

Prolegomenon

1. Reader's Guide

The purpose of this introduction book is to provide the reader with a survey of the topics that are treated in the book, as well as some supplementary information about the book itself. The purpose of these first paragraphs is to provide the reader with a guide to the introduction itself.

- (1) The first section presents the purpose of the book as well as the **rationale** for writing it. It also provides some advice about who should read the book and who should not. This section should therefore be read by all, even the casual browser in the bookstore.
- (2) The second section briefly goes through the book in a **chapter-by-chapter** fashion. Readers who have not been put completely off by the first section are encouraged to read the second section. It will enable them to decide which chapters of the book they should concentrate on and in which order.
- (3) The third and last section provides **miscellaneous information** and comments. Readers, whose curiosity is aroused by the headings should read the associated text at some time, although not necessarily before starting on the main chapters of the book.

2. Rationale

In the beginning of the 1990s the field of human reliability analysis (HRA) was in a state where there was pronounced dissatisfaction with the available methods, theories, and models, but where there as yet were no clear alternatives (Dougherty, 1990). The intention with this book is to

Prolegomenon

present such an alternative, based on the principles of cognitive systems engineering.

Throughout the 1980s there was a growing recognition in the engineering world of the role of human cognition in shaping human action -- both when it led to accidents and when it prevented them. This recognition was not felt in human reliability analysis alone, but also in the concern with man-machine systems in general, with decision support systems, with human-computer interaction etc. One consequence was that "cognitive" and "cognition" became fashionable terms for almost all aspects of man-machine interaction. As an example, the book about "Accident Sequence Modelling" by Apostolakis et al. (1988) has the following main entries:

- (1) cognitive activity,
- (2) cognitive competencies,
- (3) cognitive environment simulation,
- (4) cognitive modelling,
- (5) cognitive primitives,
- (6) cognitive processing,
- (7) cognitive reliability analysis technique
- (8) cognitive structures,
- (9) cognitive sub-elements, and
- (10) cognitive under-specification

In many cases, however, the allusion to cognition was a matter of convenience rather than a real change in orientation. Cognition, however, **is** of fundamental importance and it is consequently necessary to have adequate methods, theories, and models to address properly the role of cognition in human action -- and particularly specific issues such as the Reliability Of Cognition.

The study of human cognition has developed from experimental psychology in the 1960s and gradually grown into several distinct directions (it would probably be going too far to call them scientific disciplines). Some of these focus on basic research issues while others venture into what for academia is the *terra incognita* of applications; among the latter are cognitive science, cognitive systems engineering, and cognitive ergonomics.

Cognitive systems engineering (Hollnagel & Woods, 1983) is based on the principle that human behaviour -- in work contexts and

otherwise -- should be described in terms of joint or interacting cognitive systems.¹ A joint system where one of the parts is a cognitive system is also in itself a cognitive system. Hence all man-machine systems are by definition cognitive systems. In the classical view on man-machine systems, one could consider the man (= the operator) by himself, the machine (= the process) by itself, and add the interaction between the two. This view, however, misses the notion of integration and dependency and -- in particular -- that all activities take place in a context.

Cognitive systems engineering is obviously not the only way to look at human cognition and it cannot be proved that it is **the** correct way. It is, however, a usable basis for describing human cognition in the context of human work, i.e., it is pragmatically correct. The specific developments described in this book are focussed on the notion of how actions are controlled and on how control and reliability are related.

2.1 Credo

Better analyses of the reliability of cognition are needed for practical reasons alone. Current approaches to HRA are based on the principle of describing situations in terms of appropriate components or elementary events, e.g. as single actions. This principle of decomposition is basically a consequence of the underlying view of the human operator as a machine -- possibly a complex, cognitive machine, but a machine nevertheless.

Such approaches are, however, inadequate as a way of describing human cognition because they are not based on a clear theory of human cognition -- or even on a clearly formulated description of what human cognition is. A proper analysis or assessment of human reliability must not only acknowledge the role of cognition, but also include a theory or description of human cognition and of the reliability of cognition.

