Cognitive Engineering Laboratory

The Ecological Interface Design Experiment (2005)

Gyrd Skraaning Jr., Nathan Lau, Robin Welch, Christer Nihlwing, Gisle Andresen,Liv Hanne Brevig, Øystein Veland, Greg A. Jamieson, Catherine Burns and Jordanna Kwok

CEL 07-02

Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto
5 King's College Road, Toronto, Ontario, Canada M5S 3G8
URL: www.mie.utoronto.ca/labs/cel/

Cognitive Engineering Laboratory

Directors: Kim J. Vicente, Ph.D., P. Eng.
Greg A. Jamieson, Ph D., P. Eng.

The Cognitive Engineering Laboratory (CEL) at the University of Toronto (U of T) is located in the Department of Mechanical & Industrial Engineering, and is one of three laboratories that comprise the Human Factors Research Group. CEL was founded in 1992 and is primarily concerned with conducting basic and applied research on how to introduce information technology into complex work environments.

Current CEL Research Topics

CEL has been funded by Atomic Energy Control Board of Canada, AECL Research, Alias|Wavefront, Asea Brown Boveri Corporate Research - Heidelberg, Canadian Foundation for Innovation, Defence Research & Development Canada (formerly Defense and Civil Institute for Environmental Medicine), Honeywell Technology Center, IBM, Japan Atomic Energy Research Institute, Microsoft Corporation, Natural Sciences and Engineering Research Council of Canada, Nortel Networks, Nova Chemicals, Westinghouse Science & Technology Center, and Wright-Patterson Air Force Base. CEL also has collaborations and close contacts with the Mitsubishi Heavy Industries and Toshiba Nuclear Energy Laboratory. Recent CEL projects include:

- Developing advanced human-computer interfaces for the petrochemical and nuclear industries to enhance plant safety and productivity.

- Understanding control strategy differences between people of various levels of expertise within the context of process control systems.

- Developing safer and more efficient interfaces for computer-based medical devices.

- Creating novel measures of human performance and adaptation that can be used in experimentation with interactive, real-time, dynamic systems.

- Investigating human-machine system coordination from a dynamical systems perspective.

CEL Technical Reports

For more information about CEL, CEL technical reports, or graduate school at the University of Toronto, please contact Dr. Kim J. Vicente or Dr. Greg A. Jamieson at the address printed on the front of this technical report.

HWR-833

OECD HALDEN REACTOR PROJECT

THE ECOLOGICAL INTERFACE DESIGN EXPERIMENT (2005)

by

Gyrd Skraaning jr.Jr, Nathan Lau, Robin Welch, Christer Nihlwing, Gisle Andresen, Liv Hanne Brevig, Øystein Veland, OECD Halden Reactor Project; Greg A. Jamieson, University of Toronto; Catherine Burns, Jordanna Kwok, University of Waterloo

March 2007

FOREWORD

The experimental operation of the Halden Boiling Water Reactor and associated research programmes are sponsored through an international agreement by

- the Institutt for energiteknikk, Norway,
- the Belgian Nuclear Research Centre SCK•CEN, acting also on behalf of other public or private organisations in Belgium,
- the Risø National Laboratory in Denmark,
- the Finnish Ministry of Trade and Industry,
- the Electricité de France,
- the Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) mbH, representing a German group of companies working in agreement with the German Federal Ministry for Economics and Technology,
- the Japan Atomic Energy Agency,
- the Korean Atomic Energy Research Institute,
- the Spanish Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, representing a group of national and industry organisations in Spain,
- the Swedish Nuclear Power Inspectorate,
- the Swiss Federal Nuclear Safety Inspectorate, representing also the Swiss nuclear utilities and the Paul Scherrer Institute,
- the BNFL, representing a group of nuclear research and industry organisations in the United Kingdom, and
- the United States Nuclear Regulatory Commission,

and as associated parties:

- the Central Research Institute of Electric Power Industry (CRIEPI), representing a group of nuclear research and industry organisations in Japan,
- the Czech Nuclear Research Institute,
- the French Institut de Radioprotection et de Sûreté Nucléaire (IRSN),
- the Hungarian Academy of Sciences, KFKI Atomic Energy Research Institute,
- the Russian Research Centre “Kurchatov Institute” and “TVEL” Co.,
- the Slovakian VUJE - Nuclear Power Plant Research Institute,

and from USA:

- the Westinghouse Electric Power Company (WEC-LLC),
- the Electric Power Research Institute (EPRI), and
- the Global Nuclear Fuel - Americas.

The right to utilise information originating from the research work of the Halden Project is limited to persons and undertakings specifically given this right by one of these Project member organisations.

The present report is part of the series of Halden Work Reports (HWRs) which primarily are addressed to the Halden Programme Group (the technical steering body of the Halden Project) as a basis for its continuous review of the Project's research programmes. The HWR-series includes work reports on the different research items of the Project, reports on the findings from workshops convened under the auspices of the HPG, and other internal reports to the HPG on the findings from the Project's activities to the extent deemed desirable by the HPG or the Project staff.

Recipients are invited to use information contained in this report to the discretion normally applied to research and development programmes periodically reported upon before completion. Recipients are urged to contact the Project for further and more recent information on programme items of special interest.

Institutt for energiteknikk
OECD HALDEN REACTOR PROJECT

Title

The Ecological Interface Design Experiment (2005)

Author:

Gyrd Skraaning Jr, Nathan Lau, Robin Welch, Christer Nihlwing, Gisle Andresen, Liv Hanne Brevig, Øystein Veland, OECD Halden Reactor Project; Greg A. Jamieson, University of Toronto; Catherine Burns, Jordanna Kwok, University of Waterloo

Document ID:	Document location:	
HWR-833		

Keywords:

Ecological Interface Design, Situation Awareness, Process overview, Scenario understanding, Metacognitive accuracy

Abstract:

The objective of the Ecological Interface Design experiment (2005) was to demonstrate the presumed benefits of ecological displays during unanticipated NPP events. The study was performed in HAMMLAB on the HAMBO BWR simulator with six participating crews. Each crew consisted of one reactor operator and one turbine operator. We compared ecological displays to traditional computerized displays in the detection and mitigation phase of within design basis, and beyond design basis scenarios. The ecological displays were implemented only on the turbine side of the process. Therefore, the scenarios and the data analysis focused on the turbine operator in each crew. Even though the ecological displays provided process information according to Ecological Interface Design (EID) principles, a traditional process mimic was integrated and used for intervention and control of the system. The experiment concentrated on how the display types affected the operators' Situation Awareness (SA). A model of SA for nuclear process control was developed, extracting three dimensions of operator cognition in the control room: (a) Process overview, (b) Scenario understanding, and (c) Metacognitive accuracy (degree of realistic self-assessment). The hypothesis was that ecological displays would support the Situation Awareness of NPP operators in beyond design basis events, and during the detection phase of the scenarios. The findings suggest that the ecological displays supported Situation Awareness in the detection phase of beyond design basis scenarios. If the operators were given more training, and the ecological design elements fully supported intervention with the process, it is possible that the benefits of EID would extend to the mitigation phase of beyond design basis scenarios as well.

Issue Date:		Name	Signature	Date
2007-03-02	Prepared by:	Gyrd Skraaning Jr.	Sign	2007-02-23
Confidential grade:	Reviewed by:	Stine Strand	Sign	2007-02-27
HRP Only	Approved by:	Jon Kvalem	Sign	2007-03-02

MAIL	TELEPHONE	TELEFAX		TELEFAX	
P.O. BOX 173	+47 69 21 22 00	Administration	+ 47 69 21 22 01	Safety MTO	+ 47 69 21 24 60
NO-1751 HALDEN		Nuclear Safety	+ 47 69 21 22 01	Reactor Plant	+ 47 69 21 24 70
Norway		Purchasing Office	+ 47 69 21 24 40		

TABLE OF CONTENTS

1. INTRODUCTION

Ecological Interface Design (EID) was developed to support rapid perception and correct interpretation of real-time information in complex socio-technical systems. In the context of nuclear process control, EID is supposed to aid knowledge-based problem solving in unanticipated situations by assisting the operators' formation of an appropriate mental model of the nuclear process, and giving visual support to ensure an effective utilization of this model. In other words, EID provides *higher order information* about the nuclear process and aims to reduce cognitive effort by transforming demanding information processing into perceptual tasks. Higher order information means calculated variables that describe the system state at an abstract level, as well as thermodynamic properties, and physical laws that express the relationships between variables and their meaning in relation to system goals and safety boundaries.

The EID work at the Halden project was initiated by a research survey (Veland, 2004). Ecological display prototypes for the feedwater and steam generator systems of the FRESH PWR simulator were then developed and tested in HAMMLAB (Welch et al., 2005). Based upon the experiences from these projects, we performed a large scale EID experiment on the HAMBO BWR simulator. This experiment is reported here, and was carried out in close collaboration with the universities of Toronto and Waterloo in Canada, which have a leading role in the international research on EID. The research team also included VTT in Finland, who carried out a Contextual Assessment of Systems Usability (CASU) to supplement the statistical data analysis.

1.1 Summary of the experiment set-up

The objective of the experiment was to demonstrate the presumed benefits of Ecological displays during unanticipated NPP events. We used the HAMBO BWR simulator for this purpose, and the data were collected during December 2005 and January 2006. Six licensed NPP operating crews participated in the study. The operators had their daily work at a plant similar to the HAMBO process. Each crew was composed of one Reactor Operator (RO) and one Turbine Operator (TO). However, the EID displays were only developed for the turbine side of the process, and the scenarios therefore introduced disturbances that had to be handled mainly by the TO. Consequently, the data analysis focused on the performance of the TO. All crews were tested under all experimental conditions in six scenarios.

We compared three different display types:

- *Traditional displays* - a computerized version of conventional control room panels corresponding to the current industry standard.
- *Ecological displays* - design solution guided by categorizations of cognitive performance and an analysis of the work domain to determine the system's information requirements.
- *Advanced displays* - Traditional displays with graphical elements made possible by computer technology, such as integrated mini trends.

These display types were tested in three *Within design basis scenarios*, and three *Beyond design basis scenarios*. The Within design basis scenarios included situations that were anticipated by system designers and familiar to operators, while the Beyond design basis scenarios were unanticipated by system designers and unfamiliar to operators.

All of the scenarios had a *Detection phase* and a *Mitigation phase*. In the Detection phase, the main incident was not yet annunciated by the alarm system and the operators had to use process displays to identify the problem. During the Mitigation phase, an alarm revealed the problem that the operators had to solve.

The experiment concentrated on how the display types affected Situation Awareness (SA) under varying operating conditions. A model of SA for process control was developed, extracting three dimensions of operator cognition in the control room: (a) Process overview - the ability to separate signals from noise by detecting and acting upon unexpected changes in the process, (b) Scenario understanding - the ability to diagnose problems correctly and find effective control actions during disturbances, and (c) Metacognitive accuracy - the ability to self assess your performance level while engaged in complex tasks. Measures of SA were developed for each of these dimensions. A measure of Workload was also included in the study.

The general hypothesis was that Ecological displays would support Situation Awareness in Beyond design basis scenarios, and during the Detection phase of the scenarios.

1.2 Nature of the study

The objective of the study was to conduct a first evaluation experiment in a validation process for ecological interface forms in nuclear power plant operations.

This statement places the study in the context of an ongoing research and development program. The aim is to improve the designs by identifying their strengths and weaknesses under representative settings. Knowledge of these performance characteristics will inform redesign of the forms and subsequent evaluation experiments. Thus, the evaluation is formative, as opposed to summative. Stating this objective provides a basis for many decisions to follow in subsequent sections.

Our goals were twofold:

1. To assess the feasibility of controlling a full-scope NPP simulator using an ecological interface.
2. To evaluate the performance of that interface relative to the control performance achieved with a contemporary interface and an advanced interface.

In consideration of our long-term goals and our current resources, several compromises were required.

The anticipated benefits of the EID framework are primarily moderate long-term effects in safety, productivity, and worker health (Vicente, 1999). Such benefits are not easily tested for. A design framework evaluation study can typically only reveal large short-term effects upon introduction of an exemplar interface. For example, a usability evaluation will often assess the learnability of a new interface or task completion time or accuracy on a discrete task. Although we would be happy to see such large short-term benefits to safe and effective operation as a result of an interface intervention, we recognize that the maturity of nuclear power technologies leaves room for only marginal gains in steady-state operation. It is the possibility for repeating such gains many times over at relatively low cost that creates the opportunity for having a substantial impact on the industry. That is to say, a small and long-term effect could be of high practical significance, particularly if the intervention can be readily reproduced.

In light of the above discussion, what is to be considered a successful demonstration of a benefit for EID in this study? Two benchmarks come to mind: First, a small short-term effect would be sufficient to justify further (in particular, longer-term) investigation of the design framework.

Second, a finding of 'no effect' might still justify further investigation if there is supplementary evidence that suggests other avenues for further study.

A note of caution is warranted here. Even if a large short-term effect were observed, further study would be required for several reasons. First, the finding would need to be replicated (i.e., the risk of detecting false effects remains). Second, the robustness of the finding would need to be established (an objective that cannot be met by any single study). Third, the breath and limits of any possible advantages would have to be explored.

In summary, the combination of two types of limitations necessitates a validation approach that relies on replication of evaluation experiments. First, limits in an evaluation study's ability to detect the (anticipated) small long-term beneficial effects of an interface intervention. Second, limits on generalizing any observed effects to general NPP operations.

We further note that the possibility of observing even small effects for alternative interfaces in this study is reduced by the expertise of the participants. NPP operators have deep knowledge of their work domain and years of experience using the contemporary information systems. Moreover, the knowledge and skills possessed by operations teams extends well beyond that which is supported by the operator interface. We would therefore expect that it will be difficult to observe any design-induced performance advantages for alternative interfaces when the baseline condition is highly representative of the state-of-practice. This concern is partially mitigated in the present study because the 'baseline' interface condition, while consistent with the type of interface used in the operators' home plant, does not replicate that interface. This would tend to limit the extent to which the operators can apply their prior knowledge.

Finally, it is important to note that this study does not seek to set up a "straw man" comparison. The traditional interface to be used in this study is a highly evolved basis for comparison against the alternative displays. It is the product of many years of concerted effort by experienced process experts and human factors specialists. While the experimenters believe that the alternative interfaces can yield performance benefits over the traditional interface, the mimic-based displays are by no means expedited to "fall down".

1.3 Scope of the report

The analysis and reporting from the experiment follows three parallel paths:

- A description of the HSI-design prototypes, the design process, and user feedback from operators after working with Traditional, Ecological, and Advanced displays (HWR-847, Welch et al., 2007).
- A documentation of the experimental methodology, human performance measures, hypotheses and the statistical results – this report (HWR-833, Skraaning et al., 2007).
- The results of the Contextual Assessment of Systems Usability for the study (Norros & Salo, 2007).

A complete understanding of the experimental findings requires a combined interpretation of the three reports. A full discussion of the results is therefore beyond the scope of this text. Hence, the discussion in chapter 7 is constrained to initial speculations concerning the statistical findings, and an account for the methodological limitations of the study.

2. METHOD

2.1 Research approach

In his influential article “The earth is round ($p < .05$)”, Jacob Cohen (1994) explains the logical shortcomings of null hypothesis testing, which was the prevailing methodological paradigm in experimental psychology for decades. The American Psychological Association responded to this eye-opener by establishing a committee called the Task Force on Statistical Inference (TFSI). TFSI’s mandate was to clarify controversial methodological issues and provide new guidelines for the application of statistics in psychology (Wilkinson, 1999). The committee concluded that rigid methodological orthodoxies should be avoided, and encouraged researches to give up “practices that institutionalize thoughtless application of statistical methods” (p. 604).

Given Cohen’s convincing arguments and TFSI’s general recommendations, we accept that null hypothesis testing can be a meaningless mechanistic ritual. However, controlled experimentation is still seen as a valid and useful research method. John Stuart Mill (1843) argued that causal inferences are trustworthy if, (a) an effect is present only when the cause is present, (b) an effect is absent when the cause is absent, and (c) both of the these relationships are observed. Alternative interpretations of the covariation between cause and effect can then be ruled out. In other words, threats to valid causal inference can be eliminated by comparing matched situations where particular variables do operate, or do not operate. Experimental manipulation is therefore a powerful technique, even though null hypothesis testing is logically flawed.

To build solid experiments without the comforting formalism of null hypothesis testing can be a challenge. The following approach is taken here:

- The experimental manipulation involves a comparison of matched situations where treatments are present or absent. Classical experimental designs that were originally developed for null hypothesis testing are used to systematize the manipulations and facilitate the statistical data analysis.
- Hypotheses are written in natural language prior to the experiment and concern the anticipated effect of the manipulated variables. The researcher specifies, (a) expected main effects and/or interaction effects, (b) the direction of these effects, and (c) separate predictions for each measurement construct. Hypotheses are justified by theory and practical experience.
- The quality of the dependent variables is given special attention, since poor measurement has been a major weakness in experimental research (Pedhazur & Pedhazur Schmelkin, 1991; Cohen, 1994).
- Statistical analysis is used to evaluate whether the hypotheses are supported by the data. The analysis is not a formal test, but a goal-driven statistical exploration to aid an intelligent interpretation of the data. TFSI emphasized that statistical techniques should never be more complicated than necessary, and promotes the use of graphics to display results and assess assumptions (Wilkinson, 1999).
- Statistical techniques are used to search for unanticipated effects that can contribute to the development of new theories and the formation of hypotheses for later experiments.
- Null hypothesis testing gives the impression that meaningful theoretical generalizations are achievable without the use of inductive logic. According to Cohen (1994), this is an illusion and psychological generalizations from experiments should therefore rely on replication, as in the older sciences.

2.2 Experimental design

We applied a 3x2x2 within-subject design, i.e. a Randomized Block Factorial design (RBF-322). The purpose of *blocking* is to isolate the effect of a nuisance variable, i.e. partition out variance that would otherwise be included in the error term. This is advantageous because a smaller error term improves the F-ratio and thereby increases the statistical power for the analysis of variance. Every block is formed by a set of homogeneous experimental units, e.g. humans that share the same response tendencies. To accomplish this objective, subjects that are similar with respect to the dependent variable can be grouped into a block (matching), or the same subjects can be observed under all treatment combinations (within-subject design). The latter approach was taken here.

Six operating crews (n=6) participated in the study. Each crew consisted of a Reactor Operator (RO) and a Turbine Operator (TO). However, the data analysis mainly addressed the individual performance of the TO, since the Ecological displays - which are of primary interest here - were implemented only on the Turbine Operator workstation. For this reason, the scenarios specifically challenged the TO, even though both operators were engaged in taskwork. The RO was included in the experiment to create a realistic working environment, and because it was difficult to operate the simulated process for one operator alone. We did not facilitate teamwork particularly, but a certain degree of cooperation is necessary when a highly interconnected system is operated by two people. We solved this by looking at the RO's possible influence on TO performance as experimental noise (error variation).

An overview of the experimental manipulations is presented in Table 1.

Table 1. Experimental manipulations

Manipulated variables	Levels of manipulation
Display type	Traditional displays
Display type	Ecological displays
Display type	Advanced displays
Scenario type	Within design basis
Scenario type	Beyond design basis
Scenario period	Detection phase
Scenario period	Mitigation phase

The manipulated variables were completely crossed and therefore produced a total of 3x2x2=12 experimental conditions (3 display types, 2 scenario types, and 2 scenario periods).

2.2.1 Display type

As indicated by Table 1, three different display types were compared in the experiment. The defining characteristics of each display type are presented below in terms of the analysis on which the designs were based, and the forms by which information was represented.

There are several design factors that are pertinent to operator interface design. Navigation, screen layout, icon symbols, controls, and alarm displays are amongst many that are commonly discussed, but are <u>not</u> explored in this experiment. These factors were accounted for by having the same design for all three display types. For instance, all three interfaces had the same navigation, screen layout, icons, controls, and alarms displays. The display manipulation only changed the "format field" of the process displays (see Figure 1).

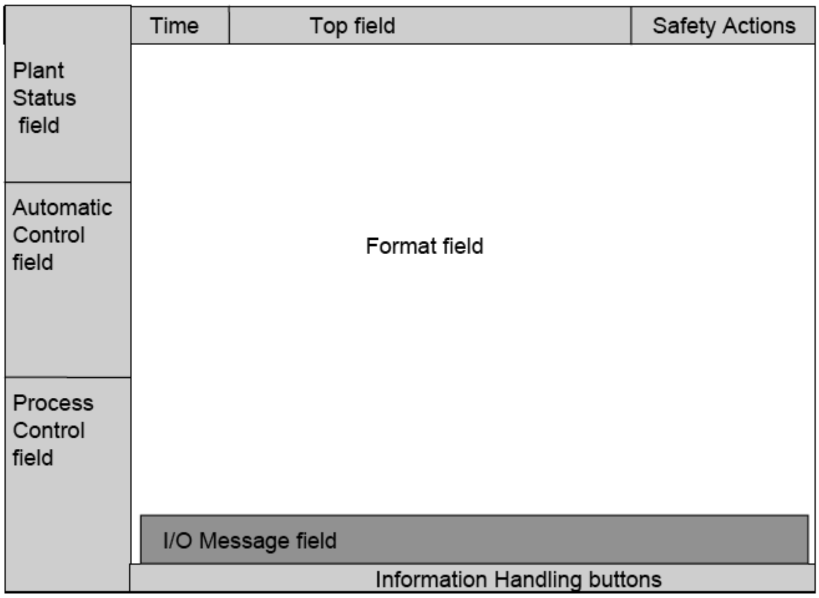


Figure 1. Screen layout of HAMBO displays (from Sørenssen et al., 1999)

Traditional displays

The traditional display type represented the computerized interfaces that are used currently, or up to the recent past by the industry. It illustrates the most classical information presentation style in control rooms of nuclear power plants – the Single Sensor-Single Indicator (SSSI) design philosophy (Bennett and Flach, 1992).

Traditional computerized interfaces typically contain information identified by task analysis. Task analysis can specify information content and structure that enable operators to run a facility efficiently in specific contexts. It is a normative analysis that defines the actions for a given set of anticipated situations. Thus, the information content is often readings of instruments and the structure is the physical plant layout that enables operators to execute a series of actions. Task analysis is often a basic requirement for the regulators (e.g., O'Hara et al., 2002). The operators also provide some design inputs and indicate relevance of information as they participate in the task analysis.

The traditional interface was essentially a computerized version of the hardwired control panels in the operators' home plant. It contained mimic-displays integrated with numerical values of instruments. In other words, the graphical forms were the mimic displays that resemble Piping and Instrumentation Diagrams of the plant. The Traditional displays were similar to the original HAMBO interface anno 2000 (Sørenssen et al., 1999; Karlsson et al., 2001).

Ecological displays

Ecological displays are <u>not</u> based on conventional displays that are commonly found in existing control rooms, i.e. they substantially deviate from the SSSI design philosophy. These displays were instead designed according to the Ecological Interface Design (EID) framework (Vicente & Rasmussen, 1992). The information content and structure of the interface are determined through Work Domain Analysis (WDA). The WDA is a formative analysis that represents the plant process in multiple levels of abstraction and decomposition. The formative nature of the analysis claims to support operators in all types of events including unanticipated situations (c.f. task analysis for Traditional and Advanced displays). The multilevel representation of the plant process is supposed to be psychologically relevant for problem solving, identifying the constraints, invariants and parameters that should be included in the interface. A detailed description of the WDA conducted to design the ecological interface can be found in Welch et al. (2007).

The Ecological displays were intended to present information content and structure identified by the WDA in a way that capitalizes on innate human perceptual capabilities. To achieve this, information at all levels of abstraction is presented graphically as well as numerically. Similar to the Traditional displays, the Ecological displays integrated mimics and indicators for controlling the process and presenting information at the lower abstraction level. Mini trends (see Advanced displays) were also available for the critical indicators as in the advanced interface, but they were integrated into, rather than presented along-side the mimic displays. Information at the higher abstraction levels was typically presented using configural displays (relationships between indicators). A detailed description of the Ecological displays is given by Welch et al. (2007).

Advanced displays

Advanced displays represent the "state-of-the-art" computerized interface for nuclear process control. We expect advanced interfaces to be implemented in controls rooms that are undergoing, or will undergo modernization in the near future. These displays can be viewed as an evolutionary development of Traditional displays that retain and present information, content and structure using mimic diagrams and indicators. The formal analysis upon which information content and structure of the Advanced displays were based, are exactly the same as for the Traditional displays.

