EXPERIMENTS IN THE HALDEN HUMAN-MACHINE LABORATORY (HAMMLAB) INVESTIGATING HUMAN RELIABILITY ISSUES.

G. Andresen
P. Øivind Braarud
A. Bye
F. Øwre

Institute for Energy Technology, OECD Halden Reactor Project
P.O.Box 173, N-1751 Halden, Norway

ABSTRACT: The OECD Halden Reactor Project aims at playing a key role in collecting data that can support the conceptual basis and improve the quantification process of so-called 2nd generation Human Reliability Analysis (HRA) methods. The Halden Project will use its laboratory facility, the Halden Human-Machine Laboratory (HAMMLAB), in this research. After presenting key characteristics of 2nd generation HRA methods, potential research issues and methodological challenges are discussed. Two examples of ongoing research are then presented. In the procedure automation experiment, the relationship between procedure automation, situation awareness and human performance are studied. Here, traditional statistical analysis is supplemented by more qualitatively oriented methods. In a series of experiments on task-complexity, the impact of various plant conditions on human performance is investigated. The plant conditions create difficulties that reflect the three dimensions of task-complexity: masking, information load and time pressure. So far, the three-factor structure of the task-complexity construct has been supported by the data. Further, the results indicate that task-complexity has some predictive power. The two studies show how the impact of the context on performance can be investigated in experiments, and illustrate what kind of HRA-data can be collected in HAMMLAB.

1 INTRODUCTION

Because of the gaps in the knowledge of human performance, the modelling of human actions in probabilistic risk/safety-assessment (PRA/PSA) is associated with too much uncertainty [1]. These uncertainties would not have been of any concern if the human's role in failure events were negligible. However, accidents, incidents reports, and human factors research, within and outside the nuclear domain, tell us differently: human actions are indeed a significant part to the risk picture. Thus, the motivation for conducting research on Human Reliability is to acquire knowledge on how complex human-machine systems should be designed in order to minimise the likelihood of human failures and to increase the precision of the quantitative estimates of human error in Human Reliability Analysis (HRA) methods as part of PRA.

Significant progress has been made in the HRA field during the last ten years. An array of so-called 2nd generation methods[1] is emerging as alternatives to the 1st generation of HRA methods[2]. However, the 2nd generation methods are still immature in the sense that their conceptual basis only to a limited extent has been tested empirically and that there is little data underpinning the error probabilities they postulate. The OECD Halden Reactor Project (henceforth referred to as the Halden Project) aims at playing a key role in exploring the conceptual basis of the 2nd generation HRA as well as providing data that improve the precision of the quantification of both 1st and 2nd generation methods.

This paper provides an introduction to the Halden Project's approach for investigating human reliability issues. Using ATHEANA as an example of a 2nd generation method, we begin the paper by describing some of ATHEANA's key characteristics and identify issues that can be investigated in the Halden Human-Machine Laboratory (HAMMLAB). Then we present some of the methodological challenges we are facing. We end the paper by presenting two ongoing studies inspired by recent thinking in HRA. These studies illustrate the kind of topics that will be investigated by the Halden Project and the types of HRA-relevant data that can be collected in HAMMLAB.

2 CENTRAL CONCEPTS AND RESEARCH ISSUES IN 2ND GENERATION HRA

We have identified two characteristics of 2nd generation HRA that will be addressed in the Halden Project's research programme. The first characteristic is the central position of the context in predicting and explaining human failures: failures are triggered by an 'error-forcing context'. The second characteristic is that the operator's cognition must be considered in order to identify what failures the context may trigger; i.e., the context may trigger 'error-mechanisms' if the human's information processing capacity is exceeded or if the strategies and methods humans apply in solving problems does not fit to the requirements of the task. In the following paragraphs, we will provide a short description of these characteristics, before we look at some possible issues to investigate in HAMMLAB.

It is beyond the scope of this paper to review existing HRA methods (See [6] and [7] for reviews). Instead, we will use the ATHEANA method as an example. The US Nuclear Regulatory Commission supported the development of both ATHEANA and THERP. The latter is the most frequently used 1st generation method.

