ended in much the same way as the analysis of "error in procedure", as shown in Figure 27:

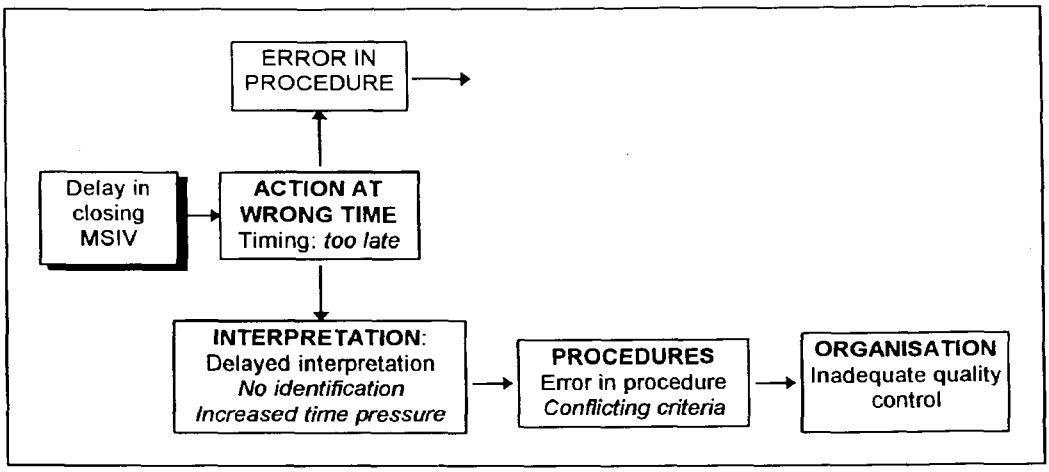

Figure 27: Analysis of Ginna example - second cause.

I.1.8 Summary Of Analysis

It is interesting in this case that the analysis of the two possible causes to "action too late" ended up by pointing to the same underlying cause, namely "error in procedure" which then, in turn, was caused by "inadequate quality control" in the organisation. One causal chain is longer than the other because it contains an additional step. The usual principle followed in scientific research is to choose the simpler of two explanations; this is generally known as Occam's razor. However, in the case of finding an explanation for a human erroneous action, the purpose is to find the most complete or reasonable explanation rather than the simplest. This means that one cannot apply Occam's razor in a mechanical fashion - but neither can the inverse principle be invoked. The determination of which explanation is the most complete or most reasonable must be based on the experience of the analyst. The analysis guidelines described here serve to facilitate the analysis and to ensure greater consistency. They cannot, and should not, be used as a complete algorithmic procedure that can be blindly used. Retrospective event analysis will always require a modicum of expertise, but the analysis should not rely on subjective expertise alone.

In the present case, the detailed description of the event actually makes it more likely that the second - and longer - explanation is the correct one. It does, however, lead to the same recommendation for remedial action, namely to improve the procedures both in terms of their content and in how they are to be interpreted and used.

APPENDIX C: ILLUSTRATION OF PERFORMANCE PREDICTION

1. PRINCIPLES OF PREDICTION

The purpose of this appendix is to illustrate in principle how the phenotype-genotype classification system can be used to make predictions of likely erroneous actions. The example used is only for the purpose of showing the principle. Furthermore, the prediction is of a qualitative nature, i.e., of the erroneous actions and not of human reliability *per se*.

As described in the main text of the report, the predictive use of the phenotype-genotype scheme goes through a number of distinct steps. The first step is an analysis of the application, usually in the form of a task analysis. A variety of Task Analysis techniques are currently available in the literature to facilitate the attainment of such a description.

The second step is to provide a description of the Common Performance Conditions. This is done in the same way as described in Appendix A. The difference between the retrospective and the predictive applications is that the latter case must refer to the expected rather than the actual performance conditions. Clearly, a prediction is made for a hypothetical situation. It is, however, important that the hypothetical situation is described in sufficient detail, and in particular that the common performance conditions are accounted for. The second step in the predictive analysis therefore helps to ensure, that the analyst provides a description as precise as possible about the scenario being considered.