Any such model -- even a very simple model of cognition -- will show that cognition must be considered as a whole and as an integrated activity that reveals itself in a context, rather than as a decomposable ordering of elementary functions and bits of knowledge. Any assessment

¹ A cognitive system (1) is goal oriented, and based on symbol manipulation, (2) is adaptive and able to view a problem in more than one way, and (3) operates using knowledge about itself and the environment and is therefore able to plan and modify its actions on the basis of that knowledge. The definition is intended to be equally applicable to men and machines.

Prolegomenon

method must start by recognizing this fact and strive to derive a description which does not conflict with that.

An alternative approach to human reliability analysis may make it less straightforward -- but also less necessary -- to provide point estimates or point probabilities of individual actions. It will, however, improve the qualitative basis for developing solutions that consider the system as a whole and which therefore contribute to the overall goal of reducing the number of unwanted consequences. An alternative approach will also make it easier to assess the overall risk or reliability of a work situation in a meaningful way.

On the other hand it will also reduce the need to collect data (estimates) for minute aspects of human performance, since such data no longer will be very important. Instead data must be sought on the level of cognitive ensembles, i.e., the practically meaningful segments of work.

2.2 The Root Cause

Risk and reliability analyses are often made on the basis of descriptions that use trees as an underlying structure: operator action trees, event trees, cause-consequence trees, etc. Since every tree has one root -- at least in the simplified graphical representations that commonly are used -- the notion of a root cause has become widespread. The root cause, of course, means the single, identifiable cause for an observed consequence, even though most practical cases show that there rarely is only one cause.

In the case of this book the root cause was a special issue of the *Journal of Reliability Engineering and System Safety* that dealt with the problems of HRA and the unhappy state of the art. The basis for the special issue was a position paper by Ed Dougherty (1990), which was followed by a number of comments (some short, some long, some agreeing and some disagreeing) from people who, in one way or another, either had experienced the problem or had an opinion on it.

I am sure that there are even more opinions than were expressed in the special issue. In fact, I was asked to contribute a comment and started to write down my views but did not finish them in time for the special issue. As luck would have it, another opportunity came at the International Conference on Probabilistic Safety Assessment and Management (PSAM), which was held in Beverly Hills, February 4-7, 1991. For this occasion I elaborated on my unfinished comments and presented them as a paper entitled "What Is a Man That He Can Be Expressed by a Number?" That paper in turn became the starting point

for this book, which can be seen as an elaboration and extension of the main theme of that paper, i.e., a long argument against viewing and describing humans in terms of numbers -- whether as reliability measures or something else.

Although the special issue of *Reliability Engineering and System Safety* mentioned above can be seen as a root cause for this book, it is certainly not the only cause. The paper by Dougherty (1990) merely expressed the concerns that many HRA practitioners had. In addition, psychologists and others had generally criticised the approach to quantitative modelling that HRA practitioners had taken. In his editorial Apostolakis (1990) rather bluntly expressed it thus: "... researchers who try to understand human behavior and to develop models for the operators have a very negative view toward the use of such quantitative models, whose foundations they consider to be unacceptable." This critical view can be found in practically all of the books and papers published during the 1980s that looked at "human error" from the behavioural or social sciences point of view (e.g. Perrow, 1984; Rasmussen et al., 1987; Reason, 1990; and Senders & Moray, 1991). It is a criticism which is amplified by the general view of cognitive systems engineering and cognitive ergonomics, as described above. The real "root cause" for this book is therefore an assortment of views and issues that gradually were developed during the 1980s by the international community of people concerned with the study of human cognition.