2.1 Error-forcing context: PSFs and plant conditions

According to ATHEANA, the error-forcing context comprises two components: the Performance Shaping Factors (PSFs), and the plant conditions. As for most HRA methods, the PSFs are broad categories such as 'procedures' and 'teamwork', which the HRA analyst should assess in terms of their adequacy. If the PSF is sub-optimal, then, multiplying the nominal failure-rate with some constant increases the estimated likelihood of human failure. So far, ATHEANA does not provide its own approach calculating the impact of PSFs (but see [8]) - instead, it borrows its methodology from another HRA method called HEART [9].

What *is* unique about ATHEANA is the way by which plant conditions are incorporated into the analysis. Starting off from a 'base case scenario', i.e., a scenario described in the PRA, alternative 'deviation scenarios' are developed. The deviation scenarios include additional events that increase the likelihood of certain error-mechanisms being triggered. ATHEANA provides a comprehensive list of possible plant conditions that may have this effect; multiple failures, slowly progressing failures etc.

The difficulties created by the plant conditions may be addressed through PSFs such as ‘Available time’ and ‘Complexity’. However, ATHEANA argues convincingly that plant conditions should be analysed separately.

2.2 Cognition and error-mechanisms

The 2nd generation HRA methods usually frames the operators cognition by means of a cognitive model. The model used by ATHEANA consists of four cognitive functions (Detection, Situation assessment, Response planning, and Response implementation), and two components representing the operators’ current understanding of the situation (situation model) and the knowledge operators use in developing their current understanding (knowledge/mental model). Other HRA methods may include other components in their models, but, in general, all models are quite simple. Of course, this does not mean that cognition during disturbances is simple; rather, it reflects a compromise between having a model being sufficiently detailed for making reasonably precise predictions and a method still being manageable for the practitioner.

ATHEANA provides additional information about cognitive factors that may affect the quality of the cognitive functions. Three classes of factors are defined: knowledge factors, processing factors, and strategic factors. Failures in the cognitive functions may arise when these factors are inadequate in some way. For example, the operator may not have the required knowledge (knowledge factor), have too many tasks to handle simultaneously (processing factor), or underestimate the risk (strategic factor).

By combining the cognitive model with the cognitive factors, ATHEANA supersedes its predecessor THERP in providing a more detailed picture of the operator’s cognition and the possible ways in which cognition may fail.

2.3 Possible issues to investigate in HAMMLAB

The cognitive model used in ATHEANA, and the error-mechanisms derived from it, are based on several decades of research on cognitive psychology and human factors. There is reasonable consensus about the existence of the psychological phenomena; however, the applicability of the concepts in HRA has yet to be validated. By using the conceptual framework of 2nd generation HRA methods as a theoretical basis for the experimental work, the research in HAMMLAB will gradually reveal strengths and weaknesses of the models, and, hopefully, indicate how the models can be improved.

Data reflecting the different components of the models is often collected in HAMMLAB experiments. In respect to the cognitive functions, it is common to analyse the operator’s performance in terms of categories similar to that used by ATHEANA (e.g., ‘Detections’ of important parameters). One way to utilise such data is to investigate if the collected data matches the categories proposed by the HRA method. Previous research on the CREAM method has shown that this is not always the case [10].

In respect to the error-mechanisms, some are less well understood than others. Because new technology changes the way operators work, opportunities for new kinds of errors are created [11]. It should be possible to predict how new technology influences performance through the cognitive model. If we cannot predict a certain error-mechanism, or if a predicted error-mechanism cannot be found in the data, then the cognitive model may have to be revised.

Although the cognitive models will be investigated, the research will primarily address the impact of various PSFs. Decisions about what factors to study will be made in cooperation with other research communities. Through the development of the Halden Project's research programme, the conduction of workshops and active participation in conferences, the Halden Project has a solid basis for identifying which factors are most safety-significant. Of course, also the context provided by HAMMLAB makes some PSFs more suitable to study than others. The system interface, computerised support systems, procedures and teamwork - PSFs at the “sharp end” - are all factors that can be manipulated in the laboratory and are thus good candidates.

3 HOW WILL HRA-DATA BE COLLECTED IN HAMMLAB?

The transition from 1st to 2nd generation HRA has not solved what must be considered a major problem in the field: the lack of empirical data supporting the quantification process. Through the Halden Project's empirical approach, we will be one supplier of the much-needed data. But how can we do this?