The evolutionary changes that lead to the Advanced displays come from operator inputs over time. Advanced displays provide visualization features enabled by the current computer and display technology, such as *mini trends* for critical indicators. In our study, mini trends were included along-side the mimic diagrams (Figure 2). The process experts predefined the indicator that should be displayed on each mini trend, but operators could also assign a different indicator for each trend. In addition to mini trends, configural display graphics were developed to depict some plant processes. Process experts in HAMMLAB designed these configural display graphics as they discovered design opportunities while working with Traditional displays. Unlike mini trends, configural displays are not distributed equally across the different parts of the plant. The Advanced displays resemble the HAMBO interface anno 2005 (written documentation of the current HAMBO interface is not yet available).

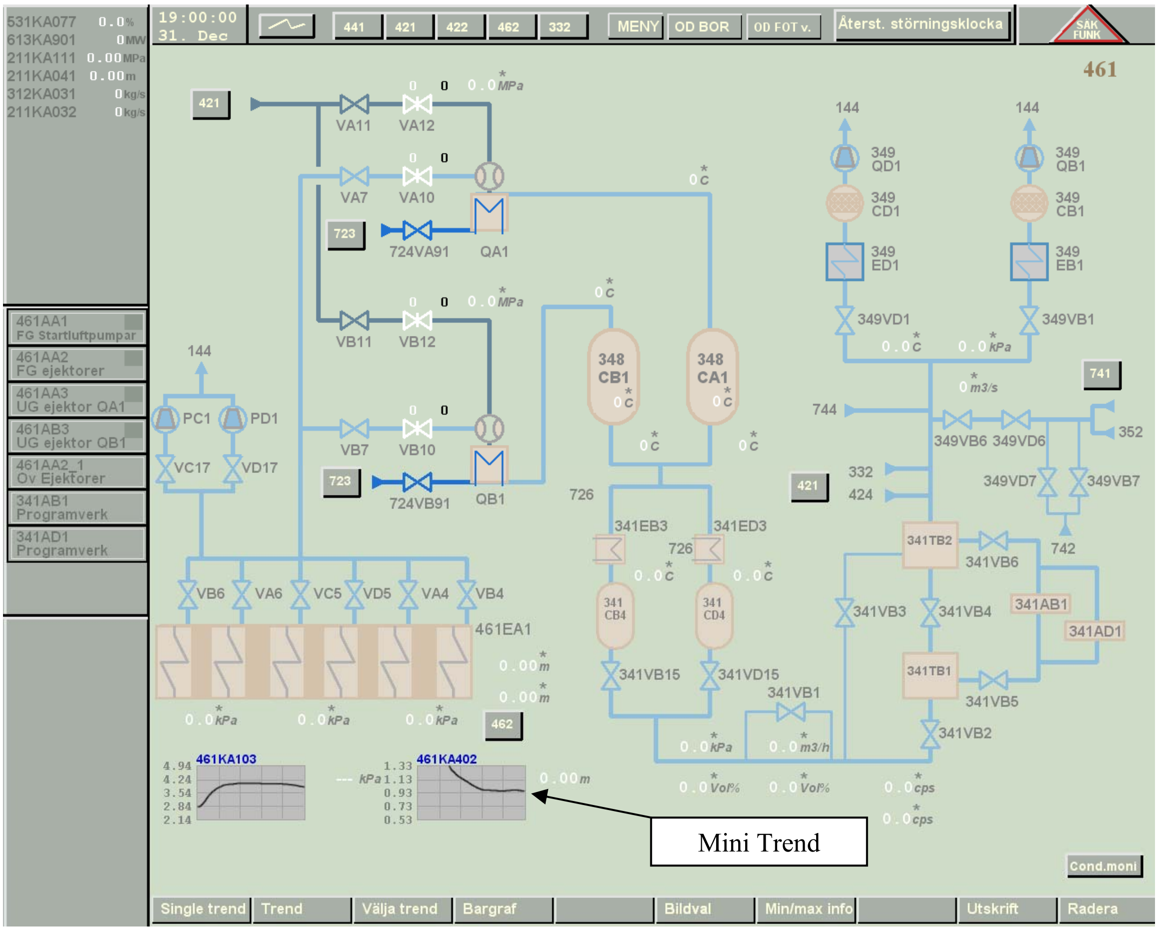


Figure 2. Example of mini trends in an advanced display (not available in Traditional displays)

2.2.2 Scenario type

As indicated by Table 1, two scenario types were tested in the experiment: Within design basis scenarios and Beyond design basis scenarios.

Within design basis scenarios

These scenarios were anticipated by system designers and familiar to operators. Procedures existed for all main incidents.

Beyond design basis scenarios

These scenarios were unanticipated by system designers and unfamiliar to operators. No procedures existed for these situations.

A total of three Within design basis scenarios, and three Beyond design basis scenarios were developed, i.e. one scenario of each kind per display type. Detailed descriptions of the scenarios, which account for the technical difference between the two scenario types, are provided in section 2.7 and Appendix 2 (chapter 11).

2.2.3 Scenario period

Each of the six scenarios had a Detection phase and a Mitigation phase.

Detection phase

In the Detection phase, the operational challenges introduced by the scenario were evolving, but this was not yet annunciated through the alarm system. In order to detect, and possibly prevent the problem, the operators had to use the Traditional, Ecological or Advanced process displays (see 2.2.1) – depending on the experimental condition. The Detection phase stopped immediately before alarms notified the operators about the developing problem. The operators could begin to diagnose the problem during the Detection phase.

Mitigation phase

The Mitigation phase started when certain prespecified alarms informed the operators explicitly about the presence, and the causes of the process deviation. The operators then mitigated the disturbance until a well-defined problem solving action was executed. The Mitigation phase involved detection, diagnosis and mitigating actions.

Figure 3 illustrates the fixed sequence of the two scenario periods, and the approximate amount of time used for each period. The simulator was frozen in a short break between the scenario periods in order to collect Situation Awareness data (see section 3.4 and 3.5).

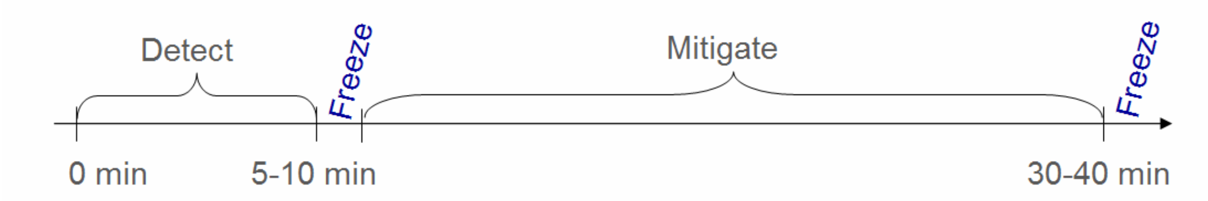


Figure 3. Scenario periods

Detailed descriptions of the scenarios, which account for the technical difference between the scenario periods, are provided in Appendix 2 (chapter 11).

2.2.4 Counterbalancing

Since all crews were observed under all experimental conditions, the presentation order of the conditions may have had systematic effects on performance. Such order effects can be caused by learning, fatigue, and/or carry-over effects (i.e. when experimental conditions contaminate each other). Order effects are compensated by counterbalancing, meaning that the presentation order of experimental conditions is varied across subjects. If the order of conditions has an effect on the dependent variable, and counterbalancing is not employed, it would be impossible to separate experimental effects from order effects. Efficient counterbalancing is therefore essential for the internal validity of within-subject designs (if the researcher is not specifically interested in the cumulative effect of preceding conditions, such as learning effects).

In a simple repeated measures design with one manipulated variable of two levels, half of the subjects would receive one presentation order (“a” before “b”), while the other half would receive the opposite order (“b” before “a”). This procedure is called complete counterbalancing, and is suitable when even numbers of experimental units can represent all possible presentation orders.

Complete counterbalancing is the ideal method for compensation of order effects, but is seldom achievable for factorial designs with a limited number of crews.

In complex experiments, where the number of crews is too small to represent all possible presentation orders, *incomplete counterbalancing* is employed. Representative sequences of conditions are then spread evenly among the experimental units. A frequently used incomplete counterbalancing technique is to randomize the order of conditions for every crew. Randomization controls order effects adequately when a large number of sequences are in use. However, unacceptable condition orders can occur. For example, two crews may receive similar or identical sequences by chance. Such effects will even out in the long run, but is problematical when there are just a few randomized sequences in an experiment. It should also be noted, that randomization can never guarantee an even distribution of carry-over effects.

Latin square counterbalancing is an incomplete counterbalancing technique where *Latin squares* are used strategically to distribute order effects evenly amongst experimental units. A Latin square has p rows and p columns with p Latin letters assigned to the cells of the square, so each letter appears once in each row and once in each column (Kirk, 1995). The rows in the Latin square represent the crews, while the columns specify the presentation order of the experimental conditions. The same square can be replicated several times if there are more crews than experimental conditions, but to achieve an even distribution, the number of crews has to be a multiple of the number of conditions. Thus, an experiment with k treatment combinations can have k×1, k×2, k×3,... k×n participating crews. When there are an odd number of conditions, two reverse Latin squares have to be constructed. The crews are then tested in every experimental condition twice; once for each of the reverse Latin squares.

Latin squares effectively control learning and fatigue effects, but carry-over effects are not compensated. In order to control immediate sequential effects between conditions that are adjacent in the order of presentation (as well as learning/fatigue), a balanced Latin squares can be used (Lewis, 1989). A balanced Latin square is a special version of the Latin square, where each condition precedes and follows every other condition equally often.

We employed a 6x6 balanced Latin square in this experiment. This was convenient since there were 6 crews and 6 experimental conditions that had to be counterbalanced (the order of the two scenario periods was given).

The experimental conditions were first coded (see Table 2).

Table 2. Coding of experimental conditions

	Traditional displays	Ecological displays	Advanced displays
Within design basis scenarios	C	A	B
Beyond design basis scenarios	F	D	E

We then inserted the codes into the 6x6 balanced Latin-square shown in Table 3 (from Lewis, 1989).

Table 3. Presentation order of experimental conditions

	Run order					
	1	2	3	4	5	6
Crew 1	A	F	B	E	C	D
Crew 2	B	A	C	F	D	E
Crew 3	C	B	D	A	E	F
Crew 4	D	C	E	B	F	A
Crew 5	E	D	F	C	A	B
Crew 6	F	E	A	D	B	C

There were three different versions of each scenario type (one for each display type), producing a total number of 3x2=6 different scenarios. *Graeco-Latin squares* can often be used to counterbalance the order effect of another variable in addition to the experimental conditions. A Graeco-Latin square consists of two superimposed orthogonal Latin squares (Kirk, 1995). Two Latin squares are *orthogonal* if each letter of one square occurs only once with each letter of the other square (ibid.). Experimental conditions may then be denoted by Latin letters and the additional variable by Greek letters. Unfortunately, a Graeco-Latin square counterbalancing solution was not achievable in this study due to the dependence among scenario types and scenario versions, i.e. three scenarios were associated with the Within design basis scenarios, while the remaining three scenarios were associated with the Beyond design basis scenarios. The presentation order of the scenario versions was therefore randomized per scenario type for each crew, resulting in the run order presented in Table 4. The letters w1, w2, and w3 denote the three versions of the Within design basis scenarios, while b1, b2, and b3 denote the three versions of the Beyond design basis scenarios.

Table 4 Presentation order of experimental conditions and scenario versions

	Run order					
	1	2	3	4	5	6
Crew 1	A w2	F b2	B w3	E b3	C w1	D b1
Crew 2	B w2	A w1	C w3	F b2	D b1	E b3
Crew 3	C w3	B w2	D b2	A w1	E b3	F b1
Crew 4	D b1	C w3	E b3	B w1	F b2	A w2
Crew 5	E b3	D b2	F b1	C w3	A w2	B w1
Crew 6	F b1	E b2	A w3	D b3	B w2	C w1

2.2.5 Power analysis

The purpose of power analysis is to estimate the probability that an experimental design with a given number of participants can produce statistically significant results for a certain effect size. This section describes how we estimated power before the study by using the Power Analysis and Sample Size software - PASS (Hintze, 1994). The analysis was based upon the following premises:

- The statistical analyses are performed on the operator level of data aggregation.
- All operators are tested individually under all experimental conditions.
- Power is estimated for the main effects produced by the RBF-322 design (3x2x2 within-subject design).

- Treatments are fixed, while blocks are random (the most common mixed ANOVA model).
- The population correlation among experimental conditions is presumed to be rho= 0.25 (which is somewhere in-between a conservative and an optimistic estimate).
- Interactions that involve the blocking variable cannot be assumed to be zero. In other words, subject-treatment interactions are expected to occur. The structural model for the design is therefore non-additive, i.e. in accordance with the conservative within-subject design model employed by most statistical analysis software.
- The adopted levels of statistical significance is $\alpha = 0.05$ and $\alpha = 0.10$, which seems reasonable in applied research settings (see section 5.2).
- Hypotheses are assumed to be non-directional, implying that two-tailed significance tests are used.
- Power is estimated for medium ($f = 0.25$, $\omega^2 = 0.059$), typical ($f = 0.33$, $\omega^2 = 0.10$), and large ($f = 0.40$, $\omega^2 = 0.138$) experimental effects. "Medium" and "large" effects are conventions developed by Cohen (1992), while a "typical" effect refers to the magnitude of the experimental effects that are normally reported in behavioural science (Rosenthal, Rosnow, and Rubin, 2000). According to Cohen (ibid.), a medium effect size represents an effect that is "likely to be visible to the naked eye of the careful observer" (p. 156). Thus, a large effect size should be rather obvious.
- Power is estimated for n=6 Turbine Operators.
- Power estimates for the multivariate approach to repeated measures and the Greenhouse-Geisser/Huynh-Feldt (GG-HF) adjustment (StatSoft, 2004) are included whenever a factor has more than two levels of manipulation (due to the sphericity and the compound symmetry assumption; see section 5.7).

The power estimates are presented in Table 5. It is generally accepted that the power should be at least 0.80. Power estimates ≥ 0.80 are therefore indicated in bold.

Table 5. Power estimates

RBF-322	n	Large effects		Typical effects		Medium effects	
		α= 0.05	α= 0.10	α= 0.05	α= 0.10	α= 0.05	α= 0.10
Display type	**6**	**0.94**	**0.98**	**0.82**	**0.91**	0.57	0.72
Display type (GG-HF adjustment)	**6**	**0.91**	**0.97**	0.77	**0.89**	0.51	0.69
Display type (multivariate)	**6**	0.67	**0.84**	0.51	0.70	0.32	0.50
Scenario type	**6**	**0.94**	**0.99**	**0.84**	**0.94**	0.63	0.79
Scenario period	**6**	**0.94**	**0.99**	**0.84**	**0.94**	0.63	0.79

According to the power analysis, the RBF-322 design with n=6 crews had sufficient statistical power to reveal typical and large experimental effects. The Greenhouse-Geisser/Huynh-Feldt (GG-HF) adjustment seemed to be preferred over the multivariate approach to repeated measures for the display type manipulation[1] (see section 5.7). Power analysis has many uncertainties and should be taken as a rough guide, i.e., it is not an exact formal tool. The population correlation among treatment combinations (rho), for example, is a conjecture based upon experiences from previous simulator experiments.

2.3 Participants

Participants for the experiment were recruited from a plant similar to the HAMBO process model. The control room at the home plant had a combination of conventional analog panels and computerized displays. Even though many of the operators had previous experience in HAMMLAB (see Table 6), it seems reasonable to assume that the fully computerized user interfaces employed in the experiment was rather unfamiliar to the participants. This was compensated to some extent by the pre-experimental training sessions (see section 2.5).

A total of six (n = 6) licensed control room operating crews, with roles ranging from supervision to reactor operation, agreed to participate on a voluntary basis. Each crew consisted of one Reactor Operator (RO) and one Turbine Operator (TO). The study examined the role of the TO in particular, but the crew was considered to be representative of a single working unit despite the absence of a shift supervisor (SS) found in a typical control room operating crew. The demographic details of the participants are presented in Table 6.

Table 6. The participants' demographic data

ID	Role	Current Position	HAMMLAB Experience (years of involvement)	Age	Gender	Licenses (Y=yes, N=no) / Years in Position		
						TO	RO	SS
1	RO	SS	0	59	Male	N / 5	Y / 3	Y / 13
2	TO	TO / RO	1	53	Male	Y / 10	Y / 5	N / 0
3	RO	SS	0	59	Male	N / 5	Y / 3	Y / 13
4	TO	RO / SS	3	44	Male	N / 0	Y / 14	Y / 0
5	RO	SS	5	46	Male	N / 2	Y / 14	Y / 9
6	TO	RO	0	38	Male	N / 5	Y / 5	Y / 2
7	RO	SS	5	46	Male	N / 0	Y / 14	Y / 9
8	TO	TO	2	37	Male	Y / 3	N / 0	N / 0
9	RO	RO / SS	2	47	Male	N / 5	Y / 10	Y / 2
10	TO	Other	0	48	Male	N / 0	N / 0	N / 0
11	RO	RO / SS	2	46	Male	N / 15	Y / 8	Y / 3
12	TO	TO / RO	0	39	Female	Y / 2.5	Y / 0.5	N / 0
Mean			1.50	45.7	9M, 1F	3 / 4.75	8 / 5.95	6 / 2.9
Standard Deviation			1.65	6.83		4.64	5.39	4.51

Due to unforeseen circumstances, two of the participants that were scheduled to act as Reactor Operators in the experiment were unable to attend. Replacing the absentees with participants from other crews was considered to be the least intrusive option given that the complete removal

[1] The display type manipulation has more than two levels. The sphericity and compound symmetry assumption may therefore not be satisfied. If the assumption is violated, we have to use the (univariate) Greenhouse-Geisser/Huynh-Feldt adjustment, or the multivariate approach to repeated measures.

of two crews would degrade the statistical power of the experiment. Specifically, participants 1 and 5 acted in place of participants 3 and 7, respectively (greyed out in Table 6). These two operators participated twice in the experiment as Reactor Operators. The Turbine Operators were 6 unique participants (one per crew). Since the scenarios challenged the Turbine Operators in particular, and the data were analysed for the Turbine Operators individually, it is believed that the repeated participation of two Reactor Operators had negligible effects. There is, however, a theoretical risk that the Turbine Operators data may have been biased.

A majority of the operators had prior experience with the HAMMLAB simulator and facilities, with an overall average of 1.5 years of involvement. In practical terms, this means that most of the operators had participated in one or two HAMMLAB experiments before. The mean age of the participants was 45.7 with a standard deviation of 6.83. Of the ten (10) unique participants, less than half possessed a Turbine Operator license, while a majority held a Reactor Operator license. As per regulatory guidelines, licenses require continual renewal in order to be valid. Half of the operators were former Turbine Operators and no longer possess a valid Turbine Operator license. The mean number of years of experience in each of the positions relevant to the study: 4.75 (TO), 5.95 (RO), and 2.9 (SS). It should be noted that participant 10 had no prior official control room operating experience, but had extensive experience as a field operator and control room instructor.

2.4 Data collection procedure

Data was collected over a period of three weeks. Three or four operators participated each week. The participants were introduced to the experiment and trained together on the first day of the experiment. We then formed a morning crew (08:00-14:00) and an evening crew (15:00-21:00) for the data collection itself. The data collection lasted for two days, during which the crews participated in six simulator runs. As explained in section 2.2.4, the presentation order of the display types and scenarios were varied across the crews. During the two first weeks of data collection, only three operators participated. The same operator then acted as RO in the morning and evening crew (see section 2.3). The following weekly schedule was used:

Day 1

1. Introduction (0.5 hours): Informed consent, demographic questionnaire, and opportunity for participant questions.
2. Training (3.0 hours): Participants trained in the operation of the HSI.
3. Lunch break (1.0 hour).
4. Training (3.0 hours): Participants trained in the operation of the HSI.

Day 2

1. Training (1.0 hour): Participants (morning) trained in the operation of the HSI.
2. Break (0.15 hours).
3. Scenario trial (1.0 hour): Participants (morning) completed the first scenario.
4. Break (0.15 hours).
5. Scenario trial (1.0 hour): Participants (morning) completed the second scenario.
6. Lunch break (1.0 hour).
7. Scenario trial (1.0 hour): Participants (morning) completed the third scenario.

8. Break (1.0 hour).
9. Training (1.0 hour): Participants (evening) trained in the operation of the HSI.
10. Break (0.15 hours).
11. Scenario trial (1.0 hour): Participants (evening) completed the first scenario.
12. Dinner (1.0 hour).
13. Scenario trial (1.0 hour): Participants (evening) completed the second scenario.
14. Break (0.15 hours).
15. Scenario trial (1.0 hour): Participants (evening) completed the third scenario.

Day 3

1. Scenario trial (1.0 hour): Participants (morning) completed the fourth scenario.
2. Break (0.15 hours).
3. Scenario trial (1.0 hour): Participants (morning) completed the fifth scenario.
4. Break (0.15 hours).
5. Scenario trial (1.0 hour): Participants (morning) completed the sixth scenario.
6. Lunch break (1.0 hour).
7. Post-test questionnaire and interview (2.00 hours). The participants (morning) responded to questionnaires and participated in the interview
8. Debriefing of morning participants (0.50 hours).
9. Scenario trial (1.0 hour): Participants (evening) completed the fourth scenario.
10. Break (0.15 hours).
11. Scenario trial (1.0 hour): Participants (evening) completed the fifth scenario.
12. Break (0.15 hours).
13. Scenario trial (1.0 hour): Participants (evening) completed the sixth scenario.
14. Post-test questionnaire and interview (2.00 hours). Participants (evening) responded to questionnaires and participated in the interview
15. Debriefing of evening participants (0.50 hours).

All participants were assigned a non-descriptive alias, which was used on personal information, performance data, and interview documents. Aliases were not made available to other participants or to the participants' employers. Data were combined and analysed in such a way that participants were unidentifiable in the results. For audio and video recordings, head mounted cameras were used and a stationary camera was placed behind each participant. The intention of the audio and video recordings was to capture the participant's interaction with the user interface. All participants were identified only by their alias in the recordings. The experiment is reported in internal HRP publications and international scientific journals, but always in such a way that the participants' anonymity is maintained. The participants gave their written consent, they were informed about their right to view their own data, and they were debriefed after the data collection.

2.5 Operator training

The operators participated in a pre-experimental training program focusing on:

- Familiarization with the human-machine interface in HAMMLAB (navigation, generic display characteristics, alarm system, trends, large screen display etc.) – 1 hour.
- Training on Traditional and Advanced displays – 1 hour.
- Differences between the operators' home plant and the simulated process model- ½ hour.
- Data collection methods and laboratory procedures – ½ hour.
- Training on Ecological displays – 3 hours.
- Repetition on the first day of the data collection – 1 hour

As explained in section 2.4, the group of operators that participated each week were trained together on the first day of the experiment. A short repetition of the HSI features and a training scenario on the second day was arranged separately for the morning and evening crews. The total training time was about seven hours per crew.

About two hours of the training was devoted to the human-machine interface in HAMMLAB, and the Traditional and Advanced displays. This training was adapted to the operators' individual needs, as many of the participants had previous experience from HAMMLAB (see Table 6).

The Ecological displays were conceptually new and unfamiliar to the participants. Effective utilization of the Ecological displays required that the operators modified and improved their mental model of the nuclear process, e.g. by forming abstract representations of energy balances (see section 2.2.1). Thus, in order to produce a meaningful experimental evaluation of EID, we had to prioritize operator training that could help the participants to use the Ecological displays as intended. All graphical elements that were specific to the Ecological displays were explained. The instruction focused on how the elements functioned, and the reasons for including them in the design. The EID features in the turbine displays, condenser display, and condensate & feedwater displays were addressed during training (see Welch et al., 2007). After the presentation of each ecological display, the operators were able to test the new features during a short training scenario. We also arranged training scenarios that covered the entire secondary side of the process. This allowed the operators to learn how the graphical elements in the different areas of the process worked together. The training scenarios were included in the total training time.

2.6 Laboratory set-up

The HAMMLAB control room environment contains three workstations and a large-screen display as shown in Figure 4. The data is presented to the operators at each workstation through interfaces on 19" LCD panels. Navigation and interaction through the interfaces is done via keyboard and mouse. A typical HAMMLAB experiment involves one Reactor Operator, one Turbine Operator, and one Shift Supervisor. In this particular study, the various interfaces and scenarios were evaluated with one Reactor Operator and one Turbine Operator (see section 2.2 and 2.3). Each operator workstation consisted of 12 individual LCD panels. The large-screen display in the middle of the room provided an overview of the plant for both operators. Note that the HAMBO version of the large-screen display was used also in the Ecological display condition.

