HUMAN
RELIABILITY
ASSOCIATES

Dr David Embrey

Methodology for Modelling and Quantification of Human Factors in the MSA Formal Safety Analysis

February 1996
R96/03

Contents

Executive Summary

1. Background

Human Factors is one of the most important aspects of the FSA, since human error is reliably estimated to account for 80% of the causes of marine accidents. For this reason, it is essential that human factors issues are systematically and comprehensively dealt with within the FSA framework. The main function of the human factors methodology presented in this document is to achieve the following objectives:

- To raise the consciousness of the hardware orientated risk analyst so that he or she is better able to take into account the critical role of human factors in marine safety.
- To provide a set of tools and techniques to support the modelling and prediction of human failures in marine systems.
- To allow the effects of policy changes on risk to be quantitatively evaluated.
- To provide specific human factors techniques to minimise risk arising from human failures.
- To provide a set of tools for quantifying the probability of human failures.

2. Qualitative modelling tools

Prior to numerically assessing the likelihood of human failures, it is first necessary to identify these failures in a systematic manner. SPEAR (Systematic Procedure for Error Analysis and Reduction) is a modelling tool which starts with the prescribed activities of people during normal operation or accident sequences and identifies human failures which can either act as initiating events for accident sequences, or can lead to failures to halt accident progression leading to loss of life. SPEAR consists of the following stages:

a) Development of a task inventory. This specifies the tasks and operations performed during various phases of systems operation, e.g. normal operations such as unloading cargo and abnormal situations such as fires or collisions.

b) Critical task identification. This is a screening process to identify those tasks or functions where severe consequences will arise if human failures occur.

c) Task analysis. This involves a more detailed breakdown and analysis of the critical tasks identified in b).

d) Performance Influencing Factor (PIF) Analysis. This identifies the factors which determine the likelihood of errors arising.

e) Error Reduction and Cost Benefit Analysis. This stage identifies methods for reducing the likelihood of the errors identified earlier. If costs can be assigned to the error reduction strategies, then cost benefit analyses allow the most effective risk reduction measures to be identified.

3. The Influence Diagram Approach (IDA).

IDA is a method of modelling the network of influences which link human failures at the operational level with their immediate causes and with company and IMO policies. In addition to providing a qualitative tool for understanding the nature of these influences, IDA can also be used as a predictive tool.

4. Quantification techniques

Once potential human errors with severe consequences have been predicted it is necessary to quantify their probability of occurrence. A range of techniques are reviewed in the report.

- Technique for Human Error Rate Prediction (THERP)
- Time dependent modelling of human reliability
- Direct Numerical Estimation
- Paired comparisons
- Human Error assessment and Reduction Technique (HEART)
- Success Likelihood Index Method (SLIM)
- Influence Diagram Approach (IDA)

It is concluded that SLIM and IDA are the most appropriate quantification techniques in the context of the FSA because they are highly flexible and allow the quantification of the effects of changes in policy on the probability of human failures. A comprehensive set of case studies to illustrate the application of these techniques is provided.

The human factors techniques described above can be applied during the various steps of the FSA as follows:

Step 1. Risk Identification. The SPEAR technique can be used to identify hazards and outcomes arising from human errors.

Step 2. Risk Analysis: SPEAR together with IDA and other quantification techniques can be used to assess the risk arising from accident sequences.

Step 3: Risk Control options. The SPEAR technique identifies methods for controlling the probability and hence the risks arising from human error.

Step 4: Cost Benefit Analysis. Since error probabilities can be expressed in SPEAR as a function of factors which can be costed, this provides a means for performing cost benefit analyses of human factors improvements.

Step 5. Decision making. The additional insights into the sources of risk provided by human factors analyses allows better decision to be made with regard to policies.

Methodology for Modelling and Quantification of Human Factors in the MSA Formal Safety Analysis

1. Introduction

This document is intended to provide inputs to the MSA FSA project regarding various aspects of the analysis and the quantification of human factors within the FSA framework.

Previous reports by the Tavistock Institute under Project 385 have provided a useful starting point for the consideration of some of the important dimensions of human factors in marine safety and cover some of the ground of this document. In order to enable these contributions to be factored into the overall conceptual and mathematical framework upon which the FSA philosophy is based, this document provides a review of some of the qualitative and quantitative methods that have been developed and used in the context of other formal safety analysis applications such as power generation, chemical processing and the aerospace and transport industries.

In the first part of the document, some of the needs for human reliability assessment within the current state of the FSA methodology will be discussed. It is emphasised that quantification cannot be performed effectively without the use of systematic qualitative techniques to identify both the negative and positive effects of the human on the initiation, progression and the magnitude of major marine accidents. A unifying modelling approach called the Influence Diagram is proposed which potentially allows the effects of policy influences on the risks of human failures to be assessed. In addition to this approach a complementary framework called SPEAR (System for Predictive Error Analysis and Reduction) is described which can be used deductively to identify human error potential in both the initiation and progression phases of accident sequence modelling. This framework is used in conjunction with quantification techniques to provide a comprehensive evaluation of risk.

The next section provides a brief overview of the main human factors quantification techniques, together with a set of criteria for evaluating their applicability to the MSA FSA project. Based on the specific requirements of the MSA FSA, two techniques, the Success Likelihood Index and the quantification method associated with the Influence Diagram are recommended. Two detailed case studies based on a marine example are provided to illustrate the methods.

Finally, the implications of the methodologies for the various stages of the FSA methodology will be discussed.

2. The Needs for Human Factors Analyses in FSA

It is useful to consider the types of human failures that could occur in the shipping context so that any analytical methods that are considered as candidates for use in the FSA human factors quantification methodology can be evaluated from the point of view of their applicability to these failures.

Human reliability specialists in general define two broad categories of human failures; errors and violations. Errors are essentially unintended in nature and can be broken down into slips and mistakes. Slips arise in well practised skill based tasks where the person has a correct understanding of what needs to be done, but fails in the execution of the necessary actions. The term mistakes, as used in human error analysis, refers to situations where a person or team has an incorrect understanding of what needs to be done (e.g. due to misdiagnosis or incorrect knowledge) and therefore performs a sequence of actions which are not necessarily incorrect in themselves, but which are inappropriate for the situation. Such errors are sometimes referred to as errors of commission, but they will also automatically give rise to an error of omission since some other action is performed instead of what is required.

Mistakes are sometimes referred to as cognitive errors, since they are usually associated with higher level 'cognitive' functions such as diagnosis or decision making. Mistakes are considered to be more serious than slips. Because mistakes are based upon an incorrect understanding of the situation, recovery is much less likely in comparison with slips, where there is usually rapid feedback that an error has occurred. In the case of mistakes, this feedback is likely to be ignored if it does not fit into the prior diagnosis of the situation. Thus people tend to hang on to their initial diagnosis of a problem, and will only change their views when the evidence becomes undeniable. Another problem is that the alternative course of action which is executed as a result of a mistake may in itself constitute the starting point for an unanticipated accident sequence.

The second broad category of human failures arise from violations. Violations can be defined as situations where people knowingly take actions which differ from those that are recommended. Violations can be further broken down into two further categories. Routine violations are those which occur on a regular basis, for example where operating practices in general do not conform to those laid down in procedures. Exceptional violations are unusual and sometimes bizarre actions which are totally different from the usual range of performance variability encountered in the system. Such violations sometimes arise with highly knowledgeable or high status individuals who believe that the rules do not apply to them. The classical example is the Chernobyl nuclear accident, where senior high status engineers visiting a provincial nuclear power plant decided that they were sufficiently knowledgeable about the system that they could disable the safety systems with impunity. Similar examples could be cited in the shipping industry; for example normal precautionary procedures are often waived on Sea Trials for reasons of technical expediency.

The classification of human failures described above is important because slips mistakes and violations are influenced by different types of performance Influencing Factors. For example, mistakes often arise from misdiagnoses which in turn are directly influenced by factors such as the quality of the information displayed on radar screens and other information sources, and the experience and competence of individuals in interpreting this information. Violations, on the other hand are more strongly affected by socio-technical factors such as the balance between risks and profit that exists at the operational level, peer pressure, and the extent to which procedures are generally complied with. The distinctions that can be made in the Performance Influencing Factors by considering these different types of failures become important in predicting the types of error likely to occur in a given situation. This is a critical stage in the modelling of different types of error that is necessary before effective quantification can take place. This will be discussed in more detail in section 3. Another reason for using this classification is that there is usually enough information available in incident reports to classify any human failures which occurred into these categories. The ability to link these categories of failure with data which may be present in existing incident reports is clearly very important, in that it provides a data source which may be useful in the quantification aspects of the human factors methodology.

The distinction between slips, mistakes and violations is essentially a psychological classification. However another classification is available which is also useful in specifying the way in which particular human failures may affect risk during the initiating event, progression or magnitude phases of an incident progression. Latent failures do not directly lead to the outcome but create conditions which combine with other conditions (e.g. mechanical stress, other human failures such as poor maintenance) to give rise to other failures which may impact on initiation, progression or magnitude of an event sequence. Latent failures commonly arise during design, construction or maintenance, as will be illustrated later. Active failures are the more familiar type of failure which leads directly to an initiating event or impacts on the likelihood of progression or magnitude of an accident sequence.

3. A Methodology for the Assessment of Human Factors within the FSA Process.

3.1 Overview

Prior to embarking on the quantitative aspects of the Human Factors methodology, it is necessary to provide a framework for modelling human factors within the overall FSA process. The first aspect of this modelling process is a global framework for quantification which essentially specifies the ways in which policy decisions made at the IMO and international levels cascade down to affect the likelihood of human failures at the operational level. This framework is the key to linking human factors considerations specifically to the risk analysis, risk reduction cost benefit analysis decision making aspects of the MSA FSA methodology.

Although this framework provides a mechanism for linking policy areas with the impact of human factors on risk for all types of scenarios that the FSA may be required to consider, it does not provide any specific guidance for identifying the types of human failures that may actually occur in various aspects of shipping operations. This aspect of the human factors methodology is necessary in order to deduce the nature of the specific human failures which could give rise to an initiating event, and which could affect the likelihood of progression of an accident sequence or the magnitude of the final consequences. It could be argued that at this stage of the methodology development the focus is primarily on the **generic ship** and hence it is not necessary to model specific human failures which could impact on risk. Nevertheless, if the methodology is to be applied to a specific case study and subsequently to evaluating a wide range of shipping operations with different types of craft, a method is required for predicting the types of human failures likely to occur prior to quantification. The second part of this section describes such a methodology.

The final part of this section provides an overview of the main types of quantification methods that are available and provides recommendations for those techniques which can be effectively applied within the overall modelling frameworks already described.

