Status and Needs on Human Reliability Assessment of Complex Systems

Oliver Straeter *, Radim Dolezal **, Marcus Arenius * & Georgios Athanassiou *

*Universität Kassel
Fachbereich Maschinenbau
Arbeits- und Organisationspsychologie
Heinrich-Plett-Strasse 40, D-34132 Kassel

** Technical University of Liberec
Department of Dependability and Risk
Studentska 2, 461 17 LIBEREC, Czech Republic

1. Introduction

For the design of technical systems planned to come to market or for the change of an existing system, there is the need to secure that the change is not affecting safety and availability of the new or amended system. Authorities often state as a criterion that a new or modified system should be at least as reliable as the system currently in operation. The successful implementation of a system-change or innovation is therefore bound by the evidence, whether the new system provides the expected safety performance. The basis for this consideration is the risk-based approach, defined as

\[ \text{Risk} = \text{Amount of damage} \times \text{Frequency of damage} \]

As the definition shows, safety is defined by the risk imposed but not by the safety generated by a change.

For complex systems, safety targets are defined, which need to be reached in order to be approved by an authority. Such so-called “Target Level of Safety” determines at what point an operating permit (license) will be issued by the authority. If the risk calculated by a safety case is above the target level, one usually must retrofit mitigations in order to get approval. This can consist of improvements in the design of the system or guidelines for test intervals for safety-critical components.

The correct calculation of the target value is an important variable - not only for regulatory approval, but also for the safety and operational planning of production and operation. A good risk assessment allows manufacturers and operators to achieve an efficient integration of safety and operational goals.

Of particular importance in complex systems is the fact that complex systems are composed of various technical and human elements. As an example, modern control systems usually contain software. This software is programmed by man, and there is the potential that the system has a negative impact on safety due to a programming error. A prominent case in this context is the crash of an Ariane 5 rocket on 4 June 1996 due to an undetected programming error in various steps of the quality assurance. Another problem of complex systems lies in the interdependencies between technical and human capabilities that lead to dependencies in the safety performance of the human (e.g. Human trust the reliance of the braking system while driving and usually fail to properly react if the breaking system is failing on demand).

Thus, contrary to a commonly held expectation, the significance of the human contribution to safety is increasing with automation (cf. Bainbridge, 1987). This change can mainly be traced back to a role-shift in the man machine interaction, as the human contribution, formerly consisting of direct system manipulation, is now shifted into the areas of design, maintenance and emergency management. Especially in emergency management of complex systems, higher cognitive demands are required by the operators, particularly if (1) the system does not behave according to user expectations, (2) the system behaviour can be interpreted ambiguously, or (3) pre-planned procedures are not available, not applicable or ineffective in the given situation.

An accurate safety case must therefore capture a complex system of human actions and interactions
between people and technology. The accuracy of a safety case in depicting the complexity of a system is highly dependent on the capability of the methods employed in the so-called probabilistic safety analyses (Probabilistic Safety Analysis - PSA) and on the extent they are able to support proper human reliability analysis (Human Reliability Analysis - HRA).

2. Requirements for Human Reliability Methods used in Safety Cases

To perform a human reliability assessment, different methods were developed, referred to as HRA methods. The key issue of HRA methods in a safety case for complex systems is the correct modelling of the situation (respectively scenario) to assess and the correct calculation of human reliability. These in turn are determined by the correct scope of an assessment, the capabilities of the selected HRA method to represent the issues within the scope completely (the better the representation the higher the quality) and the suitability of the data for the human reliability assessment (Straeter 2008; ISO / IEC 31 010). Risk management can only be valid and effective if these conditions are met successfully, a risk management can be valid and effective.

Picture 1 shows the relationship of these elements of risk management of complex systems. The steps of risk management are discussed below. Four key elements are introduced in the Picture: the safety scanning to provide proper scoping of safety issues, HRA methods to represent the safety-related performance model of human behaviour, the CAHR Method to derive appropriate data for assessment and the risk management to ensure that correct mitigations are derived. The different elements will be outlined in the following sections.

2.1 Importance of Scope of Reliability Assessments

It is not immediately obvious why an assessment of a subsystem depends on the scope chosen by the analyst. The reason for this lies in the fact that human actions are not limited to a specific function and thus may cause dependencies and reverberations in the system that can be of both positive and negative nature. These dependences are not limited to the choice of scope by an assessor of the system and may therefore affect the functioning of other parts of the overall system.

Only a functional analysis of safety cases or the observation of adverse incidents or accidents make obvious that a scope was defined too narrow and that the dependencies in human reliability are miscalculated. One can say that usually the positive contribution of human reliability is underestimated and the contribution of an automated system is overestimated in an overall system (see Straeter, 2008), if the scope is set too narrow.