Most of the research conducted in HAMMLAB is of an experimental nature. That is, the studies typically involve a systematic manipulation of some aspect of the control-room (e.g., new vs. old alarm system) whose effect on operator performance is assessed by means of measurements and statistical analysis.

Also studies of PSFs may follow an experimental approach. By defining the PSF as the independent variable, the causal relationship between the PSF and performance can be analysed in a systematic manner. To identify what aspects of the PSF the practitioner should focus on, various dimensions of the PSF can be studied individually. For example, if performance is unaltered working from hard-copy or computerised procedures, this is not a characteristic the HRA analyst should be concerned about. Further, by including various levels of the PSF in the design, it is possible to identify “how much” of a PSF is needed (e.g., how much a procedure is automated) before significant changes in performance can be observed. After the PSF of interest has been properly addressed, a possible next step is to apply a factorial design in which the combined effect of several PSFs is studied. Here, it is also possible to include ‘scenario’ as an independent variable identifying to what operational tasks the impact of the PSF can be generalised.

By using experimental methodology, the Halden Project is close to providing the HRA-data needed. There are two weaknesses, though. First, the statistical analysis used in experiments does not fit directly with the quantification used in HRA. From the statistical analysis, we can say whether conditions are different from one another, but quantifying how much a PSF influences performance does not follow from that analysis. Effect-size analysis, a method that tells how much of the variance in the dependent variable can be explained by the independent variable, may be part of the answer. Structural equation modelling and Bayesian statistics may also give useful input. We will not go into any details about these methods here, suffice to say that the Halden Project is investigating whether these methods can transform experimental results into data that can be utilised in PSF quantification.

Second, although the statistical analysis allows us to conclude, with some certainty, if a condition is influencing performance, we do not know if this relationship reflects the underlying cognitive processes proposed by the HRA method. For this reason, we must collect data that captures more details about the cognitive processes of the operators during problems solving. One way to do this is to utilise various cognitive task analysis. These methods rely more on subjective judgement and interpretation than the experimental measurement techniques, and are, therefore, more susceptive to biases. Nevertheless, the Halden Project has shown in previous research that valuable data can be obtained with these kinds of methods - data that simply cannot be captured by measurements that yields the average or the outcome of experimental trials.

4 EXAMPLES OF ONGOING HRA-RELEVANT RESEARCH

4.4 The impact of procedure automation on human performance

The procedure automation experiment was conducted between December 2002 and March 2003 in HAMMLAB. The main objective of the study was to investigate the influence of procedure automation on human performance. Three configurations of a Computerised Procedure System (CPS) were compared: low automation (the procedure steps were executed manually from within the CPS), medium automation (portions of the procedure were executed automatically), and high automation (the whole procedure was executed automatically).

By allocating the procedure-execution to the CPS, the operator was removed from the control-loop. Research has shown that performance may suffer under such conditions (e.g., [12]). One explanation for this is that when the operator is not actively engaged in the task performance, his or her understanding of the situation will diminish, making it difficult to know when and how manual intervention is required. This problem has been called ‘out-of-the-loop performance problems’ [12] and is a recurring human factors issue in control room settings were the operator acts as a system supervisor.

In respect to the cognitive model proposed by ATHEANA, the procedure automation affects all functions of the operator’s cognition. However, since the operator is supervising the CPS, situation assessment was of particular interest. Further, because the operator could freely decide whether to follow the execution of the CPS or perform other tasks, we were also interested in finding out whether the procedure automation changed the operator’s behaviour and ultimately the crew’s way of working together.

The researchers conception about how automation might influence cognition and behaviour directed the choice of scenarios and measurements. The scenarios were designed such that minor disturbances introduced during the automated procedure execution were related to more serious disturbances occurring at a later stage in the scenario. It was hypothesised that the operators’ degree of involvement in the procedure execution, while the minor deviations appeared, would influence their ability to diagnose the serious disturbance. To assess this effect, measures on situation awareness and response time were applied. Also observations, questionnaires and interview data were collected. These data were collected to get more detailed information about how the operators’ behaviour was changed, how they used the CPS and how they experienced using it.

There were several interesting findings. The statistical analysis of the situation awareness measures revealed that the operators were not affected by the automation. However, none of the crews preferred the highest level of automation. The observations and interviews gave a reasonable explanation for these contradictory results. During the trials when the CPS executed the whole procedure automatically (highest level of automation), the operators monitored the CPS more or less continuously. This was a very demanding task, however, it enabled them to handle the disturbances.

