EXPERT JUDGEMENT
OF
HUMAN RELIABILITY

Prepared by the
CSNI Task Force on
Expert Judgement on Human Reliability

JANUARY 1985

COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS
OECD NUCLEAR ENERGY AGENCY
38, boulevard Suchet, 75016 Paris, France
EXPERT JUDGEMENT
OF
HUMAN RELIABILITY

Prepared by the
CSNI Task Force on
Expert Judgement on Human Reliability

JANUARY 1985
The NEA Committee on the Safety of Nuclear Installations (CSNI) is an international committee made up of scientists and engineers who have responsibilities for nuclear safety research and nuclear licensing. The Committee was set up in 1973 to develop and co-ordinate the Nuclear Energy Agency's work in nuclear safety matters, replacing the former Committee on Reactor Safety Technology (CREST) with its more limited scope.

The Committee's purpose is to foster international co-operation in nuclear safety amongst the OECD Member countries. This is done in a number of ways. Full use is made of the traditional methods of co-operation, such as information exchanges, establishment of working groups, and organisation of conferences. Some of these arrangements are of immediate benefit to Member countries, for example by enriching the data base available to national regulatory authorities and to the scientific community at large. Other questions may be taken up by the Committee itself with the aim of achieving an international consensus wherever possible. The traditional approach to co-operation is increasingly being reinforced by the creation of co-operative (international) research projects, such as PISC and LOFT, and by a novel form of collaboration known as the international standard problem exercise, for testing the performance of computer codes, test methods, etc. used in safety assessments. These exercises are now being conducted in most sectors of the nuclear safety programme.

The greater part of the CSNI co-operative programme is concerned with safety technology for water reactors. The principal areas covered are operating experience and the human factor, reactor system response during abnormal transients, various aspects of primary circuit integrity, the phenomenology of radioactive releases in reactor accidents, and risk assessment. The Committee also studies the safety of the fuel cycle, conducts periodic surveys of reactor safety research programmes and operates an international mechanism for exchanging reports on power plant incidents.

The Committee has set up a sub-Committee on Licensing which examines a variety of nuclear regulatory problems, provides a forum for the free discussion of licensing questions and reviews the regulatory impact of the conclusions reached by CSNI.

* * * * *

A "Restricted" OECD document should not be communicated except for official purposes. The Secretariat and Member governments of the OECD are requested to take the necessary action to ensure the security of these documents.

The opinions expressed and arguments employed in this document are the responsibility of the authors and do not necessarily represent those of the OECD.

Requests for additional copies of this report should be addressed to:

Nuclear Safety Division
OECD Nuclear Energy Agency
38 boulevard Suchet
F-75016 Paris
ABBREVIATIONS

HEP  Human Error Probability
HRA  Human Reliability Assessment
IDA  Influence Diagram Approach
IDCOR Industry Degraded Core Programme
LER  Licensee Event Report
MAUD  Multi-Attribute Utility Decomposition
PRA  Probabilistic Risk Assessment
PSF  Performance Shaping Factors
SLI  Success Likelihood Index
SLIM  Success Likelihood Index Method
STAHR  Socio-Technical Assessment of Human Reliability
THERP  Technique for Human Error Rate Prediction
# TABLE OF CONTENTS

## CHAPTER 1  INTRODUCTION

## CHAPTER 2  HUMAN RELIABILITY ASSESSMENT (HRA) IN PROBABILISTIC RISK ASSESSMENT (PRA)

1. GENERAL STRUCTURE OF PRA

2. INCORPORATING HUMAN ERROR IN PRA

   2.1 Initiating Events
   2.2 Accident Sequences
   2.3 Event Tree Quantification

3. REMAINING PROBLEM AREAS

   3.1 Dependence
   3.2 Event Sequence Identification
   3.3 Human Error Data
   3.4 Validation

4. SUMMARY

## CHAPTER 3  THE USE OF EXPERT JUDGEMENT IN HUMAN RELIABILITY QUANTIFICATION

1. INTRODUCTION

2. JUSTIFICATION FOR THE USE OF JUDGEMENT BASED TECHNIQUES

   2.1 Availability of Data from Other Sources
   2.2 Capability to Quantify 'Cognitive' Errors
   2.3 Design Recommendations and Consciousness Raising
   2.4 Capability of Application at Various Levels of Decomposition

3. EXPERT JUDGEMENT BASED TECHNIQUES FOR HUMAN RELIABILITY ASSESSMENT

   3.1 Issues in Expert Judgement Techniques

      3.1.1 The validity of expert judgement based approaches
      3.1.2 The appropriate method of aggregating expert judgements
      3.1.3 Uncertainty determination
      3.1.4 Types of experts appropriate for judgement based procedures

---

(iv)
4. JUDGEMENT BASED HUMAN RELIABILITY EVALUATION TECHNIQUES.................20
   4.1 Direct Numerical Estimation........................................20
   4.2 Indirect Numerical Estimation......................................22
   4.3 Paired Comparisons..................................................22
   4.4 Ranking and Rating Procedures....................................24

5. TECHNIQUES BASED ON DECISION ANALYSIS APPROACHES..........................24
   5.1 The Success Likelihood Index Methodology (SLIM)......................25
       5.1.1 An example of the SLI procedure...............................26
       5.1.2 SLIM-MAUD: An implementation of SLIM through the use of MAUD..................30
       5.1.3 Advantages of the SLIM-MAUD approach........................32
       5.1.4 Overall evaluation of the SLIM approach.......................33
   5.2 A Socio-Technical Approach to Assessing Human Reliability: The Influence Diagram........33
       5.2.1 Introduction....................................................34
       5.2.2 The STAHRI Influence Diagram................................36
       5.2.3 Group processes................................................43
       5.2.4 Evaluation of the Influence Diagram Approach................44

6. SELECTION OF AN APPROPRIATE TECHNIQUE.....................................45
   6.1 Number of Judges Available.........................................45
   6.2 Number of HEPs to be Estimated.....................................45
   6.3 Time Available......................................................45
   6.4 Type of Experts Available..........................................46
   6.5 Physical Location of Experts.......................................46
   6.6 Specificity of the Events being Evaluated........................46
   6.7 Homogeneity of the Situations being Evaluated.....................46
   6.8 Order of Magnitude of the Events..................................46
   6.9 Availability of Calibration Data...................................47
   6.10 Resource Availability...............................................47
   6.11 Importance of Consciousness Raising and Design Recommendation Functions..................47

7. CURRENT STATUS OF JUDGEMENT BASED TECHNIQUES...............................47

8. CONCLUSIONS...............................................................48

CHAPTER 4 STATUS OF HRA ACTIVITIES IN MEMBER COUNTRIES........................51
   1. FINLAND...............................................................52
   2. FRANCE...............................................................54
   3. FEDERAL REPUBLIC OF GERMANY.....................................58
   4. UNITED KINGDOM......................................................60
   5. UNITED STATES OF AMERICA..........................................67
CHAPTER 5  CONCLUSIONS AND RECOMMENDATIONS

1. INTRODUCTION .........................................................69
2. QUALITATIVE ASPECTS.............................................69
3. QUANTITATIVE ASPECTS...........................................70
4. EXPERT JUDGEMENT APPROACHES TO HUMAN RELIABILITY
   ASSESSMENT.........................................................70
5. RECOMMENDATIONS..................................................70

BIBLIOGRAPHY ...........................................................71

APPENDIX .................................................................77
CHAPTER 1

INTRODUCTION

After the Three Mile Island accident in 1979, Human Reliability Assessment (HRA) has received more emphasis as an important part of the Probabilistic Risk Assessment (PRA). The traditional methods of quantifying human error had been mainly based on empirical data from operating plants. In the Technique for Human Error Rate Prediction (HERP), for example, the "decomposition" approach was adopted, which assumed that any task can be decomposed into a series of elementary actions with the use of an event tree method. A drawback of this approach has been, however, the difficulty in collecting data and modelling for high-level "cognitive" human functions such as diagnosis, problem solving and strategy formulation, which could be considered as critical factors in the real operation of nuclear power plants.

Various techniques based on the expert judgement approach might be able to overcome this problem. In order to explore the advantages and disadvantages of this technique, CSNI established a Task Force on Expert Judgement on Human Reliability in early 1983. Various approaches currently available for obtaining human error probabilities with the use of this technique were reviewed in comparison with other HRA activities in Member countries. This report describes the work carried out during 1983 and 1984 by this Task Force, whose members are listed in the following pages.

Chapter 2 discusses the present situation with regard to how human error is assessed with the use of PRA methods. A number of problem areas which need to be considered in the application of PRA techniques have been identified. These include the absence of a modelling approach to identify high-level "cognitive" errors, the need for a human performance model to minimise the potential for human error and the difficulties in quantifying human error.

Bearing these problems in mind, Chapter 3 discusses the use of expert judgement techniques in human reliability quantification. The justification for the use of this technique is first discussed, followed by a description and analysis of the approaches currently available for obtaining human error probabilities. It is also shown that several of these approaches are able to provide solutions to the problem areas discussed in Chapter 2.

Chapter 4 is devoted to the compilation of the descriptions regarding the current status of HRA activities in Member countries. Contributions have been received from Finland, France, the F.R.G., U.K. and U.S.A.

The report concludes in Chapter 5 by presenting several general aspects identified through the review. It is pointed out that judgement based approaches to HRA appear to be a viable alternative to existing decomposition methodologies such as HERP. Possible directions of the future activities relating to HRA are also indicated as recommendations.
MEMBERS OF THE TASK FORCE

Finland
Mr. B. Wahlström
Technical Research Centre of Finland
Electrical Engineering Laboratory (Espoo)

France
Mr. A. Cahuzac
Electricité de France (Clamart)

Mr. D. Soulatges
Ms. C. Villeroux
Commissariat à l'ÉnergieAtomique
DSN/SAF (Fontenay-aux-Roses)

Federal Republic of Germany
Mr. G. Reichart
Gesellschaft für Reaktorsicherheit
Forschungsgelände (Garching)

Italy
Mr. G. Finetti
ENEA-DISP (Rome)

Japan
Mr. T. Tobioka
Reactor Safety Research Centre
JAERI (Ibaraki)

Netherlands
Mr. R.W. van Otterloo
KEMA (Arnhem)

Spain
Mr. J.M. Figueroa Clavijo
Consejo de Seguridad Nuclear (Madrid)

Sweden
Mr. O. Nockert
Nuclear Safety Board of Swedish Utilities (Stockholm)

Switzerland
Mr. A. Voumard
Nuclear Safety Department
Hauptabteilung für die Sicherheit der Kernanlagen
(HSK) (Würenlingen)

United Kingdom
Mr. G. Ballard
UKAEA
Safety and Reliability Directorate (Warrington)

Dr. D. Embrey (Chairman)
Mr. P. Humphreys
Human Reliability Associates (Lancashire)

Mr. H. McCririck
Mr. W. Ross
Nuclear Installations Inspectorate
Health and Safety Executive (London)

United States
Mr. H.B. Clayton
Mr. H.L. Thompson
Office of Nuclear Reactor Regulation
USNRC (Washington, DC)
United States (continued)

Mr. P. Dietz
INPO (Atlanta)

Commission of the European Communities

Mr. G. Mancini
Joint Research Centre - Ispra (Italy)

OECD Nuclear Energy Agency (Secretariat)

Mr. Y. Otsuka
Mr. K. Morimoto
Nuclear Safety Division (Paris, France)
CHAPTER 2

HUMAN RELIABILITY ASSESSMENT (HRA) IN PROBABILISTIC RISK ASSESSMENT (PRA)

1. GENERAL STRUCTURE OF PRA

The role of PRA is twofold,

- to identify potential areas of significant risk and indicate how improvements can be made;

- to quantify the overall risk from a potentially hazardous plant.

To achieve these aims a PRA has a general structure which was established in WASH-1400 and has not been significantly altered. The major elements are:

a) Identify the sources of potential hazard: this is straightforward in the case of a nuclear power plant since the major hazard is a radioactive release from the core.

b) Identify initiating events that could lead to hazard: this will involve some structured logic (see fig. 1) in order that a complete identification can be made. The final event description will normally be in terms of the failure of some hardware system.

c) Determine accident sequences following from initiating events: as a result of an initiating event certain safety functions will be required to maintain the reactor in a safe condition. If some safety systems fail, the reactor plant will progress through various physical states calling upon other safety systems until the incident sequence is terminated. This incident description is usually structured by the use of an event tree (fig. 2) which characterises possible sequences in terms of a limited number of parameters. Typical parameters will be the functioning of various safety systems such as the emergency core injection and the containment.

d) Event sequence quantification: the event tree describing the accident can be quantified by a frequency of occurrence for any particular sequence. This quantification will require the frequency of the initiating event and the probability of failure on demand of the relevant safety functions. These latter figures will be determined by a reliability assessment of the appropriate systems.

e) Plant overall risk: the total risk from the plant will be a function of the frequency of all the possible accident sequences and their consequences. Any dominant contributions to risk can be identified and the reasons for their significance determined.

2. INCORPORATING HUMAN ERROR IN PRA

The discussion of PRA has focussed on the hardware aspects of the reactor plant; where does human error fit into this picture? One could take the view that all incidents and all failures are the result of human error or by operators, designers, managers, etc. This is obviously basically true but that is not a useful way to approach the subject. Therefore let us consider how we currently include human error in PRA, taking the basic stages previously identified.
Fig. 2: Standard event tree for PWR small loss-of-coolant accident.
2.1 Initiating Events

Incidents are basically initiated by some failure of hardware. The reason for the hardware failure may range from component failure to operator's error but the only number required by PRA is the overall frequency of the hardware failure which leads to the incident. To determine this overall frequency by assessment of the causes of hardware failure would require an enumeration of human error in addition to many other causes. This would be a difficult task and one which we could not claim to be complete in any sense. The alternative is to use actual operational data on the frequency of failure of particular hardware. This data will include all the appropriate contributions from causes such as operator and maintenance error. Thus, providing the system of interest is sufficiently similar to equipment on which there is operational experience, then this phase of PRA can be accomplished without any explicit human error analysis. In practice it may be desirable to reduce or minimise the occurrence of certain incidents and in this case an assessment of the causes of hardware failure would be required.

One potential problem with this past data approach concerns the application to new systems. Any new system will incorporate some design errors and it may be argued that past data cannot allow for these design failures. However, past data will incorporate some measure of past design failures, but more significantly initiating failures involve active systems which will be commissioned, tested and run, and therefore the majority of design errors should be eliminated during the early stages of life.

2.2 Accident Sequences

Given an initiating event, it is necessary to identify all systems or functions that may contribute to the course of the ensuing accident sequence. The event tree which describes the results of an initiating event has usually concentrated on particular safety system hardware and only included human actions when a specific need for an action has been identified. However, the possible range of human intervention in accident sequences is immense and in particular will include occasions where no human action had been identified as necessary. It is therefore clear that a description of event sequences only involving required human actions is unlikely to be a complete description of the potential scenarios resulting from an initiating event. The problem of identifying alternative event sequences incorporating non-required human actions requires an assessment of the human interaction with the plant. PRA has not in general attempted this type of analysis but has relied on the design of the plant to minimise the significance of such event sequences. This may be achieved by:-

(1) plant design to limit possible human intervention to specific required actions, e.g. automated processes
(11) design to ensure that human intervention produces a "safe" result
(111) information display on the state of the plant so that the operator clearly understands that progress of events and does not take inappropriate action due to faulty diagnosis
(1iv) information feedback on the result of any operator action so that incorrect actions can be identified and recovery procedures can be initiated.
All the above facets of plant should be geared to making unnecessary actions unlikely and to maximise the probability of correct action at the appropriate time.