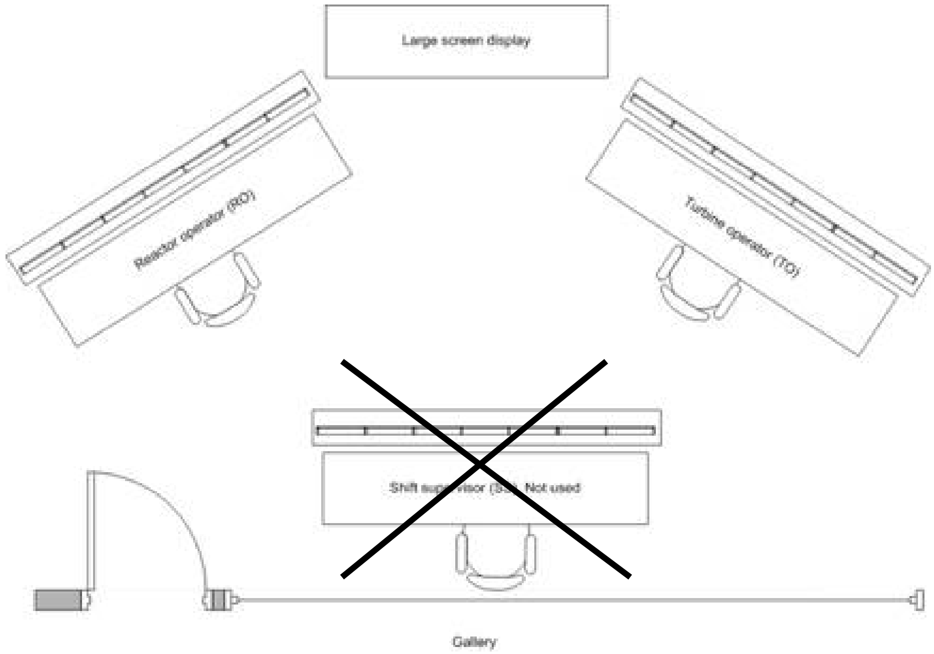


Figure 4. HAMMLAB Control Room Layout

The screen layout used in the Ecological display condition is shown in Figure 5. In the Traditional display condition, the EID displays (on the turbine side) were replaced with the original HAMBO process displays (anno 2000). Likewise, in the Advanced display condition, the EID displays were replaced by current HAMBO process displays (anno 2005). The Ecological displays were spatially dedicated to the Turbine Operator screens, while the Traditional and Advanced process formats could be displayed freely to the screens depending on the Turbine Operator's preferences. See section 2.2.1 for more information about the experimental manipulation of display type.

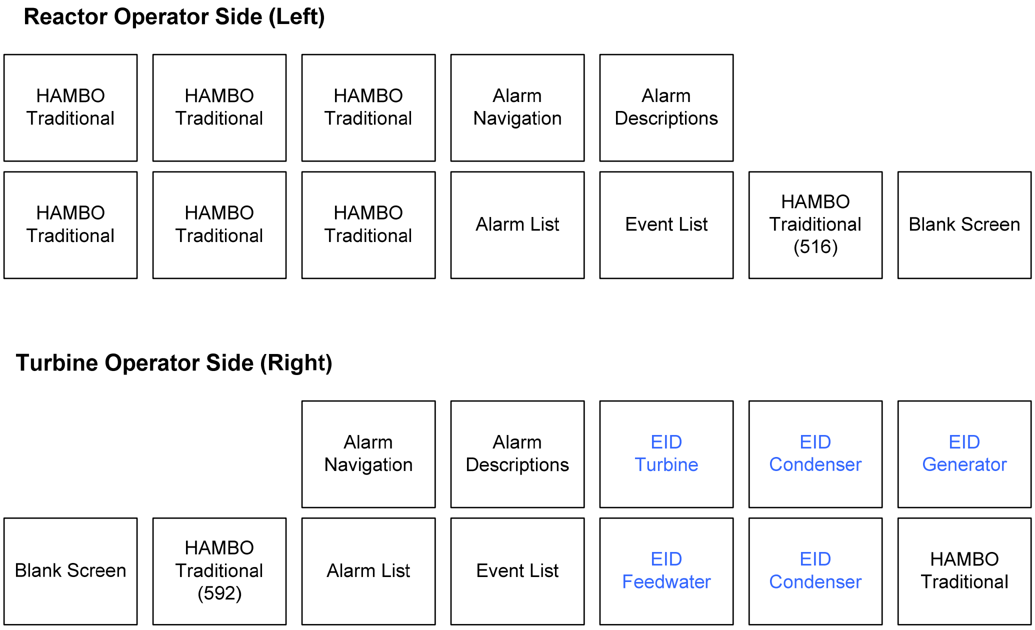


Figure 5. Screen layout in the Ecological display condition

Located behind the control room environment is an observation gallery from which the experimental team managed the simulator, scenarios, and data collection (see Figure 6). Operator interactions with the interfaces were logged in the simulator while activities within the control room environment were audio and video recorded. The experimental team consisted of an experimental leader, a process expert, a laboratory technician, and a training team. The leader was responsible for overseeing the scenario progression, acting as a representative from plant management in order to evaluate operator response, and administering the appropriate questionnaires. The process expert managed the simulator functions, acted as plant and engineering personnel, provided comments on the current state of the scenario, and evaluated operator performance. The laboratory technician handled all data recorded through the simulator, the questionnaires, and the audio-visual equipment.

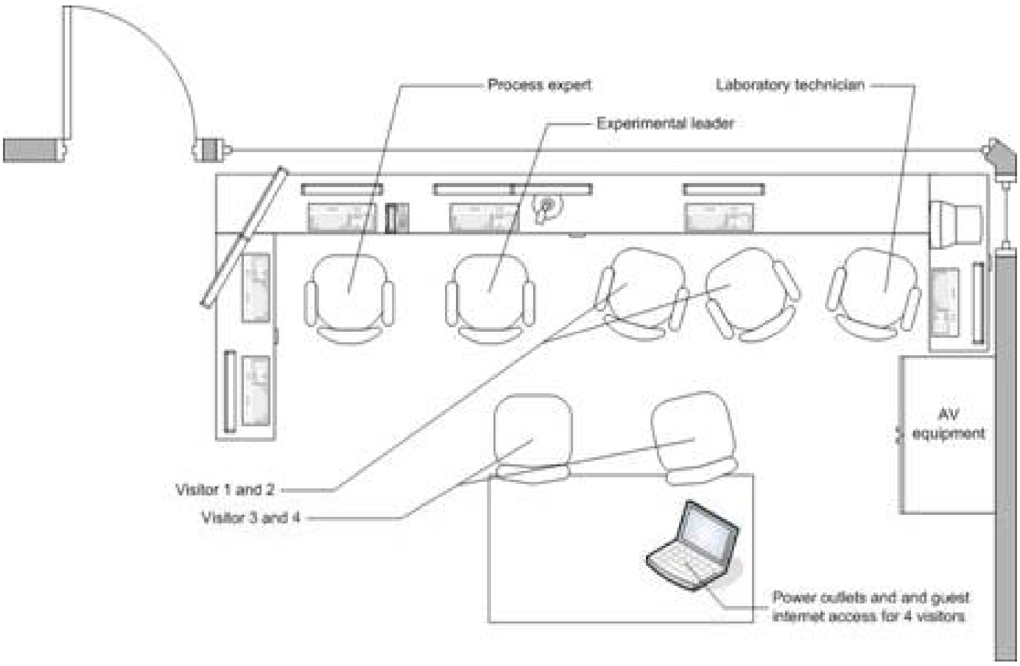


Figure 6. HAMMLAB experimenters' gallery

2.7 Scenarios

Before the experiment, the operators were told that the objective of all six scenarios was to maintain safe operation at constant power.

We developed three Within design basis scenarios, where equipment failures and job responsibilities were familiar and anticipated by the nuclear power plant organization. All failures in these scenarios could be resolved by referencing the procedures that described the occurring sound alarms. Standard procedures from the simulated plant were used during these scenarios. The following is a short description of the Within design basis scenarios:

- *Within-1* – Leakage in the intermediate superheater @ 109.3% power: There was a leakage in the immediate superheater at full power. The crew should be able to resolve the problem if they follow the procedure after the first alarm. It is particularly interesting to examine the detection speed of the leakage amongst the display types before any alarm sounds.
- *Within-2* – Drain route switch and pre-heater bypass reset @ 20% power: There were a few valve failures that prevented proper functioning of the plant. The crew should refer to the procedures for all the alarms to identify the malfunctioning equipment and report to maintenance. Identifying equipment failures and reporting them to maintenance based on alarms were the two main responsibilities of the operators.

- *Within-3* – Drift of an instrument @ 44% power: There were two independent failures occurring at the beginning of the scenario including one valve failure and one measurement drift. The measurement drift could be diagnosed easily after referencing the procedure (according to the alarms) as it indicated two other redundant measures for comparison. The operators should resolve the issue by contacting maintenance.

We developed three Beyond design basis scenarios, where equipment failures and events were unfamiliar and unanticipated by the nuclear power plant organization. Resolutions to the unstable states of the plant could not be addressed through procedures, and was based entirely on the problem solving skills of the operators. The following is a short description of the Beyond design basis scenarios:

- *Beyond-1* –Turbine trip while the generator is connected to the grid @ 109.3% power: A malfunctioning valve resulted in a turbine trip without disconnecting from the grid. In most cases, turbine trips lead to a disconnection from the electrical grid. The operators should report valves failures to the maintenance and disconnect from the grid.
- *Beyond-2* – Leakage at the condensate cleaning building @ 109.3%: There was a leakage in the condensate cleaning building that was not addressed by any procedure. In addition, there were failures of automated safety features that increased the complexity of the problem. Operators should call the maintenance to correct valve failures and operate valves manually to bypass the leakage.
- *Beyond-3* – Sudden and high increase of seawater temperature @ 109.3% power: The scenario started with one pump failure that is not addressed by any procedure. Then, the seawater temperature increased quickly from 12 to 25 °C. This may be caused by underwater volcanic activities, or even submarines heat discharges, which are not addressed by any procedure. The operator should reduce power manually.

Detailed scenario descriptions, as developed by the process expert, can be found in Appendix 2 (chapter 11).

3. PERFORMANCE MEASUREMENT

The experiment concentrated on how the Traditional, Ecological and Advanced display types affected Situation Awareness (SA) under varying operating conditions. An indicator of Workload was also included, since the effect of HSI-design on the work demand in the control room is practically important and safety relevant.

3.1 The importance of measurement

As mentioned in section 2.1, the ability to reveal experimental effects increases with the quality of measurement, and the interpretability of the experimental results depends heavily on the performance measures (Cook & Campbell, 1979). Within complex work environments, there are additional reasons for giving human performance measurement extra attention. Firstly, high-resolution raw data are gathered and aggregated, and the refinement of these data into meaningful performance indicators has to be guided by a measurement methodology that can identify suitable levels of aggregation and establish convergent validity between measures. Next, it is difficult to standardize the content of measurement in dynamic work environments, since the operator taskwork and the evolvement of the event interact. Another complicating factor is that the technical and operational complexity of the nuclear process demands the deep involvement of process experts. Finally, the many sources of measurement error in realistic settings represent a concern by itself, forcing the researcher to take measurement seriously. Thus, measurement is devoted a separate chapter in this report.

3.2 Ecological Interface Design and Situation Awareness

Ecological Interface Design (EID; Vicente & Rasmussen, 1992) and Situation Awareness (SA; Endsley, 1995a) have both been part of the cognitive engineering lexicon for over 15 years. The former is a design framework for complex systems and the latter is a theoretical construct for which various measures have been developed. Both EID and SA contribute to the development of information displays that improve operator insight into decision-making spaces. They share a mutual objective of designing for good decision making and good human performance in complex environments.

Despite the convergence in these objectives, EID and SA have evolved independently from one another. This is rather surprising considering that they must, at practical levels, overlap. An effective ecological interface should support SA, and high SA must depend on the conveyance of ecological aspects of the environment. Yet a review of the empirical EID and SA literature reveals next to no co-occurrences of the terms. EID researchers are not assessing SA in their studies and SA researchers are not using EID to achieve improved SA. A central aim of the study introduced here is to break through the conceptual divide between these concepts. We therefore decided to evaluate the first hypothesis; that good EID should support SA. Specifically, a design developed from the principles of EID should demonstrate improved SA in comparison with conventional computerized displays.

EID theory predicts directional effects of interaction on SA between ecological interfaces and types of event scenarios. EID identifies information needed to support anomaly detection, decision making and action under abnormal events. Similarly, the general theory of SA pertains to event detection and understanding. Much of the empirical literature on EID supports the theory-based prediction that the performance advantages of ecological interfaces over traditional interfaces are more pronounced in abnormal events as compared to normal events (Vicente, 2002; Jamieson, in press). The prevailing explanation for this finding is that normal events do not tax operators sufficiently to force them to rely on the additional information support provided by the ecological interface. The theory readily extends to SA as an indicator of performance.

3.3 Situation Awareness model for nuclear process control

The nature of work is increasingly cognitive or knowledge-based as advanced technology is continually elevating system complexities and replacing human operators in routine and well-defined tasks (e.g., Vicente, 1999; Durso et al., 2006). Though the pace of introducing new technology varies, the shift of responsibilities for routine or well-defined tasks from human to machine appears to be the trend in almost every domain. Under these conditions, the urgency to develop theoretical constructs and meaningful measurements for capturing the cognitive aspects of work arises. In late 1980's, Situation Awareness (SA), which was previously a jargon used by military aviators, was formalized and introduced as a theoretical construct to account for performance of cognitive work. The SA concept is obviously very well received, especially in practical settings, as it appears to account for a large portion of cognitive work and provides the expected foundation for developing measurements predictive of work performances or diagnostic of design imperfections (c.f., Durso et al., 2006).

In SA research, Endsley is, perhaps, the most influential pioneer. Her theoretical model of SA (1988, 1995a, 2000a), which originates from studying cognitive work of military pilots, is highly rational and intuitive. The model emphasizes the importance of perception, comprehension, and projection of critical signals in a situation in order to achieve system goals or mission objectives. In simple terms, desirable outcomes are much more likely to result when operators "know what is going on" (Endsley, 1995a). Endsley also developed the equally influential Situation Awareness Global Assessment Technique (SAGAT). Her model and measure are popular in the aviation community as suggested by the quantity of subsequent publications since the first introduction of her SA concept.

Endsley's model of SA, though only inspired by findings in aviation, also extends to many other domains. This is not very surprising because the abstraction or generalization of findings on cognitive work from any one domain would be of some relevance to others. After all, every operator is always in some kind of situation and has some form of awareness (or consciousness). However, such theoretical models of SA often fail to capture the intricacy of human cognition, domain specific complexities, and their interactions. Though applicable to many cases, theories are often too abstract to capture significant, unique details in every setting.

In our view, abstract models of SA are useful and necessary to communicate high-level findings across settings (c.f. Flach, 1995); however, they are not sufficient to inform meaningful, relevant SA operationalization and measurements in specific domains. Meaningful operational definitions must reflect the nature of operator work and system complexities that are often specific to domains (see Pew, 1995). Relevant operational definitions should only capture cognitive aspects pertinent to operator work. As a natural outcome of generalization, abstract theories or models of SA often *exclude* domain specific details of work and complexities, but *include* a host of cognitive aspects from multiple domains. Consequently, SA measurements that are founded upon operational definitions of abstract theories/models do not possess the desired *meaningful* and *relevant* properties. Unfortunately, much of the literature to date has been built upon studies employing SA measurements lacking such specificities to their domains.

Meaningful and relevant operational definitions are likely to be domain specific, and thus, difficult to realize solely based on abstract SA theories or models. Domain specific SA models that capture the unique aspects of human cognition, domain complexities, and their interactions are necessary to inform operationalizations in different settings, thereby, improving the usefulness of SA measurements. The present focus of SA literature is on either studies with SA measures or discussion in abstract terms, leaving a knowledge gap that renders the effective application of SA research difficult (see Figure 7).

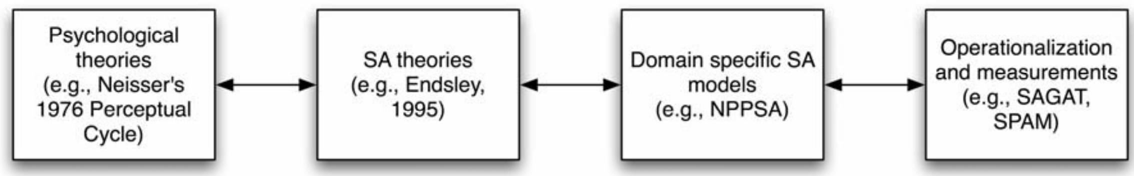


Figure 7. Generalization-specialization spectrum of Situation Awareness research.

To inform operationalization and measurement, domain specific SA models should depict: (1) the unique sets of work activities performed by the operator to acquire situation knowledge (i.e., situation assessment activities); (2) the situation knowledge gained through performing each set of work activities (i.e., SA component); and (3) the connection between cognitive science research and SA activities and knowledge. Hence, domain specific SA models are, most likely, built upon field studies which capture the richness of operator work and environment, followed by extensive reviews of the literature relevant to operator cognitive activities of situation assessment. These models are obviously restricted in application to their source domains, but they can still contribute to theories by making invariant SA phenomena across multiple domains explicit. SA literature currently lacks organized discussions at the domain level that are vital to support theoretical claims and developments.

For the nuclear power domain, the concept of SA provides a useful categorization of operator performance (c.f., Flach, 1995); thus, it would be practical to establish performance measures for those aspects of work. However, a model depicting situation assessment activities and Situation Awareness/knowledge of NPP operators in connection to cognitive science research is absent from the literature. Consequently, the nuclear industry lacks the necessary foundation upon which meaningful and relevant measures can be derived (c.f., Endsley, 1995a; Adams et al., 1995), and deprived of sound evaluation tools to validate control room design or operator competence.

The Nuclear Power Plant Situation Awareness model (NPPSA) is a domain specific SA model that provides the foundation by which operational definitions and measurements can be derived. NPPSA depicts a set of operator activities and knowledge that constitute situation assessment and Situation Awareness, respectively. It is a descriptive model constructed with a bottom-up approach, capturing the unique operator activities and knowledge in NPP control room settings. The model explicitly connects the domain specific properties to established findings in cognitive science. Hence, NPPSA illustrates the meaningful characteristics of operator work and reflects the relevant aspects of operator cognition pertaining to the generic idea of SA – “knowing what is going on”.

NPPSA is still in its early development. It is currently conceived to be composed of three unique sets of activities and three corresponding SA components (see Figure 8)[2] . First, the operator

[2] Note that each situation assessment activity can only be meaningfully discussed with respect to the corresponding situation awareness (i.e., knowledge) and vice versa (c.f., Adams et al, 1995).

engages in *monitoring* activities to acquire a *Process overview* - perception of process anomalies. The perception of anomalies directs the operator to begin *diagnosis* of the process anomalies. The knowledge resulting from diagnosis is *Scenario understanding* - the understanding of process anomalies/disturbances. Scenario understanding informs the intervention required to resolve the process anomalies. While engaging in monitoring and diagnosis, operators need to conduct periodic *self-assessment* to evaluate their performance and strategies in controlling the situation. The knowledge resulting from self-assessment is *Metacognitive accuracy* - the knowledge of one's contribution to the situation. Metacognitive accuracy regulates behaviours and strategies of monitoring and diagnosis that is crucial to appropriate operator decision-making and actions.

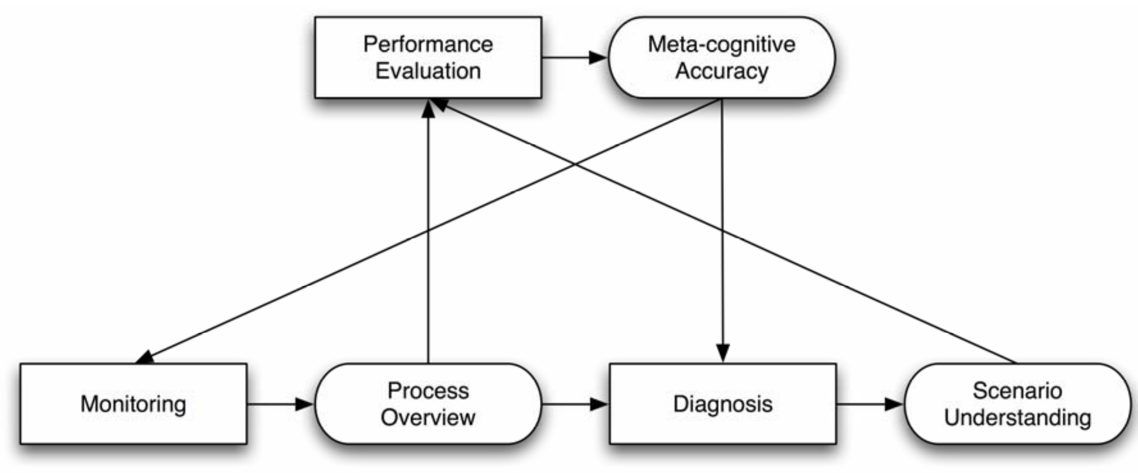


Figure 8. Nuclear Power Plant Situation Awareness (NPPSA)

3.4 Process overview

Process overview is operationally defined as *accurate detections of meaningful changes in relevant process parameters*. Relevant refers to process parameters where monitoring resources should be concentrated due to a high potential for disturbance. Meaningful refers to fluctuations that are not the results of noise, but of deviations from steady state plant processes. From a practical standpoint, accurately determining meaningful changes of relevant process parameters reflects the ability to build and update context, observe process parameters, decipher relevance, and identify 'true' process deviations. Appendix 5 (chapter 14) provides a detailed description of the practical and psychological foundation of the Process overview concept.

The Process overview measure is a modification of the Situation Awareness Global Assessment Technique, SAGAT (section 3.4.1), and the Situation Awareness Control Room Inventory, SACRI (section 3.4.2). It is developed especially for simulator studies in control room environments. The measure employs the common freeze-and-query technique. In other words, Process overview is assessed according to participant responses to the queries administered during simulation freezes at different times of the trial. The timing and length of each freeze are strategically determined according to specific characteristics of the scenario. Above all, the freezes within the trials are kept short to reduce influences on performance.

Process overview queries are constructed according to specific characteristics of the scenario. All administered Process overview queries are relevant to the scenario disturbance and development. As mentioned above, the cuing problem is resolved by strategic timing of query administration. The queries are strictly about the directional changes as opposed to specific values of process parameters. Hence, the Process overview measure is a 3 Alternative-Force-Choice (AFC) task.

The response is always a choice of: increase, remain the same, or decrease (see Appendix 3, chapter 12). The temporal range of each query is typically the time between the present and the last meaningful change or state for that parameter. Queries do not inquire about changes beyond the last meaningful state to avoid confounding with memory factors. Queries may include projection of changes in the near future; however, projection queries must be administered and interpreted with caution for two reasons. First, projection is heavily dependent on detailed reasoning of the plant processes and disturbances (i.e., Scenario understanding); thus, projection queries should only be administered after operators have some time to complete the first iteration of monitoring activities. Second, the dependence of projection on detailed reasoning suggests that accuracy of projection queries is correlated with Scenario understanding; thus, they need to be interpreted with respect to each other.

Process expert(s) determine the correct answer to each query. That is, process experts, who have full understanding of the scenario and access to simulator data, determine the answers of the queries during the trials. The scoring method is percentage correct.

The Process overview measure is an adaptation of SAGAT and SACRI. In the following, brief descriptions of SAGAT and SACRI are presented. This is followed by a comparison of the measures to illustrate the reasons for the adaptation.

3.4.1 SAGAT

SAGAT assesses SA by querying operators during freezes within each scenario trial (Endsley, 1995b, 2000b). Derived from Endsley's model of SA (1995a, 2000a), the SAGAT measure includes queries on all three levels of SA (i.e., perception, comprehension, and projection).[3] SAGAT employs "goal-directed task analysis" (GDTA) to generate a set of queries for each "job class." The queries are typically in forms that the responses are either short verbal answers or selections from multiple choices. The timing of freezes and selection of queries should be random. The correct answers to the queries are determined either by simulator data or process experts. The scoring method is percentage correct.

3.4.2 SACRI

SACRI (Hoggs et al., 1994) is a modification of SAGAT with the application of Signal Detection Theory (SDT). It is developed strictly for NPP control room settings without reference to any specific SA model. The idea of Process overview is mentioned but not fully discussed by Hoggs et al. (1994). SACRI queries are extracted from an inventory of queries that can be modified slightly for different NPP simulators (e.g., code name for specific process indicators). The inventory is created by several process experts during the initial development of the SACRI. They are always about the directional changes of process parameters in the past, present, or future; thus, the response is always a choice of: increase, remain the same, or decrease. For each trial, 18 queries on the changes of different process parameters in the past, present, and future (equally distributed) are administered during freezes within each trial. The selection of queries and timing of freezes are random. The correct answer to each query is determined through post-hoc review of simulator data. The SACRI measure includes a sensitivity and a bias score, which are calculated using two non-parametric formulas.