The third step is the specification of the initiating event for the operator actions being analysed. This denotes the start of the scenario or the event, and is typically something that takes place in the system, i.e., not an operator action. This is followed by the fourth step, which is the qualitative performance prediction. This step is described in further detail below.

After the qualitative performance prediction, the analysis can be supplemented by a more detailed analysis of critical task steps, and completed by a quantification. The latter step is necessary in order for the result of the prediction to fit into e.g. a PSA. The problem of quantification was, however, not addressed in this project, and these last steps are therefore not included in this description.

1.1 Qualitative Performance Prediction

As discussed in the main parts of this report, it is important that the performance prediction avoids the perils of a combinatorial prediction. The only way in which this can be achieved is by making the performance prediction depend on the context, i.e., to let the probable context determine the path between the classification groups. As described for the retrospective analysis, it is an essential feature of the phenotype-genotype classification system that the classification groups are **not** hierarchically ordered. The retrospective analysis creates a

specific path - or link - between the classification groups depending on the context. The predictive analysis must, in principle, do the same.

In order to achieve this it is necessary first to outline the paths or links that are possible, given the contents of the classification groups. This can be done by noting the cases where an effect of one group matches a cause of another. For instance, "equipment failure" (Appendix A, Table 12) appears as a cause in "interpretation" (Appendix A, Table 8). This means that an equipment failure may lead to a delayed interpretation, which in turn may lead to a planning failure, and so on. The basis for the performance prediction is therefore to establish the possible forward links between the classification groups, and then select from these using the context description given by the Common Performance Conditions.

In order to assist the analyst, a set of tables can be constructed which show the forward links. Thus Table 14 shows the main forward links, i.e. in terms of the group names rather than in terms of specific causes and effects. The principle of Table 14 is that the categories in the top row (the column labels) describe causes while the categories in the left column (the row labels) describe the effects. Table 14 therefore shows that a failure in planning (= a cause) can have either direct consequences for actions, shown by the marks for the error modes, or have consequences for observations (= an effect). If, in turn, observation is considered as a cause, Table 14 shows that it has possible consequences for interpretation and planning. Each of these can be taken a step further until a complete event tree has been constructed. The forward propagation clearly comes to halt only when an error mode has been reached.

Table 14: Main forward links between classification groups.

	Obs.	Interpret.	Planning	Person (temp)	Person (perm)	Tech. sys	Procedure	MMI (temp)	MMI (perm)	Comm.	Organisation	Ambient condition
Action at wrong time		⌟	⌟				⌟			⌟		
Action in wrong place		⌟	⌟	⌟			⌟			⌟		
Action of wrong type		⌟	⌟	⌟		⌟	⌟			⌟		
Action at wrong object		⌟	⌟	⌟			⌟	⌟	⌟	⌟		
Observation		⌟	⌟	⌟	⌟	⌟	⌟		⌟	⌟	⌟	
Interpretation	⌟			⌟	⌟						⌟	
Planning	⌟	⌟		⌟			⌟					
Person (temp)					⌟	⌟				⌟		⌟
Person (perm)											⌟	
Tech. system											⌟	
Procedure											⌟	
MMI (temp)						⌟						
MMI (perm)							⌟				⌟	
Communication			⌟	⌟	⌟						⌟	
Organisation												
Ambient condition											⌟	

In order to perform the predictive analysis it is, however, necessary to have a table or matrix which shows the complete forward links between causes and effects. For the purpose of illustration, a table is provided at the end of this appendix. The contents of this table

corresponds to the classification groups included in Appendix A. Clearly, if the classification groups change, the table of forward links must also change. The printed version of the table of forward links is naturally cumbersome to use, but illustrates well the principles of the predictive analysis. In practice, a simple computerised tool could make the process much easier.