2.3 Who Should Read This Book ...

I have written this book with a certain audience in mind. The audience is not defined in terms of lines of profession but rather in terms of specific interests or views on man-machine systems and human performance. In other words, there is a certain audience that I hope will find the book congenial. This audience includes:

- (1) The HRA practitioners who have found the current approaches, models, and methods lacking in one way or another.
- (2) The scientists and researchers who adhere to what can generally be called the cognitive viewpoint, i.e., who find that human cognition plays an essential role in analysing and understanding human performance.

Prolegomenon

- (3) The specialists and engineers who are practically involved with the design, management, or use of man-machine systems in all fields and who are uneasy about the impact of human performance (the human factor) on system performance.
- (4) Those people who have an interest in the practical study of human behaviour and human cognition, and who are genuinely interested in or concerned about human performance in working situations.

2.4 ... And Who Should Not!

Just as there is an intended audience, there are also several groups of people who I expect will find this book rather disagreeable, and who therefore are advised not to read it unless they want to see their views challenged. These people include:

- (1) The practitioners and risk analysts who perform human reliability analysis and who are perfectly happy with the current approaches.
- (2) The scientists and researchers who firmly believe that the study of human cognition can only be carried out with well-controlled experiments and rigorous quantitative / statistical methods. This also includes those who believe that computational models or information processing descriptions can provide perfectly adequate explanations for human performance.
- (3) The specialists and engineers who cannot understand why some people have misgivings about quantifying probabilities for human errors and why these people therefore are reluctant to provide such numbers.
- (4) Those people who think that "human error" is a perfectly good root cause, and that the solution to the problem of "human error" basically is to increase the level of automation.

Any readers who feel that they do not belong to either of these groups, for instance because they are not interested in this field at all, should probably decide for themselves whether they want to go on reading. I expect, however, that they will find this book rather boring.

3. Chapter By Chapter

The chapters in this book have been organised to express a certain flow of thought or line of argument. It may, however, not be as obvious to the reader as it is to the author. Furthermore, different readers may be looking for different things, and therefore need not read the chapters in the same order -- or indeed read all the chapters.

Chapter 1 provides a broad account of the background for the concern with the Reliability Of Cognition. It describes how the need to consider the human factor or human operator arose, and how the technological development apparently has caused a greater susceptibility to incorrect or erroneous actions. This is followed by a discussion of how accidents usually are described and what the typical responses or reactions are.

Next, the chapter opens the discussion of human reliability analysis and how it can be understood from the cognitive viewpoint. The predominant approach is to look for a specific and quantifiable cause, as exemplified by the case of the President's heart attack. The notion of "human error" is examined and the suggestion is made that it should be replaced with the concept of an erroneous action. The point is made that the concern should be to prevent or avoid unwanted consequences rather than to study human reliability and erroneous actions as separate topics.

Chapter 1 ends by a discussion of the nature of human cognition and in particular the debate about whether human cognition is inherently simple or complex. The simple view is consistent with the predominant decomposition approach in human reliability analysis. It is argued that this approach has produced two artifacts: the idea about the individual action and the performance shaping factor. Both artifacts have contributed to the problems of current HRA practice.

Chapter 2 argues for the need to have better models of the Reliability Of Cognition. It begins by developing a definition of human reliability, and continues by describing the current decomposition principle. It is argued that the current approaches are based on two assumptions about repeatability of events and similarity between situations. It is pointed out that whereas these assumptions are correct for technical systems, they are not tenable for humans. The assumptions are the result of transferring the notion of a machine to the description of humans, but this is not appropriate -- not even as the notion of a fallible machine. A human being should fundamentally be described as a

Prolegomenon

cognitive system, and this has consequences for the methods that can be used.

Chapter 2 continues the discussion of "human error" and erroneous actions by proposing a clear distinction between phenotypes (manifestations) and genotypes (causes) of erroneous actions. This is supplemented by a complete taxonomy for the phenotypes of erroneous actions. Finally, the nature of the Reliability Of Cognition is discussed in relation to the ways in which tasks and work contexts have changed. This has led to an increased dependence on tasks that involve "thinking" rather than "doing", hence on human cognition. Human performance assessments must accordingly take this dependence into account, and put greater emphasis on the context of human actions. This requires an adequate model of human cognition.