---

[3] The boundaries between the three levels of SA are not completely clear (i.e. they are subject to interpretation). This is particularly true between Level 1 and 2. This is to be anticipated, as the distinction between perception and comprehension is still quite nebulous in psychology.

3.4.3 Comparison of SAGAT, SACRI and Process overview

The adaptations for the Process overview measure offer distinguishing features and superiority over SAGAT and SACRI for assessing SA in the nuclear domain. Both SAGAT and SACRI attempt to eliminate the potential cuing effect by random selection of queries and timing of freezes (c.f., Endsley, 1995; Hoggs et al, 1994). Although administering relevant queries could cue participants towards the process areas in disturbance, irrelevant queries could mislead operators after freezes. In addition, administering queries irrelevant to the scenario is contradictory to measuring Process overview, as the ultimate aim of monitoring is to identify process areas in disturbance where the operators must devote their cognitive resources for diagnosis (see Mumaw et al., 2000). In order to eliminate such cueing effects, signals, such as alarms, are presented after each freeze for the Process overview measure. This prevents performance effects across participants due to cuing without introducing irrelevant queries.

The Process overview measure employs the query and response format of SACRI rather than SAGAT because Process overview is the knowledge generated from detecting abnormal behaviours of different parameters rather than registration of precise parameter values. Furthermore, time to access process indicators to determine the exact values of process parameters is usually available in process control settings.

In comparison to SACRI and SAGAT, the Process overview measure limits the temporal range of queries to the period from the last significant change to the time of the freeze. As mentioned, this temporal range can restrict measurement to Process overview or exclude confounding factors (e.g., memory). In addition, the subjective nature of such temporal range is natural to monitoring NPPs as anomalies could arise anytime. Queries on projecting future process parameters values (i.e., Level 3 SA) are generally not advised.

The Process overview measure also distinguishes itself by requiring process experts to determine the correct answers to queries[4]. By determining the answer in the same situation as the participants, the process experts can establish the criteria for meaningful changes more accurately than post-hoc reviews of simulator data. Hoggs et al. (1994) had expressed the difficulty in determining the answers to SACRI queries or the criteria for meaningful changes through post-hoc reviews of simulator data.

The scoring method of the Process overview measure also differs from SACRI. Although sensitivity and bias are useful indices in theory, the foundation of the SACRI scoring method is found to be questionable (see Appendix 4, chapter 13). The accepted sensitivity measure in SDT literature for 3AFC tasks as characterized by the response format of SACRI is percentage correct.

A comparison between the SA measures illustrates the reasons for the adapting SAGAT and SACRI for the Process overview measure. (Table 7 summarizes their similarities and differences.) The similarities and differences illustrate the benefits in employing the Process overview measures in assessing the situation knowledge generated from monitoring. In essence, the Process overview measure is designed to reflect the accurate detections of meaningful changes in various relevant process parameters by the operator.

[4] Note that SAGAT is not part of the discussion on determining answers to queries because its query and response format is different.

Table 7. Similarities and differences between SA measures.

Distinguishing Characteristics	SAGAT	SACRI	Process overview
Construct of interest	SA Model proposed by Endsley (1995a)	Process overview	Process overview
Elicitation Method	Queries during freezes in each trial	Queries during freezes in each trial	Queries during freezes in each trial
Query Type	Level 1, 2, 3 SA	Past, present, and future values of process parameters	Present values of process parameters relative to last change (projection of parameter values can be incorporated but not advised)
Query Development	SA Requirement Analysis	Inventory for specific power plants	Analysis of scenarios by process experts
Query Selection	Random selection based on job classes	Random selection with constraints to balance scenario relevant and irrelevant queries	Relevant parameters only
Query Administration Timing	Random selection	Random selection	Selective assignment according to scenario development
Response Format	No requirement but typically categorical choices (i.e. multiple choice)	3 AFC (increase, same, decrease)	3 AFC (increase, same, decrease)
Scoring	% Correct	Non-parametric formula for Sensitivity & Bias	% Correct

3.5 Scenario understanding

To measure Scenario understanding, we used a real-time probing technique that was developed for the purpose of the experiment. Retrospective measures of SA rely heavily on the operators' ability to recall and infer what happened during an event, i.e., reflective SA (Fracker, 1991). This is undesirable due to the risk of false inferences associated with the selection, transformation, and reconstruction of information from Long-Term Memory. Therefore, another approach to SA measurement is needed that capture the operators' concurrent understanding of the situation, i.e., momentary SA (ibid.). The most promising method in this respect is real-time probes, where verbal queries that reveal the operators understanding are posed to the operators while the scenario evolves (Durso et al., 1998; Jones & Endsley, 2004).

Real-time probes can be specific to the process state, or have an open form where the operators respond freely to general questions about the current process state and task requirements. Open probes with a free verbal response format will probably produce less reactivity than process specific probes that tend to guide the operators in a direction with respect to solving the scenario (cueing). Further, open probes are natural in the control room setting, since operators in their real work environment often provide general status updates to other plant personnel.

A frequent finding is that the accuracy of the response is close to perfect for specific real-time probes, producing ceiling effects (Durso et al., 1998; Jones & Endsley, 2004). This is not surprising, since the operators are simply asked to find and report identifiable information in the user interface. Operators are able to perform this task in most situations. The accepted and presumably sensitive indicator of SA is therefore supposed to be the response time from when a probe is presented until the operator responds. Unfortunately, this measure is deeply confounded with Workload and other factors, such as the navigation structure of process displays, which may, or may not be an integrated part of the HSI design concept under testing. Response times to real-time probes are therefore difficult to interpret. Controlling statistically for Workload and other confounding factors seems unrealistic and improvised. For open probes, it is virtually impossible to decide exactly <u>when</u> the process understanding is communicated during the operators' free verbal response. Response times to probes are therefore considered to be an unreliable SA indicator, and an alternative methodology is needed.

The probing method that we developed for this study is designed especially for simulator research in control room environments, and follows six basic steps:

1. Before the data collection, the operators are informed that the experimental leader will act as a representative from the plant management, and may call the control room for status updates in normal or deviating process situations (probes).
2. The experimental leader calls the operator of interest for status updates concurrently with the ongoing scenario. Status calls are made by internal telephone lines from the experimenters' gallery.
3. The content, format and timing of the probes are predefined for each scenario, but on-the-fly adjustments are sometimes necessary, since operator activities can influence the scenario development (dynamic problem solving)
4. The operators respond to the probes by giving free verbal interpretations of the process situation. The simulator is frozen while the operators respond to a probe.
5. For each probe, the process expert scores the operators' Scenario understanding according to an anchored and interpreted rating scale (see below)
6. The aggregated level of understanding score is calculated from the scores on several probes administered during the scenario or scenario period.

The experimental leader is responsible for giving the probes rather than the process expert, because the process expert already acts out several other roles over the telephone, such as the engineer on duty, electrician, field operators, chemist etc. These roles should not be confused with the probes. Further, this allocation of roles alleviates the already high task load on the process expert (see section 2.6).

Responding to status calls is a secondary task that may be given low priority during periods of high Workload, which would produce poor measurement. This is solved by freezing the simulator during the status call. Such "mini-freezes" avoid operator stress reactions due to competing tasks (i.e., controlling the process vs. responding to probes). In this way, the verbal response to the probes becomes the primary task for a short period of time without having any consequences for the ability to monitor and operate the system. This procedure is probably easier to defend than the simulation breaks introduced by, e.g., SAGAT (Endsley, 2000b), since the freezes last for a few seconds only - not minutes. The operators were prepared and trained to handle the mini-freezes before the data collection.

Guiding the operators (cueing) is not a serious problem with our method, since the probes are open, i.e., the operators respond freely to undirected questions. However, a possible objection to our approach is that operators are given short pauses in the simulation, and thereby the opportunity to reflect during the execution of the scenario. This may influence performance. On the other hand, operators regularly provide status updates to other plant personnel in real control room environments. The impact of reflection on performance can therefore be seen as an integrated and natural part of the operators' every day work. The number of status updates per time unit may be higher than in real life, but this issue addresses a general concern with respect to the realism of simulator studies, since simulated scenarios are often compressed representations of real events.

The probes that were given by the experimental leader during the status updates had one of the following formats:

- Why is [reference to process state]?
- What is happening with [reference to process state]?
- What are you planning to do about [reference to process state]?
- Which strategy are you following with respect to [reference to process state]?
- What are the consequences of [reference to process state]?
- Where are you in the work process with respect to [reference to process state]?
- Has [reference to process state] occurred?
- Have you taken care of [reference to problem in the process]?
- Have you come any further with respect to [reference to problem in the process]?
- Did the [reference to process event(s)] evolve as expected?
- Did the [reference to execution of task(s)] proceed as expected?
- Has the process changed since my last call?
- What has changed in the process since my last call?
- What will happen in the process now?
- Is there something I should know about this situation?
- Is everything going according to plan?
- What do you think is wrong in this situation?
- Is everything normal?
- How is everything?

These questions vary with respect to the degree of openness, reference to the current vs. future situations, focus on detections, diagnosis or mitigation etc. The idea is that the process expert should preselect the questions that are best suited to uncover the operators' level of understanding in different phases of the scenario, and whenever necessary improvise new question formulations during the scenario.

The process specific content of the probe questions is predefined for each scenario on the basis of expected events throughout the scenario and the associated performance requirements. This is comparable to a SA requirement analysis (Jones & Endsley, 2004).

Each probe should as far as possible address one fault (subgoal) in the scenario. Different faults are addressed by different probes. Otherwise, the scoring of each probe will be extremely difficult. The same fault may be addressed several times by different probes.

Follow-up questions can be asked after a probe if the process expert feels uncertain about the operator's Scenario understanding. To minimize the intrusion, one should not ask more questions than absolutely necessary. It is therefore essential to find questions on beforehand that tap efficiently into the operators' Scenario understanding.

The probes should be as open as possible, and should never guide the operators in any direction with respect to solving the scenario. It is important to be aware that operators could be guided by, (a) the process specific content of the probe, e.g., asking questions about an alarm that the operators have not detected, (b) the selection of probes, e.g., asking specifically about the execution of tasks in a situation where the operators believe that they should remain passive, or (c) implicit feedback given by the experimental leader, e.g., encouraging intonation or other indications that the provided response was good or poor. All such cues should be avoided.

The probes play an important role for the process expert's interpretation of the operators' Scenario understanding, but they are not the only source of information. Scenario understanding can also be inferred from observations made during the data collection, such as operator interventions with the process, search for information, reactions to alarms, verification of information, interface navigation, the field of attention, communication within the crew, teamwork behaviour, calls made to other plant personnel, thinking aloud etc. Hence, the probes give the operator an extra opportunity to verbalize his/her understanding explicitly, but the level of Scenario understanding is not inferred from this verbalization alone.

The goal is to probe the operators in situations where the understanding of the process situation is considered important for dealing effectively with the challenges introduced by the scenario. By probing exactly when the operators are mentally active, one can measure the momentary aspects of SA when understanding of the process situation is critical for efficient decision making and problem solving. The planned timing of the probes is predefined for each scenario, even though deviations from the plan may occur due to dynamic problem solving.

The process expert scored the operators' degree of Scenario understanding per probe on a simple scale (see Table 8).

Table 8. Scenario understanding scale

Score	Interpretation
3	Full understanding
2	Good understanding
1	Some understanding
0	Poor understanding

This scoring was done immediately after an operator had responded to a probe, i.e. when the simulator was running again after a mini-freeze. The scale should be interpreted in the following way:

Full understanding: Full understanding means that the operator has, (a) detected all process deviations that occurred before an alarm annunciation, (b) has reached a fully correct diagnosis, and/or (c) has a fully correct understanding behind the problem solving or execution of tasks.

Good understanding: Good understanding means that the operator has, (a) detected all process deviations that occurred, (b) has reached a mostly correct diagnosis, and/or (c) has a mostly correct understanding behind the problem solving or execution of tasks.

Some understanding: Some understanding means that the operator (a) has detected some process deviations that occurred, (b) has reached a partly correct diagnosis, and/or (c) has a partly correct understanding behind the problem solving or execution of tasks.

Poor understanding: Poor understanding means that the operator (a) has not detected the process deviations that occurred, or falsely detected process deviations that did not occur, (b) has not reached a diagnosis, or reached a wrong diagnosis, and/or (c) has no understanding, or an incorrect understanding behind the problem solving or execution of tasks. Thus, both ignorance and misunderstanding indicate poor Scenario understanding.

The scale does not differentiate between ignorance and misunderstanding, since this is probably impossible. Misunderstanding may sometimes lead to full understanding later in the scenario, i.e., through learning by trail and error within the limits of the system's error tolerance. Ignorance, or lack of understanding, indicates a passive approach that may correctly leave everything to automation and emergency systems in certain situations. On the other hand, ignorance or misunderstanding can lead to severe economical/safety consequences at other times. It will therefore vary whether ignorance or misunderstanding indicates the poorest Scenario understanding, and the exact range order can typically not be revealed by short scenarios.

3.6 Metacognitive accuracy

In dynamic work environments, human operators adapt their taskwork strategically to variations in the task demand, and to their own problem solving capacity (Huet, 1999). This calls for a metacognition concept. While cognitive skills are necessary to perform a task, *metacognition* refers to the understanding of how tasks are performed. We may decompose metacognition into *knowledge of cognition* and *regulation of cognition* (Schraw, 1998).

Knowledge of cognition is what a people know about their own thinking, and thinking in general. It includes declarative, procedural and conditional knowledge (ibid.). *Declarative knowledge* is knowledge <u>about</u> your learning process, memory capacity, and factors that influence your performance. *Procedural knowledge* is heuristic and strategic knowledge of <u>how</u> tasks are solved, e.g., how new information can be categorised effectively. *Conditional knowledge* refers to the temporal aspects of metacognition, such as <u>when</u> and <u>why</u> resources should be allocated to a task (ibid.).

Regulation of cognition denotes activities that help people to control performance, and includes planning, monitoring and evaluation. *Planning* is allocation of recourses and selection of strategies before tasks are executed. *Monitoring* is the awareness and comprehension of ongoing task performance. *Evaluation* is the retrospective appraisal of one's own efficiency and the outcome of the taskwork (ibid.).

Table 9 provides an overview of the dimensions of metacognition, as suggested by Schraw (1998). There are other ways to decompose the metacognition construct, but Schraw's taxonomy seems sufficient for our purpose.

Table 9. Dimensions of the metacognition construct (Schraw, 1998)

Metacognition	*Knowledge of cognition*	Declarative knowledge	Knowing "about"
		Procedural knowledge	Knowing "how"
		Conditional knowledge	Knowing "when" and "why"
	Regulation of cognition	Planning	Allocate resources/select strategies
		Monitoring	Check task performance on-line
		Evaluation	Appraise task performance

There is probably a wide range of metacognitive processes involved in the development of Situation Awareness (see section 3.3). However, the focus of this experiment is *Metacognitive accuracy,* which is the ability to correctly assess ones own level of performance effectiveness while engaged in complex tasks (Fiore, Johnston & Smith, 2004). This assessment may be concurrent or retrospective, and therefore relates to both monitoring and evaluation aspects of metacognition. We operationalized Metacognitive accuracy as the absolute difference between self-rated and Actual task performance, which expresses the extent to which operators are able to realistically assess their own performance efficiency.

In his review of Situation Awareness measures, Fracker (1991) suggested that *"…an objective assessment of SA lies in the inconsistency between subjective SA ratings and the appropriate performance criteria rather than in the ratings alone. Quantitative assessment of this inconsistency may provide a useful index of SA and may be a fruitful direction of future research"* (p. 19). This line of thinking inspired the idea of using Metacognitive accuracy as an indicator of SA.

Maki (1998) compared different measures of Metacognitive accuracy. One measure that stands out is the *bias*, i.e., the signed difference between subjective performance judgement and actual performance. This indicator is unique because it differentiates between over-confidence and under-confidence in performance, which provides useful diagnostic information in studies of, e.g., trust in automation and teamwork. However, as a measure of Situation Awareness, the scale-properties of the bias become problematic. The optimal score is zero (realistic performance estimation); degrees of over-estimated performance are expressed through positive scores, while degrees of under-estimated performance are expressed by negative scores. The meaning of a bias therefore depends on, (a) the score value, (b) the distance from the score value to zero, and (c) the direction of the difference from zero. This makes it meaningless to use the bias as a dependent variable in analysis of variance. Furthermore, correlations with typical performance scales, where a higher score means better performance, would be misleading. We therefore decided to use the absolute value of the difference between self-rated and Actual task performance as the indicator of Metacognitive accuracy (see section 3.6.3 for details about the calculation method). This measure expresses the extent to which operators are able to realistically assess their own performance. It is less informative than the bias, but complies with the implicit scale conventions for SA measurers and can be used meaningfully as a dependent variable in analysis of variance – which is essential for hypothesis testing.

The successfulness of the Metacognitive accuracy indicator depends heavily on the measurement of Self-rated task performance and Actual task performance. These two measures are therefore presented before we explain the calculation method for Metacognitive accuracy.

3.6.1 Self-rated task performance

After the completion of each scenario period, the Turbine Operator in each crew responded to a self-rating questionnaire, which was developed for the purpose of this experiment. A 4-item rating scale was used after the Detection phase and an 8-item rating scale after the Mitigation phase of the scenarios. The 4-item scale was a subset of the 8-item scale. Two versions of the scale were used because the first scenario period did not contain active interventions with the process as in the second scenario period. Some of the 8 items were then irrelevant and therefore removed. The 4-item and 8-item rating scales are shown in Table 10. The response format was a 5 point Likert scale anchored by ‘disagree’ (1) and ‘agree’ (5), where the operators rated their level of agreement with the statements expressed by each item. The unweigheted average of the 4-item or 8-item scale was used as the overall indicator of Self-rated task performance.

Table 10. 4-item and 8-item self-rating scales

4 item scale	
Item 1	I had a good a overview of the process
Item 4	I made correct diagnoses
Item 6	I utilized the displays well
Item 7	I became aware of process deviations at an early stage
8 item scale	
Item 1	I had a good a overview of the process
Item 2	I used my time efficiently
Item 3	I cooperated well with the rest of the crew
Item 4	I made correct diagnoses
Item 5	My actions steered the process in the correct direction
Item 6	I utilized the displays well
Item 7	I became aware of process deviations at an early stage
Item 8	I performed the correct actions

3.6.2 Actual task performance

We used the Operator Performance Assessment System (OPAS) to capture and quantify Actual task performance in the experiment (Skraaning, 1998, 2003). OPAS is specifically targeted at experimental evaluations of control room design solutions in research simulators, and has been used extensively in HAMMLAB since 1996. User experiences have uncovered the need for improvements of the instrument, and a revised version of OPAS is therefore presented below.

OPAS measures whether operators carry out their taskwork in accordance with scenario solutions prescribed by experts on control room operation (process experts). The purpose of the system is to assess the degree of compliance with performance expectations that remain constant across task conditions. It is thereby possible to compile scores that originate from different scenarios into one performance scale. This type of relative measurement is quite common. One may, for example, measure the proportion of available time used to complete a task. Then, the available time and the response actions can vary across scenarios, but the relative amount of available time used to complete tasks is still a psychologically meaningful performance indicator. The OPAS methodology implies that operator solutions to scenarios are decomposed and understood, but this analysis of taskwork is only a by-product of the assessment.

Experimental manipulations may depend more or less upon the task characteristics (Skraaning, 2003). When the operationalization of treatments in an experiment is accomplished through the scenario design, the manipulation can be said to be *task-dependent*. Working under high vs. low time pressure is an example of a task-dependent experimental manipulation. If a treatment is operationalized without the scenario properties taken into account, the manipulation is *task-independent*. Comparing alternative alarm systems is an example of a task-independent manipulation, since most test scenarios could be used for this purpose. This is a simplification, however, since it would be impossible to conduct the alarm experiment without scenarios that produce alarms, suggesting that treatment operationalizations often have an element of task-dependence after all. It is nevertheless useful to distinguish between experimental manipulations that depend heavily on the task characteristics, and manipulations that are more task-independent.

OPAS is well suited for experimental evaluation of control room design solutions, since such manipulations are relatively task-independent and are typically tested under varying task conditions. The ability to equate performance scores across scenarios then becomes an advantage. However, the OPAS methodology should be used carefully for task-dependent manipulations, since the selection of performance items for each scenario may then bias the measure. An effect of scenario complexity may, for example, be masked by the selection of more challenging performance items in simple scenarios, and less challenging items in difficult scenarios. Such biases are balanced out for task-independent manipulations, since each experimental condition would be tested with all scenarios, or for a sample of scenarios. The problem of defining a fair basis for comparisons of human performance across task-dependent experimental conditions is fundamental, and probably not restricted to OPAS. In general, no experimental conditions should be favoured by the selection of performance items. Simulator experiments in the area of Human Reliability Assessment (HRA) require detailed event tree analyses of operator solutions to scenarios. As indicated above, OPAS is not developed for such purposes.

A fundamental idea behind OPAS is that human task performance in complex operating environments is comprehensible only for process experts. This assumption is supported by research, indicating that the differences between novices and experts are most evident for difficult problems, and that it takes at least 10 years of full time practice to reach deep insight into a complex field (Anderson, 1985). It is improbable that non-experts should be able to grasp the meaning of human performance without a complete understanding of the process model and functional rules that underlie the operation of an advanced industrial system. Performance measurement instruments may therefore provide a general infrastructure and a set of instructions for process experts, but the content of the performance concept has to be operationalized by the experts themselves. Consequently, the task performance construct measured by OPAS varies, depending on the process experts' definition of the performance items. This "undefined" construct validity is not ideal from a psychometric point of view, but reflects the dynamic complexity of human performance in realistic operating environments. It is argued here that implicit expert knowledge about control room operation is difficult to make explicit, and that precise performance constructs would constrain the adaptability and practical value of the measure. That is, there is an inherent trade-off between the controllability and the ecological validity of the measurement content in complex environments. The experiences so far suggest that OPAS has established a reasonable balance between measurement precision and flexibility.

Expert judgment of human performance can be influenced by unwanted psychological measurement errors, such as selective memory, rigid thinking, personal hypotheses, fatigue, first impressions, implicit personality theories, stereotypes etc. (Skraaning, 1998). OPAS meets this challenge by reducing the element of human judgment during the experimental data collection, and makes decisive judgment explicit. This is achieved by splitting the measurement process into three phases:

1. *Scenario analysis.* Before the data collection, process experts develop optimal solutions to the scenarios. These solutions are subject to critique by others.
2. *Data collection.* During the data collection, process experts register the occurrence or absence of predefined control room activities in real time, concurrent with operator performance.
3. *Calculation of performance scores.* A performance index is calculated, estimating the discrepancy between the a priori expert analyses and actual operator solutions to scenarios.

Scenario analysis

In the scenario analysis, process experts develop optimal solutions to scenarios by identifying items that express the desired level of task performance. For example, a performance item may be timely detection of process information or intervention with the system. To discriminate between changing levels of task performance, the process experts should identify items that are presumed to be completed successfully by some, but not all operators, and under some, but not all task conditions. Otherwise, OPAS will produce ceiling or floor effects (Jones, 1995). It is important, however, to keep the performance items "essential", i.e., avoid peripheral activities that are unimportant for the operation of the system. Any discrete scenario event that can be verified against actual operator performance data may serve as an item in the scenario analysis.

Standardization of performance items across scenarios is not a prioritized issue in OPAS. On the contrary, the new version of OPAS utilizes the relativistic nature of the measurement approach, and gives the process experts much flexibility in the scenario analysis. The reason for this is that OPAS measures the operators' compliance with performance expectations held by process experts across task conditions, and does <u>not</u> strive towards deep analysis of specific operator solutions to scenarios.