3.2 A Global Quantification Modelling Framework: The Influence Diagram

The 'Hazards' (or 'Outcomes') identified by Stage 1 are analysed by the Stage 2 methodology and decomposed into the most basic level of failure mode (or Base Event) which can be identified from the statistical data. More fundamental investigation of the root causes, (80% or more of which may be due to human factors), must proceed by means of formal analysis, the chosen method being the Influence Diagram Approach (IDA). At this stage, the IDA will be treated as a qualitative modelling technique. However, it can easily be extended to provide quantitative measures of the impact of human factors on risk, as will be discussed in section 4. The IDA is essentially a hierarchical representation of the various levels of influences which affect the likelihood of systems failures. A generic form of the ID for all types of systems failures is provided in figure 1. As can be seen from this figure, the ID can be used to represent the network of influences, both direct and indirect, that impact on both hardware and human failures. In the context of the shipping industry within the compliance framework of the IMO and other regulatory organisations (e.g. Classification Societies), a similar generic Influence Diagram can be constructed, as shown in figure 2.

The first stage of the Influence Diagram process is to analyse each failure mode in terms of the immediate events that could give rise to them. These are subdivided into ***Slips, Mistakes*** and ***Violations***, which have been described in detail in section 2.

Slips, Mistakes and Violations can be grouped into a number of generic modes. Many of these are considered in 385D4, Revision 1 pages 19-26 (which does not use the Slip-Mistake-Violation taxonomy, although parallels will be found). Examples of generic modes might include:-

Violations

- Disdain for rules or procedures as being designed for "lowest common denominator"

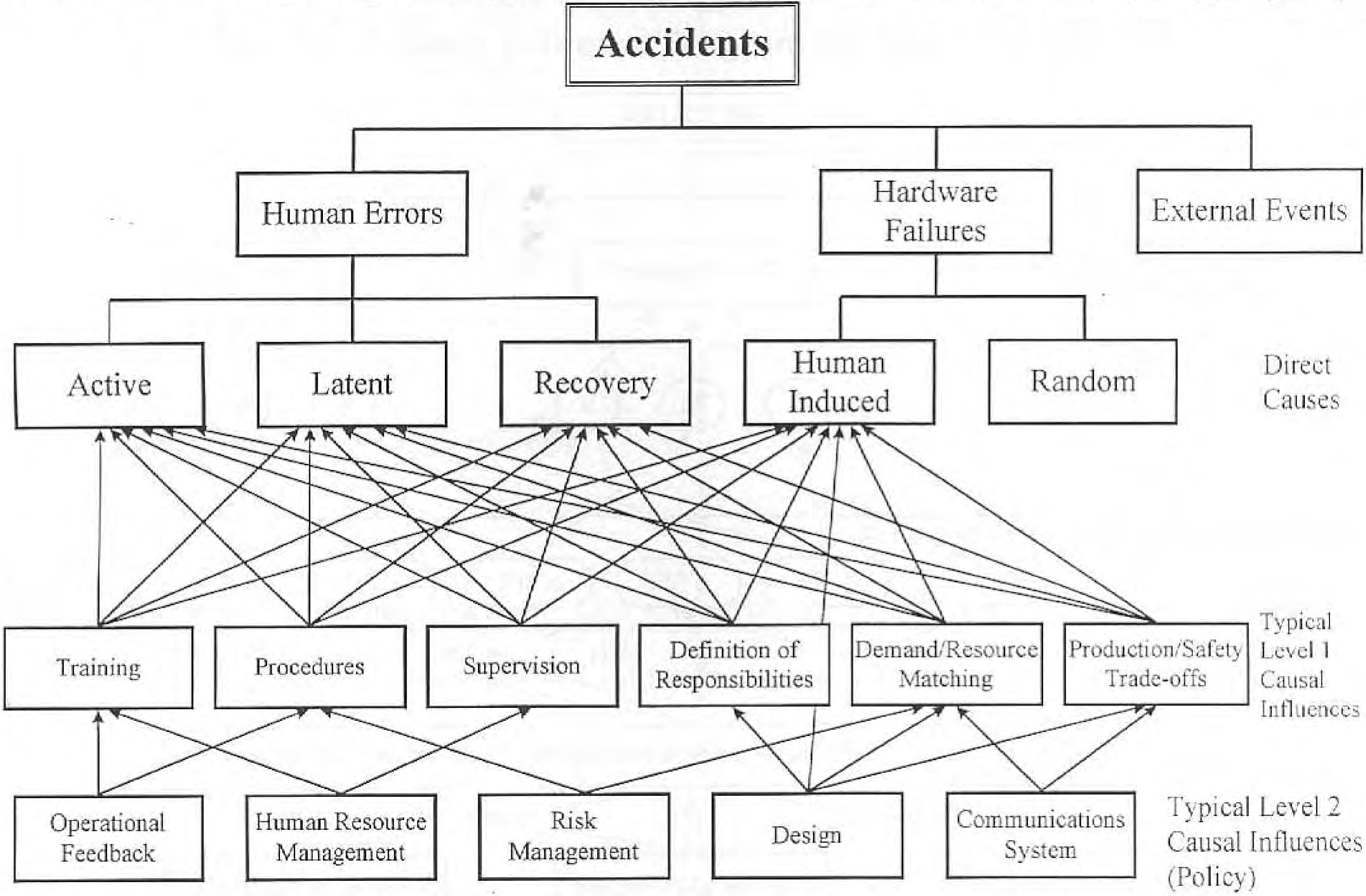
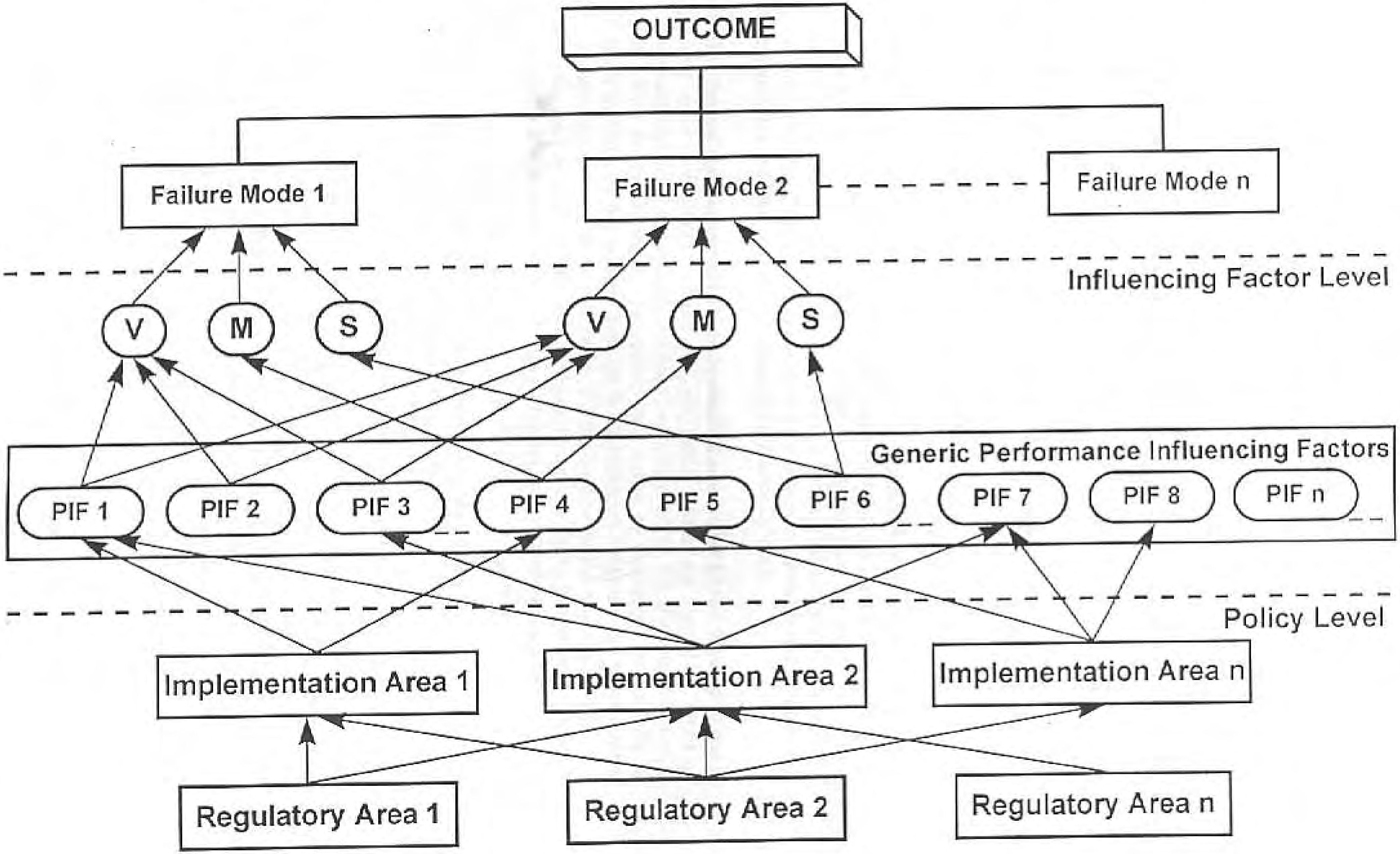


Figure 1 Generic Influence Diagram for Systems Accidents

data\1996\clients\msa\msadiag.ppt

Figure 2 Generic Influence Diagram For Marine Casualties



data\1996\clients\msa\msadiag.ppt

- Deliberately cutting corners to get the job done quicker
- Failing to seek help, for personal image or commercial reasons

Mistakes

- Failure to learn or understand rules or procedures
- Inability to use cognitive skills under pressure
- Incorrect diagnosis of problem from the data presented

Slips

- Misreading an information display (meter or monitor)
- Confusing red-green, Port-Starboard
- Failing to sound a "general alarm" under pressure

Multiple combinations of violations, mistakes and slips may give rise to each failure mode. The next step is to specify the immediate Influencing Factors (referred to elsewhere as Performance Influencing Factors) that determine the probability of each of the generic failure modes. The higher level organisational and policy factors which in turn affect the likelihood of these immediate Influencing Factors being positive or negative are analysed and traced through the Influence Diagram.

The method analyses the underlying organisational and other factors which give rise to each generic violation, mistake and slip mode. These are viewed in terms of implementation of management strategy and policy within the shipping industry, as low-level Policy Decisions. 'Implementation' relates to regulations (which may be made by IMO, by flag or port states or by Classification Societies) and procedures or business practices which are necessary for the function of the shipping industry as a whole, not just ship owning and management but the whole commercial venture of moving cargo or passengers from A to B. Typical Implementation Areas include:-

Risk Transfer (Insurance)
Technical Superintendency and Maintenance Management
Crew Selection and Personnel Management
Training, Certification and Development
Ship Procurement ("new buildings" and "sale & purchase")

These are activities over which IMO seeks to exert influence and good order by the adoption of Regulations or Codes of Practice, the analysis of the effect of which is the subject matter of the proposed FSA process. These Regulations and Codes then form the highest level of Policy Decision. However, in our generic model they are not restricted to IMO policy but may include for example global changes in Classification Society Rules and Flag State policies. It is through this high level of policy making that IMO and other bodies concerned for the safety of the industry exert influence over the management of the industry, both on board and ashore, at the working level which constitutes the Implementation Areas described above. Separation of the two policy levels is necessary for two reasons:-

1. High level policy as enacted at IMO is capable of implementation in different ways by different member states with different cultures and whose contributions to the shipping industry have differing stakeholder profiles,

2. Individual high level policies may cut across many implementation topics at industry level (e.g. the ISM Code), whereas others are "single issues" (e.g. SOLAS90).