2.3 Event Tree Quantification

The quantification of the frequency of any particular event sequence involves the frequency of the initiating event and the probability of failure of various hardware systems and identified, required human actions. In fact these probabilities are really conditional probabilities since they should properly take account of all preceding occurrences in the event sequence. The system failures will in general be quantified by data from past experience although this may be by a slightly indirect route in that the data will apply to system components and be used to synthesise a system reliability. The actual cause of system failure (or component failure) will in general not be identified but all significant contributors such as human actions should be incorporated within the data. The human actions incorporated will include design and specification errors as well as the more direct operator and maintenance errors.

It should be noted that the use of past data to incorporate presently unquantified aspects of a reliability analysis is not a new feature of assessment just applied to human error. Hardware systems also rely on past data, because in many cases the exact mechanisms of failure are difficult to predict and are not well understood. For example, even hardware failures could be viewed as human error in the sense that the failure could have been avoided by alternative design or manufacture. This method of approach therefore is the basis for most of the current probabilistic assessment, hardware or software.

A further element of the event tree quantification is the enumeration of failure of identified, required operator actions. This is perhaps the one explicit usage of human error analysis in current PRA. However, even in this case the analysis performed requires significant simplifying assumptions and is not universally accepted as valid. The operator actions assessed are limited to straightforward procedural tasks and explicitly do not include the identification/diagnosis of the incident. Under these circumstances the THERP method (NUREG/CR-1278) developed by Swain et al. at Sandia has received fairly wide usage. The basic elements of this method are a task analysis to identify the components of the procedure and a quantification via operator action fault trees using data provided in NUREG/CR-1278.

3. REMAINING PROBLEM AREAS

This summary of the present situation with regard to the incorporation of human error in PRA indicates that there are a number of remaining problem areas.

3.1 Dependence.

It was noted in the section on event tree quantification that all the probabilities used are properly conditional probabilities which allow for the preceding components of the event sequence. In practice this conditionality has rarely been recognised in PRA and independence of events has normally been assumed. The human element has a large capacity to establish a dependence between events and systems. Thus, within a safety system which has a number of redundant channels, operator and maintenance actions together with design
error significantly decrease the overall system reliability. This so-called common mode problem is recognised although current assessment methods are not adequate. Perhaps a more significant contribution of human error is the establishment of dependence between apparently separate safety systems. In the Three Mile Island accident the operator established a connection between the failure of the pressuriser relief valve and the high pressure injection (HPI) system which the risk assessment would not have incorporated on the basis of independent failure probabilities. This would be the case even though the probability of HPI failure included past failures due to human error -- the dependence on a previous event would not have been included.

3.2 Event Sequence Identification.

Current PRA clearly does not identify event sequences which include non-required operator actions. The potential for such sequences is immense and thus the current action focuses on prevention by design rather than assessment. It seems clear that this must also be the basic philosophy for the future but that the plant design should utilise a model of the operator interaction with the plant. The purpose of the model would be to aid the identification of potential problem areas and to indicate the appropriate action to minimise their effect. A typical example of this type of problem is the diagnosis of an incident. Clearly if the operator fails to correctly identify the actual condition of the plant then the potential for inappropriate action is increased. This could also be a major cause of the operator establishing a dependence between the various safety systems. The requirements of a model of human action in these circumstances would be:

(a) Identify potential sources of incorrect diagnosis
(b) Indicate action to minimise the potential problems
(c) Quantify the probability of incorrect diagnosis.

The third usage of the model (c) should be considered a long term aim -- the first two elements are the most important.

3.3 Human Error Data.

There is a current requirement for human error data to use in the quantification of the reliability of the operator in procedural tasks, but also to provide guidance for the human models that need to be developed to describe the man/machine interaction. The data in NUREG/CR-1278 and used in THERP may provide a basis, but there is some question as to its source. In any case that data is limited to simple tasks and would not provide the information to describe higher level tasks related to diagnosis and planning. A further area is that related to the so-called "performance shaping factors" (PSF) which are used to modify the base model of the man/machine interaction such as stress, supervision, plant management etc.

3.4 Validation.

Despite the considerable amount of work that has been performed to develop methods of human error assessment, there is little direct indication of the correctness of this work. It is clearly essential to attempt to obtain independent validation of the assessments made, preferably from a study of operational plant. The problems associated with the collection of human performance data are well known, but there is little value in carrying out method development unless the success of the work can be measured.
4. SUMMARY

The discussion in the preceding sections can be summarised by the following:-

Present Action

(1) Use past data to include human error in reliability assessment
(11) Use plant design to minimise the potential for inappropriate actions.

Problems

(1) Dependence -- the human is a major source of this problem area
(11) Plant design is not currently based on a model of human performance
(111) Human performance data to provide the basis for man/machine models and to attempt quantification is presently inadequate
(1iv) There is little validation of current work in the human reliability field.
CHAPTER 3

THE USE OF EXPERT JUDGEMENT IN HUMAN RELIABILITY QUANTIFICATION

1. Introduction

The discussion in chapter 2 described two aspects of the application of Human Reliability Assessment (HRA) to PRA. The first of these concerns the identification and modelling of potential human failures, particularly in accident sequences, and the second is the quantification of the likelihood of these failures for risk assessment purposes. Chapter 2 pointed out that within PRA the current approach to HRA addresses specific tasks (or sequences of human actions) that are required in connection with the various safety systems that are sequentially brought on line as an accident progresses. These tasks are highly proceduralised in nature. There is usually no attempt to identify other types of human errors such as diagnostic failures and errors of commission which could give rise to alternative accident sequences (e.g. as occurred at the Three Mile Island incident).

Three major problem areas were identified. The first of these was the absence of a modelling approach to identify the higher level 'cognitive' errors such as failures in diagnosis. The second area concerned the need for a human performance model to provide guidelines to designers with regard to how to minimise these types of error. The final problem area was concerned with quantification. In the short term there was a perceived need for better quality data for the assessment of the specified proceduralised actions associated with hardware systems, and in the longer term a requirement for methods to quantify cognitive errors.

Although this chapter will concentrate primarily on the quantification aspects of HRA, it will be shown that there is in fact a close relationship between the areas of modelling and quantification (see Embrey (1984) for a further discussion on the modelling of human errors in PRA). The primary objective will be to demonstrate that many of the present and future needs of quantification within HRA can be met by techniques which use expert judgement. To this end, the justification for the use of expert judgement based techniques will first be discussed, followed by a description and analysis of the approaches currently available for obtaining human error probabilities (HEPs) which utilise this approach. It will also be shown that several of these approaches are able to provide solutions to the other problem areas discussed in chapter 2, the modelling and quantification of cognitive failures, and the provision of design guidelines with regard to the reduction of human error. In this chapter we have not attempted to exhaustively review all of the
approaches which could be used to apply expert judgement to human error reliability evaluation. The most extensive review available is by Akersten and Wirstad (1983). Our intention has been to discuss approaches which have actually been applied in the evaluation of human reliability in nuclear power. Thus, technically interesting methods which have not yet been applied in this context (e.g. the Analytical Hierarchy Process of Saaty (1980)) have not been discussed.

2. Justification for the Use of Judgement Based Techniques.

2.1 Availability of Data from Other Sources

The primary justification for the use of judgement based techniques is the pragmatic one that there is very little data available from the field regarding human error rates, and this situation is unlikely to improve in the future. It is possible, in theory, to collect data on human errors in routine operations such as maintenance and normal plant evolutions, by setting up a data collection system similar to the U.S. licence event report (LER) reporting system or its proposed European equivalent. However, this requires considerable resources, and there are inherent classification problems which require that a comprehensive error taxonomy be applied, e.g. that proposed by the Group of Experts on Human Error Data and Assessment (Anon 1983), to allow data from disparate sources, e.g. different types of plant, to be aggregated together. This is necessary since human error probabilities are usually estimated by the ratio of the number of errors divided by the number of opportunities for error, and a reasonably large numerator and denominator are needed in order to assign reasonable uncertainty bounds to the estimates. Although it is at least feasible to collect data on routine situations by this means, the primary application area for HRA at present is to high risk, rare event scenarios such as Loss of Coolant Accidents, Steam Generator Tube Ruptures etc. It is, almost by definition, impossible to collect sufficient error frequency data from such rare events to produce credible estimates for error probabilities for the task as a whole.

The use of training simulators has been suggested as a means of accumulating human error data on individual task elements within rare event scenarios, where these are used as part of training exercises. In fact, such a study has recently been published, Beare et al. (1984), in which error probabilities for various task elements within such scenarios have been obtained by recording the frequency of erroneous actions for a range of PWR and BWR transients. However, a report by Nicolet and De Vaugelas (1983), indicated several problems with this approach. There are systematic differences between the training simulator situation and a real incident, in terms of the expectations that the operator may have with regard to the likelihood that a high risk event may occur. There are also differences in the consequences of failure, and the stress that he may experience. In addition, because of the expense of simulator runs, the
operators may be involved in the simulated incident for no more that 30 minutes, instead of the considerably longer time period that a real incident could span.

Another problem concerns the aggregation of errors which may have the same external failure mode, e.g. the omission of a step in a procedure, but which may have arisen because of a variety of different internal failure modes, Hollnagel et al. (1981). For example, the omission could have occurred because of insufficient time, because the operator did not perceive that the step was necessary, or for a variety of other reasons. Although it could be argued that the external failure mode is the only data of interest or relevance to PRA, the nature of the internal malfunctions, and hence the frequency of externally observable errors, is likely to be very different in real and simulated incidents.

Apart from these technical considerations, the cost associated with human error data collection using simulators is very high. This is because replica simulators are very expensive to run, and also a considerable amount of manual data processing is necessary in order to supplement the data recorded by automatic data acquisition systems, (Beare et al. 1984). These assume that all deviations from the standard operating procedures are errors, whereas in fact there are often alternative strategies available to the operators for achieving the required goals in a transient.

We can therefore conclude the paucity of data available from other sources is an important reason for utilising expert judgement based techniques.

2.2 Capability to Quantify 'Cognitive' Errors

Most of the existing HRA techniques are directed primarily at sequential proceduralised actions where success or failure can at least in theory be externally observed. However, there is a growing realisation that the most serious errors are those involving cognitive functions such as diagnosis, problem solving and strategy formulation. Studies of actual incidents, Pew et al., (1981), and realistic simulations of high risk scenarios, Woods et al. (1982), have emphasised the importance of such errors.

If an error occurs in diagnosis, for example, inappropriate procedures may be used, and the operating team may not realise for some time that an error has been made, because feedback from the system may be interpreted in terms of the original inappropriate hypothesis. If, on the other hand, the operator has made a correct diagnosis, or formulated an appropriate strategy to reach a safe stable state, then simple procedural errors such as valve mal-operations are likely to be rapidly detected because of feedback which indicates that the system did not change in a manner consistent with the original diagnosis.
The significance of errors which originate at a higher level of information processing means that it is important for HRA techniques to address this area. It does not seem likely that it will be possible to assess these errors using a data bank approach, because of the difficulty of observing diagnostic and other covert processes. The only feasible alternative is to use the judgements of individuals who have experienced these errors in plant or simulator situations, or who have other appropriate knowledge that bears on the assessment of the likelihood of these failures occurring.

2.3 Design Recommendations and Consciousness Raising

In addition to providing numerical probabilistic outputs which can be used to calculate expected risk values for comparison with risk criteria, another important function of PRA is to indicate ways in which risk can be reduced in the most effective manner. This is achieved partly by identifying those aspects of a system which are particularly sensitive in terms of their contribution to risk, and also by indicating the design changes which could be employed to reduce the likelihood of failure in a cost effective manner.

Directly analogous processes are carried out in HRA, and several of the judgement based techniques are able to provide numerical estimates of the relative effects of different influences such as procedures, training etc., which have a significant impact of the probability of failure.

Apart from the provision of numerical cost-benefit information relevant to design, some judgement based techniques include a specific 'consciousness raising' capability. Consciousness raising in this context refers to a structured process whereby participants e.g. designers, PRA analysts, gain an in-depth understanding of the nature of the potential human errors being investigated, their consequences, and the factors which influence their occurrence. This is a very useful capability, and in some cases may be as important, or even more important than the numerical estimates of human error probabilities which are produced by the procedure.

2.4 Capability of Application at Various Levels of Decomposition

An important advantage of the judgement based approaches is that they can be used at any level of decomposition of a task which is appropriate to the analysis. Thus in some cases it may be useful to estimate an overall probability of failure for a complex task which consists of many separate subgoals and a very large number of task elements, e.g. the probability that an operator will successfully manage a steam generator tube rupture. In other cases, interest may centre on whether or not the operator achieves certain required actions, e.g. changing to recirculation
mode in a PWR LOCA. For design purposes, it may be appropriate to consider operator functions in other ways.

For example, it may be useful to separately evaluate the likelihood of initially detecting the first indications of a transient, the interpretation of these symptoms, the probability of deciding upon and executing an appropriate set of procedures, etc. This type of analysis can be readily handled by judgement based methods.

The capability of judgement based techniques to assess human reliability at a variety of levels of decomposition means that they can be utilised for the broad screening of a number of situations, or for more detailed analyses of particularly critical human actions.

3. Expert Judgement Based Techniques for Human Reliability Assessment

The basic philosophy underlying the use of judgement based techniques is that appropriate experts, e.g. plant operators, trainers, human factors specialists, PRA analysts etc. possess knowledge which can be utilised to generate numerical estimates of human error probabilities. As discussed in section (2.3), this expertise can also be utilised to provide design information and other important qualitative outputs.

Judgement based techniques can be regarded as a means for providing data for use in decomposition approaches to quantification such as the Technique for Human Error Rate Prediction (THERP), Swain and Guttmann (1983). In these techniques complex tasks are broken down into their constituent task elements to which human error data are assigned from a data bank. The failure probability for the task as a whole is then obtained by a recombination of these elemental probabilities. However, we prefer to view judgement based approaches as stand alone techniques in their own right, which are not constrained to analyse a task at the level of procedural elements. From this point of view, judgement based techniques can be described as 'wholistic' as opposed to 'reductionist'. Even with wholistic techniques, however, it is desirable to carry out some form of prior task analysis which identifies some of the important failure modes which may occur. This can provide a starting point for the structural group discussions which are an important feature of some judgement based techniques.

3.1 Issues in Expert Judgement Techniques

A number of important issues arise in connection with the use of expert judgement based techniques, which will be discussed prior to the description of the techniques themselves.
3.1.1 The validity of expert judgement based approaches.