Process experts determine the number of items to be defined per scenario, taking their own real time scoring capacity into account (see data collection). A simple scoring system is used, where the operators earn points for completing performance items. Each performance item depicts alternative operator activities that are rewarded by 0, 1, 2 or 3 points. The experts are free to define any item that can differentiate between levels of task performance across experimental conditions. Some examples are:

- *Omissions*. If the operators fail to carry out an expected task, they receive 0 points; otherwise they get 1, 2, or 3 points depending on the overall importance of the task for solving the scenario (an implicit weighting mechanism). For example, operators may receive 2 points for detecting an alarm, and 0 points if they miss the alarm (in this example 3 and 1 points are not used).
- *Commissions*. If the operators execute an erroneous task they receive 0 points, otherwise they get 1, 2, or 3 points depending on the compliance of their task performance to the expectation and/or the overall importance of the task for solving the scenario. For example, operators may receive 3 points for correctly starting a pump and informing the field operator, 2 points for correctly starting the pump without informing the field operator, and 0 points for starting the wrong pump (in this example 1 point is not used).
- *Response times*. The operators receive points depending on the time used to complete a task. For example, operators may receive 3 points for closing a valve within 3 minutes after an

incoming alarm, 2 points for closing the valve within 5 minutes, 1 point for closing the valve within 8 minutes, and 0 points if the valve is not closed within 8 minutes.

- *Strategies*. The operators receive point depending on the acceptability and/or the observed successfulness of the selected operating strategy. For example, operators may be rewarded 3 points for stopping a leakage before they attend to another sub-disturbance, 1 point for stopping the leakage after the sub-disturbance has been handled, and 0 points for neglecting or failing to notice the leakage (in this example 2 points is not used).

In principal, any discrete performance criterion that can be verified against actual operator data may serve as an element in the scenario analysis. For complex knowledge-based scenarios, it can be sufficient to base the scenario analysis on simple performance items, such as checking whether expected actions are carried out. For easy procedural scenarios, however, it may be necessary to introduce limiting conditions in the definition of the performance items in order to avoid ceiling-effects (Jones, 1995); e.g. “perform activity A before activity B”, “perform activity A and B”, or “perform activity A if condition B or C is satisfied”. The performance items that were used for the Turbine Operators in this study are listed in Appendix 1 (chapter 10).

The performance items can be categorized in order to isolate and measure specific aspects of control room operation. Classification schemes should then fit the objective of the study. Two examples are given in Table 11.

Table 11. Example of OPAS classification schemes

“Type of activity” classification scheme		“Performance goal” classification scheme	
Categories	Description	Categories	Description
Actions	Intervening operations, active verification of information, or other control room activities, such as picking up a paper-based procedure	Safety	Activities that prevent and mitigate nuclear transients, disturbances, and accidents
Detections	Registration of information content presented in the user interface, such as alarms	Production	Activities that maintain the desired level of electricity generation
Inferences	Human reasoning, such as finding causes, combining information, diagnosing, searching for solutions etc.	Preservation	Activities that ensures that the mechanical components of the plant are not exposed to unnecessary physical stress
Teamwork	Information exchange and/or collaboration among crew members, such as requesting a task from a field operator		

Data Collection

A practical advantage of OPAS is the data collection, where a process expert registers operator activities in real time, concurrent with operator performance. This method is extremely effective compared to post-experimental analysis of simulator logs, audio-video recordings, and eye-tracking data. Comparisons of real time registration and objective data from the same scenarios showed that real time expert rating is sufficiently reliable. A reasonable level of inter-rater reliability has also been documented (Skraaning, 2003), suggesting that the OPAS data collection can be performed by a single process expert. The expert can combine several sources of observation, such as:

- *Verbalization* - through free-standing microphones in the control room and wireless microphones attached to the operators
- *Physical behaviour* - by direct view into the control room and video monitors focusing on each operator
- *Problem solving* - by slave monitors mapping activities in the human-machine interface, and video monitors showing the operators' gaze-line (when eye-tracking is applied).
- *System states* - from inspecting the large screen display, trend diagrams etc.

In order to receive a 1, 2 or 3 score for a performance item; there should be no doubt that the operator activities were correctly and fully completed - as specified. Otherwise the performance item can not be scored in real time, and must be clarified after the data collection. Clarifications can be made in the participant debriefing, from simulator logs, audio/video recordings or eye-movement-tracking. Data collection, based upon the operators' recall of their own problem solving during the debriefing should be treated carefully, since (a) introspection is a questionable methodological approach, and knowledge about our own cognitive processes is limited, (b) retrospective reports produce unwanted memory effects, and (c) humans tend to overestimate their own contribution to causal events.

Calculation of performance scores

OPAS performance scores reflect the discrepancy between an expert analysis and operator solutions to scenarios. The main performance index is simply the unweighted average of the operators' scores on the performance items. The resulting scale ranges from 0 to 3, where a higher score indicates better performance. Separate OPAS scores can be calculated for entire scenarios, limited scenario periods, different operator roles (e.g., Reactor Operator or Turbine Operator), or pre-defined categories as shown in Table 11. The exact data aggregation solution depends on the purpose of the study and the experimental design.

3.6.3 Calculation of Metacognitive accuracy scores

We have seen that metacognitive accuracy is the absolute difference between Self-rated task performance and Actual task performance, i.e., the extent to which operators are able to realistically assess their own performance. The score is obtained as follows:

1. The Self-rated task performance and Actual task performance variables are standardized to make them comparable (z-score standardization).
2. Actual task performance is subtracted from the Self-rated task performance and the resulting variable is standardized.
3. Absolute values are calculated and the variable is transformed by a $\log(X_i+1)$ function (Howell, 1997), since absolute values of a z-score standardized variable is expected to be positively skewed.
4. The variable is inverted by subtracting X_i from the maximum score, making higher scores express a more realistic level of self-assessment.

3.7 Workload

There is no commonly agreed upon definition of *Workload*, but the concept generally refers to the mental and physical strain placed on humans during task performance. One may also say that Workload expresses the cost of accomplishing the task requirements (Hart and Wickens, 1990).

In this experiment, we measured Workload by means of the *subjective task-complexity scale* (Braarud, 2000). This is a self-rating instrument that focuses on the task-related difficulties that control room operators experience while they work. The scale is developed within the OECD Halden Reactor Project and has been used successfully in a number of studies. Several psychometric evaluations have revealed that the scale is reliable and an effective predictor of task performance. In fact, the scale predicts human performance in nuclear process control far better than NASA-TLX (ibid.), which is another widely used indicator of Workload (see Hart & Staveland, 1988). An extensive description of the subjective task complexity scale can be found in Braarud and Brendryen (2001).

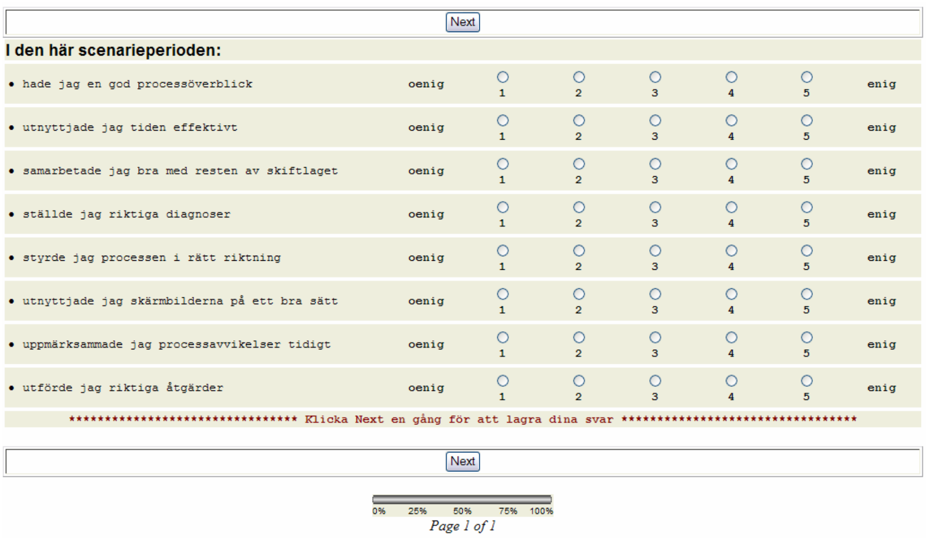
The response format of the subjective task complexity scale is a 7 point Likert scale anchored by 'very difficult' (1) and 'easy' (7), where the operators rate the items listed in Table 12. The overall Workload index was the unweighted average of the five (5) subjective task-complexity items.

Table 12. Workload items

Workload items	How difficult was this scenario period with respect to:
Item 1	Vague or ambiguous process displays, misleading or missing process information
Item 2	Ambiguous, misleading or missing feedback on operator actions
Item 3	Time for planning and controlling the work
Item 4	Many parallel tasks (several disturbances or process events) that complicated the execution of every single task
Item 5	Collection and utilization of much information to perform the work

3.8 Data collection methodology

The operators' and process expert's performance ratings were registered with the Halden Questionnaire System (Drøivoldsmo, 2003). This system has demonstrated speed and reliability in a number of studies within the OECD Halden Reactor Project. Questionnaire development, experimental setup, data collection, and preparations for analysis are done from a web browser without the need for additional software. An example of the data collection interface is presented in Figure 9. All performance data are automatically registered and stored for later analysis.


Next

I den här scenarieperioden:

• hade jag en god processöverblick	oenig	1	2	3	4	5	enig
• utnyttjade jag tiden effektivt	oenig	1	2	3	4	5	enig
• samarbetade jag bra med resten av skiftlaget	oenig	1	2	3	4	5	enig
• ställde jag riktiga diagnoser	oenig	1	2	3	4	5	enig
• styrde jag processen i rätt riktning	oenig	1	2	3	4	5	enig
• utnyttjade jag skärmbilderna på ett bra sätt	oenig	1	2	3	4	5	enig
• uppmärksammade jag processavvikelser tidigt	oenig	1	2	3	4	5	enig
• utförde jag riktiga åtgärder	oenig	1	2	3	4	5	enig

****************************** Klicka Next en gång för att lagra dina svar ******************************

Next

0% 25% 50% 75% 100%

Page 1 of 1

Figure 9. Example of Halden Questionnaire System data collection interface (in Swedish)

4. HYPOTHESES

As stated by the research approach (section 2.1), hypotheses are written in natural language for each of the measurement constructs.

4.1 Hypotheses for Situation Awareness

4.1.1 Main effect of display type

A main effect of display type is predicted. That is, we expect to observe differences in the Situation Awareness measures as a function of interface type. A directional effect of display type is also predicted. Specifically, both the Advanced interface and the Ecological interface are expected to more effectively support Situation Awareness than the Traditional interface. Moreover, Situation Awareness is expected to be superior for the Ecological interface compared to the Advanced interface.

Rationale

A wealth of laboratory data points to consistent performance advantages of EID over traditional computer interfaces (by which we mean interfaces designed based on the Single Sensor-Single Indicator (Goodstein, 1981) design philosophy for process systems (Vicente, 2002). These results were largely replicated in the one full-scope simulation of the framework conducted to date (Jamieson, in press). Separately, empirical results support the assertion that operators using advanced computer interfaces (by which we mean those that contain some configural display elements) show performance gains over those using traditional computer interfaces). We are not aware of any direct comparisons of EID and Advanced computer interfaces for process control. However, the theory behind EID would lead to the prediction that the Advanced interface would support better Situation Awareness as compared to Traditional interfaces, but not rise to the level enabled by the Ecological interface.

An effective Ecological interface identifies the constraints on effective action in a complex system and makes those constraints visible to operators. These analysis and design objectives speak directly to the *perception* and *comprehension* levels of (Endsley's (1995a) formulation of) Situation Awareness. EID should support perception by making the constraints on effective operation of the system visible through graphical forms that are consistent with the perceptual capabilities and limitations of the viewer. EID should support comprehension by communicating the purposeful structure of the system. That is, the display communicates a truthful functional model of the system to support knowledge-based reasoning about the system. These two mechanisms are not independent in the interface; rather, they speak to two different levels of Endsley's model of Situation Awareness. Does the EID framework support projection? It can be argued that it does, but to a lesser extent than it supports levels 1 and 2. The mechanism for this support is hypothesized to be the support of operator manipulations of the mental model that is externalized by the Ecological interface.

Following the logic above, the Advanced interface supports Situation Awareness, but to a lesser extent than the Ecological interface. The Advanced interface supports resource-limited perception of the process state relative to goal-relevant constraints. For example, mini-trends for key state variables offload the need to remember their historical values. The mini-trends also show alarm limits, which are expressions of purposeful constraints. Thus, the Advanced interface is expected to support perception. However, the Advanced interface is expected to be less effective in supporting comprehension. Individual state variable constraints are localized to the process flow (via the mimic display). In cases in which the sequential relationships amongst process equipment are critical to making sense of an event, this could support comprehension. However, the mimic display is available in all three interfaces and is should therefore not contribute to any systematic differences in Situation Awareness. More importantly, though, the functional coupling of constraints is not depicted in the Advanced interface. The Advanced interface also offers little support for projection. The mini-trends do support projection of the future value of a state variable, but only under the assumption that the instantaneous derivative of that variable is maintained. In summary, we would expect that the Advanced interface would be effective in supporting perception with respect to a mental model of the system that is only partially communicated in the display. The interface could also be marginally supportive of comprehension and projection, but only for physical connections (comprehension) and under constant velocity changes in state variables (projection).

Finally, the Traditional interface is expected to be largely ineffective at supporting Situation Awareness at any level. Minimal support for perception may be supported by highlighting key variables (denoted by a star in the interface). Once again, support for comprehension may be minimally afforded by the mimic display. Both of these design features are also provided by the Ecological and Advanced interfaces, however, so no differential advantage is expected.

4.1.2 Main effect of scenario type

A main effect of scenario type is predicted. That is, the two types of scenarios are expected to yield different levels of Situation Awareness. Specifically, we predict a directional effect of scenario type. Participants are expected to have better Situation Awareness on Within-design basis scenarios as compared to Beyond-design basis scenarios.

Rationale

Beyond-design basis events present operators with situations for which their information systems, procedures, and training programmes have not been designed to provide support. The impacts are expected to span the levels of Situation Awareness. With respect to perception, process variables that provide diagnostic information about an event may not be included in the displays (or may be buried in a display hierarchy). In terms of comprehension, the coupling amongst constraints may not have been emphasized in displays or in training. Finally, projecting the future state of a disturbed system may be impeded by a lack of understanding of the unusual system dynamics. Together, this incomplete support will challenge operators to engage in knowledge-based processing with an impoverished information system. Undertaking difficult and error-prone reasoning in a stressful context is expected to result in reduced Situation Awareness.

Reviews of the history of industrial accidents (e.g., Rasmussen, 1969) conclude that the most severe accidents occur when operators are faced with process states that extend beyond the design basis. There is a substantial amount of empirical research in the process control domain to support the conclusion that human and system performance during beyond-design basis events is inferior to within-design basis events.

4.1.3 Main effect of scenario period

A main effect of scenario period on Situation Awareness is predicted. However, this effect will be largely attributable to the experimental protocol.

Rationale

There is no obvious rationale for predicting that Situation Awareness would be determined by phase of operation. The Situation Awareness literature appears to apply almost exclusively to detection contexts, but the concepts can be extended to mitigation contexts as well. In the present study, however, the transition between the Detection and Mitigation phase is signalled by an alarm. This alarm will presumably affect all levels of Situation Awareness. First, it will orient the operator's attention to data that is relevant to the event, increasing their perception. Second, the alarm itself will suggest a fault condition or context that will likely improve their understanding of the event. Finally, the alarm may be associated with procedures or scripts that project the course of the event.

4.1.4 Interaction effect between display type and scenario type

A directional effect of interaction is predicted for display type and scenario type. Ecological interfaces are expected to yield better Situation Awareness in Beyond-design basis scenarios as compared to both Advanced and Traditional interfaces. This advantage is expected to be smaller or non-existent for Within-design basis scenarios.

Rationale

Ecological Interface Design was specifically conceived as a design framework that identifies information needed to support anomaly detections, decision making and action under beyond-design basis events. Although Situation Awareness arose as a construct after the theoretical foundations of EID were established, the general theory of Situation Awareness pertains to event detection and understanding. Much of the empirical literature on EID supports the theory-based prediction that the performance advantages of ecological interfaces over traditional interfaces are more pronounced in beyond-design basis events as compared to within design basis events (Vicente, 2002). This finding was largely confirmed by Jamieson (in press) in a full-scope industrial simulator. The prevailing explanation for this finding is that within design basis events do not tax operators sufficiently to force them to rely on the additional information support provided by the Ecological interface. Note that the predicted performance is still expected to be lower for beyond-design basis events compared to within-design basis events.

The directional effect of the interaction between scenario type and Advanced interface design is more nuanced. If the Advanced interface provides configural display support for the specific event in question, then we would expect the Advanced interface to be comparable to the Ecological interface. If it does not, then we would expect the Advanced interface to be comparable with the Traditional interface. This exposes the claimed benefit of EID; that it provides support for knowledge-based processing under all event types.

4.1.5 Interaction effect between display type and scenario period

A directional effect of interaction between display type and scenario period is predicted. Both the Ecological and Advanced displays are expected to support better Situation Awareness (as compared to the Traditional interface) in the Detection phase. A further advantage for the Ecological interface over the Advanced interface is hypothesized.

Rationale

The more extensive use of analogue coding in the Ecological and Advanced (particularly trends) should support faster and more effective detection of anomalies as compared to the Traditional interface. Further, the configural displays that characterize the Ecological interface are expected to support more effective pre-alarm diagnosis compared to the Advanced interface (Bennett et al., 2006). In contrast, the more clearly defined and prescribed activities of the Mitigation phase are expected to be more robust to interface effects. That is, when mitigation responses can be generated through more channels (e.g., training, procedures, creative problem solving) than just the interface, the benefit of the interface is expected to diminish.

4.1.6 Interaction between display type, scenario type, and scenario period

A three-way directional effect of interaction is predicted between display type, scenario type, and scenario period. Situation Awareness amongst operators using the Ecological interface is expected to be higher than either the Advanced or Traditional interfaces in the Detection phase of Beyond-design basis scenarios.

Rationale

The three-way interaction hypothesis is deduced from a superposition of the two 2-way interactions (see Figure 10). For Within-design basis scenarios, we do not expect any interface effects; thus a straight line. For Beyond design basis, we hypothesize an asymmetrical kinked line reflecting an advantage for the Ecological interface. For the Detection phase, we hypothesize an asymmetrical kinked line showing an advantage for EID, but a relatively flat line for the Mitigation phase.

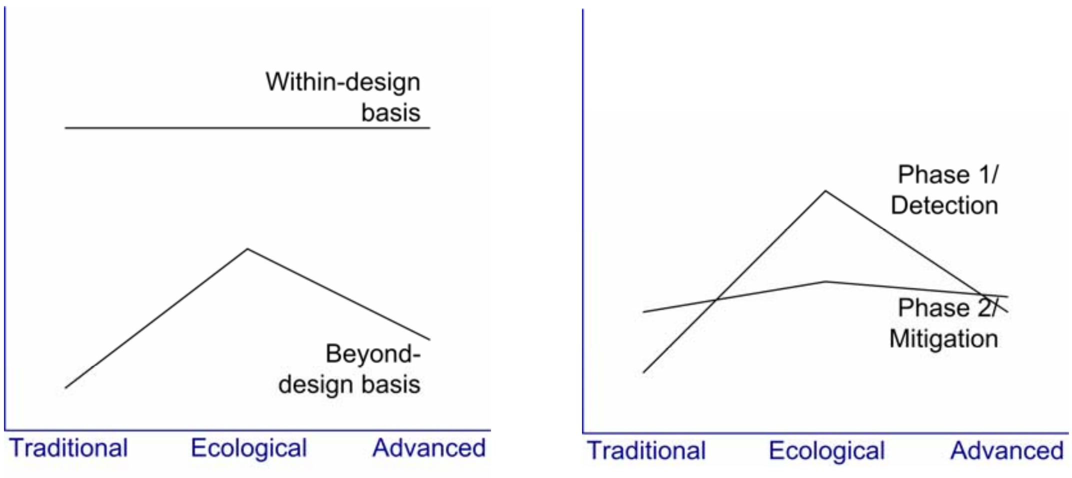


Figure 10. Two 2-way interactions

Figure 11 illustrates the expected 3-way interaction based on the two 2-way interactions. For Within-design basis in the Detection phase, we would expect a mildly kinked line given we are superimposing a straight line and a kinked line. However, for Beyond-design basis events in the Detection phase, we expected a more pronounced kink because two kinked-lines of the same pattern are superimposed.

For Within design basis events in the Mitigation phase, we also expect a mildly kinked line as in Within-design basis in the Detection phase (straight line and slightly kinked line). However, for

Beyond-design basis events in the Mitigation phase, we would expect a relatively small kinked line as we are superimposing a kinked line with a mildly kinked line.

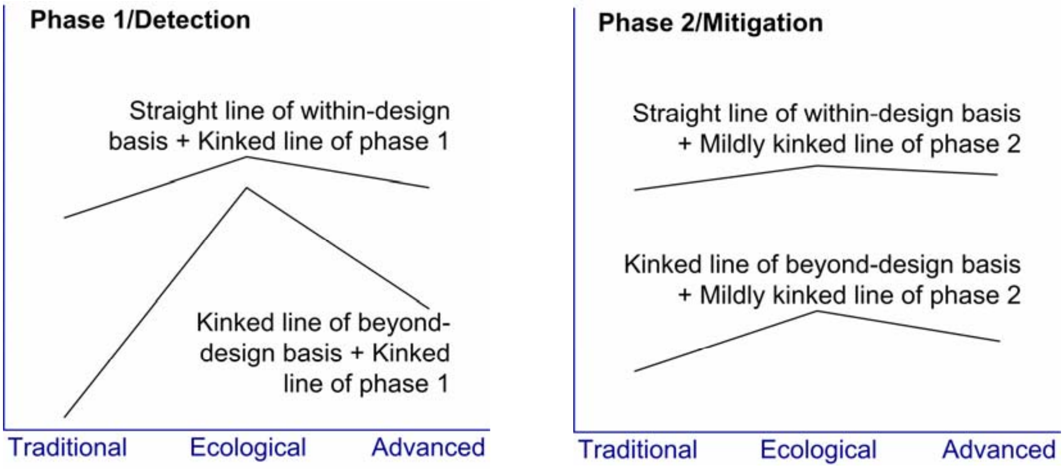


Figure 11. 3-way interaction effect by superimposing diagrams in Figure 10

In more descriptive language, we expect to observe markedly better Situation Awareness for operators using the Ecological interface to detect beyond-design basis events. In contrast, we expect at best modest advantages for the Ecological interface during the Mitigation phase, largely irrespective of event type. Overall, we would expect the three-way interaction to be weak.

Typically, 3-way interaction hypotheses are not deduced by looking at two 2-way interaction effects, but rather by looking at whether one variable would modify a 2-way interaction effect on a theoretical basis. For instance, a typical interpretation would be that the phase changes the interaction between interfaces and scenarios. In this case, the summation of the two 2-way interaction effects produces a 3-way interaction effect.

4.2 Hypotheses for Workload

4.2.1 Main effect of display type

A main effect of display type is predicted. That is, we expect to observe differences in Workload as a function of display type. A directional effect of display type is also predicted. The Ecological interface is predicted to result in lower Workload compared to both the Advanced and Traditional interfaces.

Rationale

Ecological psychology theory asserts that perception is largely effortless. In contrast, inferential reasoning is effortful. Ecological Interface Design seeks to replace reasoning (i.e., knowledge-based behaviour) with perception (i.e., skill-based behaviour) by embedding inferred relationships in perceptual-motor signals. An Ecological interface should therefore allow an operator to perform the same task at a reduced Workload because the structure of the task-environment can be perceived. Similarly, the Advanced interface makes use of analogue coding to promote perception over inferential reasoning. However, the extent of this coding is not as broad as that employed in the Ecological Interface.

Caveats: Two caveats should be noted. First, the directional effects predicted here are specific to abnormal process events. Under normal (i.e., steady-state) operation, EID theory would predict no systematic effect of interface type. Second, the operators' familiarity with the interface design frameworks would tend to reduce the effect predicted above. The Traditional interface bears resemblance to the actual interface with which the operators work on a regular basis. The Advanced interface also shares common forms with the actual operating displays. In contrast, the Ecological interface differs quite substantially from the existing displays. Thus, while the effect predicted above is consistent with theory, it does not give consideration to the likely occurrence of the operators incurring a Workload increment due to their greater unfamiliarity with the Ecological interface.

4.2.2 Main effect of scenario type

A main effect of scenario type is predicted. That is, the two types of scenarios are expected to require different levels of Workload. Specifically, we predict a directional effect of scenario type on Workload. Participants are expected to exert less Workload on Within-design basis scenarios as compared to Beyond-design basis scenarios.

Rationale

Beyond-design basis scenarios are expected to be more difficult because they require a greater degree of inferential reasoning than Within-design basis events.

4.2.3 Main effect of scenario period

A main effect of scenario period on Workload is predicted. A directional effect of period on Workload is predicted such that the Mitigation phase will require higher Workload than the Detection phase.