In order to define a proper scope for assessment, a classification of ‘levels of work’ according to Leveson (2002) is useful (see also VDI 4006-3). Leveson divides the scope possible into the following levels:

- Operational level
- Maintenance level
- Level of system design
- Organizational level
- Regulatory level

These levels of work describe the maximum scope possible. For a specific safety case, one needs to further specify the scope in order to assess the safety correctly. The so-called safety scanning (or screening) was developed in Air Traffic Management (ATM) to better define the scope of a change. The safety scanning is a structured process to go through the various levels of work and to generate a correct scope. It was developed in the context of the Single European Sky development (SESAR 2005; Straeter, 2006; SCAN TF 2009) and was applied in the railway system as well (cf. Milius & Petrek, 2011).

The Safety Scanning Technique makes use of a set of “safety fundamentals”, which are basic design criteria for safe systems. They are based on a review
of standards throughout safety relevant industries. In addition to the aviation industry, the experiences from the nuclear, petrochemical, maritime and railway industries were taken into consideration. Examples of these fundamentals are: Regulatory principles, Safety policy, Planning of safety assurance, Competence, Organisation, Transparency, Redundancy and Reliability.

Safety Scanning supports the systematic consideration of safety issues against safety fundamentals and provides input into strategy development. The objectives of Safety Scanning are to anticipate safety issues at an early stage in concept development, including both positive and negative effects on safety; to prioritise changes for more detailed safety assessment studies and to integrate safety into the decision making process. Herewith it supports

- Safety considerations in preparation for later phases of the safety assessment process;
- Reduction of project risks regarding safety and preparation of a sound safety plan;
- A safety trade-off in conjunction with cost-benefit analyses.

2.2 Assessment of Human Reliability

2.2.1 General Approach of Human Reliability Assessment

A number of different methods from different technical domains such as nuclear engineering, chemical industry, aviation, air traffic and rail traffic exist for the evaluation of human reliability. Especially in recent years, a mind shift has occurred in that newer methods have been established to better model human behaviour in complex situations.

The general approach of all procedures can be summarized in Picture 2. Every method first models human behaviour in a given safety-relevant scenario. Human behaviour is regarded as being triggered by the context and the factors influencing the performance (Performance Shaping Factors - PSF). Scenario and context are the prerequisites that the person has to manage with his behaviour. To properly cope with it, the people’s information processing capacities are essential to transform the demands into appropriate actions.

A central role for a correct assessment is the correct evaluation of the Human information processing. The question is how human information processing can be modelled in order to come to a correct prediction of the behaviour. Two assessment approaches can be distinguished:

- Task-related: Will an Operator properly perform the task required for coping with a particular scenario?
- Situation-related: Which actions will an operator take in a particular scenario given specific contextual conditions?

Figure 2: Principal Approach for the Assessment of Human Reliability

Figure-3 outlines the difference of these assessment approaches in how they proceed during a safety case (see Dolezal, 2011 for details).

The task-based perspective assumes that a task being evaluated in a scenario is the only task the person needs to accomplish and that the person seeks to deal with this task as good as possible. Other system requirements and tasks, which are present
at the same time window (e.g., parallel feedback of information; collection of information, required parallel operating activities), are merely considered as influencing factors in a task related approach - at best. As a result, the corrective actions are exclusively concentrated on the safe design of the task being evaluated in a scenario. This logic works very well in simple systems and creates safety-effective remedial action in such systems.

For complex systems, however, the situational perspective needs to be taken, as a task-based perspective may have a negative impact on the net safety performance. A typical example is a situation with two competing tasks to accomplish. The operator needs to decide which task to prefer and consequently needs to drop the other task. A task-related approach would not be able to correctly assess this situation.

2.2.2 Task-related Methods

Task-related methods appear - in comparison with situation-related methods - comparatively simple. They are usually based on simple methods of analysis of tasks, such as a human-machine systems analysis (VDI-4006, 1999; Bubb, 1993), GOMS analysis (GOMS for Goals, Operations, Methods, Selection Rules, Kieras & Polson, 1985), time course analysis (e.g. NUREG-6208, 1994) or hierarchical task analysis (Kirwan & Ainsworth, 1992).