Although the comparisons made between conditions did not reveal any significant differences, the crews varied in how quickly the disturbances were handled. The distribution ranged from very quick responses to responses performed too late. We wanted to investigate what could explain this variation, and were particularly interested in finding out whether "significant delays" could be associated with particular PSFs. The analysis did not only focus on the effect of the procedure automation (the Procedure PSF), but any element of the context that seemed to affect the crew's response time. The analysis proceeded in steps and was performed by a team consisting of two human factors researchers and one operation expert. The main findings were that delays could be associated with:

- Procedures: the crew had problems with freezing the automation, and knowing whether the CPS could have anything to do with the disturbance;
- Teamwork: the crew were not communicating efficiently or the shift supervisor did not give clear orders;
- Interface: The crew had problems with interpreting some of the process displays;
- Experience/Training: The crew was uncertain about features of the HSI and how to deal with the event;
- Workstyle: the crew prioritised getting an overview of the plant status instead of trying to diagnose the symptoms.

To sum up, the procedure automation experiment is an example of a study where the effect of new technology on performance can be analysed by means of the conceptual framework of a 2nd generation HRA method. Building on research conducted outside the nuclear domain, it was investigated if certain characteristics of the Procedure PSF - procedure automation - might degrade the situation awareness of the operators to such an extent that error-mechanisms were triggered. Different methods were used to address the topic. Through the experimental approach, the impact of the automation was measured and tested statistically. Through more qualitatively oriented methods, the variation in performance between trials was analysed. In experiments, this variation is treated as "noise" and is not of interest. However, for HRA, understanding what elements of the context are associated with performance variation can be valuable data. More details on how the procedure automation experiment was conducted to obtain HRA-relevant data can be found in [13].

4.5 The impact of task complexity on human performance

Over several years, the Halden Project has investigated what characteristics of a scenario that makes it complex and, thus, difficult for operators to handle. Two approaches for identifying such characteristics can be found in the literature: a systems approach and a task-oriented approach. The systems approach looks upon the human-machine system as a unit that can be described and analysed by means of systems theory. The task-oriented approach, on the other hand, takes the perspective of the operator, investigating the difficulties operators encounter while interacting with the system to achieve their objectives. The research at the Halden Project is task-oriented.

A central objective of the work done so far has been to identify the underlying dimensions of ‘task-complexity’ [14]. Three dimensions have been found through factor analysis of questionnaire data: ‘Masking’, ‘Information load’, and ‘Time pressure’. Masking refers to the kind of difficulties operators experience while interpreting plant conditions that are ambiguous or unclear (e.g., several faults occurring simultaneously or indication failures). Information load concerns difficulties related to the amount of information (relevant and irrelevant) operators must handle to achieve their goals. Time pressure refers to difficulties related to the time available for performing tasks.

A series of experiments addressing the effect of task complexity on risk-related human events are currently being conducted in HAMMLAB. One experiment that was completed this year investigated how task complexity affects crew performance during various loss-of-coolant (LOCA) scenarios. The preliminary results show that the crews’ performance scores were significantly lower during high complexity scenarios, compared to medium and low complexity scenarios. Also, significant differences in response implementation and teamwork activities were observed between high and low complexity scenarios. Thus, the results indicate that task complexity have some predictive power and that the three-factor structure, which the complexity of the scenarios were designed from, is supported.

Building upon the experiment just completed, a new task complexity experiment will be conducted by the end of 2003. Using ATHEANA terminology, the experimental design will consist of ‘deviation scenarios’ constructed from a ‘base case’ event sequence in the plant PRA. The base case scenario represents low or nominal task complexity. The deviation scenarios include additional plant failures that increase the task complexity. The additional failures should create difficulties reflecting the three complexity dimensions. Data on how the additional failures affect the crews’ ability detecting important pieces of information, assessing the situation, and response planning will be collected. In addition, the crew responses as defined in the PRA will be studied.

Just as for the procedure automation study, the task-complexity experiments address the effect of the context on human performance. Interestingly, though, task-complexity does not investigate a PSF (e.g., procedure) but looks at the plant conditions. ATHEANA claims that this is a central component of the error-forcing context. It is therefore important that Halden has begun collecting data on this issue.