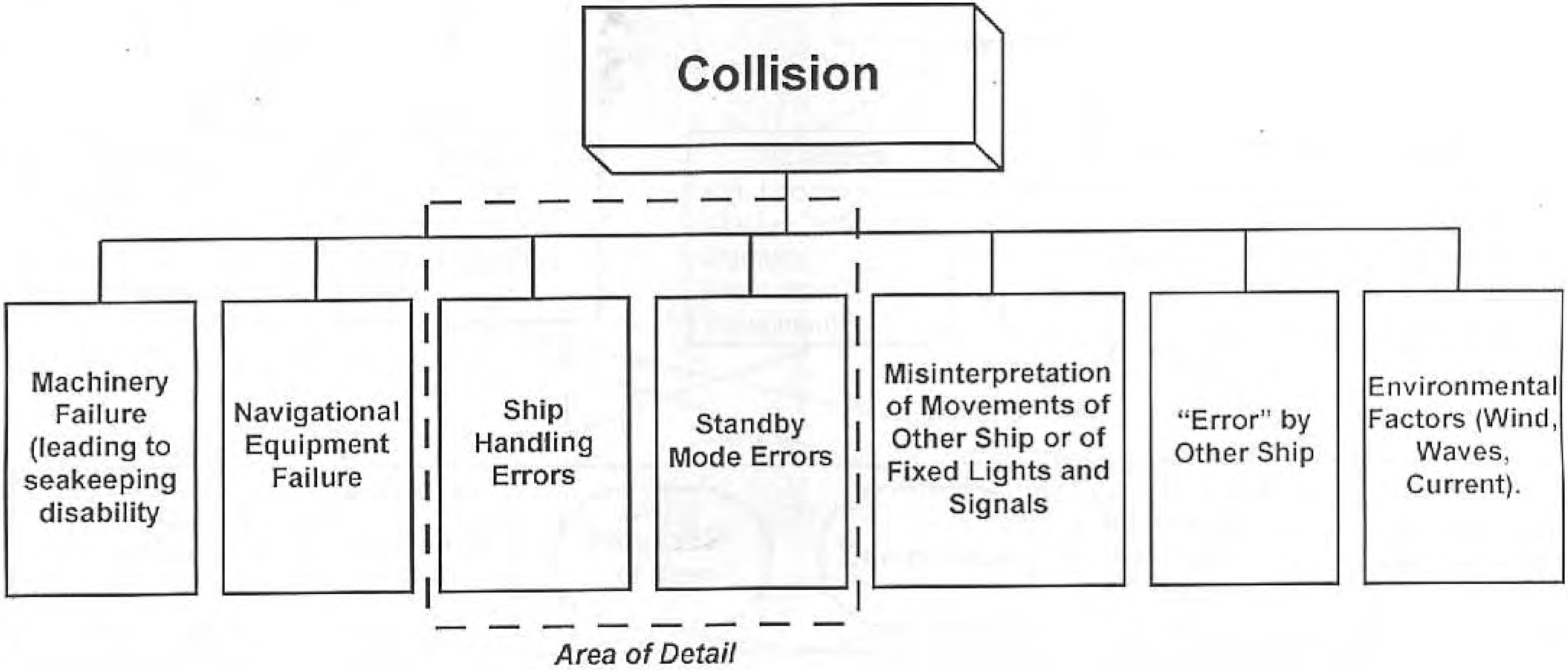
The IDA enables the underlying human factor causes of generic failure modes to be traced back to the areas of IMO policy and regulation which may influence them in a systematic and quantifiable way, thus enabling the effects of regulatory risk control options on human error to be traced through and quantified.

The application of the IDA framework to a specific shipping context will first be illustrated with reference to figures 3 and 4. These represent a collision incident, which is one of the base case scenarios considered in the MSA project, and typical human failure modes that can give rise to this scenario (the failure modes in figure 3 are not meant to be exhaustive). In figure 4, some of these failure modes are considered in more detail. In the case of ship handling errors, two of the categories of failure considered earlier, violations and slips, are modelled as possible direct causes. An example of violations might be excessive speed in a narrow channel or the use of an 'unofficial' course in order to reduce the time to reach port. Slips might arise because the crew were inexperienced in handling the ship or were subject to information overload and distractions due to having to interact with too many people at once. (a consequence of manning levels and unclear assignment of responsibilities).

Standby mode errors are defined as not putting the ship in a state of standby readiness after a potential collision has been detected (a "Close Quarters Situation") This may involve making the ship's master aware of the situation, and placing the propulsion and steering systems in a state of readiness to handle the additional power and manoeuvring requirements that may be involved in averting the collision (e.g. starting a second steering gear pumping and bringing another generator onto the Switchboard). Violations and mistakes are considered as possible failure modes in this situation. As part of the application of the IDA for modelling the causal influences, a generic set of Performance Influencing Factors (PIFs) will be provided for each of the generic failure mode types slips, violations and mistakes. In figure 4, examples of these PIFs for violations are provided. These are a regime of 'profit before safety', the existence or otherwise of a culture where people tend to adhere to procedures, and an accurate perception of risk. It can be seen that the PIF 'procedures culture' also impacts on the likelihood of violations leading to standby mode errors. In the generic case, all of these influences will be modelled and the cross linkages between failure modes defined specifically.

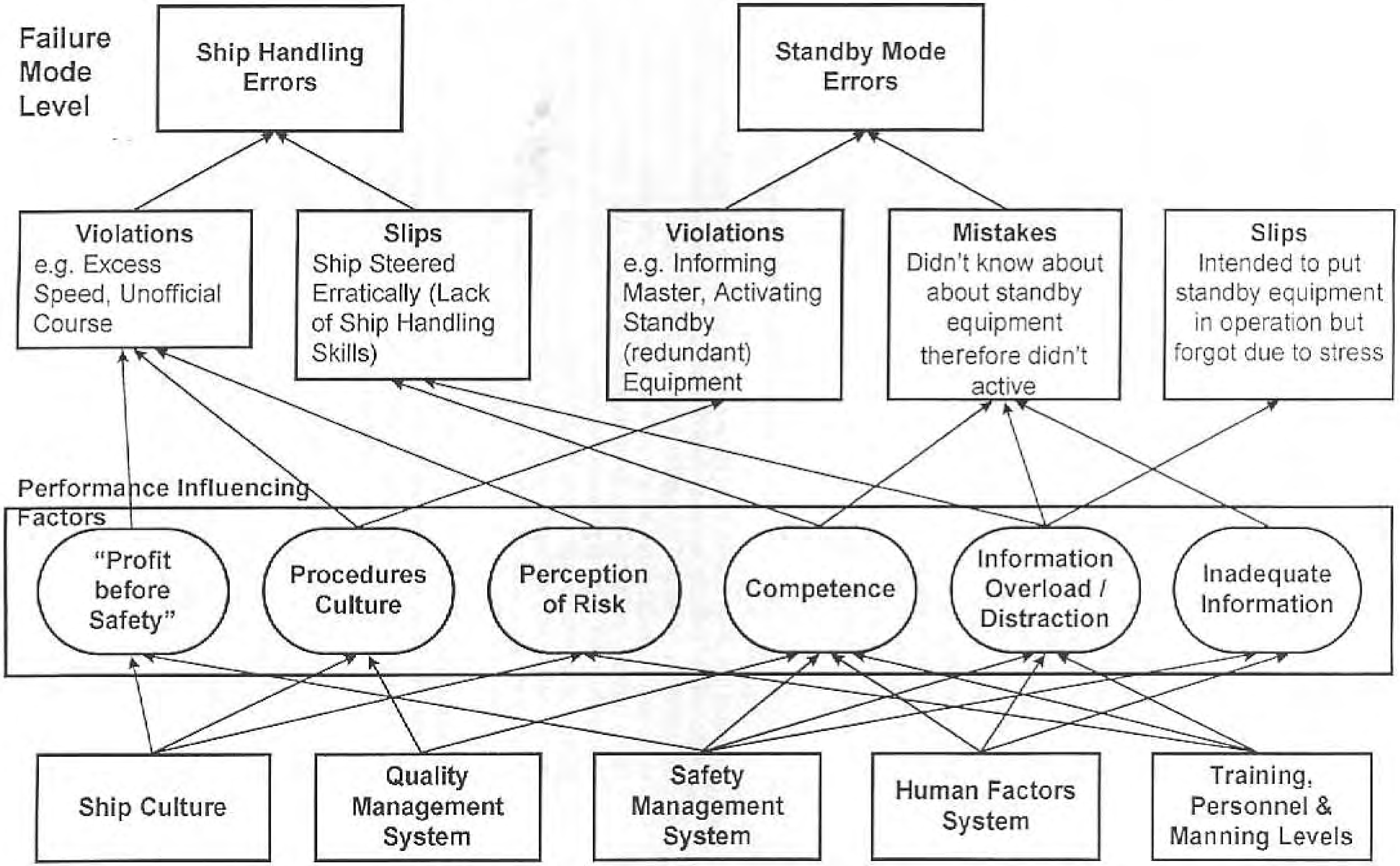
One of the strengths of the IDA is its capability to model and quantify the effects of these cross linkages. In the next stage of the process the implementation areas which impact on the quality of the immediate PIFs are set out, and the nature of the

Figure 3 Ship-to-Ship Collision Scenario



data\1996\clients\msa\msadiag.ppt

figure 4 Ship-to-Ship Collision Influence Diagram - Detail



data\1996\clients\msa\msadiag.ppt

influences defined. In the final stage (not modelled in figure 4), the regulatory influences which impact on these implementation areas are specified and the linkages defined.

3.3 System for Predictive Error Analysis and Reduction (SPEAR)

The IDA methodology provides a means for quantifying the likelihood of specific types of human failures which are known to be significant contributors to risk from the analysis of accidents that have already occurred. However, a complete human factors methodology must provide the means for an analyst to identify potential human failures which have not yet arisen, but have the potential either to create the initiating event for a major accident, or to be the main factor in whether or not the consequences of an accident are successfully averted. A technique is therefore required which provides the means to deduce possible human failures which may arise during any of the normal phases of ship operations (navigation on the high seas, cargo operations, port operations etc.) or as part of the response to potential emergency or pre-emergency conditions. Such a methodology is provided by a process called SPEAR (System for Predictive Error Analysis and Reduction) (Embrey, 1994). The various phases of SPEAR are set out in figure 5. Each of the stages of SPEAR will now be described.

3.3.1 Task Inventory Development and Screening

This is a mechanism for identifying both active and latent failures which will impact on the initiation, progression and magnitude (IP&M) stages as specified by the Stage 2 methodology. In the case of accident sequence initiation, it is necessary to evaluate the opportunities during the normal operation of the ship for human failures to give rise to initiating events. This requires the development of a task inventory, which is a high level description of the functions associated with operating the ship. This might include functions such as operation and maintenance of engineering systems, cargo related activities, navigation, loading and unloading of cargo. These functions are then broken down into the tasks associated with achieving these functions, using an analytical technique such as hierarchical task analysis.