The number of methods for the quantitative prediction of human reliability is huge. An overview of different methods is provided elsewhere (e.g. VDI 4006 and Swain, 1989). However, the following methods are established across a range of different technical domains:

- THERP (Technique of Human Error Rate Prediction; Swain, 1983) - in simplified mode known as ASEP (Accident Sequence Evaluation Program; Swain, 1987)
- SLIM (Success Likelihood Index Methodology; Embrey, 1984)
- HEART (Human Error Assessment and Reduction Technique; Williams, 1988)

The method THERP is based on a detailed task analysis with specific instructions to investigate particular aspects of the task such as dependencies between tasks and people. It can be recognized as a reliable method to depict and assess a given ergonomic solution. For the subsequent determination of the error probabilities, the method provides an extensive table with ergonomic characteristics and error probabilities. The SLIM method is expert-based and relies on subjective estimates (Edwards et al., 1977). In this process, the experts have to judge subjectively on the reliability of human actions on the basis of a number of factors (PSF - Performance Shaping Factors). The HEART method provides only a simple reference list of tasks for the selection of an appropriate assessment, which can be modified by subjective assessments of factors.

Overall, one can say that THERP is a suitable method, if specific ergonomic issues need to be assessed. HEART and SLIM on the contrast, are based on a simple task analysis and expert estimates with strong uncertainties in the assessment. These methods provide only planning information to a limited degree for corrective actions (OECD, 2000 and 2002).

2.2.3 Situation-related Methods

Task-oriented methods neglect the interpretation and decision-making behaviour of humans. As mentioned above, this will lead to neglecting important contributions to the overall system reliability. The more a system depends on several human acts, or the more complex decision situations are (e.g., due to complex automation or organization), the more those decisions are relevant to the safety performance of a system.

Compared to the task-related processes, which uses focuses on the specific task necessary to be accomplished in a scenario, situation-related methods describe systematically the entire situation and examine possible errors in the context of the scenario, including a situational analysis and the analysis of all possible procedures in that situation (Cooper et al, 1996). The context is defined as the set of all scenarios in which a person must act to achieve safety. Situation-related methods need to model the human information processing accurately in order to come to an assessment of the situation. Methods developments are:

- ATHEANA (A Technique for Human Error Analysis, NUREG-1624, 2000)
- CAHR (Connectionism Assessment of Human Reliability; Straeter 1997)
In contrast to the task-oriented methods these are often referred to as the second generation HRA methods (2nd generation HRA) in literature. According to Straeter (2005) there are situational factors that affect the application of methods. Second generation HRA methods should be made mandatory if one of the following conditions exist in the scenario to assess:

**Regarding the timing of the situation:**
- Suddenness of the onset of the occurrence or fault scenario
- Gradual process development
- Long uneventful phase fault before admission
- Misleading or faulty sequence of alarms / signaling (priming)
- Positive experience with the system (over-reliance)
- Shortly before the completion of a task (home base syndrome)

**Regarding the system states in the situation:**
- Additional, upcoming tasks in the system (e.g. preference of economic aspects rather than safety)
- Multiple technical faults or problems
- Dominant disorder and additional disturbance, partial failure of the system
- Lack of transparency of dependencies in the system (chaining, meshing)
- Ambivalent symptoms that have multiple conclusions
- Masking of relevant information through a combination of system messages
- Deferred (not visible) errors in the system
- Additional operating costs, normal system operation is not available

Situational factors describe the load which an operator (or group of operators) need to cope with during a scenario. Their information processing capabilities describe the coping capabilities they have to deal with the situation. Given such influences, task-related assessments lead to a misinterpretation of the human reliability.

Different second generation HRA methods have different concepts to represent the loading situational factors. As an example, ATHEANA introduces the concept of the Error Forcing Context (PSFs + Plant Conditions) while MERMOS uses the concept of CICA (Important Configuration of Emergency Operation or, Configuration Importante de la Conduite Accidentelle). Also CREAM, CAHR and MERMOS have different concepts. Table 1 outlines the different concepts used in various second generation HRA methods.

However the differences in the situational factors used in second generation HRA methods are differences in terminology rather than differences in content of the concepts for modelling the load. The same holds for the information processing capabilities to describe the coping of an operator.

<table>
<thead>
<tr>
<th>Concept Method</th>
<th>Load</th>
<th>Cope</th>
<th>Quantification Base</th>
<th>Mathematical Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATHEANA</td>
<td>Error Forcing Context (PSFs + Plant Conditions)</td>
<td>Expert Judgment</td>
<td>Expert Calibration</td>
<td></td>
</tr>
<tr>
<td>MERMOS</td>
<td>CICA</td>
<td>Expert Judgment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CREAM</td>
<td>Cognitive Condition</td>
<td>THERP Database</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAHR</td>
<td>Cognitive Coupling &amp; Barrier Interaction</td>
<td>Real Operation Data</td>
<td>Calibration</td>
<td></td>
</tr>
<tr>
<td>CESA</td>
<td>Applicable Actions &amp; Impairments &amp; adverse conditions</td>
<td>Cognitive tendencies</td>
<td>Bayesian Methods</td>
<td></td>
</tr>
</tbody>
</table>

2.2.4 Data for the evaluation of human reliability

Essential for a valid proof of safety in addition to suitable methods are the underlying data for evaluation. Data generally relate to the past, but are forward-looking. The fact that data can be used for prediction is, in general, the causal chain, which is
shown in Picture 4. The general causal chain of human behaviour is identical in the causal analysis (the retrospective assessment of human actions) as well as in the prediction of human error (the prospective evaluation of human actions).