Chapter 3 gives a critical account of human reliability assessment as it is currently practiced. The consequences of the decomposition principle are further elaborated by characterising the atomistic and the mechanistic assumptions. The predominantly quantitative approaches are exemplified by discussing the difference between curve-fitting and model identification. There is a need for better models to support the assessment of human reliability. However, the effort to quantify such assessments define a paradox: in order to have quantification it is necessary first to have a proper qualitative description or model. In other words, it is necessary to specify the data that are needed before they can be sought.

The practical problems of analysing the reliability of performance are discussed by presenting a comprehensive view on the nature of data. This explains the coupling between data and the underlying concepts, and how the notion of objective raw data is an illusion. It is followed by an overview of the different types of data and associated methods that are used in human reliability analysis: empirical data, data from simulations, and expert judgments. The chapter ends by summarising a major Human Factors Reliability Benchmark Exercise and by comparing the existing methods to a so-called "ideal" method.

Chapter 4 begins the description of the model of cognition that will be used as a basis for the method. It starts by recapitulating the three main approaches to the modelling of cognition: the S-O-R, the information processing approach and the cognitive viewpoint. This is followed by a characterisation of two major classes of models, called procedural prototypes and contextual control models. The former express the view that performance can be seen as variations of a pre-defined

sequence (the prototype); an example of that is found in the typical decision making model. In contrast to that the contextual control models emphasise that the sequence of actions is the result of an active choice. This choice depends on the current context, and the emphasis therefore should be put on how this choice is made.

A contextual control model has two parts: the competence model which describes which actions and plans are possible, and the control model which describes how the choice of next action is controlled. A distinction is made between several levels of control, exemplified by four distinct control modes called scrambled, opportunistic, tactical, and strategic. This is further developed in terms of a specific instance of the contextual control model called COCOM. The COCOM is described in terms of the main parameters that determine the performance characteristics on each level of control and the ways in which control can change from level to level.

Chapter 5 describes the new approach to human reliability assessment, called the Dependent Differentiation Method (DDM). The basis for the method is a systematic description of the tasks, derived by a Goals-Means Task Analysis. This task analysis method is explained in detail and the procedure is illustrated by an example. The DDM uses a characterisation of the common features of the task, named the Common Performance Modes (CPM). The CPMs are a convenient way of describing the impact of the context on the control of actions. The CPMs can be determined from the outcome of the Goals-means Task Analysis. Through an iteration procedure the DDM establishes the likely levels of the CPMs and thereby also the probable control modes. The further characterisation of the performance is based on refining the description of the control modes and how they influence the choice of actions. In cases where specific actions are known to be critical for the system, they can be analysed in detail using the same principles.

The conclusion is that it is not the Reliability Of Cognition which is important *per se*, but rather how it influences performance. The DDM therefore does not strive to produce a measure of the Reliability Of Cognition, but rather of the reliability of performance as a whole. This can be done in a qualitative fashion and improved, e.g. by using fuzzy set descriptions. It may also ultimately be turned into a quantitative description, but this should only be done if the numbers can be given a meaningful interpretation.

Prolegomenon

Chapter 6, finally, discusses a number of issues that are affected by the model and method developed in Chapters 4 and 5. It is pointed out that accident analysis is possible because the context is known and that performance prediction consequently should serve to describe the likely context as a prerequisite to describing individual actions. The consequences of the contextual control view are discussed as they apply to the design of man-machine systems and human-computer interaction. The practical problems in carrying out a human reliability analysis are addressed, and the prospects of providing computer support for the method are outlined. Following that, a framework is proposed to compare various methods for human reliability analysis.