Objections are often raised to the use of expert judgement based approaches because of the biases present in probability judgements which have been documented by Tversky and Kahneman (1981, 1974), Fischhoff and MacGregor (1982) and a number of other workers. However, the existence of these biases is not universally accepted. Phillips (1979) argues that many of the apparent biases are an artifact of artificial laboratory based situations that probability assessors are typically exposed to in psychological research. A review by Stillwell et al. (1982) found that experts are generally able to provide accurate likelihood assessments when compared with objective likelihoods (e.g. relative frequencies). Further evidence for the ability of experts to produce well calibrated probability judgements comes from Lichtenstein et al. (1981). This level of performance is contingent on certain conditions being observed. For example, the experts should be well motivated and experienced in quantitative thinking, and well trained. They should be aware of potential biases, and the task to be evaluated should be carefully and completely defined and structured. An important requirement for well calibrated judgements is that the judges have real expertise in the area where they are making probability estimates.

Most of the approaches to be described in subsequent sections do not use absolute probability estimation directly, so that some of these biases are less important to these techniques. However, several techniques rely on the use of calibration probabilities (see sections 4.3, 5.1), and where these are unavailable from empirical sources, it may be necessary to employ absolute probability judgement. In this case, the precautions discussed in this section will need to be applied. A comprehensive description of techniques for facilitating probability judgements is contained in Stillwell et al. (op. cit.) and Seaver and Stillwell (1983).

3.1.2 The appropriate method for aggregating expert judgements.

In the application of judgement based techniques, it is usual to utilise a number of different judges with different perspectives on the situation being assessed. The two main approaches to combining the judgements are either to use a mathematical aggregation approach, e.g. simple averaging, or to use a consensus process where the judges come to an agreement regarding the appropriate values for the quantities being assessed. A composite approach is also possible, where judges first make individual evaluations, these are discussed in public and the judges then make separate private re-evaluations of their estimates. These are then combined mathematically using simple averaging. (This is the Nominal Group technique, Delbecq et al., 1975).
The major argument for mathematical aggregation is a logistical one: it may be expensive or unpracticable to bring groups of experts together to engage in a consensus process. In addition a group leader is required, which may involve additional expense. There is also some experimental evidence which suggests that there is little to be gained in terms of increased accuracy if judges are allowed to interact, Seaver (1978).

On the other hand there are many valid arguments for the use of consensus techniques. Mathematical aggregation leads to a loss of information through regression effects, and in evaluating relatively ill-defined situations (as is the case in PRA), it is critical to have a broad base of expertise involved in the judgement process. Research also suggests that judges are more confident of probability estimates generated by a consensus process. This is very important if the data generated by judgement processes is to be acceptable to the PRA community as a whole.

Despite the increased resources required for the generation of judgements via an interactive group process, our opinion is that the benefits outweigh the costs. The interaction process is aided by the existence of a structure provided by some of the HRA techniques, particularly the SLIM-MAUD approach and the Influence Diagram (see sections 5.1, 5.2). It should be noted that certain techniques, e.g. the Paired Comparison procedure (section 4.3) need individual judgements and cannot be exercised in a consensus mode.

When individual judgements are aggregated together mathematically, it is desirable to calculate some measure of interjudge consistency. Detailed procedures for obtaining such measures are provided for a wide range of techniques in Seaver and Stillwell (op.cit.).

3.1.3 Uncertainty determination

For many purposes in PRA, e.g. carrying out sensitivity analyses, or bounding calculations in fault tree analyses, it is useful to have some indication of the uncertainty bounds associated with error probability estimates generated by judgement processes. Various methods can be applied for achieving this, and they will be discussed in the context of the specific HRA techniques reviewed in subsequent sections.

3.1.4 Types of experts appropriate for judgement based procedures.

The most important requirement for any group of experts involved in the assessment of human reliability is that all participants must have some relevant information or insights regarding the events being evaluated. Some of the experts, e.g. plant operators, supervisors, will probably have situation-specific knowledge relating to the actual plant or scenario being
evaluated. This is important because there are often considerable differences between the detailed practices at different plants even within the same utility with the same reactor type and control room configuration.

It is also desirable that the assessment team includes at least one individual with ergonomics or human factors knowledge, particularly in the area of human reliability assessment. Such an individual will be able to assess the effects of factors such as stress, inadequate man-machine interface design and other variables which will impact upon the likelihood of error. An important requirement for effective quantification is a detailed knowledge of the nature of the situation that the operator will have to deal with. This requires inputs from individuals such as design engineers and thermohydraulics specialists, who may be able to delineate in some detail the expected behaviour of the plant, particularly if the incident being considered has been analysed using one of the thermohydraulic simulation codes such as RELAP.

In addition to the above judges, it is desirable that the assessment team includes a PRA specialist. Such an individual is likely to have a wide experience of carrying out numerical assessments and will ensure that the exercise provides data which is appropriate for the needs of the PRA as a whole.

The extent to which there is likely to be consistency between judgements made by different expert groups depends on the extent to which the groups have access to the same information about the situation being assessed or possess the same expertise. This needs to be verified by empirical research.


In subsequent sections, descriptions of the important HRA techniques which explicitly employ expert judgement will be given. An excellent review already exists, Seaver and Stillwell (op. cit.), which provides a detailed analysis of a number of these techniques, together with explicit procedures for their use in a nuclear PRA context. The approaches considered in that review will therefore only be discussed in summary form here. Additional techniques which were not fully described or which were omitted from the Seaver and Stillwell review will be considered in more detail in this chapter.

4.1. Direct Numerical Estimation

In Direct Numerical Estimation approaches, judges assign numerical failure likelihoods directly to the tasks being assessed. In order to avoid some of the biases discussed in previous sections, it is recommended that a logarithmic odds response scale is used, as is shown in figure 3. Judges simply
mark the scale at the appropriate position corresponding to the task being assessed.

This end of the scale is for incorrect action with a high likelihood of occurrence.

<table>
<thead>
<tr>
<th>Probability</th>
<th>Chance of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1 Chance in 1</td>
</tr>
<tr>
<td>.5</td>
<td>1 Chance in 2</td>
</tr>
<tr>
<td>.2</td>
<td>1 Chance in 5</td>
</tr>
<tr>
<td>.1</td>
<td>1 Chance in 10</td>
</tr>
<tr>
<td>.05</td>
<td>1 Chance in 20</td>
</tr>
<tr>
<td>.02</td>
<td>1 Chance in 50</td>
</tr>
<tr>
<td>.01</td>
<td>1 Chance in 100</td>
</tr>
<tr>
<td>.005</td>
<td>1 Chance in 200</td>
</tr>
<tr>
<td>.002</td>
<td>1 Chance in 333</td>
</tr>
<tr>
<td>.001</td>
<td>1 Chance in 500</td>
</tr>
<tr>
<td>.0005</td>
<td>1 Chance in 1,000</td>
</tr>
<tr>
<td>.0002</td>
<td>1 Chance in 2,000</td>
</tr>
<tr>
<td>.00005</td>
<td>1 Chance in 10,000</td>
</tr>
<tr>
<td>.00002</td>
<td>1 Chance in 20,000</td>
</tr>
<tr>
<td>.00001</td>
<td>1 Chance in 50,000</td>
</tr>
<tr>
<td>.000005</td>
<td>1 Chance in 100,000</td>
</tr>
<tr>
<td>.000002</td>
<td>1 Chance in 200,000</td>
</tr>
<tr>
<td>.000001</td>
<td>1 Chance in 500,000</td>
</tr>
<tr>
<td>.0000005</td>
<td>1 Chance in 1,000,000</td>
</tr>
<tr>
<td>.0000002</td>
<td>1 Chance in 2,000,000</td>
</tr>
<tr>
<td>.0000001</td>
<td>1 Chance in 5,000,000</td>
</tr>
<tr>
<td></td>
<td>1 Chance in 10,000,000</td>
</tr>
</tbody>
</table>

This end of the scale is for incorrect actions with a low likelihood of occurrence.

**Figure 3.** Logarithmic probability/odds scale for obtaining direct estimates of upper and lower bounds of SLIM produced HEP estimates.
The steps in exercising the technique are as follows:

1. Obtain odds judgements.
2. Aggregate individual judgements to provide a single odds judgement (or use a consensus or composite approach as discussed in section (3.1.2)).
3. Transform odds estimates into HEPs.
4. Determine inter-judge consistency.
5. Estimate uncertainty bounds.

The transformation of the odds into HEPs is accomplished using the formula:

\[
\text{HEP} = \frac{\text{odds}}{1 + \text{odds}}
\]

Inter-judge consistency measures can only be calculated if mathematical aggregation is used. In this case, Seaver and Stillwell (op.cit.) propose the use of an analysis of variance across experts. The calculation of uncertainty bounds in the mathematical aggregation case can be achieved by obtaining the variance of the experts' individual log HEP estimates, and from this the standard error and hence the 95 percent uncertainty bounds.

Where a consensus approach is adopted, uncertainty bounds can be estimated judgementally by using the same form of response scale as was used for the HEP estimation (figure 3). The experts are simply asked to mark the upper and lower uncertainty bounds within which they are 95 percent certain that the HEP estimate falls.


4.2 Indirect Numerical Estimation

In this procedure, the experts compare pairs of tasks and make ratio judgements of the form: 'task A is 10 times more likely than task B'. All the events in the set being evaluated are compared with one another to produce an interlinked set of judgements. If the probability of one event is known, then the other probabilities can be calculated from the ratio judgements. The reference probability can either be obtained empirically, by Direct Numerical Estimation, or by the use of the Influence Diagram approach described in section (5.2). As with Direct Numerical Estimation, individual judge's estimates can be aggregated mathematically, or a consensus approach can be adopted.

4.3. Paired Comparisons

The Paired Comparisons approach, in common with several other techniques, does not evaluate the probability of success of a group of tasks directly, but instead uses a procedure to position
each task on a 'likelihood of success' scale. This scale is then calibrated by including tasks with known HEPs in the original task set, or by obtaining absolute probability estimates (possibly using Direct Numerical Estimation) for at least two tasks. Once the scale has been calibrated, the positions of all the tasks on the scale can be converted to probabilities of success (or failure).

The Paired Comparisons approach was first used for human reliability evaluation by Blanchard et al. (1966), and subsequently by Rigby and Edelman (1968), Hunns and Daniels (1980), and Embrey and Kirwan (1983). The basic premise of the approach is that judges are better at making relative judgements of the form 'task A has a higher likelihood of success than task B' than assigning absolute probabilities. To use the technique pairs of tasks are selected randomly from the set of tasks being evaluated, and the judges are asked to state which task in each pair has the highest probability of success. This process is repeated for all possible pairings and is carried out independently by each judge.

The model underlying the Paired Comparison approach assumes that each task has an associated subjective distribution of perceived likelihood of success, and that this distribution is normal. When comparing the likelihoods of two tasks, a magnitude is selected from each of the associated distributions and the task with the greater magnitude is reported to have the higher likelihood of success. This process, when repeated across judges, allows the proportion of times that a task A is judged to be more likely to succeed than task B to be transformed into a normal deviate. The mean of all such deviates for a particular event, in comparison with all other events, is then taken as the position of that event on the likelihood of success scale.

It is assumed that there is a consistent, monotonic relationship between the likelihood of success scale and a probability of success scale. There is empirical evidence that this relationship is logarithmic, of the form:

$$\log \text{HEP} = as + b$$

where s is the mean scale rank value judged by the assessors, and a and b are constants (Pontecorvo, 1966; Embrey and Kirwan, 1983). Once the scale values have been assigned for the two tasks with known HEPs, the constants a and b can be evaluated and the HEPs corresponding to the scale values for the remaining tasks obtained from the resulting calibration equation. Detailed procedures for obtaining the judgements, calculating consistency measures and uncertainty bounds, are available in Seaver and Stillwell (op.cit.)

The main disadvantage of the paired comparisons procedure is the large number of judgements required, i.e. n(n-1)/2 where n is the number of tasks. Thus for 20 tasks 190 comparisons are necessary. However, procedures are available for reducing the
number of these comparisons, as discussed in Seaver and Stillwell (op.cit.) A more fundamental disadvantage of Paired Comparisons is that it is not a particularly scrutable approach, i.e. it is not possible to determine why events are placed in a particular position in the likelihood of success scale. In particular, the parameters which the judges perceive as being the determinants of the likelihood of success are not made explicit in this procedure.

The Paired Comparison approach has been applied to a series of nuclear power plant tasks in a study by Comer et al., (op. cit.).

4.4. Ranking and Rating Procedures

These procedures are considered together because they are conceptually very similar. In the ranking procedure, the tasks being assessed are simply ranked in order of likelihood of success. The rating technique involves the judges assigning a numerical value on a scale (e.g. from one to ten) of likelihood of success. Since each rank can be considered as a different rating, the methods are essentially equivalent.

The underlying psychological model used for these approaches is similar to that employed for paired comparisons. The boundaries between different rankings or rating categories are assumed to produce a normal distribution of subjective magnitude of likelihood of success (or failure). The boundaries, and hence the tasks defined within each boundary, are scaled in a similar manner to the approach used in Paired Comparison. These scale values are converted to probabilities using calibration tasks as before.

5. Techniques Based on Decision Analysis Approaches.

Both of the techniques discussed in this section are derived from methodologies which have been developed by decision analysts primarily in the context of military and business decision making. As discussed in Humphreys (1982), decision analysis techniques perform two important functions, 'bootstrapping' and 'consciousness raising', both of which are of relevance when these approaches are applied to the problem of evaluating human reliability. Consciousness raising has already been referred to in section (2.3) and is the process whereby the expert group gains an increased understanding of the nature of the situation being assessed, through an explicit structuring procedure which forms part of the technique. The consciousness raising aspects of decision analysis based HRA techniques means that they tend to require increased resources in terms of time, in comparison with simpler techniques such as Direct Numerical Estimation. However, if necessary, this aspect of the techniques can be reduced by developing generic structures which can be applied to broad classes of situations. This strategy would be employed in situations where the straightforward numerical assessment of HEPs
was the primary objective, and detailed analyses to obtain design recommendations etc. were of lower priority.

The term 'bootstrapping' refers to the process whereby the assessments to be made are decomposed into simpler assessments which are then recomposed using a formal composition rule to obtain the desired result. This approach grew out of work on probabilistic information processing models (Edwards, 1966; Slovic and Lichtenstein, 1971) and multi-attribute utility models (Huber, 1974; Humphreys, 1977). This indicated that in assessment situations where complicated information processing was involved (as is the case with HRA), the results obtained by using a normative composition rule to recompose the judges' decomposed assessments accorded better with an external criterion than the use of unaided judgement alone.

It is important to make a distinction between the form of decomposition utilised in decision analysis based HRA techniques and that employed in methodologies such as THERP. In THERP the structure of the decomposition is mainly externally specified in terms of the procedural steps prescribed by the engineering requirements of the situation that the operator has to handle, e.g. observe level A, operate valve B etc. In the case of the decision analysis based techniques, the decomposition is not task element related. With the Success Likelihood Index Methodology, for example, the decomposition is in terms of the characteristics of the task (or of the operators carrying out the task) which are likely to influence the probability of successful completion. In addition, the structure via which the decomposed judgements are combined is derived from the judges themselves, rather than being determined by external constraints such as specified procedures, or a data bank containing HEP values for a limited range of task elements.

5.1 The Success Likelihood Index Methodology (SLIM)

A detailed technical description of SLIM is available in a number of publications, e.g. Embrey (1983 a,b,c), Embrey et al. (1984 a,b). The basic version of SLIM will first be considered and in subsequent sections an improved, computer based implementation of the technique will be described.