A similar process is used in the case of the human role in the progression and magnitude aspects of a scenario. In this case, the task analysis is used to elaborate on the nature of the human functions that are required to limit the progression of a sequence or mitigate the severity of the outcomes.

In all of these three types of analyses, a criticality analysis is applied to the task inventory to rank the tasks in terms of their risk potential. In this context, risk potential is based primarily on the severity of the consequences if a failure occurs, and the frequency of task performance, which determines the exposure to the risk. The probability of such failures cannot be established until later in the analysis and hence is not usually used during the criticality analysis. The results of this Criticality Analysis will be a subset of tasks which will have a major impact if failures occur, weighted by the exposure to the risk.

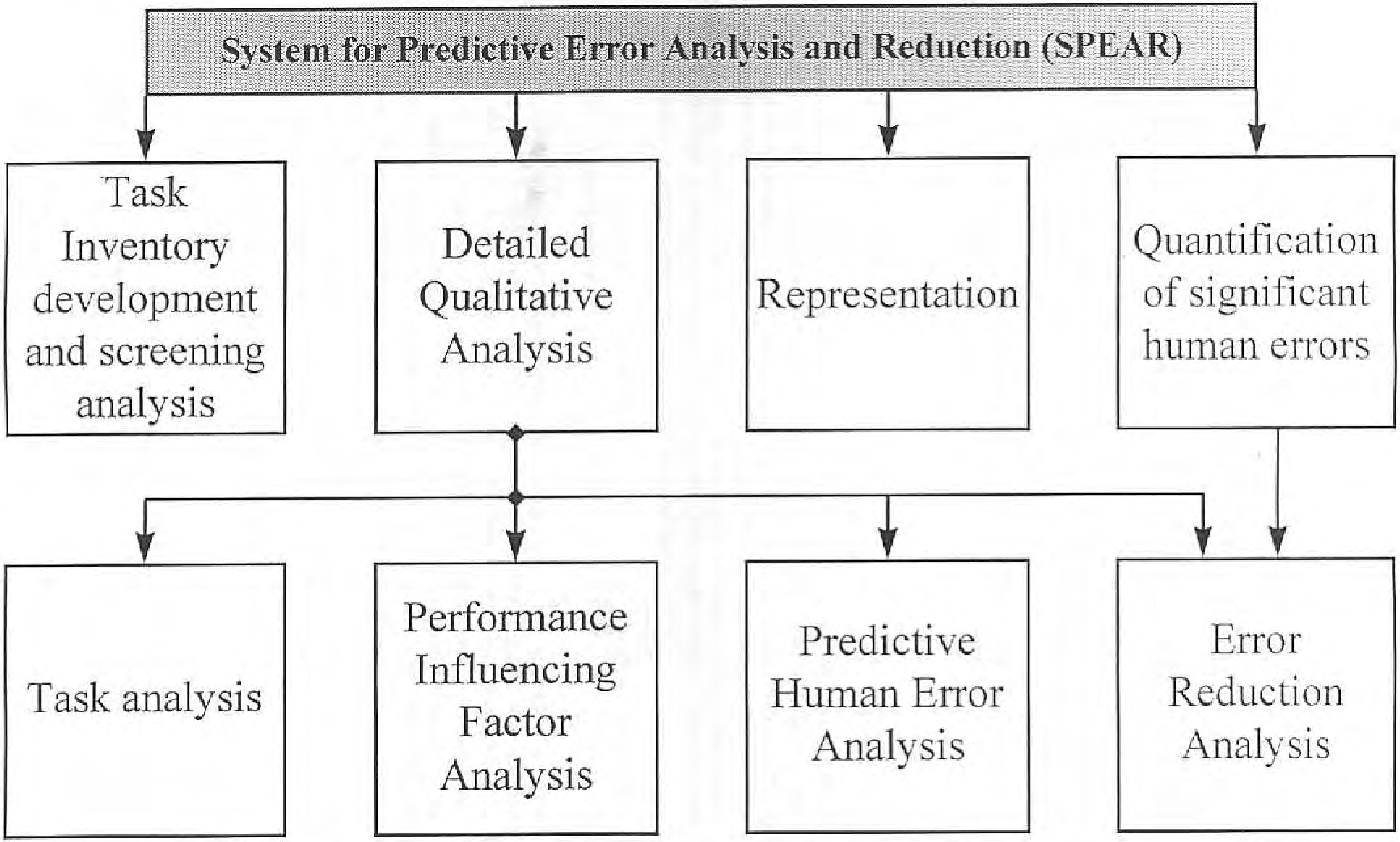


Figure 5 System for Predictive Error Analysis and Reduction (SPEAR)

data\1996\clients\msa\msadiag.ppt

3.3.2 Detailed analysis

Tasks with a sufficiently high risk potential are then subjected to a more detailed analysis to identify the specific failures which could arise which could be initiators for accident sequences or could be involved in the mitigation stage. This is achieved using a technique called Predictive Human Error Analysis (PHEA). PHEA identifies the failures which could arise as a function of the Performance Influencing Factors (PIFs) in the situation. As discussed in the context of IDA, PIFs are the direct influences which determine the probability or frequency of the failure. Identification and evaluation of the PIFs is the key to relating the process to risk reduction measures and also to the impact of higher level policy issues, as discussed in section 3.2. Once potential failures with severe consequences have been identified, the appropriate error reduction (or risk reduction) methods can be specified. However, before this can be done, the impact of the various failure modes developed during the detailed analysis on the consequences of the scenario need to be assessed. This is done during the next stage of SPEAR.

3.3.3 Representation

Representation is the process of representing the effects of combinations of failures in such a way as to evaluate their total impact on the scenario being considered. This involves developing a failure model which specifies exactly how the various types of failures which could occur can actually impact on the final consequences of the accident scenario. The usual forms of representation are event trees or fault trees. Event trees are usually preferred for human reliability work because they allow the temporal sequence of the scenario to be modelled explicitly.

3.3.4 Quantification

The final stage involves quantifying the probability of the failures which are represented in the model. Various techniques for performing the quantification process will be described in the next section. The overall result of the quantification process is to determine the overall likelihood of the possible outcomes of the scenario, (e.g. number of fatalities), taking into account the role of human factors in the initiation, progression and magnitude of these outcomes. Since the quantification process should allow the calculation of the contribution of the various Performance Influencing Factors to the likelihood of these outcomes, this information can be used as an input to both the risk reduction and cost benefit stages of the FSA.

4. Approaches to Human Factors Quantification

4.1 Introduction

This section is intended to provide an overview of the main technical approaches to the quantification of human reliability. The review begins with a brief description of the historical development of human factors quantification. The next section introduces the issues of human reliability modelling and its implications for quantification. A set of criteria are then introduced which can be used to provide the basis for selecting between techniques.

The bulk of the review describes the four main approaches to human reliability quantification with examples of the most important techniques from each approach. This review is not intended to be exhaustive, since there are several comprehensive reviews which together cover most of the large number of techniques that have been developed in this area. Many of these techniques have never been applied in practice, and therefore it was decided to limit the scope of the review to the more important approaches.

4.2 Historical Development of Human Factors Quantification

Techniques for quantifying the human factors aspects of risk analysis, usually referred to as human reliability assessment techniques, were first developed in the context of military and aerospace systems in the late fifties, where human error was found to be a cause of a large proportion of system failures. Early approaches to human reliability analysed the human in the same way as a hardware component, and the primary emphasis was on quantification of human error probabilities for insertion in fault trees as part of a larger system reliability assessment. More recently, approaches derived from cognitive psychology have analysed the human as a goal directed processor of information, and have emphasised the importance of his or her role in higher level functions such as problem solving and decision making. There has also been an increased interest in the organisational and the managerial variables that influence errors and which have been implicated in major disasters such as the Challenger space shuttle, the Herald of Free Enterprise and the Piper Alpha disaster.

4.3 The role of modelling in Human Factors Assessment.

The quantification of human reliability cannot be carried out in isolation from the modelling of human errors. Modelling in this context refers to the comprehensive identification of the nature of the errors that could lead to a particular function not being achieved, or to some alternative action being carried out which has a negative impact on the likelihood of an accident occurring. For example if an Officer of the Watch makes an incorrect diagnosis of a situation by misinterpreting a radar display indicating an oncoming vessel, then he may assume that there is no risk and fail to take necessary corrective actions. Alternatively he may take an incorrect evasive action which, alone or in combination with errors made on the other ship, leads directly to a collision. These possibilities need to be included in the modelling of a situation in order to calculate a correct overall error probability. Since some of the alternative human actions may have severe consequences, they also need to be identified to ensure that all risks to the system have been assessed.

In other industries where human factors quantification has been applied, the issue of modelling has tended to receive little attention because the human factor is primarily evaluated from the point of view of the role of the individual in mitigating the progression or magnitude of an accident sequence. From this perspective, since there are a number of specific actions or functions that the person will be required to perform as part of the sequence, then modelling human failures is simply seen as determining the probability that a required function (usually defined in the emergency operating procedures) is performed correctly. However, as discussed above, it is equally important to include failures to perform required functions because of mistakes or violations, particularly as these may lead to other severe accident scenarios that have not been anticipated by the analyst within the FSA. The SPEAR approach described in section 3 provides a framework for searching for potential errors within a structured framework. However, the question of deciding why these errors might have occurred requires the analyst to be aware of distinctions such as slips, mistakes and violations in order to postulate underlying causes and therefore specify appropriate remedial risk control measures.

4.4 At what level should human factors be quantified?

The issue of the level of detail that human failures should be quantified at often arises. If an analyst is working with a generic description of a ship, it may be convenient to quantify human functions at a similar level. These might include functions such as 'maintain propulsion systems' 'unload petrochemical products at ports' or 'handle potential collision situations correctly'. The question arises whether the analyst should attempt to break down the human functions into smaller elements or whether quantification should be performed at a the function level. The traditional approach to human reliability quantification has been to decompose tasks down to the level of detail at which data is available for the resulting task elements. The advantage of this approach is that, at least in theory, it allows any specific situation to be synthesised from a finite database of component probabilities by means of the familiar logic modelling techniques. This is essentially identical to the approach adopted in hardware risk assessment where the data available is usually at the level of individual hardware components such as pumps or valves, and this is the level at which combinations of failures are modelled using fault trees or event trees.

If an existing database is to be used, the human factors analysis is similarly obliged to break down a task to the level of the data available. Another reason for decomposing a task into its elements is that it allows a more comprehensive evaluation of the ways in which a function can fail. For example for a function such as 'handle potential collision situations correctly' the various ways in which such a function can fail can only really be established qualitatively if the task of handling potential collisions is broken down to its component elements. The level to which this decomposition needs to proceed for the purposes of predicting failure modes which have severe consequences depends on the overall hazard potential of the situation. This is why a criticality analysis is usually performed to establish which of the many ways in which people can interact with a system need to be considered in detail. A detailed breakdown is then only necessary for critical tasks.