The results from the first experiment showed that task-complexity can predict crew performance. This is an important finding. With the second experiment, the aim is to tie the underlying dimensions of task-complexity closely to specific process events. This will, hopefully, increase the precision of the predictions and make the data more useful for HRA.

5 CONCLUSION

The Halden Project will in the coming years conduct research aimed at collecting HRA-relevant data. This research will primarily address the impact of the context on human reliability, but it will also, directly or indirectly, test the conceptual basis of the 2nd generation HRA methods. In this effort, experimental methodology will be supplemented by more qualitatively oriented methods. The reason for this is that various methods provide different kinds of data, and that all methods have their strong and weak points. An eclectic approach is most likely to utilise the full potential of HAMMLAB.

6 REFERENCES

[1] E. Lois, N. Siu, M. Cunningham, E. Thornsbury, and J. Cai. “Narrowing the Uncertainties in Human Reliability Analysis,” Nuclear Safety Research Conference, Washington DC, October 28-30, 2002.

[2] USNRC, “Technical basis and implementation Guidelines for a technique for human event analysis (ATHEANA),” U.S. Nuclear Regulatory Commission, Washington, DC, NUREG-1624, 1998.

[3] E. Hollnagel, *Cognitive Reliability and Error Analysis Method CREAM*. New York: Elsevier, 1998.

[4] USNRC, ”SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgement,” U.S. Nuclear Regulatory Commission, Washington, DC, NUREG/CR-3518, 1984.

[5] A.D. Swain and H.E. Guttman, “Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications,” U.S. Nuclear Regulatory Commission, Washington, DC, NUREG/CR-1278, 1983.

[6] P. Pyy, “Human reliability analysis methods for probabilistic safety assessment,” Ph.D. dissertation, Dept. Automation, Technical Research Centre of Finland, Espoo, Finland, 2000.

[7] J. C. Byers, D. I. Gertman, S. G. Hill, C. D. Blackman, C. D. Gentillon, B. P. Hallbert, and L. N. Haney, “Revision of the 1994 ASP HRA methodology (draft),” Idaho National Engineering and Environmental Laboratory, Bechtel, Idaho, INEEL/EXT-99-00041, Jan., 1999.

[8] J. Forester, D. Bley, S. Cooper, A. Kolaczkowski, N. Siu, E. Thornsbury and J. Wreathall, “Quantification and treatment of uncertainty in human reliability analysis based on ATHEANA,” International Conference on Probabilistic Safety Assessment and Management*,* San Juan, Poerto Rico, June 23-28, 2002.

[9] J. C. Williams, “A Data-based Method for Assessing and Reducing Human Error to Improve Operational Performance,” in Proceedings of the 4th IEEE Conference on Human Factors and Power Plants, 1988.

[10] S. Collier and G. Andresen, “A Simulator Study of CREAM to Predict Cognitive Errors,” presented at the International Workshop on Building the New HRA: Errors of Commission from Research to Application, Washington, DC, 2001.

[11] L. Bainbridge, “Ironies of automation,” in J. Rasmussen, K.D. Duncan & J. Leplat (Eds.): *New Technology and Human Error*. John Wiley & Sons, 1987.

[12] M. R. Endsley, and D. B. Kaber, “Level of automation effects on performance, situation awareness and workload in a dynamic control task,” *Ergonomics*, vol. 42, 462-492, 1999.

[13] G. Andresen, S. Collier and S. Nilsen, “Experimental Studies of Potential Safety-Significant Factors in the Application of Computer-Based Procedure Systems,” in Proceedings of the 7th IEEE Conference on Human Factors and Nuclear Power Plants, Phoenix, Arizona, 2002.

[14] P.Ø. Braarud, “Simulator experiments as empirical basis for performing shaping factors in HRA,” 6th International Conference on Probabilistic Safety Assessment and Management, San Juan, Poerto Rico, June 23-28, 2002.

(Footnotes)

[1] For example ATHEANA (A Technique for Human Event ANAlysis) [2] and CREAM (Contextual Reliability and Error Analysis Method) [3].

[2] For example SLIM (Succes-Likelihood Index Method) [4] and THERP (Technique for Human Error Rate Prediction) [5].