The use of the complete table can be illustrated by considering the consequences of e.g. a decision error. If we use the main links shown in Table 14, we can see that a decision error - which is part of the "interpretation" group - can propagate forwards as shown in Figure 28. Since the classification groups are not ordered hierarchically, loops will occur which mean that the propagation can go through a very large number of steps. This corresponds to a chain of cognitive functions that form a cascade, for instance when the consequences of a misinterpretation show themselves in later tasks.

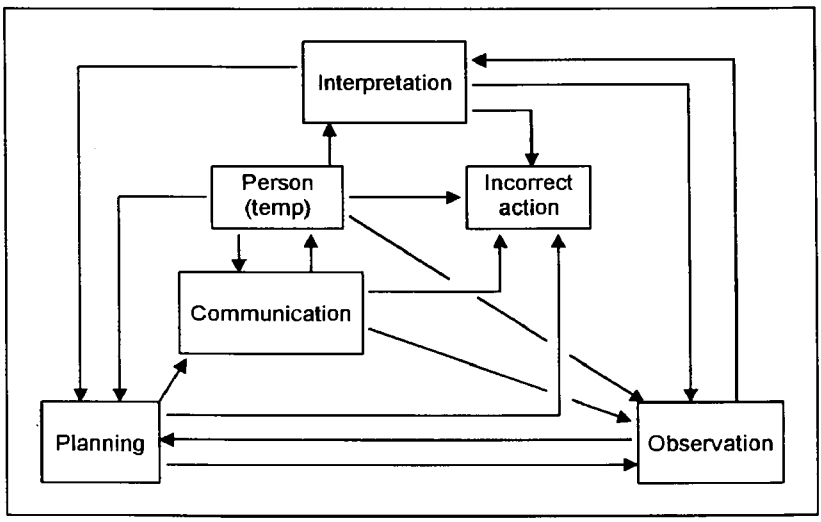

Figure 28: Forward propagation of "decision error".

For the purpose of making a prediction, the unrestricted cascading is, of course, of limited value. It corresponds to the mechanical combination of categories which -0 in the worst case - leads to a prediction of all outcomes. This is clearly not a desirable result. It can be avoided by taking the context into account. Knowing - or assuming - what the performance conditions will be, it is possible to consider only those effects that are consistent with the situation. This dramatically restricts the forwards propagation of causes and effects, and serves to focus the prediction.

As an example, consider that an assessment of the expected common performance conditions leads to a likelihood of failure as shown in Table 15. This could represent a situation where operators are likely to make incorrect interpretations/diagnoses, where the procedures are difficult to follow, where hardware / components may fail, and where the organisational support is inadequate. Activities involving planning and communication may also be prone to failure, although to a lesser degree. The remaining types of causes may, for all intents and purposes, be considered as unlikely for the analysis.

Table 15: Likelihood of failure for general causes.

Likelihood of failing			General causes
Hi	Med	Low	
		☑	Observation
☑			Interpretation
	☑		Planning
		☑	General person related functions (temporary)
		☑	General person related functions (permanent)
☑			Procedures
☑			Components / technology
		☑	Interface (temporary)
		☑	Interface (permanent)
	☑		Communication
☑			Organisation
		☑	Ambient conditions

If the contents of Table 15 is combined with the propagation paths shown in Figure 28, the result may look as shown in Figure 29. Here the thickness of the arrows indicate the likelihood that a failure may propagate through the path. The main consequence of the hypothetical situation shown in Figure 29 is that a failure in interpretation may likely lead to an incorrect action either directly or through a failure of communication. The number of possible paths has thereby been considerably reduced; in particular, there are no longer a possibility of repetitive loops, basically because it is assumed that observations will be correctly made.