The chapter ends by bringing forward an important concept of human cognition: attention. Attention is considered in relation to the reliability Of Cognition and in relation to the COCOM. It is argued that attention is a concomitant rather than a direct aspect of the contextual control view, and that it comprises several of the parameters that were described for the model. The effect of (a lack of) attention can best be seen by describing how it affects the choice of actions. The possible effects of a lack of attention depend on the relative task demands and on the possibilities for unwanted consequences to manifest themselves -- both of which can be understood in terms of the contextual control model and determined by the DDM.

4. Miscellanea

4.1 Model Multiplicity

The notion of models of cognition is widespread and is used in many different ways. The need for models can, however, be made clearer if a distinction is made between different instances of models:

- (1) **Scientific:** the primary purpose is here to aid understanding of something (a phenomena, a system). A scientific model explains the phenomenon in question and provides an account of the mechanisms or functions (the causal or functional architecture) that underlie the phenomenon.

- (2) **Engineering:** the primary purpose is to develop a representation of a system which can be used to calculate or predict future developments. The model is a translation of essential functional relationships and dependencies into a form which enables controlled manipulation of the independent parameters (including the environment). An engineering model can serve its purpose without necessarily constituting an explanation.
- (3) **Cybernetic:** the primary purpose is to provide the representation necessary to control a system. This usage is based on the Law of Requisite Variety, which can be interpreted as saying that a regulator of a system must be a model of that system. Control implies a certain amount of prediction, but the needs for precision and details are quite different from the engineering use of models. Similarly, a cybernetic model is not always useful as an explanation.

In the field of human reliability analysis a distinction is often made between engineering models and rigorous models. This book takes neither route, but instead proposes a pragmatic (read: cybernetic) model. This model has originally been developed to help in controlling a simulation of man-machine interaction, and can therefore easily be used to describe how actions are controlled. In this way it can serve as the basis for developing a method to analyse the reliability of human performance. It does not try to fulfil the need for engineering or rigorous models that is expressed by current HRA; the view is rather that this need is an artifact of the dominating approaches, hence that it disappears if an alternative solution can be developed.

4.2 Scope of the Model and the Method

The book develops both a specific model and a specific method. The obvious question is how general these are. The detailed example is taken from the field of nuclear power plants; since this field has had an influence on cognitive engineering which is disproportionately large -- due mainly to a limited number of widely published accidents -- it is not unreasonable to ask whether the model and the method, unintentionally, is limited to this area.

The answer is that both the model and the method have been developed to be applicable to a wide range of fields. The model is about

Prolegomenon

how actions are chosen and controlled; there is nothing in the model itself which favours one particular field of application. The restriction is rather that the model is concerned with human actions in the context of work with dynamic processes; this may possibly exclude other areas, such as information retrieval or text processing, although this is far from certain. Anyway, it is a limitation that is not unacceptable.

The method is designed to identify the influences from the context where the actions take place and to find the possible ways in which unwanted consequences can occur. This is predicated on a view of human action as purposeful activities carried out in a complex environment which is only partly known -- and only partly knowable. It will therefore not be surprising if this particular method of analysis is inappropriate or even inadequate for other purposes. In order to be useful a method must be of limited scope -- it must trade breadth for depth. However, within the field of work with dynamic processes I believe that the method can be of general use to determine the possible effects of limited human reliability. Neither the model nor the method are limited to specific fields such as nuclear power plants or aviation.

4.3 Terminology: An Apology to the Sensitive Reader

A small, but important, issue is which term should be used to describe the combination of people and machines that provide the context for the contents of this book. Until the mid-1970s the preferred term was man-machine system (e.g. Singleton, 1974) and no one seemed to have any problems with that. Due to the growing tendency to avoid so-called sexist language, the term man-machine system fell somewhat into disrepute and was replaced with terms like person-machine system or human-machine system. In the 1980s the developments in the study of how people interact with computers produced two new candidate terms: human-machine interaction (HCI; in Europe) and computer-human interaction (CHI; in the US). HCI / CHI, however, only deal with a subset of the problems that are addressed by the study of man-machine systems, and can therefore not be used as substitutes.