The basic rationale underlying SLIM is that the likelihood of an error occurring in a particular situation depends on the combined effects of a relatively small set of performance shaping factors (PSFs). In brief, PSFs include both human traits and conditions of the work setting that are likely to influence an individual's performance. Examples of human traits that 'shape' performance might include the competence of an operator (as determined by training and experience), his/her morale and motivation etc. Conditions of the work setting which might affect performance include the time available to complete a task, task performance
aids, etc. It is assumed that an expert judge (or judges) is able to assess the relative importance (or weight) of each PSF with regard to its effect on reliability in the task being evaluated. It is also assumed that, independent of the assessment of relative importance, the judge(s) can make a numerical rating of how good or how bad the PSFs are in the task under consideration, (e.g., achieving recirculation in a pressurised water reactor (PWR) loss-of-coolant accident (LOCA)) where 'good' or 'bad' mean that the PSFs will either enhance or degrade reliability.

Having obtained the relative importance weights and ratings, these are multiplied together for each PSF and the resulting products are then summed to give the Success Likelihood Index (SLI). The SLI is a quantity which represents the overall belief of the judge(s) regarding the positive or negative effects of the PSFs on the likelihood of success for the task under consideration. If we can assume that as a result of their knowledge and experience the judge(s) have a correct idea of the effects of the PSFs on the likelihood of success, then we would expect the SLI to be related to the probability of success that would be observed in the long run in the situation of interest (i.e., the actuarially determined probability). The calculation formula for the SLI is as follows:

$$SLI_j = \Sigma W_i . R_{ij}$$

where \( SLI_j \) = SLI for the \( j \)th task

\( W_i \) = normalised importance weight for the \( i \)th PSF (\( \Sigma W_i = 1 \))

\( R_{ij} \) = scale value (rating) of the \( j \)th task on the \( i \)th PSF.

A major assumption of the SLIM approach is that a SLI generated by this process bears a consistent relationship to the expected long-term probability of success and can be converted to it in a simple manner. It is assumed that a similar log relationship exists as has been proposed for earlier HRA techniques using scaling i.e. log human error probability (HEP) = a SLI + b. (See section 4.3). Support for the use of the log relationship comes from Embrey and Kirwan (op.cit.).

5.1.1 An example of the SLI procedure

The concepts described above can best be illustrated by a simple worked example. This section will also provide a detailed description of the practical application of the basic form of SLIM. Suppose it was desired to evaluate the probability that an operator will correctly diagnose the state of a nuclear power plant, and initiate manual intervention when a failure occurs in an emergency feedwater pump during a transient. The following steps would be carried out:

**Step 1 : Modelling and Specification of PSF**

During this first step, the judges thoroughly discuss the task to be evaluated, with particular attention being paid to identifying the various ways in which errors of omission and commission could
occur (error modes) and the PSFs which could impact on these error modes. The various forms of task analysis which are available may be employed here, together with the documentation of emergency operating procedures, photographs of the control room etc. Operator input is very important at this stage. The modelling should be as exhaustive as possible and the results documented to indicate the error modes that the judges have in mind when making their assessment. This is necessary so that the procedure can be subsequently audited if required. At the end of the modelling phase, all credible error modes will have been considered and the PSFs which have a significant effect on these errors will have been identified.

In our example, we will assume that the judges have decided that the following PSFs are the major factors influencing success in the task being evaluated:

- Quality of the information available to the operator from the control panel.
- Quality of the procedures.
- Time available to diagnose the situation and carry out the appropriate actions.
- Degree of operator training.

The documentation for this phase should include some description of exactly what is meant by each of these PSFs as used in the modelling session. The process of documenting the sessions will be monitored by the facilitator, the individual who leads the exercise. In some cases a pre-defined set of PSFs may be used if these can be shown to be applicable to the situations being assessed. This considerably reduces the time required to run a session.

Step 2. Weighting the PSFs
The determination of the relative importance of the PSFs can be accomplished by several procedures. In the recommended approach judges are first asked to consider the task being assessed and to visualise a situation where all the PSFs are as bad as they could credibly be in a real plant. They are then asked to decide which single PSF would have the most significant effect on enhancing the probability of success if it were improved. This is assigned a weight of 100. The PSF which would have the next most significant effect on success is then chosen and a weight is assigned to it relative to the most significant PSF. Thus, if the second PSF were judged to be half as important as the first in terms of its effect on success likelihood, it would be given a weight of 50. This process is then repeated for all the PSFs. The results for our example might be as shown in Table 1.

The normalised weights are obtained by dividing each individual weight by the sum of the weights. The normalised weights sum to
one and represent the relative importance of each PSF in terms of how strongly it influences the likelihood of success.

Table 1 PSF Weights

<table>
<thead>
<tr>
<th>PSF</th>
<th>Assigned Weight</th>
<th>Normalised Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Information</td>
<td>100</td>
<td>100/200 = 0.50</td>
</tr>
<tr>
<td>Training</td>
<td>50</td>
<td>50/200 = 0.25</td>
</tr>
<tr>
<td>Time available</td>
<td>30</td>
<td>30/200 = 0.15</td>
</tr>
<tr>
<td>Procedures</td>
<td>20</td>
<td>20/200 = 0.10</td>
</tr>
<tr>
<td></td>
<td>= 200</td>
<td>= 1.00</td>
</tr>
</tbody>
</table>

Step 3. Rating the Task

The rating procedure is carried out next with the judge(s) directly assigning a numerical value to each PSF on a scale of 0-100, where zero indicates that the PSF is as poor as is credibly likely, and where 100 indicates that it is as good as is credibly likely in a real plant, in terms of its effect on the likelihood of success. At this point, it is important to differentiate between importance weights and ratings. The ratings are independent of the weights assigned to the PSFs. The weights indicate the relative importance of the PSFs in terms of their overall effect on the success likelihood, and are, therefore, not independent of one another. The ratings essentially represent the experts' opinions regarding the actual situation in the nuclear power plant for the task being assessed. The rating assigned to each PSF is independent of all the others in the set of PSFs being assessed.

In our example, we will assume that the following ratings have been assigned: quality of information, 70; training, 20; time available, 10; procedures, 50. These ratings might arise from the following situation: the operator has a wide variety of information available which is much better than average, but not as good as in the best plants. The operator's training for this particular situation is inadequate and the time available to perform the action is so short that it will negatively impact on the likelihood of success. The procedures are about average for the nuclear industry.

Step 4. Calculation of SLI

The calculation procedure for the SLI is shown in Table 2 below. As can be seen from Table 2 the process of calculating the SLI involves simply forming the products of the normalised weights and the ratings for each PSF and then summing the results. The SLI can range from 0 to 100, where 0 indicates that the task has
a high probability of failing, and 100 where it has a high probability of success.

Table 2 Calculating the SLI

<table>
<thead>
<tr>
<th>PSF</th>
<th>Normalised Weight (From Table 1)</th>
<th>Rating</th>
<th>Product Weight x Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of information</td>
<td>0.50</td>
<td>70</td>
<td>35.0</td>
</tr>
<tr>
<td>Training</td>
<td>0.25</td>
<td>20</td>
<td>5.0</td>
</tr>
<tr>
<td>Time available</td>
<td>0.15</td>
<td>10</td>
<td>1.5</td>
</tr>
<tr>
<td>Procedures</td>
<td>0.10</td>
<td>50</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>= 1.00</td>
<td></td>
<td>SLI = 46.5</td>
</tr>
</tbody>
</table>

Step 5 Conversion of the SLI to Probabilities
Transforming the SLI to a probability estimate can be achieved by several procedures. For the purpose of this example, the approach employed requires the availability of at least two tasks for which the probabilities of success (or failure) are known. In this case, let the tasks be Task A, with a known failure probability of $10^{-3}$ (0.001) and Task B, with a failure probability $10^{-2}$ (0.01). Assume that the judges assigned SLI values of 80 to Task A and 20 to Task B, using the same procedure as has been outlined for the original task.

These values are substituted into the calibration equation given earlier, i.e.:

$$\log \text{HEP} = a \text{SLI} + b$$

this produces two simultaneous equations which can be solved for a and b to give:

$$\log \text{HEP} = -0.16667\text{SLI} - 1.6667$$

This is a general calibration equation for the group of tasks evaluated by this particular set of judges. We can therefore substitute the SLI value of 46.5 into this equation to obtain the probability of success for the specific task in our example. If we do this, we obtain a log HEP of -2.4417. This is equivalent to a success probability of about 0.994, i.e. a failure probability of $3.617 \times 10^{-3}$. In other words, the operator might be expected to fail to correctly diagnose the situation and perform the appropriate actions on about three to four occasions out of every thousand times that this action is required.
In this example, the SLIs were converted to probabilities by using two tasks (A and B) for which the probabilities were assumed to be known. However, reference tasks with known probabilities may not always be available. In such instances, it may be necessary for judges to make absolute probability estimates for two tasks which would then serve as reference tasks for converting the SLIs to probabilities. The Direct Numerical Estimation procedures discussed in section 1 can be used for this purpose, or the Influence Diagram (section 5.2) can be used. A number of other calibration approaches are discussed in Embrey et al. (1984b).

Step 6. Calculation of Uncertainty Bounds
There are several approaches to generating uncertainty bounds around SLIM produced HEPS. In the first of these procedures, judges are asked to make a direct estimate of the upper and lower bounds for each HEP estimate produced by SLIM, using the logarithmic probability/odds scale already shown in figure 1, together with a question of the form:

"For this event, what are the upper and lower bounds of the HEP that make you 95% certain that the true HEP falls between these bounds?"

When SLIM is conducted as a consensus process, the uncertainty bounds should be arrived at consensually. When judges independently estimate HEPS with SLIM, calculation of uncertainty bounds is a straightforward application of statistical theory to the problem of estimating probabilities. Confidence limits, or error bounds in this application, are placed around HEP estimates on the basis of the standard deviation computed from the variability in HEP estimates by the individual judges. Specific procedures for accomplishing this are discussed in detail in Seaver and Stillwell (op.cit.).

In many instances upper and lower uncertainty bounds will be available for the reference tasks used to produce the logarithmic calibration equation. If these bounds are available, they can be used to derive separate calibration equations for generating uncertainty bounds for all tasks being assessed in the same way as the point estimate itself is calculated.

5.1.2 SLIM-MAUD: An implementation of SLIM through the use of MAUD.

MAUD (multi-attribute utility decomposition) is a flexible, interactive computer based system which has been used to implement SLIM. MAUD was originally developed by the Decision Analysis Unit of the London School of Economics and Political Science for use in decision analysis problems. It runs on a variety of microcomputers (e.g. the IBM PC, North Star Horizon and Advantage, Epson QX10) which have the CP/M or PC-DOS operating system and at least 64K of memory and two disk drives.
The use of MAUD represents a more sophisticated way of eliciting from judges the rating and weighting information utilised in the basic SLIM approach. Furthermore, the elicitation procedures of MAUD are in closer accord with the theoretical assumptions underlying the SLI methodology than is the case for the basic procedure described in the previous section. The MAUD-based implementation has the additional advantage of being able to deal with the evaluation of up to 10 tasks in the same session. It employs MAUD's built-in checks to monitor any dependencies between PSFs which may be present. The MAUD system is fully interactive and is sufficiently 'user friendly' such that it can be used unsupervised by individuals or groups of judges with minimal training in computer-based techniques. An example of the dialogue used in SLIM-MAUD and a detailed technical description of the technique is given in Embrey et al. (1984b). An important feature of MAUD is the fact that it allows the judges to constantly modify their assessments as their understanding of the situation develops.

In a typical MAUD session, the system first asks the judge(s) to name the various tasks for which HEPs are required. It is assumed that the SLIs for all the tasks being assessed in a particular session can be determined by the same PSFs with the same relative weights. At least two reference tasks for which HEPs are available need to be included in the session for calibration purposes. SLIM-MAUD then interactively elicits the PSFs which are relevant in determining the probability of success. MAUD performs a comprehensive set of consistency checks on the judges' use of these PSFs in assessing the tasks under consideration. This process is repeated with the various combinations of tasks to generate a series of factors which are equivalent to the PSFs that are elicited directly in the basic SLIM technique.

With SLIM-MAUD, judges first rate tasks and then weight them, thus reversing the order used in the other SLIM elicitation technique. Judge(s) are first asked to rate each of the tasks on nine-point scales and define their 'ideal' point on each scale, i.e. the rating scale value which would be optimal in promoting success. MAUD uses this information to re-scale the PSFs so that increasing scale values always indicate increasing likelihood of success. This is necessary, because with some PSFs, e.g. stress, high or low values degrade the probability of success, whereas moderate values increase it.

The next step in SLIM-MAUD develops the PSF weights by comparing pairs of tasks which have different values on two of the PSF scales. SLIM-MAUD asks the judges which of the two tasks would be most likely to succeed, and then iteratively 'degrades' one of the PSF ratings of the task judged most likely to succeed and improves one of those of the task less likely to succeed. This process is repeated until the judges' opinions reverse themselves with respect to which of the two tasks is most likely to succeed. By repeating this process for a range of PSFs, SLIM-MAUD is able to determine the relative weights of the various PSFs for the
task set under consideration, as perceived by the judge(s). From the weights and ratings SLIM-MAUD then calculates the SLIs for each task.

A separate computer program is then used to convert the SLI values into probabilities using the calibration equation derived from the two reference tasks.

5.1.3 Advantages of the SLIM-MAUD approach

In comparison with the basic SLIM approach, the advantages of SLIM-MAUD can be summarised as follows:

a) Effectiveness of the elicitation process.
The SLIM-MAUD procedure uses a particularly effective method for eliciting the PSFs relevant for assessing the set of tasks under consideration. This is an interactive process which allows the PSFs to be modified at any time during the SLIM-MAUD session.

b) User friendliness.
The comprehensive interactive software provided by MAUD reduces the need for a facilitator, and enables the methodology to be easily used in-house after a relatively small amount of training.

c) On-line checking of the assessment procedure.
One of the important assumptions of the underlying SLIM model is that the ratings assigned to each PSF are independent. MAUD checks the degree of shared variance between the ratings generated by the judges, and provides feedback if the ratings on two PSFs appear to be correlated. The assessors are then given the opportunity to delete the correlated PSFs and define a new PSF which expresses the shared meaning of the original PSFs.

d) Use of theoretically optimal scaling procedures.
The technique used within SLIM-MAUD to place the tasks on a likelihood of success scale conforms with the theoretical model more closely than the approach used for the basic SLIM technique.

e) Capability to handle several tasks simultaneously.
Up to ten tasks can be evaluated within a single SLIM-MAUD session, which greatly reduces the time required to carry out assessments.

In view of the advantages of the MAUD implementation of SLIM described above, this is the recommended approach in future applications of the SLIM methodology.
5.1.4 Overall evaluation of the SLIM approach

A major advantage of SLIM compared with the other judgement-based methodologies considered up to now is that it explicitly identifies the factors (PSFs) which are judged to be major determinants of the probability of error in the tasks being assessed.

The weights which are assigned to these factors can be used to provide design recommendations with regard to which changes will have the greatest effect in reducing the likelihood of error. For example, if the analysis suggested that training was twice as important as procedures in terms of its effect on the probability of success, then it would be logical to assign resources to the former rather than the latter area. It is also possible to conduct 'what if' or sensitivity analyses, in which the effects of postulated changes in PSFs on the overall expected likelihood of success can be evaluated. This is done by simply varying the ratings for the PSFs of interest.