The question of what is the appropriate level for quantification therefore depends on the purposes for which quantification is required and the nature of the data available. If quantification is required for high level evaluation of broad human functions, then some of the techniques describe in the following section can be applied without the need to break the function down to a fine level of decomposition. These are techniques which do not depend on the existence of a database. However, even if it is not intended to use a low level detailed database such as that used by THERP (see below), it maybe necessary to break down critical tasks to their constituent subtasks in order to ensure that modelling of potential failures is complete. Another reason for a detailed breakdown in these cases is that the analyst may want to separately quantify the probability of failure of discrete human functions such as detecting signals, making diagnoses or performing actions. This may be necessary in order to specify risk control measures for failures arising from these functions.

4.5 Criteria for using techniques within the FSA process

In order to define the types of quantification techniques that will be applicable to the MSA FSA process, the following criteria need to be applied:

4.5.1 Capability to address the types of human failures found in marine operations

Because the FSA process is intended, amongst other things, to address the sources of risk, the human factors methodology which is embedded within it needs to be able to identify and quantify the likelihood of all types of human failures that are likely to impact upon risk. As described in section 2, the three broad categories that need to be considered are slips, mistakes and violations.

4.5.2 Able to express probabilities of failure as a function of Performance Influencing Factors

As discussed in section 2, Performance Influencing Factors are the characteristics of the person or situation which directly determine the probability of a human failure occurring. For example, a highly trained, well motivated person in a situation where distractions and stress are low, will have a lower probability of failure (e.g. misinterpreting the position of an oncoming vessel) than if these conditions are at the negative end of their range. The type of human failure that may occur in a particular situation is dependent upon the Performance Influencing Factors which are present and the type of human performance required. For example, the factors which affect the likelihood of slips are very different from those which affect violations and mistakes. It is therefore essential that these factors are addressed during both the qualitative and quantitative aspect of incorporating the human element in the FSA process. One of the criteria that needs to be applied, therefore, is the extent to which the technique being considered is able to handle these factors in a systematic manner.

4.5.3 Able to address the effects of organisational and policy decisions on the likelihood of human failures.

Because the FSA process is directed at assisting policy makers in identifying the policy directions which will have the greatest impact on risk, it is therefore essential that any human factors quantification methodology is able to make explicit links between changes in policies and the impact of these changes at an operational level. Ideally, these links should be at least semi-quantitative, so that the effects of policy changes can be translated into their effects on the risks of the various failure categories for the generic ship being considered in the MSA project.

4.5.4 Able to use the data sources likely to be available and applicable to marine operations.

In addition to techniques which address the quantification of the human factors aspects of shipping, relevant data sources which can be utilised by these techniques are also required. The quantification techniques that are chosen therefore need to be able to make use of the types of data likely to be available.

4.6 Overview of Quantification Techniques

This section provides a brief overview of the major groups of quantification techniques and builds on the preliminary review presented by the Tavistock Institute in 385D4. An assessment will then be made of those techniques to be used in the context of the MSA FSA process which meet the criteria defined in section 4.4. Those techniques which satisfy these criteria will then be described in more detail in section 4.6.

Four major approaches to human reliability quantification exist: These are decomposition methods, time based modelling of human error, direct numerical estimation, scaling methods, and Influence Diagram methods.

4.6.1. Decomposition Approaches

These approaches are analogous to hardware techniques. The task or function to be performed is broken down into a series of elements such as subtasks or task steps (usually down to the level of the data which is available), and a failure model (usually in the form of an event tree) is drawn up which defines the various failures or failure paths which could occur and how they might be recovered. A screening process is performed where nominal (worst case probabilities) of failure are inserted into the failure model to evaluate if the overall failure probability is sufficiently high to require more detailed assessment of the task as a whole. If further analysis is justified, appropriate nominal human error probabilities from a data bank are inserted for each of the task elements in the failure model. These data are then usually modified to reflect the specific conditions under which the task is performed. They may be further modified to reflect dependencies between task elements. These probabilities are then combined on the basis of the failure model using the standard mathematical rules to give an overall probability of failure for the task as a whole.

The most extensively developed decomposition technique is THERP (Technique for Human Error Rate Prediction) which was originally applied mainly in the nuclear power industry, although it has been subsequently used in a variety of industries. THERP follows all of the stages of the typical decomposition method described above. Its distinctive feature is the existence of a comprehensive database of human error probabilities. These are mainly concerned with human errors particularly associated with tasks and equipment found in the process and power generation industries. For example, human error probabilities for choosing a valve operating switch from a number of similar switches on a control panel are provided. Although comprehensive lists of Performance Influencing Factors are provided, only a limited number are treated explicitly in terms of their effects on failure probabilities. This is done by providing tables of data which give the probabilities in the database as a function of factors such as the level of proceduralisation, presence of time stress, etc. A model is provided for adjusting baseline probabilities as a function of the levels of stress likely to be present. THERP also includes a method for adjusting probabilities to take into account the degree of dependence between the probabilities assigned to individual steps in the failure model. Higher level factors such as the effects of policies and organisational variables are not considered in THERP.

4.6.2 Time Dependent Modelling of Human Reliability

The main approach for quantifying human reliability using time dependent modelling is the HCR (Human Cognitive Response) technique. This is based upon the idea that there is usually a time window within which a particular response must be made to some initiating event (e.g. a radar alarm of an impending collision) In this technique a relationship between the time available for a person to respond and the probability that the response will be made within that time window is established empirically by means of simulator experiments. Different relationships are derived for skill, rule and knowledge based behaviour. To use the method, tasks are first classified into one of these categories and an appropriate time/non-response probability curve selected. The time available for the response is then evaluated and is then used to read off a non-response probability for the selected curve.

This method was most extensively developed in the nuclear industry where training simulators were available to develop the required curves to define the relationship between time available and non-response probabilities for the three categories of behaviour which were addressed. However, the method has not been widely applied because of the need to develop different time/reliability curves for each installation. There have also been a number of technical objections to the assumption that the probability of not responding is only a function of the time available. Subsequent research has also failed to confirm the major assumption of the method that response data from simulators can be fitted by smooth curves belonging to the skill, rule and knowledge based categories of behaviour.

4.6.3 Direct Numerical Estimation (DNE)

DNE techniques involve the direct estimation of probabilities by expert judges. Various techniques are available for minimising the biases inherent in direct estimation of probabilities. Usually, group elicitation methods are employed to enable measures of inter-judge agreement to be calculated. As part of these methods, precautions are taken to minimise the known biases that exist with regard to the estimation of probabilities. Of particular importance is the need to ensure that all of the judges involved in the probability elicitation process have some direct and relevant knowledge of the domain for which the probabilities are required. Another requirement is that all participants have a correct understanding of the scenario being considered. If a group elicitation session is involved, it is also very important that the group dynamics are managed effectively by a facilitator, in order to avoid the domination of the exercise by strong personalities. It is also very important that the underlying rationale for the probabilities are recorded.

The main advantages of direct expert judgement techniques are that they have been well researched and are generally simple to perform. However, they may require some resources in terms of time and availability of experts.

4.6.4 Scaling techniques

Scaling techniques involve some form of explicit extrapolation process from a set of tasks with known failure probabilities. Some of the techniques are based on the principle of developing a scale of likelihood of success which is not initially based on probabilities. This scale is then transformed to probabilities by using some form of calibration relationship which is developed by including some tasks with known probabilities in the set being assessed. Other techniques provide explicit multipliers to perform the extrapolation process.

Paired Comparisons

Paired comparisons involve judges making pairwise comparisons of the form : "Is task A more likely to fail than task B?" for a set of tasks for which quantification is required. If these comparisons are made for all possible combinations of tasks in the set, a likelihood of success scale can be developed on which each task has a scale value. If at least two tasks with known error probabilities are included in the set being evaluated, and a mathematical relationship between scale values and error probabilities can be assumed, (usually a logarithmic transformation is used) then the scale can be calibrated and a human error probability derived for each task, based upon its position on the likelihood of success scale.

Human error Assessment and Reduction Technique (HEART)

The HEART technique provides a set of nine generic task types with associated baseline probabilities. These task types are broad descriptions of tasks for which data is available from experimental or other sources. By providing a generalised description, the intention is that the user can match the task being evaluated with one of those in the list to specify the baseline probability. This is the starting point for the analysis. HEART then provides a list of error producing conditions (EPC) with associated multipliers which can be applied to the baseline probabilities if they are associated with the tasks. For example the EPC: 'Unfamiliarity with a situation which is potentially important but only occurs infrequently or is novel' has an associated multiplier of x17. This means that going from the least favourable conditions for this EPC, e.g. total unfamiliarity, to the best possible conditions, i.e. complete familiarity, will mean that the baseline probability will change by a factor of 17.

To actually use the technique, the analyst, having decided upon which of the generic task types matches the situation being assessed most closely, then decides which of the EPCs are present. The next stage is to decide what proportion of each of the factors is present. For example if the specific task is about half way between the best and the worst conditions, then the multiplier which is applied to the baseline probability will be 0.5 X 17 (this multiplication is adjusted so that it does not lead to a decrease in error probability). Similar products are obtained for all the EPCs deemed to be present and the baseline probability is multiplied by each of these products to obtain an error probability adjusted by the influencing factors that may be present.

HEART has been widely used by engineering risk analysts in the nuclear power and other industries. Its major attraction is that it is relatively easy to use. Essentially, the user only has to make only three types of decision: which generic task type most closely matches the situation being assessed, which EPCs are relevant, and what proportion of each EPC is present. In addition all of the information required to generate numbers is available as part of the technique (i.e. generic task type error probabilities, EPCs and their associated multipliers), and hence no external database is required. In terms of its technical basis, HEART can be defended on the basis that the EPCs chosen are those which the human factors literature supports and that the baseline task error probabilities and the EPC multipliers are also derived from the literature.

Critics of HEART have pointed to a number of potential problem areas in the technique. Many of these problems stem from the very ease of use of the technique. Firstly, because the technique provides a set of generic tasks as its starting point, it is very tempting for the analyst to try to fit the tasks he is assessing into one of these task types, even though this may not be appropriate. Similar considerations apply to the EPCs. These appear to have been included mainly on the basis that they are the types of factor which have been studied by applied psychologists and for which generic multiplying factors can be derived by aggregating a number of studies. However, the existence of these EPCs constrains the analyst's choice and encourages him or her to use these factors whether or not they are appropriate in a specific situation. From a technical standpoint, the HEART method of combining together the effects of more than one influencing factor that may be present in a situation takes no account of possible interactions between factors. For example the effects of lack of experience and shortage of time would certainly appear to be greater if both of these factors are present than if they occur in isolation or in combination with other factors for which there is no negative interaction. From the perspective of the MSA application, a major disadvantage of HEART is that it does not provide any explicit modelling or quantification of the effects of higher level policy decision on human failures at the operational level.

Success Likelihood Index Methodology (SLIM)

In SLIM, a model is developed in which the Performance Influencing Factors relevant to a set of tasks being evaluated are first defined, and the tasks are then rated on these factors. Rating in this context means that the conditions under which each of the tasks is performed is evaluated on each of the factors. Weights which define the importance of each factor are then derived, and the ratings and weights then combined to give a Success Likelihood Index (SLI) for each task. If at least two tasks with known human error probabilities are included in the set being evaluated, this allows the SLI scale to be calibrated, such that the SLI for each task can be converted to a probability of error.