I shall, in this book, continue to use the term man-machine system, abbreviated as MMS. There are several reasons for that. Firstly, one meaning of the word man (and usually the first entry in dictionaries) is human (being), and the man in MMS it is to be understood in this sense rather than as a synonym for male. Secondly, although the term MMS may offend some academics, it is well entrenched in the applied fields.

One of the most prestigious journals is called "International Journal of Man-Machine Studies"; and practitioners routinely refer to MMS and MMI (meaning either man-machine interaction or an-machine interface). Changing the term to e.g. human-machine system would also require that the widely used acronyms were changed to HMS and HMI. Since this would probably cause a lot of unnecessary confusion, I have decided to stick with the usage of man-machine system and MMS. I hope that readers will not be offended by this.

A related, but less contentious issue, is the choice of a term to refer to the people or persons who work with the machines. The more frequently used candidates are "operator", "user", "person", and "agent." I have decided to use the term person throughout the book. In cases where it is necessary to use a personal pronoun, I have opted for "he". This is not for sexist reasons, but purely for convenience and conformity with the tradition. Finally, most of the people who work in industrial settings such as power plants and cockpits are undeniably male. So using the pronoun "he" could also be defended on the grounds of the *a priori* distribution in the population.

4.4 Acknowledgements

It is customary to acknowledge intellectual debts in the writing of a book like this, and I am indeed very happy to do so. I will, however, not produce a long list of names. Rather I will acknowledge my intellectual debts to what is sometimes known as the "cognitive circus" -- the group of people from a broad range of countries who for the last 10-15 years regularly have met (in subsets) for various occasions and among whom the cognitive viewpoint gradually has matured. Many of the ideas described in this book have developed during the meetings of the "cognitive circus" -- in presentations or through discussions. Since it is impossible to attribute every idea to a specific source, I refrain from doing it altogether. The book is both an expression of the views of the "cognitive circus", as of myself.

I would, however, like to mention a few people without whom this book might not have been realised. Firstly, Ed Dougherty who started the whole thing by his (in)famous paper in 1990. Ed was supportive of the idea of writing this book from the very start, and has been willing to provide me with his view on many things as the chapters gradually emerged; in particular, he suggested the "Feed and Bleed" as a good

Prolegomenon

example to use and provided me with many insights on that particular event.

Secondly, much of the theory presented here has been developed as part of the work in two projects, the Human Reliability Analysis Method, sponsored by the European Space Agency, and the System Response Generator, sponsored by the CEC. I have learned a lot through many discussions with my colleagues in these projects, as I am sure they can see throughout the book. I have in particular enjoyed many hours of discussion with Robert Taylor and particularly (standing, sitting, walking, and running!) with Carlo Cacciabue. During the later phases of writing I have received many useful comments and criticisms from Lisanne Bainbridge, Paul Booth, Yushi Fujita, John Hammer, Jacques Leplat, and Neville Moray. The latter in particular did his best to correct the worst abuses of the English language. At last, I have to thank Dave Woods; although he has not been closely involved with the writing of this book we are twin brothers of the mind and our irregular collaboration over the last decade or so has helped in cementing the foundations of cognitive systems engineering -- and therefore also the views expressed in this book.

Finally, and most of all, I must thank my wife Agnes for her unwavering patience and support during the many evenings and weekends that I have spent time writing and rewriting chapter upon chapter instead of being with her. In addition, her common sense has often forced me to make clear what I was writing about -- expressing it without overly use of technical jargon, and not writing for the initiated and converted.

Needless to say, despite my discussions and loans from others (at times incompletely acknowledged), the responsibility for the final result is mine. If there is any merit or value in what I have written, I gladly claim the honour. But neither will I shy away from anything that is incorrectly or wrongly put. I have tried to avoid mistakes, but if there are any the blame is certainly mine.