A third advantage of the SLIM approach is that it is highly scrutinizable, i.e. the means via which the final result is arrived at is accessible to external auditing and review. In the other techniques reviewed up to now, external review is not possible, since the process via which the judges arrive at their assessments is covert.

Most of the disadvantages of SLIM are confined to the basic form of the technique, where it is not easy to ensure that the PSFs are independent, and the judgements required can be time consuming. As with all scaling techniques, there may be problems associated with obtaining data for the calibration events. However, if necessary, Direct Numerical Estimation techniques can be used to generate these probabilities, or alternatively the Influence Diagram, to be described in the next section, can be employed.

5.2 A Socio-Technical Approach to Assessing Human Reliability: The Influence Diagram

The description of the Influence Diagram methodology in this section is based on a paper by Phillips, Humphreys and Embrey (1983). A key feature of this approach is that it draws on two fields of study: decision theory and group processes. Decision theory provides the form of the model that allows the desired error rates to be determined, while the model inputs are generated through the group interaction of experts who are knowledgeable about the factors influencing the event whose error rate is being assessed. The different perspectives of these experts, if managed effectively by the group, can lead to informed, useful inputs to the model. Thus, the validity of any error rates that are produced by the model depends not only on the technical model itself, but also on the social processes that help to generate model inputs. Thus, there are both social and technical components to this approach. To help keep this in
mind, and to provide an acronym which will identify the approach, the methodology is referred to as the "socio-technical assessment of human reliability", (STAHR). Many of the aspects of group processes which will be considered also apply to the consensus aspects of SLIM.

5.2.1 Introduction

Influence Diagrams were developed in the mid-70's at the Stanford Research Institute (Miller et al. 1976), then applied and further developed at Decisions and Designs, Inc. (Selvidge, 1976) for intelligence analysis. Howard and Matheson (1980) extended the theory and showed that any event tree can be represented as an Influence Diagram, but not all Influence Diagrams can be turned into event trees unless certain allowable logical transformations are performed on the linkages between the influencing events.

Application of the Influence Diagram technology to HRA is quite straightforward, so some simple illustrations will be used to establish the key principles. Figure 4 shows the simplest kind of influence. Here event A is influenced by event B, that is, the probabilities that one would assign to the occurrence or non-occurrence of event A are conditional on whether or not B has occurred. An equivalent event tree representation, where events A and B are assumed to have only two outcomes, A and not A, B and not B, is also shown. In the event tree the probability of B occurring is given by p₁ while the probability of A given B is shown by p₂, and the probability of A given the non-occurrence of B is given by p₃. The point here is that p₂ is not equal to p₃. If they were equal, then the Influence Diagram would show two circles unconnected by any influencing link.

A slightly more complex Influence Diagram is shown in Figure 5. Here, A is influenced by both B and C. The comparable event tree consists of three tiers because the probability assigned to A at the extreme right depends upon the occurrence and non-occurrence of both B and C, previously. These probabilities for A, conditional on previous events, are shown by p₃ through p₆. Note that p₃ appears in two places, indicating that the probability assigned to B is the same whether or not C occurs.

Finally, Figure 6 shows the same influences on A as in Figure 5, but now C influences B. Note that the event tree has the same structure as in Figure 5, but the probability assigned to B conditional on C is no longer the same as the probability of B conditional on not C. Thus, there are six different probabilities in Figure 5, and seven different probabilities in Figure 6. It is easy to see that the Influence Diagram representation is not only compact, but also contains more information than the structure of the event trees without any probability assignments.

In practical situations where an influence diagram has many nodes, it is typical for the actual number of influencing paths
Figure 4: Event A is influenced by event B.

Figure 5: Event A is influenced by events B and C.

Figure 6: Event A is influenced by events B and C, and B is influenced by C.
to be far fewer than the maximum that could occur if every node were linked to every other node. Any assessment procedure based on the Influence Diagram will require only the minimum number of probability assignments. For example, in Figure 5, an influence diagram procedure would require only six probabilities and would recognise that the same probability is assigned to B whether or not C occurs. In an event tree representation of the same problem, dependencies between events are not obvious until probabilities have been associated with each branch, and keeping track of independent events within a large tree can be a tedious housekeeping chore.

In applying Influence Diagram technology, A is taken as the target event, and assessments are made of only the necessary and sufficient conditional probabilities that enable the unconditional probability of the target event outcomes to be calculated. For example, in Figure 4, the probability of A is given by calculating the joint probabilities of all paths on which an A occurs and then summing the joint probabilities, i.e. \( P_1 P_2 + (1-P_1) P_3 \). For more complex Influence Diagrams successive application of the addition and multiplication laws of probability are sufficient to enable the unconditional probability of the targeted event to be calculated.

It is, of course, important to recognise that no probability is ever unconditional. All events shown on an Influence Diagram occur within some context, and it is this context that establishes conditioning events that are not usually shown in the notation on the Influence Diagram. Thus, in applying this technology, it will be important to establish at the start of every assessment procedure what these common conditioning events are.

The Influence Diagram provides the technical means for organising the conditional probability assessments that are required for calculating the unconditional probability of the target event. The Influence Diagram is the technical component of the STAHHR approach. The Influence Diagram itself comes from experts working in groups and this is the 'socio' aspect of the STAHHR approach. The group process involved will be discussed in detail in section (5.2.3).

5.2.2 The STAHHR Influence Diagram

After several revisions, the Influence Diagram as of May 1984 for events that are influenced by operator actions in nuclear power stations is shown in Figure 7. It is not yet known whether this Influence Diagram is generic in the sense that it can handle all events in which operators are expected to take actions. Possibly parts of the diagram are generic and others need to be developed to fit the specific situation. The STAHHR approach is sufficiently flexible that modifications to Figure 7 can be made
to suit the circumstances, or indeed, entirely different Influence Diagrams could be drawn.

The top node in Figure 7 indicates the target event. For example, if an alarm in the control room signals that some malfunction has occurred and the operator attempts to correct the malfunction, one target event might be that the operator correctly diagnoses the event and carries out the appropriate procedures. The Influence Diagram shows three major influences on the target event. One is the quality of information available to the operator, a second is the extent to which the organisation of the nuclear power station contributes to getting the work done effectively, and the third influence is the impact of personal and psychological factors pertaining to the operators themselves. Another way of saying this is that the effective performance of the target event depends on the physical environment, the social environment and personal factors.

Each of these three major factors is itself influenced by other factors. The quality of information is largely a matter of good design of the control room and of the presence of meaningful procedures. The organisation is requisite, i.e., it facilitates getting the required work done effectively, if the operations department has a primary role at the power station, and if the organisation at the power station allows the effective formation of teams. Personal factors will contribute to effective performance of the target event if the level of stress experienced by operators is helpful, if morale and motivation of the operators is good, and if the operators are highly competent.

These seven 'bottom level' influences actually describe the power station, its organisation and its operators. Each of these seven influences is defined in some detail in Table 3 (Appendix). For example, design is a matter of displays, operator involvement and automation of routine functions. The design of a particular power station would be judged good if all the descriptors in the column headed 'good' were characteristic of the station, and would be judged 'poor' if the right column descriptions were true. Of course, most power stations fall between these extremes; they are mixtures of good and poor, or of degrees of 'goodness' and 'poorness'.

Using the Influence Diagram involves applying the following ten steps:

1. Describe all relevant conditioning events.
2. Define the target event.
3. Choose a middle level event and assess the weight of evidence for each of the bottom influences leading into this middle event.
4. Assess the weight of evidence for this middle-level influence conditional on the bottom-level influences.
5. Repeat steps 3 and 4 for the remaining middle and bottom influences.
6. Assess probabilities of the target event conditional on the middle level influences.
Figure 7: The STARR influence diagram at June 1983.
7. Calculate the unconditional probability of the target event and the unconditional weight of evidence of the middle influences.

8. Compare these results to the wholistic judgements of the assessors. Revise the assessments as necessary to reduce the discrepancy between wholistic judgements and model results.

9. Iterate through the above steps as necessary until the assessors have finished refining their judgements.

10. Do sensitivity analyses on any remaining group disagreements. Report either point estimates if disagreements are of no consequence, or ranges if disagreements are substantial.

Step 1
In step 1, participants would describe the general setting in which the target event might occur as well as all conditions leading up to the target event. Assessors are reminded that this description and statement of initial conditions form a context for their subsequent assessments, and that these assessments are conditional on this context.

Step 2
In the second step, the target event is defined in such a way that its occurrence or non-occurrence is capable, at least theoretically, of confirmation without additional information. Thus, 'rain tomorrow' is a poorly-defined event, whereas 'less than 0.1mm of precipitation falls in a rain gauge located at weather station x' is a well-defined event.

Step 3
In carrying out step 3, the assessors might begin by focussing attention on the left-most middle node, 'quality of information', and assess weights of evidence for the bottom influences 'design' and procedures'. This is done with reference to the definitions of these bottom influences. For example, with respect to the design influence, the group of assessors must decide whether on balance the design of the particular power station is more similar to the descriptions under the 'good' column or the 'poor' column. The assessors may find it helpful to imagine a continuous dimension between 'good' and 'poor' and then try to determine where on this dimension this particular power station lies with respect to the event in question. In short, the assessors are judging numbers that reflect the relative weight of evidence as between the poles of the design influence. The weight of evidence would also be judged for the next bottom node, 'meaningfulness of procedures', only here six different factors, from 'realism' to 'format', must be taken into account in making the judgement.

The weights of evidence placed on the poles of each dimension are assigned as numbers that sum to 1. Thus, letting $W_1$ represent the weight of evidence on display being good and $W_2$ representing the weight of evidence on procedures being meaningful, then the
assessments for these two bottom nodes can be represented as follows:

<table>
<thead>
<tr>
<th>DESIGN</th>
<th>Good</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$W_1$</td>
<td>$1-W_1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROCEDURES</th>
<th>Meaningful</th>
<th>Not meaningful</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$W_2$</td>
<td>$1-W_2$</td>
</tr>
</tbody>
</table>

Step 4 requires the assessment of probabilities for the quality of information, a middle level influence, conditional on the lower level influences. The poles of the two bottom-level influences combine to make four different combinations, good design and meaningful procedures, good design and not-meaningful procedures, poor design and meaningful procedures, and poor design and not-meaningful procedures. Each of these four combinations describes a hypothetical power station of the sort under consideration, and these hypothetical stations are kept in mind by the assessors when they determine the weight of evidence for the quality of information. This can be set out as follows:

<table>
<thead>
<tr>
<th>DESIGN &amp; PROCEDURES is</th>
<th>QUALITY OF INFORMATION is</th>
<th>JOINT WEIGHTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIGH</td>
<td>LOW</td>
</tr>
<tr>
<td>Good Meaningful</td>
<td>$W_3$</td>
<td>$1-W_3$</td>
</tr>
<tr>
<td>Good Not meaningful</td>
<td>$W_4$</td>
<td>$1-W_4$</td>
</tr>
<tr>
<td>Poor Meaningful</td>
<td>$W_5$</td>
<td>$1-W_5$</td>
</tr>
<tr>
<td>Poor Not meaningful</td>
<td>$W_6$</td>
<td>$1-W_6$</td>
</tr>
</tbody>
</table>

For example, $W_3$ is the weight of evidence that the quality of information is high, given that design is good and procedures are meaningful. Here 'high' quality of information does not mean an ideally perfect power station, instead it is meant to represent a power station where both design and procedures are of a high, yet practically-realisable standard. Also, 'low' quality of information does not mean some abysmally bad standard but rather a standard that is minimally licensable. The assessments $W_3$
through $W_5$ capture possible interactions between design and procedures. This is a key feature of the influence diagram technology and experience to date suggests that it is an important feature for human reliability assessment. For example, in some power stations good design may compensate to some extent for procedures that are not very meaningful, whereas if design were poor the additional burden of procedures that are not meaningful could be very serious indeed.

It is now possible to illustrate the calculations that are involved in using Influence Diagrams. The weights are assessed in such a way that they are assumed to follow the probability calculus. Thus, the overall weights of evidence that would be assigned to the four hypothetical stations described at step 4 can be obtained by multiplying the two relevant weights of evidence. For example, the weight of evidence assigned to the actual power station under consideration being both good in design and meaningful in procedures is given by the product of $W_1$ and $W_2$. These are shown above as joint weights. Note that the product rule for probabilities is applied. The next stage in the calculation is to multiply these four joint weights by the weights $W_3$ through $W_6$ and then to add these four products to obtain the overall weight of evidence that quality of information is high for the power station under consideration. That is,

$$W(\text{HIGH}) = W_3W_1W_2 + W_4W_1(1-W_2) + W_5(1-W_1)W_2 + W_6(1-W_1)(1-W_2).$$

Note that this calculation makes use of both the product and the addition laws of probability. It is the repeated application of these two laws that allows unconditional weights at higher nodes to be determined. The unconditional weights now determined for the quality of information will serve as weights on the rows of the matrix for the next higher level event, and the types of calculations just illustrated are repeated to obtain the unconditional probabilities for the target event.

**Step 5**

Returning now to the ten-step procedure, step 5 requires that steps 3 and 4 be repeated for the rest of the middle- and bottom-level influences. Thus, weights of evidence would be assessed for the role of operations and for teams, then a matrix of conditional probabilities would be assessed for the organisational influence conditional on the lower level influences. The same procedures would then be followed in making the necessary assessments for the personal factors.

**Step 6**

Step 6 requires, for the first time, assessments of probabilities. However, these probabilities are for the target event conditional on the middle-level influences. In a sense, what is being assessed is conditional error rates, that is, assessors are giving their judgements about what error rates would be under the assumption of particular patterns of influences. Since the quality of information can be either high or low, the organisation can either be requisite or not and
personal factors can be favourable or unfavourable, there are eight possible combinations of these influences. A separate error rate associated with the target event is assessed for each of those eight combinations. This is not a particularly easy job for assessors because they must keep in mind three different influences as well as their possible interaction. Favourable personal factors, for example, may well save the day even if the organisation is not requisite, and may even compensate to some extent for low quality of information. Insofar as the middle-level influences interact, this stage in the assessment process is important because it allows assessors to express the effect on error rates of these interactions.

Step 7
Step 7 is best carried out by a computer which can apply the multiplication and addition laws of probability to determine the unconditional probability of the target event as well as the next-lower influences.

Step 8
In step 8, the unconditional probabilities and weights of evidence for the middle-level influences are given to the group of assessors who then compare these results to their own holistic judgements. Discrepancies are usually discussed in the group and revisions made as necessary to any assessment.

Step 9
Step 9 indicates that iteration through the first 8 steps may occur as individual assessors share their perceptions of the problem with each other, develop new intuitions about the problem and revise their assessments. Eventually, when the sense of unease created by discrepancies between current model results and holistic judgements disappear, and when no new intuitions arise about the problem, model development is at an end and the model can be considered requisite.