Rationale

The Mitigation phase calls for more overt action on the part of the operator. Whereas in the Detection phase he/she will be largely concerned with monitoring and some amount of display navigation, in the Mitigation phase he/she will be concerned with these activities in addition to planning and executing of control actions.

4.2.4 Interaction effect between display type and scenario type

A directional effect of interaction is predicted for display type and scenario type. Ecological interfaces are expected to allow for reduced Workload in Beyond-design basis scenarios as compared to both Advanced and Traditional interfaces. This advantage is expected to be smaller or non-existent for Within-design basis scenarios.

Rationale

As noted above, the EID framework is specifically conceived as a means of supporting operators in the detection and mitigation of beyond-design basis events. The exposure of a functional model of the process in the interface is expected to reduce the degree of inferential reasoning required in responding to process disturbances.

Caveat: The same caveats noted for the main effect of interface also apply here. It is possible that the brief training period available for the Ecological interface will not be sufficient to familiarize the operators with the Workload offsetting aspects of the design.

4.2.5 Interaction effect between display type and scenario period

A directional effect of interaction between display type and scenario period is predicted. Both the Ecological and Advanced interfaces are expected to have lower Workload as compared to the Traditional interface in the Detection phase. Moreover, the Ecological interface should result in lower Workload than the advanced interface. In contrast, no differences between interface types are expected in the Mitigation phase.

Rationale

The more extensive use of analogue coding in the Ecological and Advanced interface (particularly trends) should support detection of anomalies at low Workload as compared to the Traditional interface. Further, the configural displays that characterize the Ecological interface are expected to support more effective pre-alarm diagnosis compared to the Advanced interface (Bennet et al., 2006). In contrast, the more clearly defined and prescribed activities of the Mitigation phase are expected to be more robust to interface effects on Workload.

4.2.6 Interaction effect between scenario type and scenario period

A directional effect of interaction is expected between scenario type and scenario period. Specifically, the increase in Workload from the Detection to the Mitigation phases is expected to be smaller for Within-design basis events than the increase for Beyond-design basis events.

Rationale

Detecting violations of constraints is seen as being largely indifferent to whether those violations are anticipated in the design basis. In contrast, mitigating beyond-design basis disturbances is expected to require greater Workload than mitigating within-design basis disturbances. Thus, a Workload difference between scenario types is expected to emerge in transitioning from the detection to Mitigation phases.

5. DATA ANALYSIS APPROACH

This section summarizes the statistical data analysis approach for the experiment. There are no exact answers to the methodological issues raised below, and the solutions that we found reasonable in this study do not exclude other acceptable analysis approaches.

5.1 Parametric vs. non-parametric statistics

Parametric statistical tests presume an interval or ratio level of measurement, and normal distribution of the population from which the sample data are drawn. However, many statisticians argue that parametric tests are sufficiently robust to make non-parametric tests unnecessary. This view is implicitly reflected by the strong emphasis on parametric methods in statistical text books and software packages. Furthermore, a number of studies have documented the robustness of ANOVA (Maxwell & Delaney, 2000), and non-parametric tests do not even exist for complex factorial experimental designs. However, some statisticians still dispute the general suitability of parametric tests. Since the mathematicians disagree on this matter (see Howell, 1997), it is difficult for applied researchers to judge about the appropriateness of the two approaches.

A following pragmatic approach is taken here:

- Parametric tests are used unless distributions are asymmetrical or skewed in opposite directions across treatment populations. ANOVA is not robust to these particular violations of the normality assumption (Kirk, 1995), and whenever these deviations occur, the data may have to be transformed or treated non-parametrically (Howell, 1997). Thus, we are not testing the normality of the distributions as such, but search actively for pre-defined distribution patterns that are known to destruct ANOVA. For a within subject design with six experimental units, it is meaningless to examine distributions on the cell level (there are only six observations per cell). We therefore decided to aggregate the data across all experimental conditions, and per level of each manipulated variable. It can not be claimed that this approach protects against all possible threats to the validity of ANOVA, but it seems to be the best compromise given the experimental design and the low number of subjects. Distributions are evaluated through histograms, normal probability plots, and the Shapiro-Wilk W test (StatSoft, 2004).
- Parametric tests are used unless the dependent variables are clearly on the nominal or ordinal level of measurement. Performance indicators are otherwise assumed to be in the "gray" area between the ordinal and interval level of measurement, and are therefore treated as if they conform to interval scales (Pedhazur & Pedhazur Schmelkin, 1991).

5.2 Statistical significance testing

According to the Task Force on Statistical Inference (TFSI), it is always better to report actual p-values than to make dichotomous accept-reject decisions with a predefined significance level (Wilkinson, 1999). Even though p-values directly express the likelihood that statistical results have occurred by chance, the convention presented in Table 13 was used to guide the interpretation.

Table 13. Interpretation of p-values

p-values	Interpretation
$p \leq .01$	Strong indication that the effect is systematic
$.01 < p \leq .05$	Indicates that the effect is systematic
$.05 < p \leq .10$	Suggests that the effect may be systematic
$p > .10$	Indicates that the effect occurred by chance

5.3 One-tailed vs. two-tailed significance tests

When the manipulated variables are expected to produce an effect in a predefined direction, the hypothesis is said to be directional. One may then apply a one-tailed test of statistical significance. If the researcher expects experimental effects, but in either direction, the hypothesis is non-directional. A two-tailed test of statistical significance should then be used. One-tailed tests produce more statistical power than two-tailed tests, but are highly controversial. Even though many statistical textbooks include one-tailed significance testing as an option, such tests are often regarded speculative and therefore ignored by statistical software packages. Kimmel (1957) developed three criteria that should be fulfilled, in order to use a one-tailed test:

1. Observed difference in the unpredicted direction should be psychologically meaningless.
2. Results in the unpredicted direction should never be used differently than results that occur by chance.
3. The directional hypothesis should be deducible from psychological theory, but results in the opposite direction should not be supported by coexisting theory.

Statistical textbooks are often liberal and leave it up to the researcher's judgment to choose between directional and non-directional significance tests. To be on the safe side, we used two-tailed tests of statistical significance if Kimmel's criteria were not satisfied.

5.4 Effect size

TFSI encourage researchers to present effect sizes for primary outcomes (Wilkinson, 1999). Effect size indicators reveal the practical significance of the findings, and inform future power analyses and meta-analyses (ibid.). However, the interpretation of effect sizes is not straight forward. Fern and Monroe (1996) identify the following problematic issues:

- The impact of an effect is a function of its size, but also the perceived value of a unit change on the dependent variable. For example, a 1 percent reduction in the death rate for cancer is a small effect of a new drug, but would still be regarded scientifically noteworthy due to the magnitude of the cancer disease and the importance of saving lives. Thus, practical significance is more than effect size.
- The levels of a fixed treatment are arbitrarily determined by the experimenter, and do not represent the population of all possible treatment levels. It is therefore impossible to estimate the population component of variance in the dependent variable attributable to fixed treatments (StatSoft, 2004). Thus, effect sizes for fixed effects can only be generalized across studies that have chosen identical levels of experimental manipulation on the fixed treatments.

- The total variance estimate in the denominator of the effect size is strongly affected by subjective judgment. That is, researchers aggregate the data according to their needs and calculate partial effect size indicators (see Cohen, 1973).
- Comparison of effect sizes across experimental designs with different error terms, sample sizes, and/or number of levels on the manipulated variables can be misleading.
- The effect size is affected by the type of study. It is, for example, difficult to develop strong experimental manipulations in a controlled laboratory environment (O'Grady, 1982).

Fern and Monroe (1996) still conclude that it is informative to report effect sizes along with the level of statistical significance for primary results. However, the effect size must be interpreted within specific research contexts and in light of the weaknesses described above. TFSI is probably making the same point by stating that effect sizes should be reported in the context of previously reported effects (Wilkinson, 1999).

The most appropriate effect size indicator for fixed effects is omega-squared (ω^2) (Howell, 1997). For within-subject designs it makes sense to ignore the subject effect (block) in the total variance estimate of the effect size denominator, and calculate the partial omega-squared (Kirk, 1995)[5]. Effects sizes for random factors are estimated by the intraclass correlation (ρ), or the partial intraclass correlation for within-subject designs (ibid.). Omega-squared and the intraclass correlation are both explained variance indicators of effect size, and express the amount of variance in the dependent variable that is accounted for by the experimental manipulation (Fern & Monroe, 1996). Standardized mean difference indicators of effect size, such as Cohen's d, have a less intuitive interpretation and can be calculated directly from omega-squared (Kirk, 1995). Such measures of effect size are therefore not reported.

5.5 Confidence intervals

According to TFSI, interval estimates should be given for principal effects. It is also recommended to include interval estimates in graphical representations whenever possible (Wilkinson, 1999). The rationale for this seems to be that interval estimates give information about precision as well as statistical significance. The most popular interval estimate is the confidence interval, and no other alternatives will be considered here.

As for effect sizes, the interpretation of confidence intervals can be problematic. A common misunderstanding is that the population mean lies within the limits of the 95% confidence interval. The correct interpretation is that 95% of the confidence intervals would correctly bracket the true value of the mean if we repeatedly calculated the interval from many independent random samples of the same size (Pedhazur & Pedhazur Schmelkin, 1991; StatSoft, 2004).

Further, it is not advisable to make judgments about statistical significance from graphical inspection of overlaps between confidence intervals (Schenker & Gentleman, 2001; Payton, Greenstone & Schenker, 2003; Cumming & Finch, 2005). If researches use the separation of 95% confidence intervals as a significance criterion, effects that occurred by random are declared

---

[5] The calculation method used here was based upon the observed F-value, the number of levels on each manipulated variable (p, q, r) and the number of blocks/operators (n). A full account of the formulas can be found in Kirk (1995), p. 399 and p. 460. The formulas were expanded to include three manipulated variables instead of two variables as described by Kirk. Formula for main effects: $\omega^2_p= ((p-1)(F_p-1))/((p-1)(F_p-1)+npqr)$. Formula for two-way interaction effects: $\omega^2_{pq}=((p-1)(q-1)(F_{pq}-1))/ ((p-1)(q-1)(F_{pq}-1)+npqr)$. Formula for three-way interaction effects: $\omega^2_{pqr}=((p-1)(q-1)(r-1)(F_{pqr}-1))/ ((p-1)(q-1)(r-1)(F_{pqr}-1)+npqr)$.

systematic more often than desired, and many systematic effects remain undetected (Schenker & Gentleman, 2001).

For within-subject designs, confidence intervals around the cell means are irrelevant for inferences about mean differences (Loftus & Masson, 1994; Cumming & Finch, 2005). This is because the variance among subjects affects the size of the confidence interval, even though subject differences are removed from the error term in the repeated measures analysis of variance. Belia et al. (2004) showed that the majority of researchers misinterpret graphical representations of confidence intervals for within-subject effects by making incorrect inferences about mean differences.

It is concluded that interval estimates should be given for principal effects, as recommended by TFSI. Interval estimates provide valuable information regarding the precision of findings, but inferences about statistical significance may be speculative. TFSI suggest that graphical representations of results should include interval estimates whenever possible. Since most researchers misinterpret figures that include confidence intervals for within-subject effects, such graphical representations are probably not advisable. Interval estimates for the cell means of within-subject effects is therefore reported in tables below the graphical display of effects.

5.6 Additive vs. non-additive model

A within subject-design is a type of randomized block design (Kirk, 1995; Howell, 1997; Montgomery, 1997). The purpose of *blocking* is to isolate the effect of a nuisance variable, i.e. partition out variance that would otherwise be included in the error term (see section 2.2). This is advantageous because smaller error terms provide more statistical power in the analysis of variance. For within-subject designs, each subject (operator or crew) form a block. Measurements on all treatment levels are then obtained per block by repeated testing of the same subjects. The subjects' individual response tendencies can thereby be partitioned out from the error term. It is essential that the units within a block are assigned randomly to the experimental conditions. This is achieved by counterbalancing the presentation order of treatment levels across subjects (see section 2.2.4).

A Randomized Block Factorial design is denoted RBF-*pq* (assuming two treatments), where p and q represent the number of levels on each treatment. As explained in section 2.2, this kind of experimental design was used in the current study.

The simplest model for a RBF-*pq* design is:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \pi_i + e_{ijk}$$

Where Y_{ijk} is the score in the *i*th block and *jk*th treatment combination; μ is the grand mean of the population means; α_j is the treatment effect for population *j*, β_k is the treatment effect for population *k*; $(\alpha\beta)_{jk}$ is the joint effect of treatment levels *j* and *k*; π_i is the block effect for population *i*; and e_{ijk} is the residual error (Kirk, 1995, Howell, 1997). The model presumes no interactions between the blocking variable and the treatments, and is therefore an *additive* model. The residual error for the additive model is the portion of a score that remains when all other terms in the model are subtracted. Thus, several sources of error variance are pooled into one error term.

An alternative model for a RBF-*pq* design is:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \pi_i + (\alpha\pi)_{ji} + (\beta\pi)_{ki} + (\alpha\beta\pi)_{jki} + e_{ijk}$$

Where *$(\alpha\pi)_{ji}$* is the joint effect of treatment level *j* and block *i*, *$(\beta\pi)_{ki}$* is the joint effect of treatment level *k* and block *i*, and *$(\alpha\beta\pi)_{jki}$* is the joint effect of treatment combination *jk* and block *i*. Denotations for the remaining structural terms are identical to the additive-model. However, this is a *non-additive* model because interactions between the blocking variable and the treatments are presumed. There is no independent estimate of the residual error for the non-additive model. It can be shown, however, that the block-treatment interaction for each individual effect can serve as the error term in the analysis of variance (Howell, 1997).

There is controversy among statisticians on the structural model for randomized block designs. The advantages of the additive model over the non-additive model are more degrees of freedom and the absence of interaction terms in the residual error, which reduces the error term (Kirk, 1995; Howell, 1997). However, the assumption that the blocking variable and treatments do not interact is usually unrealistic in psychological research (Howell, 1997). This is especially true for human-machine interaction studies, where the operating crews may follow very different solution paths. Some statisticians propose mathematical transformations in order to remove interaction effects between blocks and treatment levels, but unfortunately, this is often impossible (Montgomery, 1997). Others have suggested that the researcher should conduct a test of non-additivity and thereby identify the appropriate model (Kirk, 1995). The most common approach is to employ the non-additive model under all circumstances, even though the model is conservative in certain situations (Howell, 1997). In fact, commercially available statistical software packages apply the non-additive model by default and seldom include tests of non-additivity.

Since subject-by-treatment interactions are common in psychological experiments, it seems reasonable to employ the non-additive model as a starting point. The appropriateness of the more powerful additive model can then be evaluated if the research hypotheses are not clearly supported by the non-additive tests.

5.7 Sphericity and compound symmetry assumption

When a treatment has more than two levels of experimental manipulation in a within-subject design, statistical assumptions about sphericity and compound symmetry enter the picture (Kirk, 1995; Howell, 1997). Compound symmetry means that the pooled variances within each treatment and the covariances among the treatments within blocks are homogeneous. Sphericity refers to a general condition where the components of the model are orthogonal. Thus, compound symmetry will fulfil the sphericity requirement (ibid.).

Assumptions about sphericity and compound symmetry rarely hold for complex experiments in realistic laboratory environments. Unfortunately, analysis of variance is <u>not</u> robust to violations of the assumptions either. It is therefore particularly important to test these assumptions and apply alternative statistical procedures when violations occur. One solution is to always employ the multivariate approach to repeated measures whenever there are more than two levels on a manipulated variable (StatSoft, 2004). However, the multivariate approach has low statistical power in studies with a small number of observations (Howell, 1997). In such situations, the (univariate) Greenhouse-Geisser or Huynh-Feldt adjustments (ibid.) may therefore be helpful when the sphericity and compound symmetry assumptions are violated. On the other hand, Monte Carlo studies have demonstrated that the power of the multivariate approach increases with the degree of violation of the compound symmetry assumption (Mendoza et al., 1974). It is therefore not given that studies with few observations will gain power by employing the univariate adjustments instead of the multivariate approach.

The pragmatic strategy chosen here is to employ the multivariate approach to repeated measures first. If this test indicates a systematic effect ($p \leq .05$), we know that the sphericity and compound symmetry assumption is no longer an issue, and that the power of the multivariate approach was sufficient. Further univariate testing is therefore unnecessary. If, however, the multivariate test suggests that an effect occurred by chance ($p > .10$) or might be systematic ($.05 < p \leq .10$), the potentially more powerful univariate test is carried out along with a test of sphericity (StatSoft, 2004). If the sphericity assumption is violated, the Greenhouse-Geisser and Huynh-Feldt univariate adjustments are calculated.

5.8 Comparison of means

According to TFSI, pairwise multiple comparison methods, such as the Tukey Honestly Significance Difference (HSD) test, are too conservative when preceded by an omnibus F test in a stagewise testing procedure (Wilkinson, 1999). The committee further points out that comparisons of all possible means may restrict researchers to uninteresting hypotheses, and are unnecessary for the understanding of experimental results (ibid.). Thus, comparison of means should be restricted to a few meaningful contrasts, trend analysis or other investigations of anticipated structure. The familywise error rate for planned comparisons should be controlled, even though this is not always the normal procedure (Howell, 2006). However, conducting even a few meaningful contrasts will dissect the experimental data set and reduce the number of sample points, and thus, marginalize the statistical power for the contrast analyses. In addition, multiple comparisons for within-subject designs is risky business due to the lack of protection against the sphericity assumption (ibid.). When there are more than two levels of manipulation on a within-subject factor, comparisons of means may therefore be speculative.

Given this situation, it makes sense to limit the comparison of means to graphical inspection of the effects, if there are no particular reasons that justify further statistical testing. This is described by Howell (2006) as the "minimalist school", which is a legitimate position, since there are no absolute criteria for when the statistical data analysis should stop and the interpretation of results begins. The following stepwise procedure was followed in this experiment:

1. We checked whether the overall effect was statistically significant - disregarding possible interpretations of the observed means (direction, trends, interaction pattern etc.).
2. We checked whether the effect was practically significant (effect size) and/or theoretically interesting.
3. If the observed effect was systematic (see 1 above) <u>and</u> of theoretical/practical interest (see 2 above), we went further to interpret the effect as described below.
4. The interpretation was limited to graphical inspection of the effect in combination with a search for "reasonable findings". A reasonable finding fits into a logical pattern and is corroborated by independent facts, such as (a) when a hypothesised effect is reproduced by statistically unrelated dimensions of one theoretical construct, (b) a finding is consistent with previous results (replication) or accepted theory, or (c) when effects go in opposite directions for measures that reflect opposite phenomena.

6. RESULTS

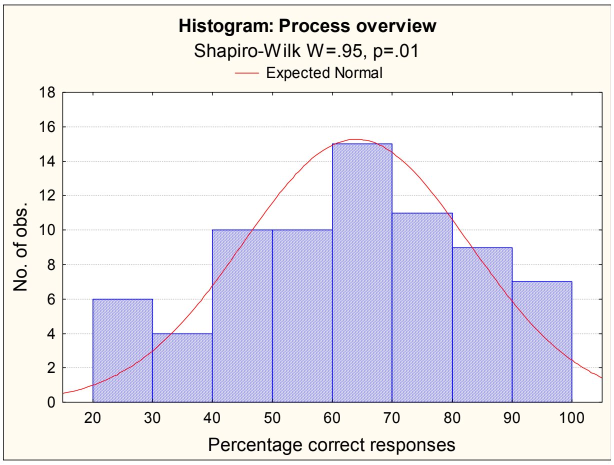
6.1 Data aggregation and descriptive statistics

6.1.1 Process overview

After the completion of each scenario period, the Turbine Operator in each crew indicated whether important process parameters had increased, decreased, or remained stable over the last minutes. The simulator was frozen, and the user interface unavailable to the operators while they responded to the questions. In parallel, the process expert scored the actual drift of the process parameters by inspecting trend diagrams, event lists and other simulator data. The operators were given the score one (1) whenever there was agreement between the process expert and the operator rating (e.g., both said that the pressure in the condenser had decreased). When the operators disagreed with the process expert, the score zero (0) was given (e.g., the operator believed that the reactor temperature had increased, while the process expert concluded that the temperature was stable). The number of parameters that were scored per administration varied between 9 and 12 depending on the scenario characteristics; see Appendix 3 (chapter 12). As explained in section 3.4, the Process overview score was calculated as the percentage of correct operator responses per administration.

The Process overview items originally included "global" process parameters that addressed the overall state of the system, but also "local" parameters related to the scenario development. However, Process overview was calculated for all parameters together, since the validity of the global vs. local distinction may be questioned in a highly interconnected system, and because separation of the scores failed to reveal interesting differences between the two types of parameters.

The distribution of the Process overview variable for all subjects and treatment combinations is presented in Figure 12 along with basic descriptive statistics. According to the Shapiro-Wilk test, the variable is <u>not</u> normally distributed (W=.95, p=.01). However, the distribution is reasonably symmetrical and probably poses no threat to the robustness of ANOVA (see section 5.1).



Variable	Descriptive Statistics Valid N	Mean	Minimum	Maximum	Std.Dev.
process overview	72	63.96	30.00	100.00	18.81

Figure 12. Distribution of Process overview

6.1.2 Scenario understanding

During the scenario periods, the process expert rated the Turbine Operators' level of Scenario understanding as explained in section 3.5. As already mentioned, a simple scale was used for this purpose, where zero (0) means 'poor understanding', one (1) 'some understanding', two (2) 'good understanding', and three (3) 'full understanding'. The operators were probed and scored twice per scenario period, and the unweighted average of the two scores expressed the degree of Scenario understanding for each period. The first probe occurred during the Detection phase (mini-freeze), the second probe initiated the break in-between the Detection and Mitigation phase (full freeze), the third probe occurred during the Mitigation phase (mini-freeze), and the fourth probe concluded the scenario after the Mitigation phase (see Figure 13). Aggregated scores ranged from 0 to 3, where a higher score indicated a better understanding of the scenario.

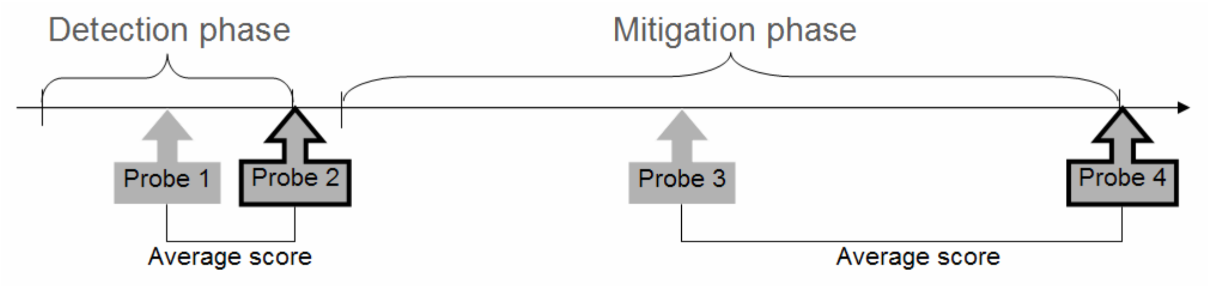
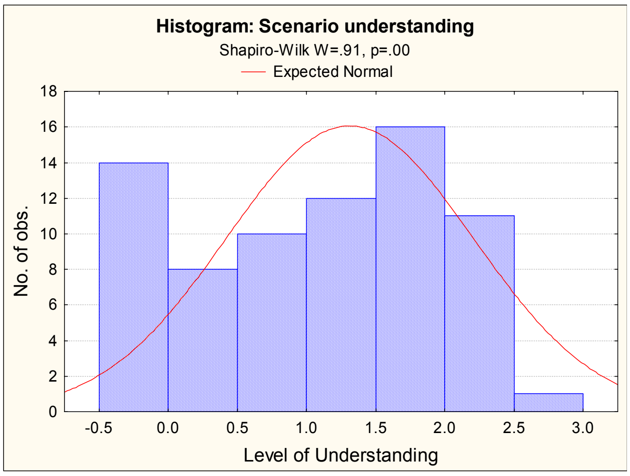


Figure 13. Timing of Scenario understanding probes

The distribution of the Scenario understanding variable for all subjects and experimental conditions is presented in Figure 14 along with basic descriptive statistics. According to the Shapiro-Wilk test (W=.91, p=.00), the variable is not normally distributed. There seems to be a floor effect towards the lower end of the scale, which suggests that poor levels of understanding are not sufficiently differentiated. Examination of the raw data revealed that 14 of the 72 scores per scenario period were zero (0). Transformation of the variable was therefore not an option. The shape of the distribution is still rather symmetrical, with a peak around the mean. ANOVA will therefore be used, but a possible reservation is made with respect to the robustness of the test.