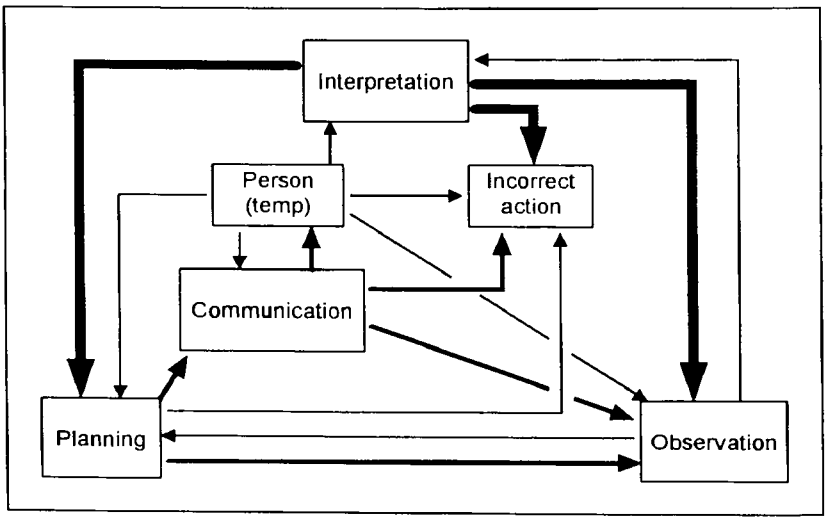

Figure 29: Constrained forward propagation of "decision error".

With the overall propagation links in Figure 29 as a guideline, we can now chart the more specific paths using the complete table of forward links. This yields the Table 16.

Table 16: Reduced forward cause-effect links.

Step 0	Step 1	Step 2	Step 3
Decision error	**Action at wrong time.** **Action in wrong place.** **Action of wrong type.** **Action at wrong object.**		
	Observation error *Wrong identification*		
	Faulty planning	**Action at wrong time.** **Action in wrong place.** **Action of wrong type.** **Action at wrong object.**	
		Observation error	
	Priority error	**Action in wrong place**	
		Failure to share information	**Incorrect action**
			Person (temp)
	Inappropriate scheduling	**Incorrect action**	

Entries shown with **boldface** are terminals, i.e. error modes.
Entries shown in *italics* stop the propagation, because the likelihood of a failure is assumed to be insignificant.

Table 16 illustrates how the effects of the context can be used to curtail the forward propagation of the cause-effect links and thereby avoid the consequences of a combinatorial performance prediction. The context is described by means of the Common Performance Conditions, which in turn can be used to assign qualitative (or fuzzy) likelihood to the various causes, as shown in Table 15. Ultimately, any performance prediction will end with some of the error modes. The interesting part is, however, how the error modes are reached, and how an initial failure (say, of reasoning) can have consequences for other cause-effect links.

The reader should remember, however, that this example is used only to show the **principles** of performance prediction. Furthermore, that this method should never be used in a mechanical fashion, without understanding fully the situation that is being analysed. In formal terms, the correctness of the predictions depend on the correctness of the classification groups, and the appropriateness of the Common Performance Conditions and the specific values they have been assigned. In both cases practical experience plays an important role. Thus, for a given application and scenario, the classification groups may have to be modified to reflect the distinct features of the system. Similarly, the evaluation of the Common performance Conditions require a good deal of experience and understanding. While computerised tools may go some way towards facilitating performance prediction, the process can never be automated as a whole.

To us, the main advantage of the phenotype-genotype classification system is that the same principles can be used for retrospective and predictive analyses. The use of the classification system for retrospective analysis - event analysis - will gradually lead to a refinement of the categories and of the potential links between the groups. This will probably have to be done separately for each domain or application, although a more general set of classification groups may also emerge. This coupling between event analysis and prediction, mediated by the classification groups, is of utmost importance for the predictions, since it provides the best possible assurance that the predictions reflect realistic assumptions about cause-effect relationships. The report has presented the current state of development and indicated how the classification system can be used. There is clearly much work to be done, particularly in the

predictive applications, and in applying the phenotype-genotype principles to a proper quantitative human reliability analysis. We hope to remain part of this work.