Step 10
Since individual experts may still disagree about certain assessments, it is worthwhile as the tenth step to do sensitivity analyses to determine the extent to which these disagreements influence the unconditional probability of the target event. An easy, but not entirely satisfactory, way to do this is to put in first all those assessments that would lead to the lowest probability for the target event and see what its unconditional value is, then put in all assessments that would lead to the largest probability, thus determining a range of possible results. The difficulty with this is that no individual in the group is likely to believe all of the most pessimistic or all of the most optimistic assessments, so the range established by this approach to sensitivity analysis is unduly large. It should not be too difficult, however, to develop easy and effective procedures for establishing realistic ranges for the probability of the target event, that accommodate the actual variation of opinion in the group.
This has been only a very brief description of the stages that appear to be necessary for applying the influence diagram technology. As experience is gained in the STAHR approach, these steps will be modified and elaborated. The steps are certainly not intended as a rigid procedure to be followed without deviation. Instead, they should be thought of as an agenda that will guide the work of the group.

5.2.3 Group processes

So far, little has been said about the group processes that form the second component to the STAHR approach. A key assumption here is that many heads are better than one where probability assessment is concerned. In the area of human reliability assessment in complex systems there is unlikely to be any single individual with an unbiased perspective on the problem. Although each individual may be biased in his view, the other side of the coin is that each person has something worthwhile to contribute to the overall assessment. It is within the context of the group that different perspectives of the problem can most effectively be revealed and shared with others so that the group can take as its main function the generation of assessments that take account of these different perspectives.

To ensure that all perspectives on the problem are fairly represented, it is important that a climate be established within the group such that information is seen as a neutral commodity to be shared by all regardless of status, or investment in the problem. To help create this climate it is important to establish the role of group consultant. This individual needs to be conversant with the technical aspects of Influence Diagrams and with probability assessment, and needs a working knowledge of group processes. The consultant should be seen by the group as an impartial facilitator of the work of the group, as someone who is providing structure to help the group think about the problem but is not providing any specific content. Although the group consultant needs some minimal acquaintance with the principles of nuclear power generation and with the key components in the plant itself, it is probably desirable that he not be a specialist in nuclear power, otherwise he might find it more difficult to maintain a neutral, task-oriented climate in the group. Thus, a major role for the group consultant is not to tell people what to think about the problem but how to think about it.

The other major role for the group consultant is to attend to the group processes and intervene to help the group maintain its task orientation. The group can easily become distracted from its main task because viewpoints in the group will often be divergent. The cognitive maps that a design engineer and a reactor operator have of the same system may be quite different, yet each will at times insist on the validity of his particular viewpoint. The group consultant must help the group to legitimise each of these viewpoints and to explore them in generating useful assessments.
To a certain extent, adversarial processes may even operate in these groups. Operators will openly criticise certain aspects of design, and design engineers may well be contemptuous of procedures that they deem to be unnecessary "if only people would operate the system properly". Trainers may be somewhat sceptical of the optimistic 'can-do' attitude of the operators, while operators may feel that anyone who has not had 'hands-on' experience in the real control room rather than a simulator is out-of-date at best and simply out-of-touch at worst. Unless the group consultant manages the group processes effectively, minor squabbles can easily turn into major confrontations that seriously divert the group from its effective work.

This discussion is not meant to imply that the group should be composed so as to reduce adversarial processes. On the contrary, an underlying assumption of the STAH approach is that diversity of viewpoint is needed if good assessments are to be generated. Differences are to be confronted openly in the group, and taken seriously regardless of the status of the holder of the viewpoint. Thus, diversity of viewpoint is a key criterion in composing the groups. As yet, there is some uncertainty about the roles that should be represented: group consultant, technical moderator to help direct the discussion on technical issues, trainer of nuclear power station operators, reliability and systems analyst, thermo-hydraulics engineer, possibly one or two other engineers with specialised knowledge of the power station, and, of course, reactor operators. Further work is needed to establish exactly who the 'problem owners' are for these human reliability assessments.

5.2.4 Evaluation of the Influence Diagram Approach

One of the major advantages of the Influence Diagram technique is that it allows a comprehensive evaluation of the interrelationships between the influences which impact on human reliability. In this way, it can be regarded as the most effective of all the techniques considered in this review with regard to the consciousness raising function. Experience with the technique in assessment situations has indicated that the participants gain very deep insights into the nature of the scenario being evaluated. It is also able to provide comprehensive outputs for use by designers. It has very flexible capabilities for conducting 'what if' analyses, since the effects on the overall probability of failure of changing the balance of evidence estimates at various levels in the Influence Diagram can be readily evaluated. The structure provided by the Influence Diagram provides the optimal conditions for the absolute probability assessments required.

The primary disadvantages of the approach are the resources required in terms of time and personnel to use the technique. Our experience has been that 1-2 days of training are probably necessary before the assessment group are able to effectively evaluate scenarios. However, this generally includes very
comprehensive discussions of both the hardware and the human reliability aspects of the situation being analysed.

6. Selection of an Appropriate Technique.

Although we have provided a brief analysis of the advantages and disadvantages of most of the techniques considered in this review, we have not attempted to be prescriptive with regard to which technique is 'best' because this will depend on a large number of situational factors. The development of systematic procedures for the selection of appropriate human reliability techniques for different applications is an area which requires further research. Kneppreth et al. (1978) describe possible approaches that could be employed. However, it is possible even at this stage to delineate some of the constraints which will influence the choice of an appropriate technique. Seaver and Stillwell (1983) have identified some of these constraints and we present an extended version of their list below.

6.1 Number of Judges Available.

This constraint is particularly important for techniques such as paired comparisons and ranking/rating which require statistical measures from the mathematical aggregation of individual judgements. Seaver and Stillwell suggest a lower limit of eight judges although fewer judges may be used under some circumstances. However, in this case there may be adverse effects on reliability of the estimates. SLIM and STAHR, which are essentially consensus approaches, do not require any specific number of judges, as long as the judges that are used have appropriate expertise.

6.2 Number of HEPs to be Estimated.

For Paired Comparisons and ranking/rating, the larger the number of HEPs evaluated together, the greater the reliability of the estimates. However, the length of time or the number of experts required will become unmanageably large in the case of Paired Comparisons, if a large number of HEPs are evaluated.

6.3 Time Available.

If time constraints exist, the direct estimation procedures such as Direct Numerical Estimation are likely to take less time. However the quality of the judgements may not be as great as when more sophisticated approaches are used. There has, as yet, been insufficient field experience with most of the techniques to make definitive estimates of the time they require to exercise.
6.4 Type of Experts Available.

The review by Stillwell et al. (op. cit.) suggested that experts who are probabilistically unsophisticated will need to receive extensive training for Direct Numerical Estimation and Indirect Numerical Estimation techniques. This will be less important for the approaches that do not require absolute probability judgements such as Paired Comparisons, ratio methods and SLIM.

6.5 Physical Location of Experts.

Paired Comparisons and ranking/rating do not require the judges to interact. This may be a significant advantage in terms of resources if the judges are widely dispersed. The considerable advantages that accrue from the use of consensus with the other approaches, in terms of the improved quality of the judgements, probably outweighs the increased resources required.

6.6 Specificity of the Events Being Evaluated.

Obviously, the greater the amount of detailed information that is available concerning the situations being evaluated, the more reliable and accurate the judgements are likely to be. It is probable that the non-numerical procedures such as Paired Comparisons are better able to tolerate reduced information than the other approaches. However, approaches such as SLIM, and particularly STAHR, are better able to extract from the judges any relevant information they may possess.

6.7 Homogeneity of the Situations Being Evaluated.

If there are considerable differences in the nature of the events being assessed, this will create particular problems for the Paired Comparison approach, since it is very difficult to make pairwise comparisons for very different situations. SLIM also requires that all events being assessed in a particular session are sensitive to the same set of PSFs, weighted in a similar way. The solution to this problem is the development of a classification scheme or taxonomy, which enables tasks with similar perceived characteristics to be grouped together. Such a classification is being developed as part of an on-going research programme associated with the further development of SLIM. (See Embrey et al., 1984,b).

6.8 Order of Magnitude of the Events.

In PRA, many of the events of interest are likely to have a very low probability of error. These are likely to be less well estimated by the direct approaches, in comparison with the other techniques, because of the known difficulties experienced by judges in estimating very low probabilities.
6.9 Availability of Calibration Data.

The less calibration data that is available, the greater preference there will be for Direct Numerical Estimation and STAHR as opposed to the other approaches, which require one or more calibration tasks.

6.10 Resource Availability.

If resource constraints are severe it is likely that Direct Numerical Estimation will be used. This is largely the current situation, although this is usually due to a lack of knowledge of the alternative techniques available, rather than to resource limitations. As the importance of effective HRA evaluation in PRA becomes increasingly realised, it seems likely that more resources will become available for this area. In any event, the resources required are likely to remain considerably less than those which are applied to the PRA as a whole. If a reasonable budget is available, it is likely that more effective approaches such as Paired Comparisons, SLIM or STAHR will be applied.

6.11 Importance of Consciousness Raising and Design Recommendation Functions.

If the function of the human reliability analysis is not simply to derive numerical HEPS for explicitly pre-defined tasks, but is also intended to throw light on the types of error likely to occur and the design changes which can reduce these errors, it is likely that SLIM or STAHR will be the preferred methodologies. These are structured to allow a much more comprehensive exploration of the situations being evaluated than the other techniques, and are better able to provide design recommendations and cost-benefit analyses. They are also the only techniques which allow the facility of effective 'what-if' and sensitivity analyses.

7. Current Status of Judgement Based Techniques

Judgement based techniques are currently the focus of considerable interest and development efforts. Subsequent to the reviews of Stillwell et al. (1982) and Seaver and Stillwell (1983), supported by the U.S. NRC, the Direct Numerical Estimation and Paired Comparison approaches are presently being applied to the evaluation of HEPS using experts drawn from a commercial U.S. BWR plant, Comer et al., (op.cit.). This is part of a continuing programme monitored by Sandia National Laboratory. Task elements at the level used in THERP, and described in NUREG CR/1278, Swain and Guttmann (1983), are being evaluated, together with 'whole tasks' such as the probability of successfully managing a steam generator tube rupture.
The results of early work (originally carried out in 1980) on SLIM were published in Embrey (1983a). Subsequent research supported by the NRC through Brookhaven National Laboratory has investigated the logarithmic assumption in the calibration equation, together with the development of SLIM-MAUD, Embrey et al. (1984 a,b). This work was carried out by Human Reliability Associates Ltd. in conjunction with the Decision Analysis Unit of the London School of Economics and Political Science (LSE).

The basic SLIM technique was used to evaluate human actions in degraded core situations as part of the U.S. Industry Degraded Core programme (IDCOR) supported by the U.S. Atomic Industrial Forum. A field test of SLIM-MAUD is currently being carried out by Brookhaven National Laboratory, using the same tasks as those employed by the Comer et al. study described earlier. The SLIM technique is also being employed by the U.S. Institute for Nuclear Power Operations (INPO) as part of their Sequence Risk Analysis programme, to quantify human errors with a significant "cognitive" content.

The STAHR/Influence Diagram approach which was also developed jointly by Human Reliability Associates Ltd. and LSE, has been utilised by Oak Ridge National Laboratory to evaluate human actions in Pressurised Thermal Shock scenarios for two U.S. PWRs. It is anticipated that further resources will become available to complete the development of this technique.

8. Conclusions

This chapter has attempted to review the broad spectrum of judgement based techniques that are currently being researched and applied. We have tried to emphasise that these offer a viable alternative to the data bank based methodologies such as THERP.

In fact the dichotomy between THERP and judgement based approaches is only apparent, since in the continuing absence of empirical data from operating plants, most of the data contained in the THERP data bank (Swain and Guttmann 1983) is subjectively derived. The procedures for extrapolating THERP data to particular applications, and the modelling used to take into account task element dependencies are also largely judgementally based. One advantage of the techniques reviewed in this chapter is that they make the judgemental aspects of the process explicit, and therefore more accessible to external review. Some of the approaches also have considerable additional capabilities for consciousness raising, the development of design recommendations to reduce error, sensitivity analyses, and cost benefit assessments. These are generally considered to be an important aspect of PRA in general, and are certainly a necessary component of HRA.
We have not attempted to be prescriptive with regard to which technique is 'best' in any particular application. It seems likely that the HRA analyst will need a repertoire of several techniques together with a methodology for specifying the appropriate approach for a particular application. The development of this specification methodology is seen as an important goal for future research. Also necessary is an extensive programme of field work in which some of the techniques which have been described in this review are deployed in actual PRA assessments so that their capabilities can be assessed. This will provide feedback so that they can be refined to become validated and tested tools for use in PRA in the future.
CHAPTER 4

STATUS OF HRA ACTIVITIES IN MEMBER COUNTRIES

In conjunction with the review of expert judgement based techniques for human reliability assessment (HRA), the current status of the HRA activities in Member countries was surveyed by the Task Force. Written contributions have been received from Finland, France, the Federal Republic of Germany, the United Kingdom and the United States of America. The interests of the participating Member countries identified through the survey are briefly summarised in Chapter 5.
1. Status of HRA activities in Finland

The activities in the field of human reliability assessment (HRA) in Finland are mainly performed at the Electrical Engineering Laboratory of the Technical Research Centre of Finland (VTT). The ongoing activities could be divided into the following main areas:

- the Nordic cooperation
- research and development projects for the Finnish authorities and utility companies.

The Nordic cooperation is taking place in many different areas which have relevance to nuclear safety. The present program of cooperation was initiated 1981 as a continuation of the earlier programme and will be running until the end of 1984. Actions have already been taken to formulate a new four year programme to be started in 1985. The Nordic cooperation is a joint effort between organizations and companies in the four Nordic countries, Denmark, Finland, Norway and Sweden. The following subprojects within the Nordic cooperation are of interest with regard to the HRA field:

- probabilistic risk assessment and licencing (SXK-1)
- human reliability in test and calibration (LIT-1)
- safety oriented organizations and human reliability (LIT-2)
- planning and evaluation of operator training (LIT-4).

The SXK-1 project is associated with the development of methods for PRA studies. It is the intent to include also some account of human errors. Cases from the Swedish Barsebäck plant have been selected to get realistic scenarios to be studied.

The LIT-1 project is investigating human errors made outside the control room especially during test and calibration activities. Systematic search procedures are developed for identifying possible chains of events and actions which could lead to unsafe plant conditions. A number of case stories have been prepared from Finnish and Swedish BWR plants.
The LIT-2 project has been devoted to the study of the organization as a cause of human errors and the demands to be placed on a safety oriented organization. Different cases from the dispatching and the nuclear industry have been investigated. The project will produce a set of organizational guidelines and checklists.

The LIT-4 project is a Nordic continuation of an earlier international cooperation on International Evaluation of Operational Practices (IEOP). As a part of the project a scheme will be initiated to collect human errors during simulator training.

The research and development for the national needs have usually been performed in projects with specific goals. Among the different projects the following could be mentioned
- assessing the alarm presentation for a PWR plant
- a control room design review for a BWR plant
- evaluation of process computer functions viable for inclusion in future PWR process computers
- development of training methods for simulator training
- development of approaches for emergency operating procedures.