4.6.5 Influence Diagrams

As described in section 3, in the Influence Diagram, a model is set up which defines the factors which directly influence error probability for a task (or a broad category of tasks) and the interrelationships between these factors. Higher level factors which influence the direct factors are then defined and this network of influences constitutes the Influence Diagram.

In order to quantify the qualitative model defined during the first stage of IDA, numerical assessments are made of the strength of influence of the higher level factors on the lower level direct factors. Assessment of probabilities of success or failure of the task are then made for various combinations of direct factors, and these probabilities are weighted by the effects of the higher level factors. Finally these assessments are combined together by the model to give the overall probability of success or failure for the task or category of tasks under consideration. The Influence Diagram is particularly useful for evaluating the efforts of socio-technical variables such as organisational structures and management policies.

4.7 Conclusions

The SLIM and Influence Diagram approach (IDA) are recommended as the primary approaches to be used for the human factors quantification requirements of the MSA FSA process. This is because these are the only techniques which satisfy the criteria set out in section 4.4. In particular, they allow the quantification of the effects of Performance Influencing Factors and organisational and policy variables on the human failures identified as important to risk. A similar preliminary conclusion was reached by the Tavistock Institute.

5. Examples

In this section two case studies will be described to illustrate the application of the SLIM and IDA methodologies to the quantification of human factors in a specific scenario. An important feature of this case study is that it shows how SLIM and IDA can be used together. SLIM is first applied to quantify human failure probabilities as a function of the immediate Performance Influencing Factors (PIFs) affecting the likelihood of failure. By transferring these results into the Influence Diagram, and quantifying the effects of higher level policy variables on the direct Performance Influencing Factors, the original probabilities are modified to reflect these variables. It is then possible to use the Influence Diagram structure to assess how the operational level probabilities of failure change as a function of changes in policy.

5.1 Scenario description

In order to provide a vehicle for the case studies, the following marine scenario is provided.

Grounding at river bar scenario

A Containership is approaching a tidal estuary, making for a port which is 10 hours Standby sailing time up-river, with the prospect of only a few hours in port working cargo before returning and continuing with a deep sea passage. There is a shallow tidal bar at the mouth of the estuary; but local regulations, custom and practice do not require the Ship to take a Pilot until she is several miles inside the bar (in severe weather, Pilot Boats are at risk outside the bar). There is an anchorage area outside the bar but it is exposed to the weather and considered by seafarers to be not good holding ground. Company Standing Orders instruct Masters not to use this anchorage unless absolutely necessary.

The vessel has only 4 deck officers, the Master and Chief, Second and Third Mate. At the time of approaching the Bar, it is the Third Mate's watch and the Second Mate has been off watch for an hour and is asleep. The Master and Chief Mate intend to rest for as long as possible in view of the prospect of a long double standby with a period of intense cargo work in between.

The Master therefore turns in and instructs the Third Mate to awaken him when the Pilot Boat is approaching, but not before unless absolutely necessary. He is under great pressure to make port both because he has received a fax from Head Office emphasising the importance of arriving on time and because his productivity bonus would be affected if he does not arrive on time. He has computed the water under the keel at the time of passing over the bar and knows it is marginal, but he is an old hand at approaching this river, has taken a chance on the state of the tide a dozen times before and is confident of getting away with it. The Third Mate is aware of the risks also but is reluctant to speak out against the authority of the Master and peer pressure inhibits him from being seen to have an over-cautious perception of risk.

In fact the Master has taken a chance once too often and while he is asleep, the Third Mate steers a little too near the shore of the estuary and puts the vessel aground at the bar of the river.

5.2 SLIM Case Study

The SLIM technique can be applied to quantify groups of tasks which are influenced by the same PIFs, or to assess how the probability of the same task or task segment varies as a function of changes in these PIFs. The latter mode will be illustrated in this case study, but exactly the same calculations would apply to a group of separate tasks or failure modes.

The particular failure mode being considered in this example is the likelihood of a ship to ship collision arising from a ship handling error which in turn arises from a violation (see figure 4). If the overall probability or frequency was required, then similar calculations would be performed for the same failure mode arising from slips or violations, and these three estimates would then be added to give the total failure probability.

Of course it is likely that several of the PIFs such as competence and information overload would influence both slips and mistakes. The PIFs for violations are, however, likely to be different from those which influence slips and mistakes. In the following case study, the probability of a violation will be assessed as a function of a range of situations, represented as different ratings of the PIFs in the different situations. The SLIM process consists of a series of stages each of which will be described in turn below. It is worth noting that although the calculations may appear to be tedious, computer programs are available which greatly facilitate the process. Alternatively, the calculations can be implemented quite simply using a spreadsheet.

Step 1 Group tasks or error modes into homogeneous sets

The purpose of this stage is to ensure that all tasks or errors in a particular group are all influenced by the same PIFs. Since the first stage of SLIM can be interpreted as developing a series of specific models for different types of errors connecting variations in PIFs with variations in failure probabilities, then it is important that all tasks or errors under consideration are correctly assigned to an appropriate group. When a SLIM tool is developed for a particular industry, then all of the failure types of interest are usually grouped together in a single exercise, and instructions are generated to assist analysts in assigning new tasks or errors to an appropriate category. In this particular case study, we are evaluating the same failure mode under different conditions and hence it is not necessary to perform the grouping process.

Step 2 Decide on the relevant PIFs

The PIFs which are deemed to influence violations leading to ship handling errors are as follows:

1. Existence of 'profit before safety' attitudes on the ship
2. Extent which procedures are respected and adhered to (a 'procedures culture')
3. Degree of accuracy of perception of risk
4. Extent to which the ships master will follow his own judgement to take advantage of conditions that may confer commercial advantages

It should be emphasised that these are merely meant to be illustrations. In practice it is desirable that an expert group develops these factors and defines exactly what they mean. This means developing numerical scales which will allow situations to be rated consistently. In this case, the following numerical scales could be defined with descriptive 'anchors' to facilitate the assessment process. In the table below, the end point of the scales are defined explicitly. The numbers which are appended to each description refer to the ideal point for each scale. This is defined as the value of the scale which represents the best conditions in terms of minimising the likelihood of failure. With some scales the maximum value of the scale (9) and other scales the minimum (1) is used to represent the ideal conditions. This is because some scales are easier to think of in terms of increasing values giving rise to better conditions, whereas in other cases the opposite is the case. For example, increasing distractions will probably lead to increasing errors, therefore the ideal value on distractions scale would be 1, representing minimum distractions. On the other hand increasing competence will lead to fewer errors, therefore the highest levels of competence should be assigned a rating of 9. This is handled by assigning an ideal value to the scale (which does not have to be at the end point) which represents the best case. The arithmetic of the calculations in SLIM then uses the absolute distance from the ideal point as its input.

Best case	Worst case
Safety always comes before profit (9)	Profit always comes before safety (1)
Procedures always followed because crew helped to develop them and only useful procedures retained (9)	Procedures rarely followed because they are impractical and crew do not understand them(1)
Crew have a clear understanding of the nature of risks (9)	Crew have no understanding of the nature of the risks (1)
Ship's master accurately considers the risks when taking advantage of local conditions (9)	Ships master usually misunderstands the risks when taking advantage of local conditions (1)

Table 1 Rating scales for PIFs

Step 3 Rate the task conditions on the PIFs

At this stage numerical ratings are made of each task or error condition on each of the scales. This is represented in Table 2 for the four task conditions being considered in this analysis.

Case	Ships master's risk perception (weight=0.40)	Crew risk perception (weight=0.10)	Adherence to procedures (weight=0.30)	Balance between profit & safety (weight=0.20)
1	6	8	3	6
2	2	8	5	6
3	2	7	6	2
4	2	8	6	2

Table 2 PIF ratings and weights for four situations

In verbal terms the task conditions can be expressed as follows:

Case 1
The Masters risk perception is above average, the crew risk perception is very accurate (close to the ideal value of 9), adherence to procedures is poor (3) and the balance between profit and safety is above average (6)

Case 2
In this case, the Masters risk perception is poor, the crew risk perception as before, adherence to procedures is average and the balance between profit and safety as before.

Case 3
Masters risk perception is as the previous condition, crew risk perception slightly poorer but still good, adherence to procedures is slightly better, but profit versus safety is poor (i.e. profit is likely to dominate)

Case 4
Masters risk perception is as in 3, crew risk perception slightly improved to very good, adherence to procedures is the same and profit versus safety unchanged.

Step 4 Assign weights

The SLIM procedure provides a facility to allow the factors which determine the error probability to be differentially weighted to reflect the differing generic influence of each PIF in terms of its strength of effect on the probability of failure or success. If in general, all of the factors have an equal influence then all the weights are equal. If on the other hand, changes in the ratings of some factors have a greater affect on the failure probability than changes in other factors, then this can be included in the model by assigning differential weights to each PIF. It is important to understand the distinction between ratings and weights. Ratings are the specific states of each PIF for a particular situation or task. Weights are generic multipliers that apply to all tasks or situations in a set. In Table 2 some weights have been assigned to the PIFs to illustrate the process. The fact that the weight of the Master's risk perception PIF has four times the weight of the crews risk perception means that changes in the former have four times the effect of changes in the latter. This does not necessarily lead to the same magnitude of change in the error probability, as will be discussed later.

Weights are normally assigned by the use of models of error or from inputs by experienced personnel. The weights do not normally change once they have been assigned to a particular category of tasks. If there is insufficient information to assign weights it is better to assume that they are equal.

Step 6 Calculate Success likelihood Index

The SLI is given by the following expression:

$SLI_j = \Sigma R_{ij}W_i$

where SLI_j = SLI for task $_j$

W_i = normalized importance weight for the ith PIF (weights sum to 1)

R_{ij} = Rating of task on the ith PIF.

The SLI for each task is the weighted sum of the ratings for each task on each PIF.

In order to calculate the SLIs, ratings such as those in Table 2 may need to be to be rescaled to take into account the fact that the some of the ideal points are at different ends of the rating scales. Rescaling also converts the range of the ratings from 1 to 9 to 0 to 1. Although this step is not strictly necessary, it can be used as a formula in a spreadsheet or computer program to convert different rating scales with different ideal points to a common scale to facilitate checking the calculation. The following formula converts the original ratings to rescaled ratings:

$RR = 1 - ABS (R - IP) / 4 + ABS (5 - IP)$

Where RR = rescaled rating

R = original rating

IP = ideal value for scale on which the rating is made.

The accuracy of this formula can be verified by substituting the values 1 and 9 for scales where the ideal point is either 1 or 9. The formula converts the original ratings to 0.0 or 1.0 as appropriate. Values of ratings between 1 and 9 are converted in the same way.