Variable	Descriptive Statistics Valid N	Mean	Minimum	Maximum	Std.Dev.
scenario understanding	72	1.31	0.00	3.00	0.89

Figure 14. Distribution of Scenario understanding

6.1.3 Metacognitive accuracy

Metacognitive accuracy is the absolute difference between self-rated and Actual task performance. The measures and the calculation method are explained in section 3.6.3. In the following, we will describe the results for Self-rated task performance and Actual task performance separately before the examination of the Metacognitive accuracy variable itself.

Self-rated task performance

After the completion of each scenario period, the Turbine Operator in each crew responded to a self-rating questionnaire. A 4-item rating scale was used after the Detection phase and an 8-item rating scale after the mitigation period (see section 3.6.1).

The 4-item rating scale had an inter-item reliability of α=.81. There were no strong ceiling or floor effects for any of the items, and the means and standard deviations across items were rather similar (see Figure 15). Removing items had no beneficial effect on the reliability coefficient. A factor analysis revealed only a single factor with an eigenvalue greater than 1 (Kaiser's criterion), which confirms that the 4-item scale measured only <u>one</u> construct (StatSoft, 2004).

The 8-item rating scale had an inter-item reliability of α=.89. There were no strong ceiling or floor effects for any of the items, and the means and standard deviations across items were rather similar (see Figure 15). Item 3 is a possible exception in this respect, as the Turbine Operators seem to rate their cooperation with the rest of the crew somewhat higher than the other items. This makes sense, since the scenarios were not designed to challenge teamwork. However, removing items had no beneficial effect on the reliability coefficient and there was therefore no basis for elimination of item 3. A factor analysis revealed only a single factor with an eigenvalue greater than 1 (Kaiser's criterion), which confirms that the 8-item scale measured only <u>one</u> construct (StatSoft, 2004).

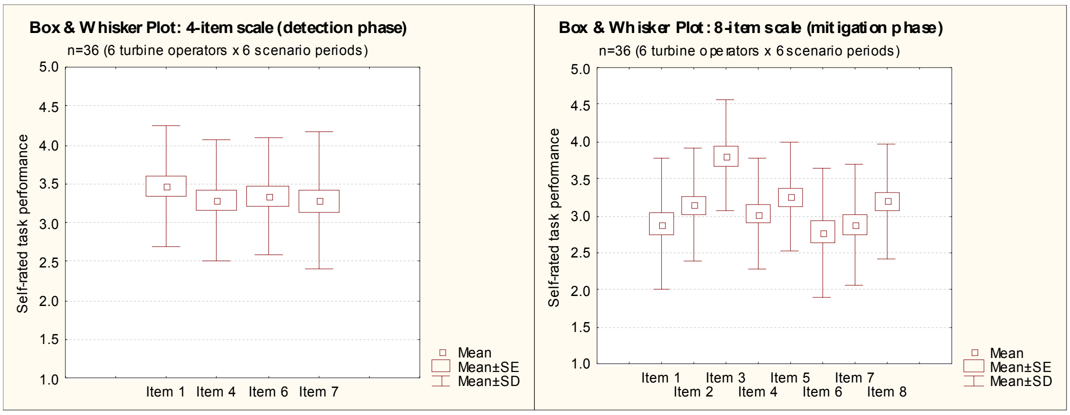
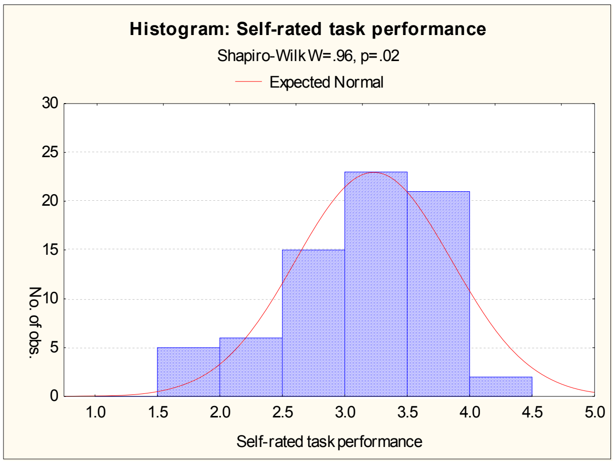


Figure 15. Box & Whiskers plot of 4-item and 8-item self-rating scales

It was concluded that the Self-rated task performance scale was reliable. The unweighted average of the items was used as a total score for both the 4-item, and 8-item scale.

The distribution of the aggregated Self-rated task performance score for all subjects and treatment combinations is presented in Figure 16 along with basic descriptive statistics. According to the Shapiro-Wilk test (W=.96, p=.02), the variable is <u>not</u> normally distributed, but somewhat skewed to the left.



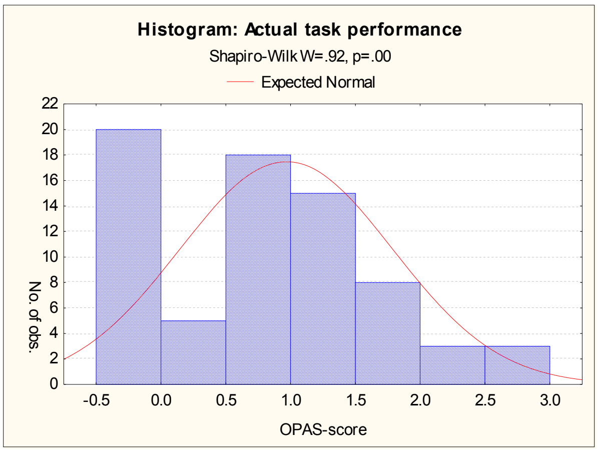
Variable	Descriptive Statistics Valid N	Mean	Minimum	Maximum	Std.Dev.
Self-rated task performance	72	3.23	1.75	4.50	0.63

Figure 16. Distribution of Self-rated task performance

Actual task performance

The Operator Performance Assessment System (OPAS) was used to measure Actual task performance (Skraaning, 1998; Skraaning, 2003). As described in section 3.6.2, OPAS measures the extent to which operators carry out their taskwork in accordance with scenario solutions that are prescribed by process experts. A simple scoring system is used, where the operators earn points for completing prescribed performance items. Each performance item depicts alternative operator activities that are rewarded by 0, 1, 2 or 3 points. A higher score indicates better performance. The main performance index is the unweighted average of the operators' scores on all performance items per scenario period. This experiment did not calculate separate scores for different performance categories, such as detections or actions, i.e., the Actual task performance was expressed as a general score that incorporated all aspects of operator performance.

The distribution of the Actual task performance variable for all subjects and experimental conditions is presented in Figure 17 along with basic descriptive statistics. According to the Shapiro-Wilk test (W=.92, p=.00), the variable is not normally distributed. There seems to be a floor effect towards the lower end of the scale, which suggests that poor levels of performance are not sufficiently differentiated. Examination of the raw data revealed that 20 of the 72 OPAS scores were zero (0). Transformation of the variable was therefore not an option. Except for the floor effect, the shape of the distribution is relatively symmetrical with a peak around the mean.



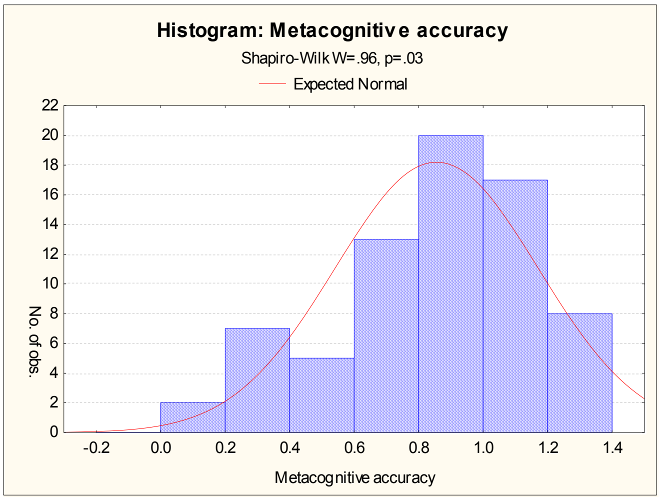
Variable	Descriptive Statistics				
	Valid N	Mean	Minimum	Maximum	Std.Dev.
actual task performance	72	0.97	0.00	3.00	0.82

Figure 17. Distribution of Actual task performance

Metacognitive accuracy

Metacognitive accuracy expresses the extent to which operators are able to realistically assess their own performance. The score is obtained as described in section 3.6.3.

The distribution of the Metacognitive accuracy variable for all subjects and treatment combinations is presented in Figure 18 along with basic descriptive statistics. According to the Shapiro-Wilk test (W=.96, p=.03), the variable is <u>not</u> normally distributed. This is probably not a threat to the robustness of ANOVA, since the distribution seems to be rather symmetrical with a peak around the mean.



Variable	Descriptive Statistics				
	Valid N	Mean	Minimum	Maximum	Std.Dev.
Metacognitive accuracy	72	0.86	0.00	1.37	0.32

Figure 18. Distribution of Metacognitive accuracy

6.1.4 Workload

After the completion of each scenario period, the Turbine Operator in each crew responded to a Workload questionnaire. A 5-item self-rated task complexity scale was used for this purpose (see section 3.7).

The scale had an inter-item reliability of α=.89. There were no strong ceiling or floor effects for any of the items, and the means and standard deviations across items were rather similar (see Figure 19). Removing items had no beneficial effect on the reliability coefficient. A factor analysis revealed only a single factor with an eigenvalue greater than 1 (Kaiser's criterion), which confirms that the scale measured only <u>one</u> construct (StatSoft, 2004). It was concluded that the self-rated task complexity scale was reliable, and the unweighted average of the 5 items were used as the total Workload score. The scale was inverted to let higher scores indicate higher levels of Workload.

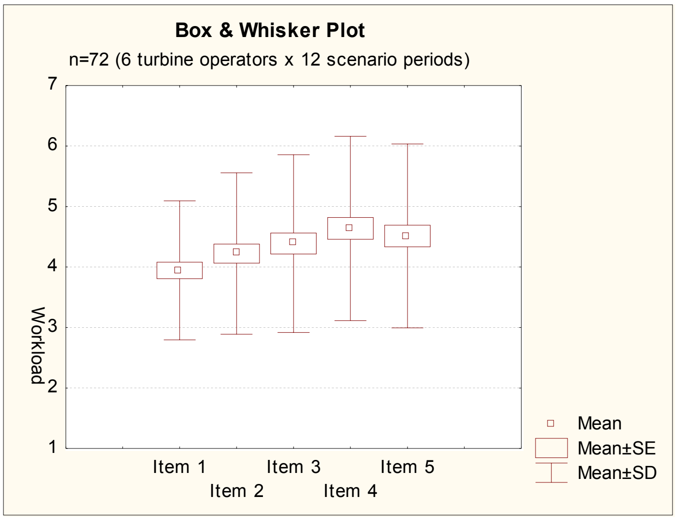
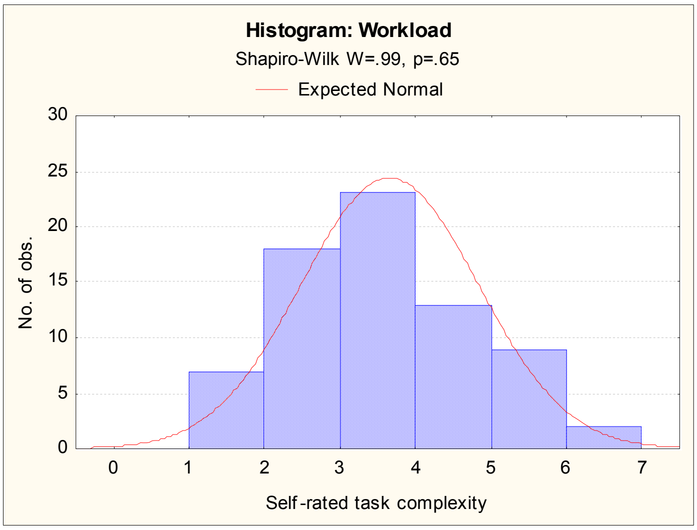


Figure 19. Box & Whiskers plot of Workload scale

The distribution of the Workload variable for all subjects and treatment combinations is presented in Figure 20 along with basic descriptive statistics. According to the Shapiro-Wilk test (W=.99, p=.65), the Workload variable is normally distributed.



Variable	Descriptive Statistics				
	Valid N	Mean	Minimum	Maximum	Std.Dev.
Workload	72	3.66	1.20	6.40	1.18

Figure 20. Distribution of Workload

6.2 Relationship among dependent variables

The Process overview, Scenario understanding, and Metacognitive accuracy variables are meant to be complementary indicators of Situation Awareness (SA), i.e., the measures are supposed to express different aspects of the underlying construct. The amount of shared variance (r^2) among the Situation Awareness indicators turned out to be low as expected (see Table 14). Even though two of the correlations among the Situation Awareness variables are systematic when judged by their associated p-values, the amount of shared variance never exceeds 6% for any of the variable pairs. It is therefore reasonable to maintain the assumption that the three Situation Awareness measures represent rather independent dimensions of SA. Knowing that MANOVA is disadvantageous for unrelated dependent variables and small sample sizes (Bray and Maxwell, 1985), the preferred analysis approach is to employ separate ANOVAs for each Situation Awareness measure.

Table 14. Correlations among Situation Awareness measures

Var. X & Var. Y	Correlations						
	Mean	Std.Dv.	r(X,Y)	r²	t	p	N
Process overview	63.96	18.81					
Scenario understanding	1.31	0.89	-0.03	0.00	-0.29	0.78	72
Process overview	63.96	18.81					
Metacognitive accuracy	0.86	0.32	0.22	0.05	1.93	0.06	72
Scenario understanding	1.31	0.89					
Metacognitive accuracy	0.86	0.32	0.25	0.06	2.19	0.03	72

Since Metacognitive accuracy is the absolute difference between Self-rated task performance and Actual task performance, the relationship between these two variables and Metacognitive accuracy is mathematically given. Further empirical investigations are therefore meaningless in this respect.

We found no correlation between Self-rated task performance and Process overview or Scenario understanding. Actual task performance and Process overview were also unrelated, but a noteworthy positive correlation appeared between Scenario understanding and Actual task performance (r=.71, r^2=.52, p=.00). The relationship is not moderated by interface type, scenario type or scenario period, and is consistent across operators. This finding may suggest that the Scenario understanding variable has criterion validity (predicts performance), but it could also mean that the Actual task performance items included cognitive elements. A post-experimental classification of the items revealed that ‘detections’ and ‘inferences’ were important cognitive dimensions of the task performance requirements (see Figure 21). Consequently, the observed correlation is not a convincing demonstration of criterion validity for the Scenario understanding variable. Multi-categorization was used to classify items, since most of the items were complex and included more than one performance dimension (e.g. both inferences and actions). It was therefore impossible to isolate the cognitive elements of the measure to establish an independent task performance criterion based upon behaviour alone.

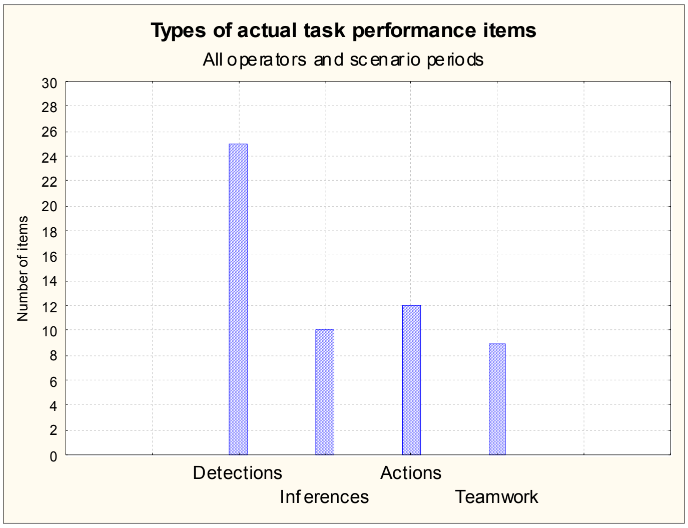


Figure 21. Types of actual performance items

Table 15 shows the correlations among Workload and the other dependent variables (including Self-rated task performance and Actual task performance, which were used to calculate Metacognitive accuracy). Workload and Process overview correlate negatively, meaning that higher levels of Workload are associated with reduced Process overview. The amount of shared variation is 16% (r=-.40, r^2=.16, p=.00), which is a moderately strong relationship. Even though Workload and Scenario understanding are systematically related, the amount of shared variation is only 7% (r=.27, r^2=.07, p=.02), meaning that the relationship is weak and probably not practically significant. The direction of the correlation is also counterintuitive (higher Workload is associated with better Scenario understanding). There is no systematic relationship between Workload and Metacognitive accuracy.

Workload and Self-rated task performance are negatively correlated, meaning that higher Workload is associated with lower Self-rated task performance. The amount of shared variation is 41% (r=-.64, r^2=.41, p=.00) which is a relatively strong relationship. There is a systematic relationship between Workload and Actual task performance, but the amount of shared variation is only 9% (r=.31, r^2=.09, p=.01) and may not have practical significance. The direction of the correlation is counterintuitive (higher Workload is associated with better task performance), which is consistent with the correlation between Workload and Scenario understanding. This is not surprising given the strong correlation between Scenario understanding and Actual task performance (r=.71, r^2=.52, p=.00).

Table 15. Correlations involving Workload

Var. X & Var. Y	Correlations Mean	Std.Dv.	r(X,Y)	r²	t	p
Workload	3.66	1.18				
Process overview	63.96	18.81	-0.40	0.16	-3.62	0.00
Workload	3.66	1.18				
Scenario understanding	1.31	0.89	0.27	0.07	2.37	0.02
Workload	3.66	1.18				
Metacognitive accuracy	0.86	0.32	0.03	0.00	0.25	0.80
Workload	3.66	1.18				
Self-rated task performance	3.23	0.63	-0.64	0.41	-6.92	0.00
Workload	3.66	1.18				
Actual task performance	0.97	0.82	0.31	0.09	2.68	0.01

6.3 Hypothesis testing

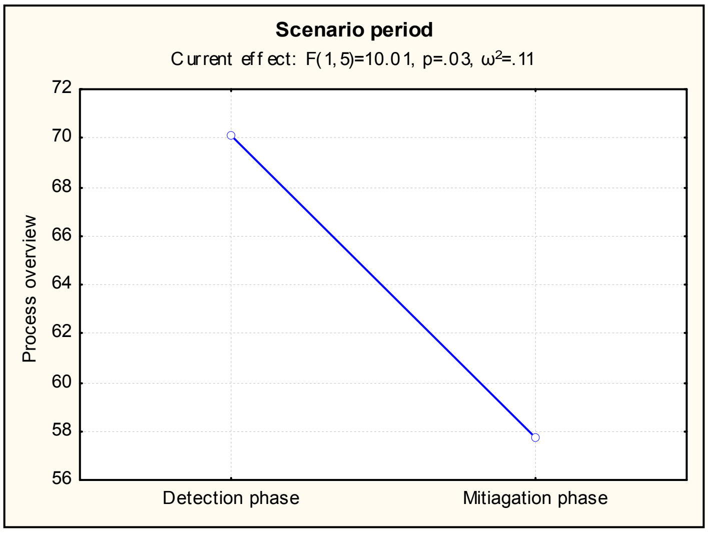
6.3.1 Process overview

The multivariate approach to repeated measures turned out to be sufficiently powerful to test whether the hypothesized effects on Process overview are supported by the data. Table 16 shows the outcome of this test for all effects.

Table 16. ANOVA results for Process overview

Effect	Multivariate tests for Process overview Sigma-restricted parameterization Effective hypothesis decomposition Test	Value	F	Effect df	Error df	p
Interface type	Wilks	0.95	0.10	2	4	0.91
Scenario type	Wilks	0.94	0.33	1	5	0.59
Scenario period	Wilks	0.33	10.01	1	5	0.03
Interface type*Scenario type	Wilks	0.09	20.32	2	4	0.01
Interface type*Scenario period	Wilks	0.79	0.53	2	4	0.62
Scenario type*Scenario period	Wilks	0.61	3.22	1	5	0.13
Interface type*Scenario type*Scenario period	Wilks	0.18	9.41	2	4	0.03

The main effect of scenario period is systematic ($F(1,5)=10.01$, p=.03), and is illustrated in Figure 22. It can be seen from the graph that the operators' have better Process overview in the detection than the mitigation phase. The effect explains 11 % of the total variance (ω^2=.11), which is a typical, and probably medium, effect size for experiments in realistic work environments. The distributions for the Detection phase and the Mitigation phase are rather symmetrical, and are not skewed in opposite directions. ANOVA is therefore presumed to be robust against violations of the normality assumption for this effect (see section 5.1).



	Confidence intervals			
	Mean	Std.Err.	-95.00%	+95.00%
Detection	70.14	1.33	66.72	73.56
Mitigation	57.78	3.15	49.69	65.87

Figure 22. Main effect of scenario period on Process overview

There is a systematic interaction effect between interface type and scenario type (F(2,4)=20.32, *p*=.01), and between interface type, scenario type and scenario period (F(2,4)=9.41, *p*=.03). Graphical inspection of these interaction effects made it obvious that the two-way interaction provides redundant and limited information compared to the three-way interaction. It was therefore decided to disregard the two-way interaction and present graphical results only for the three-way interaction.

The distributions for Traditional and Advanced displays were skewed slightly in opposite directions, but probably not to an extent that raises a concern with respect to the robustness of ANOVA (see Figure 23). All other distributions are symmetrical, or skewed consistently in one direction per manipulated variable. The overall conclusion is that ANOVA is robust against violations of the normality assumption for the three-way interaction (see section 5.1).

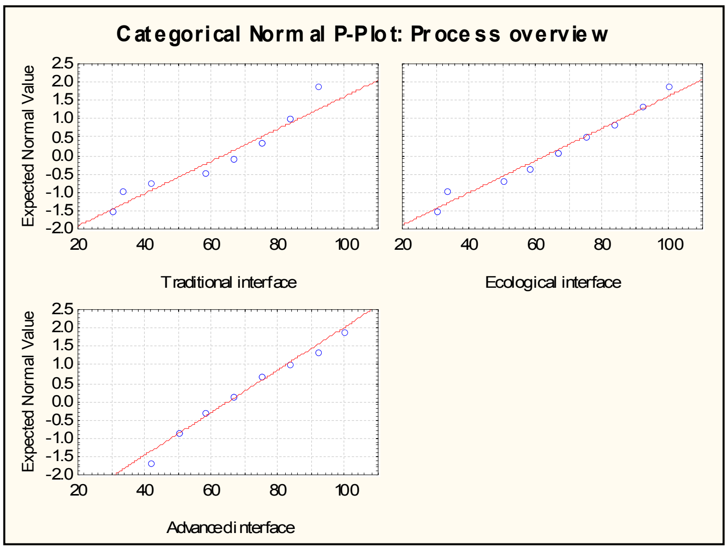
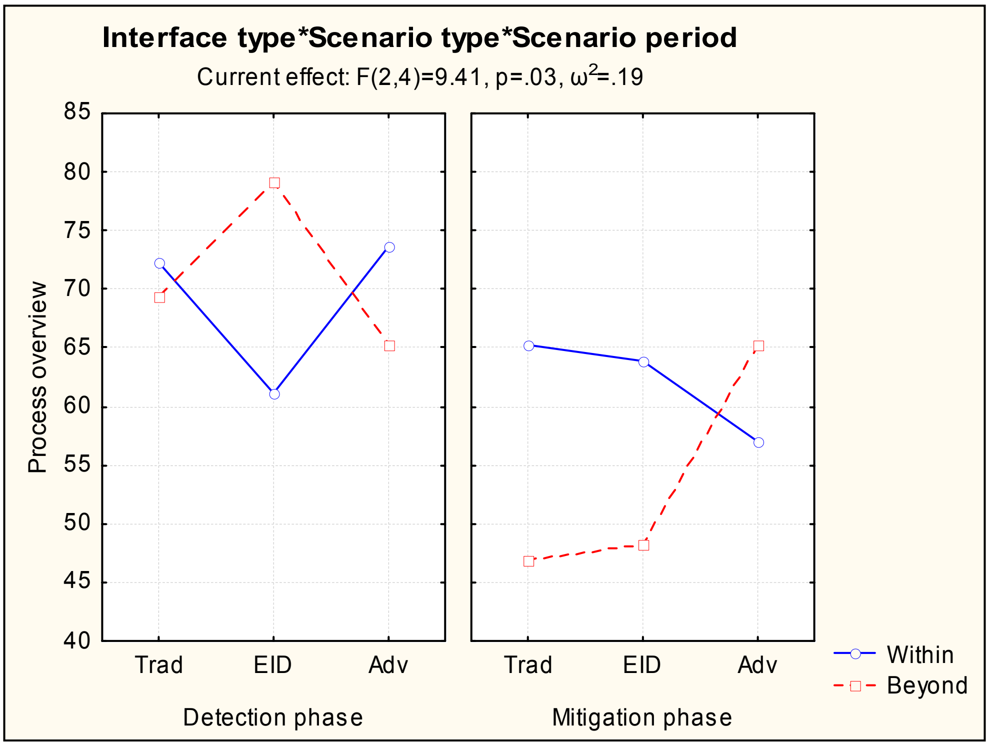


Figure 23. Categorical normal p-plot for Process overview

The interaction between display type, scenario type and scenario period on Process overview is presented in Figure 24. The effect explains 19 % of the total variance (ω^2=.19), which is a relatively large effect for experiments in realistic work environments.