In addition to the earlier activities an other recently completed project could be mentioned. The project has been performed as a cooperation between Combustion Engineering (CE), Windsor, USA, the OECD Halden Reactor Project, Halden, Norway, IVO (the Finnish utility company) and VTT. The project was validating the SPDS concept of CE using the Loviisa full scope simulator in a series of experiments carried through during the fall of 1982. During the experiments all twelve crews of the Loviisa station was running through two very complicated loss of coolant transients. The projects have been reported at the EHPG meeting in Loen, Norway the 24-27 May 1983. The data collected during the experiments will be further analyzed during 1983.

Reference:
2. POSITION OF ELECTRICITE DE FRANCE WITH REGARD TO
THE ESTIMATION OF HUMAN RELIABILITY

Everybody at EDF is convinced of the importance of human error as a
cfactor contributing to failures in nuclear power stations. Everybody
therefore makes an effort to take account of them. If this presentation is to
be concerned only with the action taken by EDF to quantify this type of error,
it will be limited to the following remarks:

-- Probabilistic risk assessments (PRA) have been made for certain safety
systems that may require human intervention (e.g. ASG at Paluel). Other
selective measures will be undertaken using traditional risk probability
calculation methods -- FMEA, use of the data in the Rasmussen report. But for
the moment there is no question of undertaking a "complete" PRA, i.e. covering
the whole of a nuclear power station, given the present state of the art and
the lack of reliable data on human behaviour.

It is not impossible that such a measure will be undertaken in the
future, to the extent that PRA may become an additional argument supporting
the export of nuclear power stations.

-- It may also be considered that the system of collecting data on human
factors (3) established at EDF in liaison with INPO may lead to the production
of numerical data on human errors, but this is certainly not the prime object
of this action, which is aimed at giving a better understanding of incident
sequences and the behaviour of operators from the qualitative standpoint.

More generally, the action undertaken at EDF concerning the treatment
of human errors could be described as follows:

Without trying to quantify what is perhaps not quantifiable in view of
the very nature of man, the majority of actions undertaken at EDF have
been and will continue to be aimed at:

Ta. 9719
-- eliminating the causes of breakdowns in the man/machine interface
    -- what might be called the preventive aspect;

-- mitigating the effects of these breakdowns -- the curative aspect.

A) Preventive action

Studies on the behaviour of operators, individuals and teams, will be
developed to give a better understanding of "human mechanics" and to draw any
appropriate conclusions with regard to work organisation and training. The
word "team" is used in a very broad sense and is not limited to control room
staff (9).

-- Feedback from experience, as discussed above, is an essential
    component in this type of action (it may further the understanding of
certain incidents (3) and hence improve our knowledge of all nuclear
power stations, given their standardisation in France) in the same way
as time spent working on simulators -- function simulators or
full-scale -- or computer-assisted training (4). But these last two
types of activity are aimed above all at improving the human side of
the man/machine interface in the same way as the specific safety
courses given, for example, to the radio-protection safety engineers
discussed below.

Obviously, action is also being taken and will continue to be taken to
improve "the machines". The effort here is concentrated on the choice of the
optimum level of automation and the processing and presentation of information
and alarms (synthesising, ranking, etc.).

The various aids provided for the operator are other ways of reducing
the probability of human error (but still without calculating it!). Examples
are the new presentation of the operating procedures (5) and the installation
of safety panels (6). As will be seen below, the safety panel may be used to
rectify an error made by the operators.
8) Curative action

In the same way as the majority of physical systems are duplicated or triplicated in order to reduce the probability of failure and hence the unavailability of a system, the operator team is backed up by the radio-protection safety engineer who is superior to the shift leader and delegates his authority to the latter during normal operation. By means of the safety panel and the information it provides, the safety engineer can monitor the correctness of the shift leader's judgement using the "approach by states" (7). His engineering training, his two years of training centred on reactor safety, his position in the organisation are so many guarantees of a low probability of human error. What is more, operator redundancy is not limited to the safety engineer, but also extends to local and national crisis teams (2, 8). It should be noted that in the final analysis it is the manager of the plant who is responsible for the safety of his installation.

A further example of curative action is one of the principles of power station control room organisation:

Bringing of control consoles into line with the operators' mental image of the installation should enable operators to take faster and more appropriate decisions in the case of an incident (7).

This clearly implies that studies have to be made of the mental image of the operator (9) and of the team, if it turns out that the team as such possesses such a mental image.

In this brief presentation we have not spoken of the particular attention that needs to be given to the problems of maintenance and design, the latter having to be centred on the problems of operation.

In conclusion, we repeat that human error is mainly taken into account in a qualitative fashion, but this does not preclude the development of certain "traditional" PRA procedures for certain systems, and action could eventually develop in the direction of a complete PRA.
Bibliography

1. "La fiabilité humaine et l'interface homme-machine: actions entreprises à EDF" by Messrs. LLORY and CHAUBARON (EDF-1981)

2. "Actions menées par EDF pour minimiser les erreurs humaines" by Mr. MIRA (EDF-1983)

3. "Recueils de données sur les facteurs humains à EDF" by Ms. CRIFONDON-FOUCO and Ms. GHERTMAN (EDF-1983)

4. "Rapport sur l'enseignement assisté par ordinateur au Service de la Production Thermique" by Mr. FELGINES (EDF-1982)

5. "Design of test and emergency procedures to improve operator behaviour in French Nuclear Power Plants" by Ms. CRIFONDON-FOUCO and Mr. GOMOLINSKI (EDF-CEA)

6. "Panneaux de Sûreté dans les centrales PWR 900" by Messrs. FERAUD and GARCIN (R.E.A.M. - EDF) and Mr. MARCILLE (EDF - 1983)

7. "L'approche par états - Une nouvelle conception des procédures post accidentelles" by Mr. DEPONDE SUREAU (EDF) and Mr. OLIOT (FRAMATOME) - 1983

8. "Organisation des plans d'urgence en cas d'accident dans une centrale nucléaire PWR" by Mr. LAVERIE (MRI) and Mr. BERTRON (EDF) 1983.

9. "Quantification des erreurs humaines -- nécessité d'une nouvelle approche" by Ms. CARNINO, Messrs. MORLAT, LLORY, CAHUZAC (EDF)

3. STATUS OF HRA ACTIVITIES FOR NPPS IN GERMANY

Human reliability analysis for nuclear power plants were used first in Germany within the German Risk Study /1/. The methodology of HRA in this study follows quite closely the methods and the data of the HRA in WASH-1400. However, in some cases data were modified, to adapt to specific plant conditions. For a specific task (initiation of the emergency system) the time required was taken from the plant experience with operator training. The corresponding error probab. was calculated on basis of a method described in /2/.

In the German Risk Study, only planned manual interventions were evaluated as in WASH-1400. These are understood to include activities which are carried out either in accordance with written instructions (operating handbook) or are trained during operation (e.g. resetting of \( \Delta p/\Delta t \) signal) and also activities indicated by an unambiguous alarm. Activities were considered unplanned when the necessity for carrying them out can only be recognised after due consideration, even in the presence of such alarms.

It was assumed that no unplanned manual interventions are made. Unplanned manual interventions, which can have both positive and negative effects, were therefore not quantified. Also malevolent behaviour of plant personnel (e.g. sabotage) was not taken into consideration.

As the basis for the quantification of human error relevant performance shaping factors like stress and control room layout as well as personnel redundancy were taken into account.

This approach was also used in reliability analyses for licensing processes and in the Risk Study for the German fast breeding reactor SNR-300.

For the Phase B of the German Risk Study, which has already started, it is planned to extend the analysis to unplanned actions.
In a safety study /3/, performed mainly by the KfA Jülich, concerning the concept of the German high temperature reactor HTR 1250 a similar approach was taken as in /2/. In this study all actions were evaluated by using the method described in /2/, for which the time needed by the operators was known. For all other actions the WASH-1400 data was applied.

Summarizing the mentioned efforts, it can be seen that currently HRA's in Germany generally follow the approach taken in WASH-1400 and the recommendations in the Handbook of Human Reliability.

With respect to human reliability data some efforts have been made to gain information from field experience by the analysis of transients. This analysis has shown the relatively high impact of human error on the frequency and on the further development of a transient, however, it was not possible to describe human error data.

References

/1/ Gesellschaft für Reaktorsicherheit:
    Deutsche Risikostudie Kernkraftwerke - Fachband 2
    Zuverlässigkeitsanalyse Hrsg.:
    Der Bundesminister für Forschung und Technologie,
    Verlag TÜV Rheinland, Köln, 1980

/2/ General Atomic:
    HTGR Accident Initiation and Progress Analysis,
    Status Report
    GA-A13617, Oktober 1975

/3/ Kernforschungsanlage Jülich und Gesellschaft für Reaktorsicherheit:
    Sicherheitsstudie für HTR-Konzepte unter deutschen Standortbedingungen, Hauptband zur Phase 1B
    Jül-Spez-136, Bd. 1, Dez. 81
4. List of Current Research Activities in the Human Factors Field

<table>
<thead>
<tr>
<th>Institution</th>
<th>Activity*</th>
<th>Classification*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aston University</td>
<td>Analysis of control room and operator related safety procedures in AGRs (R)</td>
<td>A,B</td>
</tr>
<tr>
<td></td>
<td>Mental models in the control of fossil fuel power stations (R)</td>
<td>A,C</td>
</tr>
<tr>
<td>Mr J Hallam</td>
<td>Team functioning in high information load processing tasks (R)</td>
<td>C,D</td>
</tr>
<tr>
<td>Dr R B Stammers</td>
<td>Simulator requirements for nuclear power plant training (R)</td>
<td>A,B</td>
</tr>
<tr>
<td>Mr M Carey</td>
<td>Designs of man-computer dialogues (R)</td>
<td>A,B,C</td>
</tr>
<tr>
<td>Occupational Safety</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr A I Glendon</td>
<td>Perception of risk (R)</td>
<td>C,D</td>
</tr>
<tr>
<td>Dr A Hale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr D Lihou</td>
<td>Human reliability in maintenance and operation of chemical plant (I)</td>
<td>A,B</td>
</tr>
<tr>
<td>Dr L K Ainsworth</td>
<td>Methods of performance measurement (R)</td>
<td>A,D</td>
</tr>
<tr>
<td>Dr L Bellamy</td>
<td>Social and Affective Influences on Human Reliability (R)</td>
<td>C,B</td>
</tr>
<tr>
<td>Loughborough University of Technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dept of Human Sciences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof B Shackel</td>
<td>Dept of visual display (R) unit formats</td>
<td>A,B</td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof F Lees</td>
<td>Process control alarm systems computer assisted diagnosis (R)</td>
<td>A,B</td>
</tr>
<tr>
<td>Dr P Andow</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* See page 66 as regards the classification scheme used.
<table>
<thead>
<tr>
<th>Institution</th>
<th>Activity</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof J Reason</td>
<td>Human error (R)</td>
<td>C,D</td>
</tr>
<tr>
<td>Psychology Dept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manchester University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manchester M13 9PL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr I Tanner</td>
<td>Management systems (I)</td>
<td>B</td>
</tr>
<tr>
<td>Business School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Booth Street West</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manchester 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr Lianne Bainbridge</td>
<td>Process control skills (R)</td>
<td>A,B,C</td>
</tr>
<tr>
<td>University College</td>
<td></td>
<td></td>
</tr>
<tr>
<td>London University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 Bedford Way</td>
<td></td>
<td></td>
</tr>
<tr>
<td>London WC1H OAP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof P Colquhoun</td>
<td>Vigilance and monitoring;</td>
<td>C,D</td>
</tr>
<tr>
<td>Dr A Craig</td>
<td>circadian rhythms (R)</td>
<td></td>
</tr>
<tr>
<td>Dr S Folkard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRC Perceptual and Cognitive Performance Unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory of Experimental Psychology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Sussex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brighton BN1 9QG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr L D Philips</td>
<td>Decision Analysis (R) and subjective judgement</td>
<td>E,C</td>
</tr>
<tr>
<td>Brunel University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Analysis Unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kingston Law</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uxbridge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middlesex UB8 3PH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr R J Strutt</td>
<td>Human reliability in maintenance (I)</td>
<td>A,B,C</td>
</tr>
<tr>
<td>Mechanical Engineering Dept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cranfield Institute of Technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cranfield</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedford</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof K Duncan</td>
<td>Task analysis for in-chemical plant operation (R)</td>
<td>C,D</td>
</tr>
<tr>
<td>Dept Applied Psychology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Wales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institute of Science and Technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Llwyn-y-Grant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penylan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiff CF3 7UX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Activity</td>
<td>Classification</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
</tbody>
</table>
| Dr I G Umbers  
Control Division  
Warren Spring Laboratory  
PO Box 20  
Gunnels Wood Road  
Stevenage  
Herts SG1 2BX | Design of computer driven displays. Process control skills (R) | A,B,C |
| Mr R H Pope  
Central Electricity Generating Board  
General Development & Construction Division  
Barnett Way  
Barnwood  
Gloucestershire GL4 7RS | Process control skills (R)  
Nuclear power station interface design (R)  
Training, human reliability (I) | A,B,C |
| Mr J Hamilton  
South of Scotland Electricity Board  
Cathcart House  
Inverlair Avenue  
Glasgow S4 | As above (I) | |
| Mr J C Williams  
United Kingdom Atomic Energy Authority  
SRD Wigshaw Lane  
Culcheth  
Warrington WA3 4NE | Human reliability assessment (R)  
Human factors aspects of design maintenance and testing (I)  
Effects of management systems on system reliability (I)  
Data collection use of subjective techniques (R) | A,B,C,D |
| Mr D M Humna  
Nuclear Installations Inspectorate  
Silk House Court  
Tythebarn Street  
Liverpool | Human reliability in safety-related activities (R) | A,D,D |
| Mr G Hensley  
British Nuclear Fuels Ltd  
Risley  
Warrington WA3 6AS | Human reliability assessment (I)  
Data collection (R) | D |
<table>
<thead>
<tr>
<th>Institution</th>
<th>Activity</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr E Johnson</td>
<td>Human reliability assessment (I)</td>
<td>D</td>
</tr>
<tr>
<td>Imperial Chemical Industries, Mond Division P O Box 14 The Heath Runcorn, Cheshire WA7 4QG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr B W Eddershaw</td>
<td>Human reliability assessment (I)</td>
<td>D</td>
</tr>
<tr>
<td>Imperial Chemical Industries Petrochemicals and Plastics Division P O Box 90, Wilton Middlesbrough Cleveland TS6 8JE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr A A Clarke</td>
<td>Interface design (R) &amp; (I)</td>
<td>A, B</td>
</tr>
<tr>
<td>Plessey Radar Ltd Oakcroft Road Chessington Surrey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr J R Hughes</td>
<td>Interface Design (R)</td>
<td>A, B</td>
</tr>
<tr>
<td>IBM United Kingdom Laboratories Mail Point 157 Hursley Park Winchester Hampshire SO21 2JN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr K Webley</td>
<td>Human reliability assessment (R)</td>
<td>C; D</td>
</tr>
<tr>
<td>Rolls Royce &amp; Associates Ltd P O Box No 31 Derby DE2 8BJ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA3 Branch Ministry of Defence RARDE Fort Halstead Sevenoaks Kent</td>
<td>Man-computer dialogues (R)</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Man/machine system modelling (R)</td>
<td>A, C</td>
</tr>
<tr>
<td>R Astley Ergonomics Section Central Industrial Engineering Pilkington Bros Ltd Canal Street St Helens, Merseyside</td>
<td>Process control (I)</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Process interfaces (R)</td>
<td>A</td>
</tr>
<tr>
<td>Institution</td>
<td>Activity</td>
<td>Classification</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>M Brunt Redifon Simulation Ltd Flight Simulation Div Gatwick Road Crawley West Sussex</td>
<td>Simulator design and research (R) Simulator fidelity (R)</td>
<td>A B</td>
</tr>
<tr>
<td>Dr M Tainsh Applied Psychology Unit Admiralty Marine Tech Establishment Queen's Road Teddington Middlesex</td>
<td>Target acquisition (R) equipment design Training research (R)</td>
<td>A D</td>
</tr>
<tr>
<td>Mr G Brander Admiralty Surface Weapons Establishment XCO8 Box 3 MOD AWRE Portsdown Cosham Portsmouth Hants PO6 4AA</td>
<td>Man/computer modelling (R) Design Standards (I)</td>
<td>A,C B,D</td>
</tr>
<tr>
<td>Dr G Gillies Army Personnel Research Establishment c/o Royal Aircraft Establishment Farnborough, Hants</td>
<td>Man/computer dialogues (R) Human performance (R)</td>
<td>A B</td>
</tr>
<tr>
<td>Miss J A Hopkinson Easams Ltd Lyon Way Frimley Road Frimley, Camberley Surrey</td>
<td>Human performance (I) modelling in large-scale systems Design of man/machine (I) systems for hazardous situations, and hazard assessment</td>
<td>C A</td>
</tr>
<tr>
<td>R Blyth 1D/HF Dept Rank Xerox Eng Group Bessemer Road Welwyn Garden City Hertfordshire</td>
<td>Human performance testing (R) Design of microprocessor-driven displays (R)</td>
<td>B A</td>
</tr>
<tr>
<td>Institution</td>
<td>Activity</td>
<td>Classification</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>P Michael Human Factors Dept British Aerospace (GW) Ltd. Filton House Filton, Bristol</td>
<td>Perceptual modelling (R) Man/computer interaction (I)</td>
<td>C B</td>
</tr>
<tr>
<td>Mr D O'Brien Human Factors Group Police Scientific Dev. Branch, Home Office Horseferry House Dean Ryle Street London SW1</td>
<td>Man/computer interaction</td>
<td>A C</td>
</tr>
<tr>
<td>Mr M Ballantine Birkbeck College Dept of Occupational Psychology, Market Street London WC1E 7HX</td>
<td>Man/computer interaction(R) Stereotypes (R)</td>
<td>A C D</td>
</tr>
<tr>
<td>Mr J N Clare Scicon Consultancy International Ltd Sanderson House 49 Beners Street London W1P 4AQ</td>
<td>Man/computer interface</td>
<td>A C</td>
</tr>
<tr>
<td>Dr J Berisham Polytechnic of South Bank Borough Road London SE1 0AA</td>
<td>Man/machine systems (R) Selection and Training (R)</td>
<td>A C D</td>
</tr>
<tr>
<td>G. Peters Open University Systems Group Walton Hall Hilton Keynes</td>
<td>Systems Failures (R)</td>
<td>A D</td>
</tr>
<tr>
<td>Dr I D Brown Dr P Wright Medical Research Council Applied Psychology Unit</td>
<td>Human Error, Mental Load (R) Presentation of information (R)</td>
<td>C;B;D C;D</td>
</tr>
</tbody>
</table>
Classification Scheme Indicated on Pages 60 to 65