Using this formula on the ratings in Table 2 produces Table 3 which contains the rescaled ratings, the assigned weights for the PIFs and the calculated Success Likelihood Indices for each task:

Task conditions	PIFs				SLI
	Master risk perception	Crew risk perception	Adherence to procedures	Balance between profit & safety	
1	0.63	0.88	0.25	0.63	0.54
2	0.13	0.88	0.50	0.63	0.41
3	0.13	0.75	0.63	0.13	0.34
4	0.13	0.88	0.63	0.13	0.35
Weights	0.4	0.1	0.3	0.2	

Table 3: Re-scaled ratings and SLIs

- Convert the Success Likelihood Indices to Probabilities

The SLIs represent a relative measure of the likelihood of success or failure, relative to one another. In order to convert the SLI Scale to a probability scale, it is necessary to calibrate it. If a reasonably large number of operations in the set being evaluated have known probabilities (for example as a result of incident data having been collected over a long period of time), then it is possible to perform a regression analysis which will find the line of best fit between the SLI values and their corresponding error probabilities. The resulting regression equation can then be used to calculate the error probabilities for the other operations in the group by substituting the SLIs into the regression equation.

If, as is usually the case, there are insufficient data to allow the calculation of an empirical relationship between the SLIs and error probabilities, then a mathematical relationship has to be assumed. The usual form of the assumed relationship is log-linear, as shown below:

$$\log(\text{HEP}) = A\ \text{SLI} + B \quad \text{——} \quad (1)$$

where HEP = human error probability
A, B are constants

This assumption is based partly on experimental evidence which shows a log-linear relationship between the evaluation of the factors affecting performance on maintenance tasks, and actual performance on these tasks, Pontecorvo (1965). In order to calculate the constants A and B in equation, at least two tasks with known SLIs and error probabilities must be available in the set of tasks being evaluated.

In the example under discussion, it was estimated by a group of expert seafarers that given the conditions for the first (best) case, the probability of a vessel running aground would be very low but not zero. Since over the 20 year time period considered, there would be 10,000 opportunities for this event to occur on a world wide basis, the expected frequency of grounding was set at 1 in 20 years. For the worst case scenario, case 3, the corresponding frequency was estimated to be 100 events. On the basis of this evidence, the following probabilities were assigned to these errors:

Case 1 a ship will run aground once in 20 years	$p= 1 \times 10^{-4}$
Case 3 a ship will run aground 100 times in 20 years	$p = 1 \times 10^{-2}$

These values, and the corresponding SLIs for these tasks (from Table 3), are substituted in the general equation (1). The resulting simultaneous equations can be used to calculate the constants A and B. These are substituted in the general equation (1) to produce the following calibration equation:

$$\text{Log (HEP)} = -2.303\ \text{SLI} + 3.166 \text{———} \quad (2)$$

If the SLI values from Table 3 for the other two conditions in the set are substituted in this equation, the resulting frequencies and error probabilities are as follows:

Case 1 a ship will run aground 18 times in 20 years	$p= 1.8 \times 10^{-3}$
Case 4 a ship will run aground 75 times in 20 years	$p= 7.5 \times 10^{-3}$

Step 7 Perform Sensitivity Analysis

The nature of the SLIM technique renders it very suitable for 'what if' analyses to investigate the effects of changing some of the PIF values on the resulting error probabilities. For example, as shown in Table 2, the Master's risk perception is poor for both of the above situations (rating of master's risk perception = 2, best value = 9). The effects of improving risk perception can be investigated by assigning a rating of 9 for each situation instead of 2. This changes the SLI, and if the new SLI value is substituted in equation (2) the probabilities change as follows:

Case 1 a ship will run aground about 0.5 times in 20 years	$= 5.6 \times 10^{-5}$
Case 3 a ship will run aground about 2.4 times in 20 years	$= 2.4 \times 10^{-4}$

Conclusions

The SLIM technique is a highly flexible method which allows considerable freedom in performing what-if analyses. In common with most human reliability quantification techniques, it requires defensible calibration data, preferably from incident reporting systems, to be effective. In the absence of such data, the calibration values have to be generated by expert judgments made by experienced seafarers.

5.3 Influence Diagram case study

In this section, the previous case study used in the SLIM example will be extended to show how the effects of organisational variables can be factored into the analysis.

Step 1 Develop the Influence Diagram

The Influence diagram is developed by means of structured with a group of marine experts who are familiar with both operational realities and policy level issues and who are prepared to develop a structural model to describe the influencing relationships between these two domains. In this case study, an illustrative model is developed based upon the earlier case study. It should be pointed out that this case study was developed purely to illustrate the principles involved, and is a highly simplified version of the actual interrelationships.

It is easier to develop the ID starting with the direct PIFs that influence the probabilities of failure at the operational level. In this case study these are the PIFs which were used in the SLIM case study in the last section. The example ID that was developed to illustrate the IDA quantification method is shown in Figure 6. In this diagram, the failure mode level is represented by the specific error mode under consideration. At the Influencing Factor level, for simplicity only three of the four PIFs considered in the SLIM case study have been included. At the implementation level, two influences which affect the risk perception of the Master have been postulated. These are his knowledge of the situation, as gained from operational experience and the standing orders that exist. The risk perception of the crew is influenced by the ship's culture in the form of peer pressure not to be over cautious, and also by the training, certification and manning policies which exist at the policy level. The profit before safety pressures arise from the implementation level (e.g. the shipping company) influences staffing levels, the instructions from head office to arrive at port on time and the potential for losing productivity bonuses. Staffing levels are also influenced by policy level directives.

Step 2 Quantify the Influences

The quantification of the ID is an intensive iterative process which requires inputs from knowledgeable participants in a group elicitation setting. An experienced facilitator is essential for this exercise. Although some resources will be necessary to generate the ID, in a real setting, it is probable that a much more generic ID than the one being considered in this example will be generated. Once this has been done it can be applied to a wide variety of scenarios.

The quantification process begins by evaluating the lowest level factors in the ID, in this case these are the policy and implementation level factors. In the evaluation process the term 'balance of evidence is use to denote the numerical assessment of the group that the specific conditions or combination of conditions are either good or bad in terms of their influence on the next level in the ID. The first set of evaluations made are simple unconditional assessments, since the first level influences are not influenced by any lower level factors. Thus, the first set of questions might be:

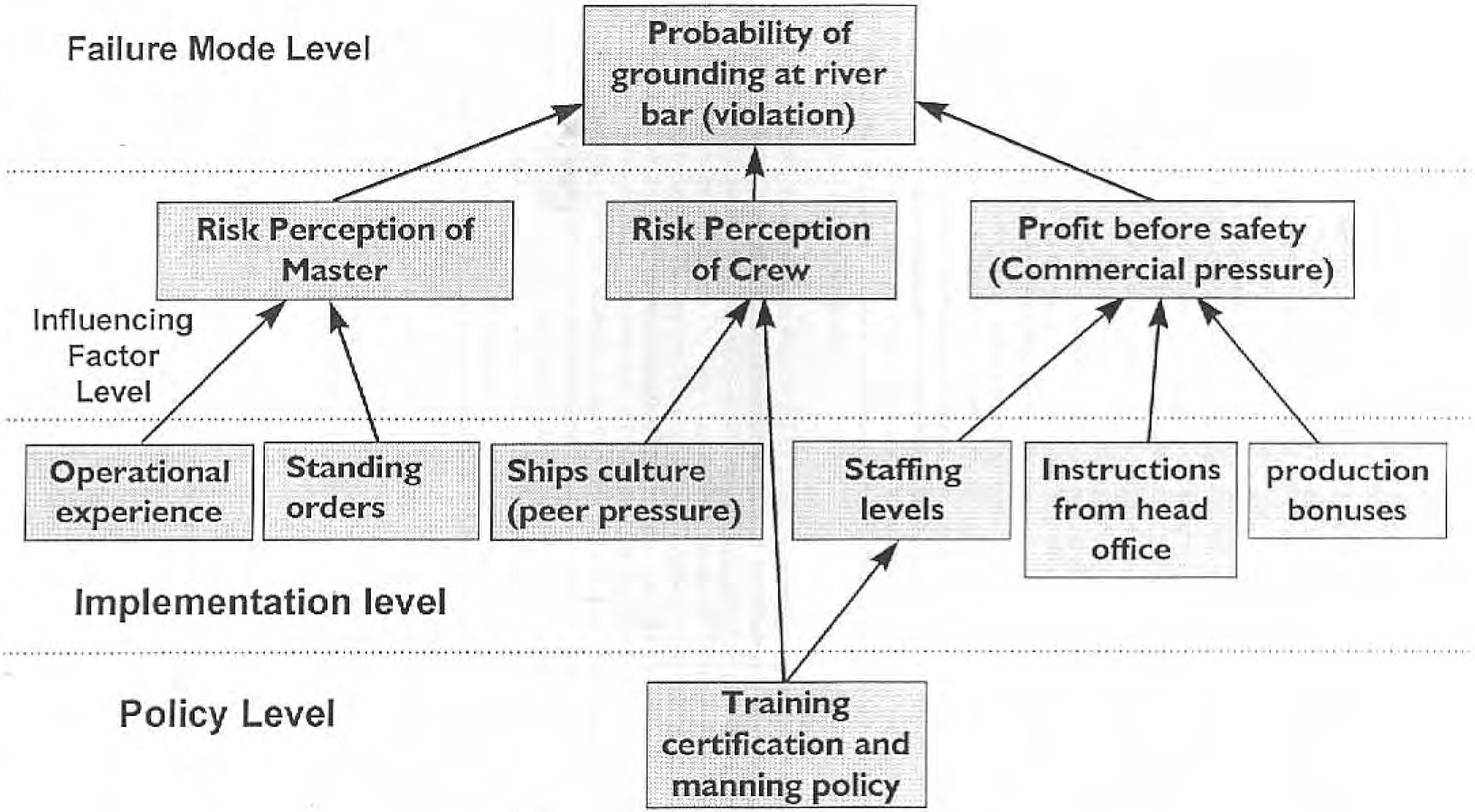


Figure 6 Influence Diagram

data\1996\clients\msa\msadiag.pct

"What is the balance of evidence that operational experience is having a positive effect on the risk perception of ships masters?" If the consensus was that 80% of the evidence indicated that operational experience led to a less accurate perception of risk by ships masters (i.e. to a greater degree of risk taking), and only 20% indicated that experience had a positive influence, then this information would be recorded in the tabular form shown in Table 1 in the Appendix. This process is then be repeated for the other bottom level influences and this information is recorded in tables 1, 2 4, 5 7, and 8 shown in the Appendix. Once the evaluations of the lowest level influences have been completed, the next stage is to begin the assessment of any combined influences. The first of these is the effect of training certification and manning policy on staffing levels. The form of questioning at this level is indicated in table 9 in the Appendix. In this case the questions are of the form: "If manning policies are effective, then what is the balance of evidence that staffing levels will be appropriate?" A question of this form is necessary because there is no guarantee that effective manning policies will automatically lead to perfect staffing levels. This is because there may be other factors involved, which have a greater effect on staffing levels. The results of the questions at this level are then weighted by the results from the previous level of questioning, i.e. the balance of evidence that manning level policy was effective in its effects on staffing levels. This process is continued as described in the Appendix until a complete set of combined weights is built up which can be used to modify the absolute probabilities derived from another source such as SLIM or Direct Numerical Evaluation (see table 11 column 5 and 6 in the Appendix).