![Figure 4: Retrospective and prospective evaluation of human actions in a) the causal chain of an accident b) the process of prediction in safety assessments](image)

From the picture one can infer that event data and simulator data from different technical domains can be used on the basis of this argument for the evaluation of human reliability. Event data provides in terms of data quality, the so-called external validity (that is the accurate description of reality) by revealing those factors relevant to reliability in reality. Simulator data provides the data quality (the so-called statistical reliability), which represents the exactness of the data to determine average values and uncertainty bands.

The CAHR method allows supporting HRA with appropriate data in order to develop risk matrices and reliability figures in a unified approach (Ghamdi & Straeter, 2010; VDI 4006 Sheet 3).

2.3 Risk Management

Improving safety means both improving the operational system regarding safety performance and also improving the methods to define, identify, understand and assess safety issues. The former one may be called safety performance of the system while the latter one may be called safety understanding. Risk Management needs to consider both paths of learning as every safety assessment will reveal open issues of the assessment methods applied. However, in practice, safety approaches focus too much on the safety performance aspect and not the safety understanding.

2.3.1 Mitigation related Risk Management

Mitigation related Risk Management is depending fully on the HRA method applied to the assessment. If the method allows depicting ergonomic problems, the application of the method will definitely reveal ergonomic factors of improvement. If the method allows representing organisational aspects, mitigation might fall into organisational procedures. However, still the higher level mitigations are hardly represented in the HRA methods though these are the ones most critical for safety (see Reason, 1997; Hollnagel et al., 2005).

Many different methods are proposed to assess human reliability and to implement safety management systems. All approaches have their advantages and disadvantages but unfortunately these are often not discussed openly. On the other hand all claim to assess safety somehow. The questions are: Are they different? Are some better than other? Is the set complete?

A systematic decomposition of the capabilities of methods according to the Picture 5 would help to better assess commonalities and differences of HRA methods. The logic would work as follows: Safety starts with the construct of safety. We cannot observe safety directly but build models of what we think is safe. Several of such models were developed. Amongst them are safety maturity ratings, safety principles, safety indicators, safety culture, but also safety assessments in safety cases or probabilistic safety assessments (PSA). Each model reflects certain parts of the construct safety. As the picture also

![Figure 5: The general overall systematic assumed for assuring safety](image)
indicates it is obvious that neither of the models can be considered as complete. A benchmarking regarding the scientific basis of the safety models, measures and measurements would allow a good comparison of the methods. However, most of the times benchmarking only concentrates on measurements not on the models.

2.3.2 Safety method management

Woods (2003) described in his paper the classic patterns for the Challenger disaster, which also can be seen in other accidents and research results as follows:

- Drift toward failure as defences erode in the face of production pressure.
- An organization that takes past success as a reason for confidence instead of investing in anticipating the changing potential for failure.
- Fragmented problem solving process that clouds the big picture.
- Failure to revise assessments as new evidence accumulates.
- Breakdowns at the boundaries of organizational units that impedes communication and coordination.

These classic patterns could easily serve as indicators for being on a bad safety path. The patterns described by Woods can also be observed in the providers of safety culture methods. That means the process of providing safety methods is likely to be not inherently safe.

However, from the safety point of view, the entire community (from licensee to regulator to the scientific community) is responsible to achieve, assure and improve safety. An approach for a SMMS (Safety Method Management System) could be a solution that takes the principles of SMS and applies them in the management of safety issues and in the development of the regulatory framework throughout the development process of safety methods (Picture 6). A SMMS would serve as a continuous improvement-process to assure that safety methods are dealing with the right safety issues.

An initial approach towards a SMMS was undertaken by Everdij & Balk (2008). They developed a safety method maturity rating in order to assess how fit for purpose safety methods are for a particular change in question in Air Traffic Management.

Outlook

This article outlined the state of discussion on the assessment of human reliability. The importance of the proper scope of consideration of a safety case has been discussed and different methods for assessment of human reliability have been addressed. Regarding the task-related human reliability assessment processes, the method THERP is still the best analysis for the assessment of ergonomic flaws. Situational methods allow a better analysis of the cognitive aspects of human reliability and allow describing in particular the positive contribution of humans to the overall reliability. This property of the methods is of particular importance for the long term reliability of complex systems, because a purely task-based reliability analysis leads to a deterioration of the robustness of the system to respond appropriately to unexpected situations.

References