The four-way classification scheme (A,B,C and D) in the column "Classification" represents the following:

<table>
<thead>
<tr>
<th>Models</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface</td>
<td>A</td>
</tr>
<tr>
<td>Man</td>
<td>C</td>
</tr>
</tbody>
</table>

Classification Scheme

The 'interface' dimension refers to the boundary across which information flows from the hardware of the system to the man. The 'man' dimension indicates work which is primarily concerned with individual factors, i.e. characteristics of the individual which are likely to affect performance. The models section of the classification refers to various schemes which have been produced to analyse, predict or quantify operator performance in man-machine systems. 'Data' are qualitative or quantitative information collected in experimental or field situations, or derived by subjective methods.

Furthermore, the letters "R" and "I" appearing in the column "Activities" should be interpreted as follows:

"R": current research programmes or active research interests

"I": specific interest, even though research is not necessarily in progress.
U.S. Nuclear Regulatory Commission (NUREG-1080)

4.7 Human Reliability

This research involves analysis of nuclear power plant (NPP) operations and maintenance personnel errors and their contributions to man-machine safety system failures. Human error assessment methods and data (rates/probabilities) emerging from this research will support Section II.C of the TMI Action Plan (NUREG-0660) and NRC reliability evaluation programs, including PRA and complex man-machine safety systems design and evaluation.

4.7.1 Major Regulatory Needs and Their Justifications

1. Valid human error data and methodologies and techniques for qualitative and quantitative assessment of NPP operator and maintenance personnel reliability, especially for control room personnel, to determine their contribution to risk, for use in support of PRA (1985-1988).

   Justification: Human error in the operation and maintenance of safety systems and equipment has been identified as a major contributing factor to NPP unreliability and risk. Human reliability research is therefore directed toward the development of valid, reliable human error data (rates/probabilities) and techniques for applying these data to the human reliability analysis segment of PRAs and for developing insights as to what can be done to reduce/eliminate human error.


   Justification: Human performance is a primary basis for assessing the utility of products developed under all human factors research, e.g., staffing, training, procedures, and organization and management. Human reliability research is directed toward the development of human performance criterion measures from available human error data sets to support rigorous evaluations of products emerging from other human factor research.

3. Guidelines developed through analysis and modeling of NPP operations and maintenance functions crucial to safety, for use in identifying additional human reliability research needs and for use as regulatory design requirements for advanced man-machine safety systems (1985-1989).

   Justification: Human error (omission, commission, extraneous acts, sequential, time) is of critical importance in identifying immediate and future human reliability research needs and in developing design requirements for advanced man-machine safety systems. A long-range goal of the human reliability program is the transformation of human error data into specific research needs and design guidelines for advanced man-machine systems so as to improve both plant safety and public safety.

4.7.2 Research Program Description

The strategy for the research is to develop a technical basis for supporting complete and accurate NPP human performance reliability analysis programs (e.g., PRA). Its objective is to develop (1) baseline human error probability
statistics from data obtained from operating plants, nuclear power industry training simulators, performance modeling, and expert judgment; (2) a human reliability data bank for compiling, collating, and storing human error data from all the above media; (3) performance aids, e.g., handbook, workbook, human reliability models, to assist the PRA specialist in conducting human reliability analyses of NPP safety-related events; and (4) technical data for use in conducting near- and long-term human reliability research and in developing technical design criteria for advanced man-machine safety systems for NPPs. This research will be coordinated with INPO, EPRI, utilities, user groups, other Government agencies, and foreign countries.

The major research products will be:


b. Probabilistic risk assessment specialist aids (e.g., handbook, workbook, analytic models) for analyzing a variety of crucial normal, transient, and accident precursor sequences involving human action (1987).


b. Human error probability data from operating plants, industry training simulators, and expert judgment to support a wide variety of operations and maintenance reliability analyses (1985-1987).

c. Human reliability data bank implementation plan (1986).

d. Human reliability data bank combining varied human error data acquisition media and automated storage and retrieval techniques (1986).


b. Regulatory requirements and design criteria for advanced man-machine safety systems (1989).

Chapter 5
Conclusions and Recommendations

1. Introduction

The original orientation of Task Force 2 was towards the application of expert judgement approaches to human reliability assessment, particularly in the context of PRA. However, the considerably broader interests of PHG 1 (Operating Experience and Human Factors) has provided a much wider context within which human reliability assessment has been considered. The survey of HRA activities in Member countries will first be briefly summarised to indicate the scope of these interests.

2. Qualitative Aspects

Several qualitative areas have been discussed in relation to HRA in the participants' contributions. Within PRA there is a perceived need to identify in a systematic manner the types of error that could occur, particularly in abnormal situations, in order to indicate threats to system safety which originate in human failures. There is also interest, particularly from France, in ways in which human errors can be recovered. A number of countries have programmes which involve the collection of data from simulators, including information relevant to understanding the nature of human errors, particularly in accident sequences.

There seems to be a general acceptance that in abnormal situations, the most serious errors are those associated with failures of 'high level' operator functions such as diagnosis, decision-making and strategy formulation. Methods are required to model and assess these so-called 'cognitive errors'. This is an aspect of the general need for more comprehensive predictive models of human functioning in nuclear power stations.

There is a widespread interest in the relationship between human reliability and design. It seems to be generally accepted that human reliability techniques should have the capability to identify the sources of failure and to assist in the development of design recommendations which will lead to the cost-effective reduction of human error.

In addition to the technical causes of human failures, there is also interest, particularly from the Scandinavian countries, in the characteristics of organisations which can influence the occurrence of errors at various levels in a system.

Although the major focus for research on the qualitative aspects of human error continues to be abnormal situations and human actions in the control room, there is also considerable work being undertaken in the area of errors during maintenance and testing, and their potential significance for the reliability of safety systems.

These areas, together with a number of other programmes not specifically oriented towards human reliability, are all providing an increased understanding of the qualitative aspects of the subject.
3. **Quantitative Aspects**

Most participating countries seem to be actively interested in collecting quantitative human reliability data for use in PRA, either from operational experience or from simulators. The commonest method of quantifying human reliability is the THERP technique. Only the U.S.A. appears to be actively supporting research into the development of new quantification techniques.

4. **Expert Judgement Approaches to Human Reliability Assessment**

Chapter 3 provides an extensive review of judgement based approaches to HRA, and concludes that these appear to be a viable alternative to existing decomposition methodologies such as THERP. It is pointed out that there is likely to be a continuing need for judgement based approaches, particularly for the assessment of failures of diagnosis and other high level functions. Expert judgement methods seem to be particularly useful for consciousness raising, i.e. for assisting judges in gaining an in-depth understanding of the situation being assessed, in order to facilitate the identification of likely failure modes. The capability of providing quantitative design recommendations for error reduction is also felt to be an important feature of these techniques.

5. **Recommendations**

With regard to technical recommendations, it is clear from the review in Chapter 3 that there are now a number of HRA approaches available which could be subjected to a field evaluation. The results could be compared with the predictions of established techniques such as THERP. A suitable vehicle for such a study would be to use data obtained from a simulator. The techniques should be evaluated from the standpoint of the extent to which they accurately predict both the nature and the probability of the human errors which occur. A proposal for a small scale pilot study was made to the present Task Force but received little support.

The considerable interest in the modelling and design aspects of HRA indicates that these areas should receive greater attention in any future Task Force concerned with HRA. It is therefore recommended that diagnosis and other cognitive aspects of the management of abnormal events should be considered by the Task Force, together with the use of HRA techniques to generate design recommendations for error reduction. A further important area is the topic of error recovery. This would provide a specific application for the two previous topics.
BIBLIOGRAPHY


Delbecq, A., Van de Ven, A., and Gustafson, D., (1975), Group techniques for program planning. Scott, Foresman; Glenview, IL. USA.


Miller, A.C., Merkhofer, M.M., and Howard, R.A. (1976), Development of automated aids for decision analysis. Stanford Research Institute, Menlo Park, , California, USA.


Seaver, D.A., Assessment of group preferences and group uncertainty for decision making. SSRI Report 76-4. Los Angeles:
University of Southern California, Social Science Research Institute.


## APPENDIX

### TABLE 3: Definitions of Lowest-Level Influences in Influence Diagram

1) **Design**

#### Good

1) **Displays**
   - Easy to read and understand and accessible;
   - Make sense; easy to relate to controls;
   - Alarms discriminable, relevant, coded;
   - Mimic display;
   - Displays re event are present, clear, unambiguous.

#### Poor
   - Hard to read, difficult to interpret; inaccessible;
   - Confusing; not directly related to controls;
   - Alarms confusing, irrelevant, not coded;
   - Non-representational display;
   - Displays re event are not present, unclear or ambiguous.

11) **Operator involvement**
   - Operators have say in modifications.
   - Prompt confirmation of action.

   **Poor**
   - Little or no say.
   - No confirming information.

111) **Automation of routine functions**
   - Highly automated - operators act as systems managers.

   **Poor**
   - Low level of automation - operators perform many routine functions.

2) **Meaningfulness of procedures**

   **Meaningful**
   - Realistic; especially the way things are done.

   **Not meaningful**
   - Unrealistic; not the way things are done.

11) **Location aids**
   - Location aids provided.

   **Poor**
   - Few or no location aids.

111) **Scrutibility**
   - Procedures keep operators in touch with plant.

   **Poor**
   - Procedures do not.
1v) **Operator involvement**

Operators involved in developing procedures. Not involved.

---

**v) Diagnostic**

Allow unambiguous determination of event in progress. Allow inappropriate diagnosis.

---

**vi) Format**

Clear, consistent easily read format. Confused, difficult to read.

---

### 3. Role of operations

<table>
<thead>
<tr>
<th><strong>Primary</strong></th>
<th><strong>Not primary</strong></th>
</tr>
</thead>
</table>

1) **Accountability**

All other functions report to operations supervisor. Only operations staff report to operations supervisor.

---

1i) **Relationship to maintenance and other functions**

Good relations. Antagonism.

---

1ii) **Paperwork**

About right Excessive

---

1v) **Operator involvement**

Operators have a say in how the place is run. No say.

---

### 4. Teams

<table>
<thead>
<tr>
<th><strong>Present</strong></th>
<th><strong>Absent</strong></th>
</tr>
</thead>
</table>

1) **Shifts**

Allow teams to stay together. Prohibit team formation.

---

1i) **Roles**

Well defined accountabilities. Poorly defined accountabilities.
### 5. Stress

<table>
<thead>
<tr>
<th><strong>Helpful</strong></th>
<th><strong>Level not helpful</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1) Shifts</strong></td>
<td></td>
</tr>
<tr>
<td>No jet lag.</td>
<td>&quot;Permanent jet lag&quot;.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Time available</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequate.</td>
<td>Too little.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Operating objectives</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No conflict.</td>
<td>Conflict.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Transient related stress</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Little or none.</td>
<td>Overstressed.</td>
</tr>
</tbody>
</table>

### 6. Morale/motivation

<table>
<thead>
<tr>
<th><strong>Good</strong></th>
<th><strong>Poor</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1) Status of operators</strong></td>
<td></td>
</tr>
<tr>
<td>Treated as professionals.</td>
<td>Treated as labourers.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Career structure</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Operators can find best level in organisation.</td>
<td>Peter Principle operates.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Physical/mental well being</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Operators physically and mentally capable of performing job.</td>
<td>Job performance adversely affected by physical and/or mental impairment.</td>
</tr>
</tbody>
</table>
7. **Competence**

<table>
<thead>
<tr>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1) Training</strong></td>
<td></td>
</tr>
<tr>
<td>Operators generally well trained in emergency procedures.</td>
<td>Poorly trained in emergency procedures.</td>
</tr>
</tbody>
</table>

| **11) Certification** |  |
| Peer review is used. | No peer review. |

| **111) Performance feedback** | |
| Operators given periodic feedback on performance. | No feedback. |

| **1v) Experience** | |
| Operators experience in dealing with target event. | Operators inexperienced. |