These probabilities are derived from SLIM by using the conditions specified by each line of the table. For example, the probability in table 11 for which risk perception of the master is accurate, risk perception of the crew is accurate and profit before safety is not a pressure, represents the best case situation where all of these PIFs were optimal. Although this fully optimal situation was not evaluated in the SLIM exercise set out in section 5.1, an approximation to this best case was provided by the sensitivity analysis which indicated that this corresponded to a probability of 1 x 10^{-4}. This probability would be inserted in the first row of table 11 in the Appendix. A similar process would be used for deriving the other absolute probabilities in columns 5 and 6 of this table. In each case appropriate numerical ratings corresponding to the conditions implied in columns 1-3 of table 11 are inserted into the SLIM model to derive the absolute probabilities required in columns 5 and 6. These probabilities are then multiplied by the corresponding weights in column 7 to take into account the effects of the various influencing factors that have been evaluated by the questions asked previously in the ID session. Finally, the products of the weights and absolute probabilities are then summed to obtain an overall probability. In the example described in detail in the Appendix, some of the numerical values from the SLIM case study are inserted into the table to illustrate how the calculation would be completed.

6. Conclusions

In this section the application of the methodologies presented in this report to the five stages of the FSA process will be briefly described.

6.1 Step 1 Identification of hazards

An important potential application would be the use of the SPEAR process to assist in the identification of human initiated causal sequences which may not be revealed by the brainstorming approaches that have been adopted up to now. The proposed process would comprise the following steps:

a) Beginning with the generic ship, a functional breakdown would be produced which would specify the various domains of operation that are common to all ships. These might include high level functional descriptions such as manage propulsion systems, navigate the high seas, load and unload passengers and cargo, design build and commission ships etc.

b)These high level functions are then decomposed further to a point where human interactions with the system can be explicitly identified. At this stage the techniques contained within SPEAR can be used to identify critical human-hardware interfaces and to evaluate possible human failures which could be the initiators of failure sequences. A similar process can be used to identify failure modes in the progression and magnitude reduction phases Since the majority of accidents have a substantial human contribution, it seems logical to ensure that process designed to identify significant hazards should explicitly apply a methodology to search for human failure modes that could act as initiators.

6.2 Step 2 Risk assessment

Both the SPEAR methodology and the Influence Diagram can play a major role in this stage of the FSA process. A specific application area is in providing a qualitative approach to allow the sources of risk to be identified arising from potential failures. It is anticipated that human factors methodologies will also have considerable relevance to identifying and representing causal modelling of failure and recovery sequences associated with the scenarios which have been identified in Step 1. Another benefit is that the IDA and SLIM techniques for quantifying the likelihood of human errors can also be used to generate quantified failure probabilities as a function of the Performance Influencing Factors in a situation. This supports the identification of risk reduction strategies which will be passed on to Step 3.

6.3 Step 3 Risk Control Options

The fact that the IDA technique can provide a link with policy factors means that the policy implications of the risk reduction strategies carried forward from Step 2 can be evaluated. The risk reduction options available at the human factors level are implicit in the PIFs which determine the failure probabilities in the recommended SLIM and IDA methodologies. This greatly facilitates the specification of these options.

6.4 Step 4 Cost-Benefit analysis

By using the SLIM and IDA methodologies it is possible to express human failure probabilities as a function of factors such as procedures and training quality. If improvements in these factors can be costed, then the corresponding benefits in terms of reduced human error probabilities and hence reduced risks can be calculated. This allows a systematic approach to be adopted by using cost benefit criteria from Step 4 to assist in the choice of risk control options in Step 3.

6.5 Step 5 Decision Making

A major application of the IDA methodology at the decision making stage is to provide an easily understandable visual summary of the implications of changes in policies for operational level risks. The ID structure can therefore provide a basis for communication about FSA results and their policy implications within the IMO. This should facilitate the decision making processes and provide a means of integrating the outputs from the other steps of the FSA into the IMO level decision making processes.

Appendix

Influence Diagrams

Commentary on the Calculations

This commentary is provided to clarify the calculations in the following tables. In Table 1, the assessment team is asked to evaluate the evidence that operational experience is effective in enhancing the risk perception of the Master. In order to make this evaluation, they will be provided with an 'indicator' in the form of a scale specifying the nature of the evidence that should be taken into account. For example, the end of the scale defining the ideal situation would include conditions such as: 'results of grounding incidents and near misses regularly fed back to ship's masters'. The other end of the scale would describe the worst case situation, for example 'No feedback from operational experience fed back to the master.'. In the example cited, the evidence strongly indicates that feedback is not used effectively to improve perception of risk.

Table 2 contains a similar assessment to Table 1 but for the effectiveness of standing orders. As illustrated in Table 3, the assessment team is then asked to evaluate the weight of evidence that the accuracy of the Master's risk perception will be high (or low) given various combinations of the influencing factors operational experience and standing orders. Of course, such evaluations are difficult to make. However, they utilize whatever expert knowledge is possessed by the evaluation team, and factor this into the analysis. They also allow the assessors to factor into their evaluations any interactions between factors. For example, the combined effects of inadequate feedback of operational experience and unclear standing orders may degrade the accuracy of risk perception more strongly than either factor considered in isolation . Each of the conditional assessments is therefore weighted by the results of stages 1 and 2 and the products added together to give an estimate of the unconditional probability that the risk perception is adequate.

Similar assessments are performed to evaluate the probability that the risk perception of the crew is adequate (Table 6) that staffing levels are adequate (Table 9) and that commercial pressures (profit or safety) will be high or low (Table 10). In this latter case, since three influences impact upon profit before safety, eight joint assessments need to be made.

Although these combined assessments are arduous, it should be noted that the evaluations of the effects of combinations of influences may be regarded as applicable across a range of systems, and hence would only need to be performed once for a generic model. The system specific evaluations would then be the simpler level 2 assessments set out in Tables 1, 2, 4, 5, 7 and 8. As discussed earlier, guidance for performing these assessments could be provided by the use of PIF scales delineating the conditions for the least and most favorable ends of the scales. Similar scales can be used to make direct evaluations of the level 1 influences, if the assessments described earlier are judged to be too difficult. Even if the full assessments are made, it is useful to compare these with the indirect assessments to check convergence.

The final stage of the procedure is to generate an overall unconditional probability of human error (Table 11). This is achieved by assigning probabilities of error to combinations of the three first level influences of risk perception of the master, risk perception of the crew and profit before safety. These conditional probabilities are generic, in that they could apply to any system. They are made specific to the situation under consideration by multiplying them by the assessed probabilities of the level 1 influences, as derived from the earlier analyses. These products are then summed to give the overall unconditional probability of error occurrence in the situation being evaluated.

Influence Diagram Calculations

1 What is the weight of evidence for availability of feedback from operational experience as an aid to enhancing risk perception of the Master?	
Good	Poor
.20	.80

2 What is the weight of evidence for the value of standing orders as a means of enhancing risk perception of the Master?	
Used	Not Used
.20	.80

3 For Risk perception of the Master

If	and	then	*weight of evidence that risk perception of the Master is*		Joint Weight (feedback x Standing orders)
feedback from operating experience is:	*standing orders are clear*		*accurate is:*	*inaccurate is:*	
Good	Used		.95	.05	.04= (.20 x .20)
Good	Not Used		.80	.20	.16 =(.20 x .80)
Poor	Used		.15	.85	.16 =(.80 x .20)
Poor	Not Used		.10	.90	.64 =(.80 x .80)
Unconditional Probability (weighted sum) that risk perception of master is accurate versus inaccurate is:			.254	.746	

4 What is the weight of evidence that ship's culture (peer pressure)is:	
Positive	Negative
.30	.70

5 What is the weight of evidence that Training certification and manning policies are:	
Effective	Ineffective
.10	.90

6			**For risk perception of crew**		
If **ship's culture is**	**and** *Training certification & manning policies are*	**then**	*weight of evidence that Risk perception of crew is*		Joint Weight (Culture x Training, certification & manning policies)
			Positive	*Negative*	
Positive	Effective		.90	.10	.03 =(.30 x .10)
Positive	Ineffective		.60	.40	.27 =(.30 x .90)
Negative	Effective		.50	.50	.07 =(.70 x .10)
Negative	Ineffective		.05	.95	.63 =(.70 x .90)
Unconditional Probability (weighted sum) that Risk perception of the crew is accurate vs. inaccurate is:			.255	.744	

7 What is the weight of evidence for Instructions from head office being	
Helpful	Unhelpful
.50	.50

8 What is the weight of evidence for Production Bonuses being	
Positive	Negative
.60	.40

9 **For Staffing Levels**				
If Training, certification and manning policy	**then**	*weight of evidence that Staffing Levels are*		Weight (training, certification & manning policy) (from 5)
is:		*adequate is:*	*inadequate is:*	
Effective		.60	.40	.10
Ineffective		.20	.80	.90
Unconditional Probability (weighted sum) that Staffing Levels are adequate vs. inadequate is:		.24	.76	

10 **For Profit before safety (Commercial pressure)**

If *Staffing levels are:*	and *Instructions from Head Office are*	and *Training, certification and manning policy is:*	then	*weight of evidence for profit before safety is:* *high is:*	*low is:*	Joint Weight (staffing levels x instructions from head office x training, certification & manning policies
Adequate	clear	effective		.95	.05	.072 (.24 x .50 x .60)
Adequate	clear	ineffective		.30	.70	.048 (.24 x .50 x .40)
Adequate	unclear	effective		.90	.10	.072 (.24 x .50 x .60)
Adequate	unclear	ineffective		.25	.75	.048 (.24 x .50 x .40)
Inadequate	clear	effective		.50	.50	.023 (.76 x .50 x .60)
Inadequate	clear	ineffective		.20	.80	.015 (.76 x .50 x .40)
Inadequate	unclear	effective		.40	.60	.023 (.76 x .50 x .60)
Inadequate	unclear	ineffective		.01	.99	.015 (.76 x .50 x .40)
Unconditional Probability (weighted sum) that profit before safety is high vs. low is:				.3981	.6019	

11 **For probability of grounding at river bar**

IF **Risk perception of master is:**	**and** **Risk perception of crew is:**	**and Profit before safety is:**	**then**	*the probability of* *success is:*	*failure is:*	Joint Probabilities (risk perception of master x risk perception of crew x profit before safety)
Accurate	Accurate	Low		0.9999	0.0001	.0390 (.25 x .26 x .60)
Accurate	Accurate	High				.0258 (.25 x .26 x .40)
Accurate	inaccurate	Low				.1137 (.25 x .74 x .60)
Accurate	Inaccurate	High				.0752 (.25 x .74 x .40)
Inaccurate	Accurate	Low				.1145 (.75 x .26 x .40)
Inaccurate	Accurate	High				.076 (.75 x .26 x .40)
Inaccurate	Inaccurate	Low				.3341 (.75 x .74 x .60)
Inaccurate	Inaccurate	High		0.99	0.01	.2209 (.75 x .74 x .40)
Assessed Unconditional Probability of success vs. failure is:				.99778	.0022	