Cell No.	Confidence intervals Interface type	Scenario type	Scenario period	Mean	Std.Err.	-95.00%	+95.00%
1	Traditional	Within	Detection	72.22	1.76	67.71	76.74
2	Traditional	Within	Mitigation	65.28	7.88	45.02	85.54
3	Traditional	Beyond	Detection	69.44	7.03	51.38	87.51
4	Traditional	Beyond	Mitigation	46.94	10.70	19.44	74.45
5	Ecological	Within	Detection	61.11	7.95	40.66	81.56
6	Ecological	Within	Mitigation	63.89	8.24	42.71	85.07
7	Ecological	Beyond	Detection	79.17	5.99	63.77	94.56
8	Ecological	Beyond	Mitigation	48.33	8.48	26.54	70.12
9	Advanced	Within	Detection	73.61	5.86	58.55	88.67
10	Advanced	Within	Mitigation	56.94	5.01	44.07	69.82
11	Advanced	Beyond	Detection	65.28	6.24	49.23	81.32
12	Advanced	Beyond	Mitigation	65.28	8.17	44.28	86.28

Figure 24. Three-way interaction effect for Process overview

It can be seen from the graph that Ecological displays enhance Process overview in the Detection phase of Beyond design basis scenarios, but impairs the overview in the Detection phase of Within design basis scenarios. Advanced displays improve the Process overview in the Mitigation phase of Beyond design basis scenarios, while Traditional and Ecological displays seem to facilitate Process overview in the Mitigation phase of Within design basis scenarios.

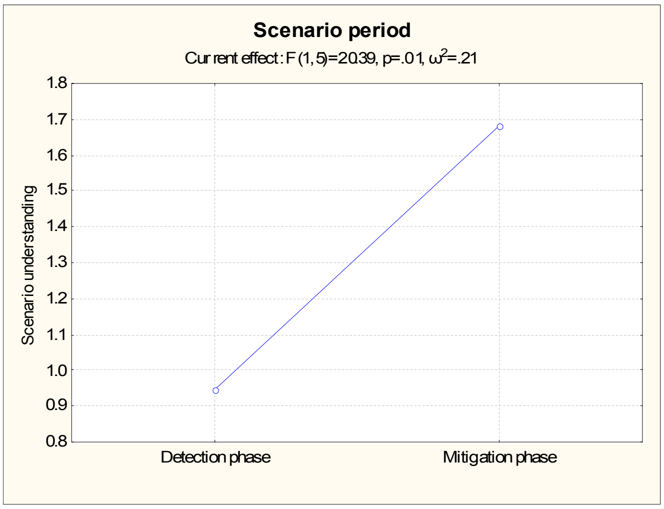
6.3.2 Scenario understanding

The multivariate approach to repeated measures turned out to be sufficiently powerful to test whether the hypothesized effects on Scenario understanding are supported by the data. Table 17 shows the outcome of this test for all effects.

Table 17. ANOVA results for Scenario understanding

Effect	Test	Value	F	Effect df	Error df	p
	Multivariate tests for scenario understanding Sigma-restricted parameterization Effective hypothesis decomposition					
Interface type	Wilks	0.83	0.40	2	4	0.70
Scenario type	Wilks	0.99	0.03	1	5	0.87
Scenario period	Wilks	0.20	20.38	1	5	0.01
Interface type*Scenario type	Wilks	0.17	9.98	2	4	0.03
Interface type*Scenario period	Wilks	0.42	2.81	2	4	0.17
Scenario type*Scenario period	Wilks	0.94	0.32	1	5	0.60
Interface type*Scenario type*Scenario period	Wilks	0.06	31.07	2	4	0.00

The main effect of scenario period is systematic (F(1,5)=20.38, *p*=.01), and is illustrated by Figure 25. It can be seen from the graph that the operators' have better Scenario Understanding in the Mitigation phase than the Detection phase (the effect of scenario period on Process overview went in the opposite direction). The effect explains 21 % of the total variance (ω^2=.21), which is a large effect for experiments in realistic work environments.



	Confidence intervals			
	Mean	Std.Err.	-95.00%	+95.00%
Detection	0.94	0.24	0.33	1.55
Mitgation	1.68	0.13	1.35	2.01

Figure 25. Min effect of scenario period on Scenario understanding

The distribution for the Detection phase is not symmetrical, and the distributions for the detection and Mitigation phase are skewed slightly in opposite directions (see Figure 26). ANOVA may therefore <u>not</u> be robust against violations of the normality assumption for this effect (see section 5.1).

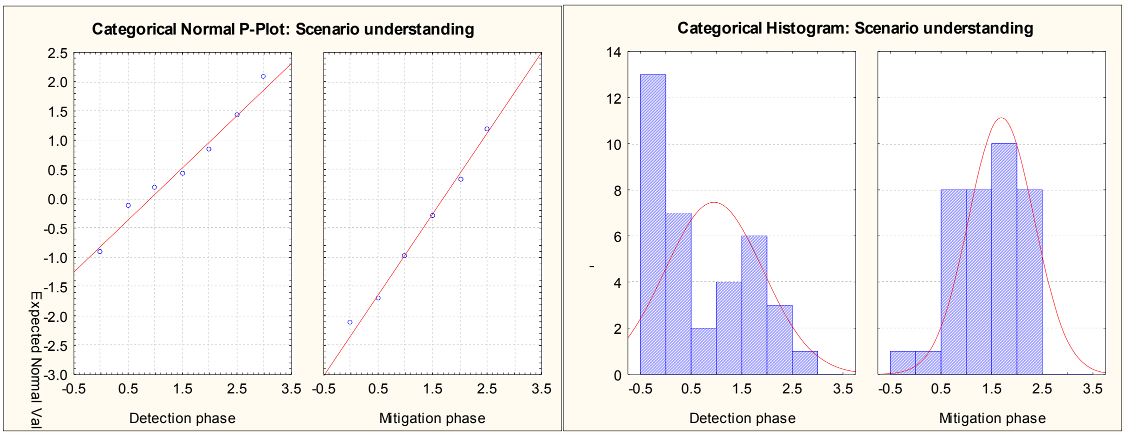


Figure 26. Distribution of Scenario understanding per scenario period

A non-parametric test for the full RBF-322 design is not available, but a Friedman one-way ANOVA for dependent samples can be performed for the scenario period variable alone (StatSoft, 2004). The outcome of this test is Chi square (n=36, df=1) = 12.45, *p*=.00, which indicates that the parametric test was robust after all.

There is a systematic interaction effect between interface type and scenario type (F(2,4)=9.98, *p*=.03), and between interface type, scenario type and scenario period (F(2,4)=31.07, *p*=.00). Graphical inspection of these interaction effects suggested that the two-way interaction provides redundant and limited information compared to the three-way interaction. It was therefore decided to disregard the two-way interaction and present graphical results only for the three-way interaction.

The distribution for Traditional displays (see Figure 27) is asymmetrical. Except for the period variable, which has already been discussed, the other distributions are symmetrical, or skewed consistently in one direction per manipulated variable. Since nonparametric test of higher order interactions for within-subject designs are unavailable (StatSoft, 2004), it is possible that ANOVA is sensitive to violations of the normality assumption for the three-way interaction (see section 5.1). We should therefore interpret the data with some care.

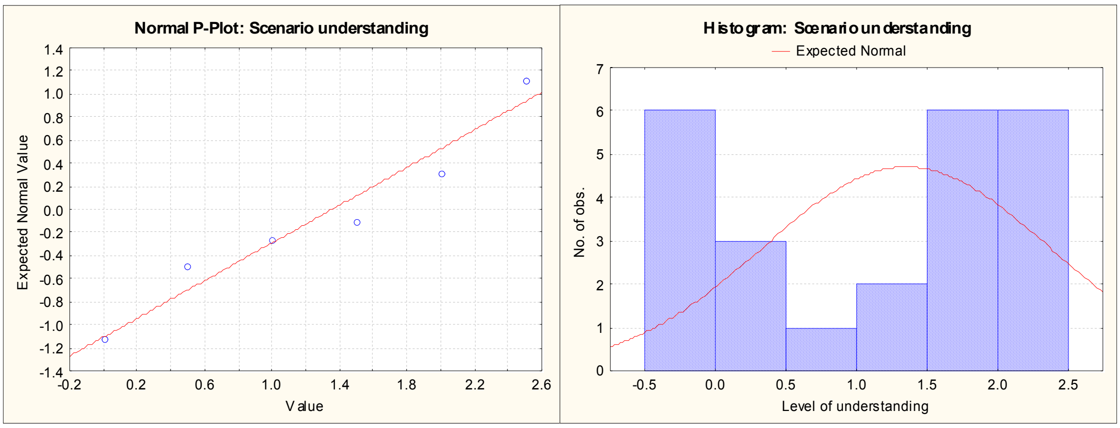
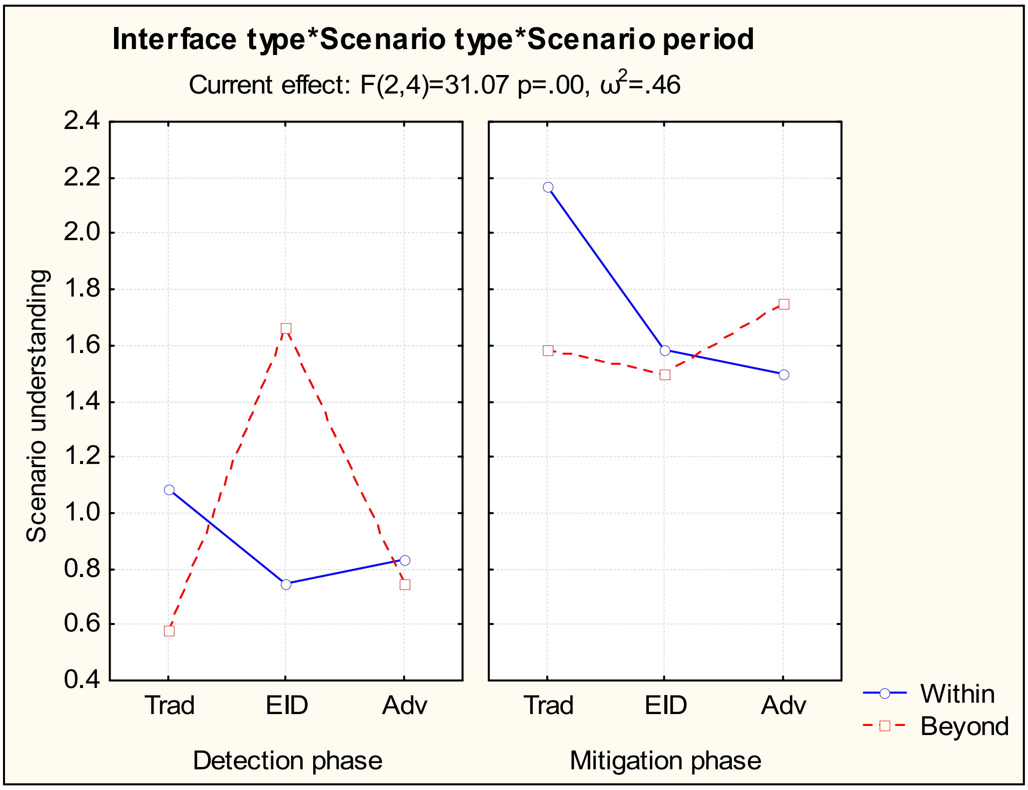


Figure 27. Distribution of Scenario understanding for Traditional displays

The interaction between display type, scenario type and scenario period on Scenario understanding is presented in Figure 28. The effect explains 46% of the total variance (ω^2=.46), which is a very large effect for experiments in realistic work environments.



Cell No.	Confidence intervals Interface type	Scenario type	Scenario period	Mean	Std.Err.	-95.00%	+95.00%
1	Traditional	Within	Detection	1.08	0.42	0.01	2.15
2	Traditional	Within	Mitigation	2.17	0.17	1.74	2.60
3	Traditional	Beyond	Detection	0.58	0.40	-0.44	1.60
4	Traditional	Beyond	Mitigation	1.58	0.40	0.56	2.60
5	Ecological	Within	Detection	0.75	0.28	0.03	1.47
6	Ecological	Within	Mitigation	1.58	0.24	0.97	2.20
7	Ecological	Beyond	Detection	1.67	0.44	0.53	2.80
8	Ecological	Beyond	Mitigation	1.50	0.29	0.76	2.24
9	Advanced	Within	Detection	0.83	0.46	-0.35	2.01
10	Advanced	Within	Mitigation	1.50	0.26	0.84	2.16
11	Advanced	Beyond	Detection	0.75	0.34	-0.11	1.61
12	Advanced	Beyond	Mitgation	1.75	0.17	1.31	2.19

Figure 28. Three-way interaction effect for Scenario understanding

It can be seen from the graph that Ecological displays enhance the Scenario understanding in the Detection phase of Beyond design basis scenarios. Traditional displays seem to improve the Scenario understanding in the Mitigation phase of Within design basis scenarios. These tendencies are consistent with the effects on Process overview (Scenario understanding and Process overview are uncorrelated dependent variables).

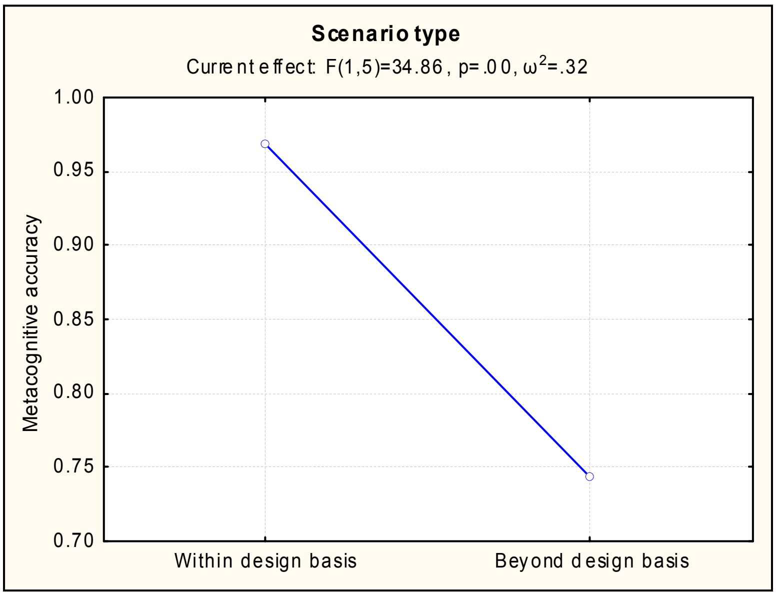
6.3.3 Metacognitive accuracy

Except for the effect of interface type, the multivariate approach to repeated measures turned out to be sufficiently powerful to test whether the hypothesized effects on Metacognitive accuracy were supported by the data. Table 18 shows the outcome of this test for all effects.

Table 18. ANOVA results for Metacognitive accuracy

Effect	Multivariate tests for metacognitive accuracy Sigma-restricted parameterization Effective hypothesis decomposition Test	Value	F	Effect df	Error df	p
Interface type	Wilks	0.35	3.69	2	4	0.12
Scenario type	Wilks	0.13	34.86	1	5	0.00
Scenario period	Wilks	0.60	3.29	1	5	0.13
Interface type*Scenario type	Wilks	0.76	0.64	2	4	0.58
Interface type*Scenario type	Wilks	0.49	2.07	2	4	0.24
Scenario type*Scenario period	Wilks	0.95	0.28	1	5	0.62
Interface type*Senario type*Scenario period	Wilks	0.56	1.60	2	4	0.31

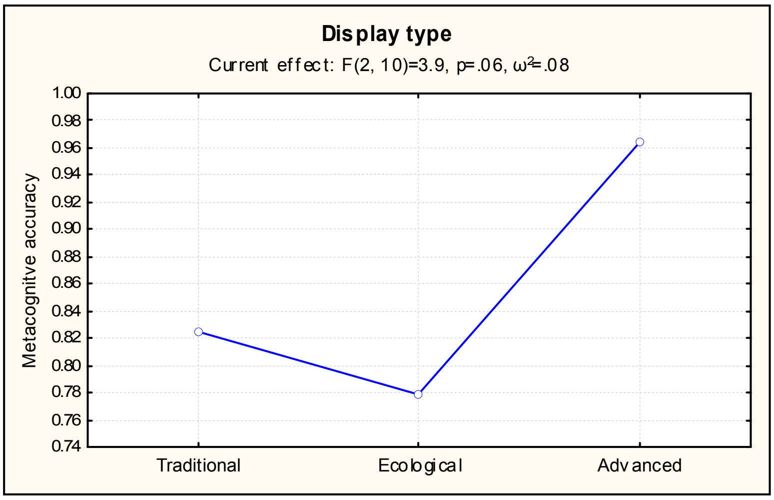
The main effect of scenario type is systematic (F(1,5)=34.86, *p*=.00), and is illustrated in Figure 29. It can be seen from the graph that the operators' have better Metacognitive accuracy during Within design basis scenarios than Beyond design basis scenarios. The effect explains 32% of the total variance (ω^2=.32), which is a very large effect size for experiments in realistic work environments. The normality assumption was not violated for this effect.



	Confidence intervals Mean	Std.Err.	-95.00%	+95.00%
Within design basis	0.97	0.03	0.88	1.06
Beyond design basis	0.74	0.04	0.63	0.86

Figure 29. Main effect of scenario type on Metacognitive accuracy

Following the analysis strategy described in section 5.7, the more powerful univariate test revealed a possible systematic main effect of interface type on Metacognitive accuracy F(2, 10)=3.9, p=.06. The sphericity and normality assumptions were not violated for this effect. The effect explains 8% of the variance (ω^2=.08), which is a medium, or rather small effect size for experiments in realistic work environments. As illustrated in Figure 30, the operators seem to assess their own performance more accurately when they work with the advanced display.



	Confidence intervals			
	Mean	Std.Err.	-95.00%	+95.00%
Traditional displays	0.82	0.07	0.65	1.00
Ecological displays	0.78	0.03	0.71	0.84
Advanced displays	0.96	0.06	0.82	1.11

Figure 30. Main effect of display type on Metacognitive accuracy

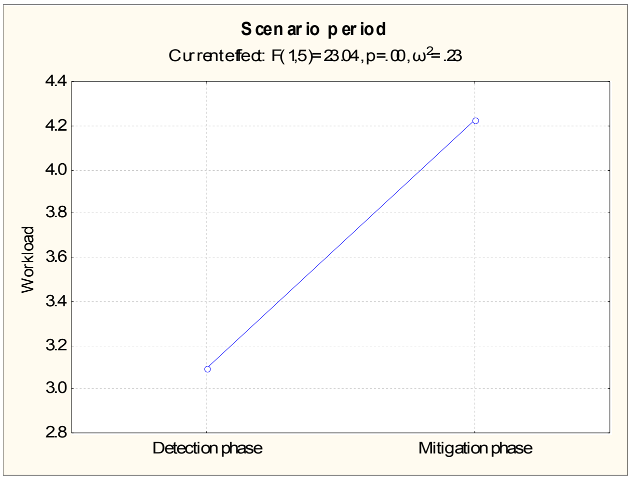
6.3.4 Workload

The multivariate approach to repeated measures turned out to be sufficiently powerful to test the effects of the experimental manipulations on Workload. Table 19 shows the outcome of this test for all effects.

Table 19. ANOVA results for Workload

	Multivariate tests for workload Sigma-restricted parameterization Effective hypothesis decomposition					
Effect	Test	Value	F	Effect df	Error df	p
Interface type	Wilks	0.82	0.44	2	4	0.67
Scenario type	Wilks	0.84	0.92	1	5	0.38
Scenario period	Wilks	0.18	23.03	1	5	0.00
Interface type*Scenario type	Wilks	0.89	0.25	2	4	0.79
Interface type*Scenario period	Wilks	0.16	10.61	2	4	0.03
Scenario type*Scenario period	Wilks	0.37	8.59	1	5	0.03
Interface type*Scenario type*Scenario period	Wilks	0.55	1.61	2	4	0.31

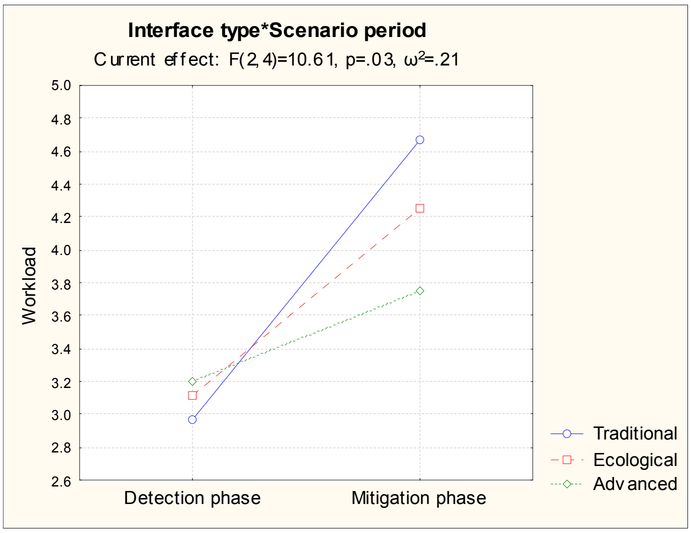
The main effect of scenario period is systematic ($F(1,5)=23.03$, $p=.00$), and is illustrated by Figure 31. It can be seen from the graph that the operators' experience more Workload in the Mitigation phase than the Detection phase. The effect explains 23 % of the total variance ($\omega^2=.23$), which is a large effect for experiments in realistic work environments. The normality assumption was not violated for this effect.



	Confidence intervals			
	Mean	Std.Err.	-95.00%	+95.00%
Detection	3.09	0.35	2.20	3.99
Mitigation	4.22	0.16	3.81	4.63

Figure 31. Main effect of scenario period on Workload

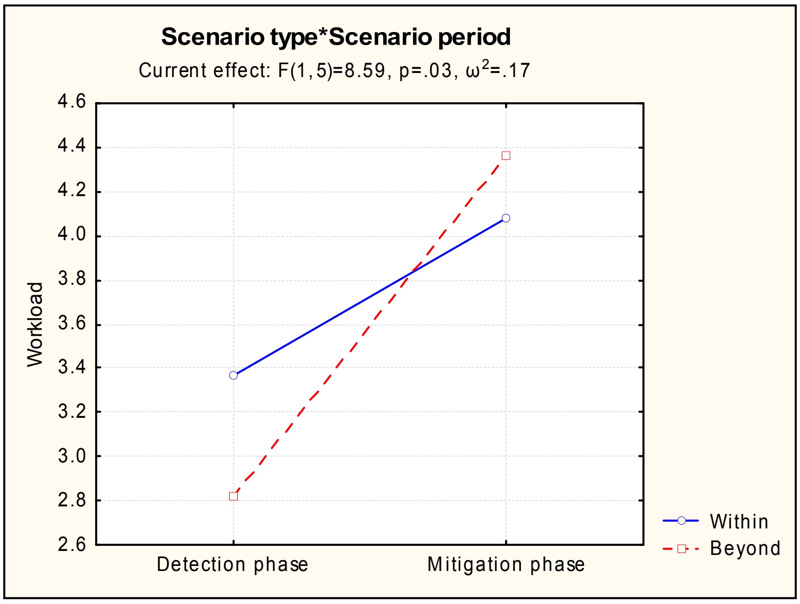
The systematic interaction between display type and scenario period ($F(2,4)=10.61$, $p=.03$) is illustrated in Figure 32. It can be seen from the graph that the operators experience most Workload in the Mitigation phase when they use Traditional displays. Advanced displays produce least Workload during the Mitigation phase. There are no differences between the display types with respect to Workload in the Detection phase. This interaction effect explains 21 % of the total variance ($\omega^2=.21$), which is a large effect for experiments in realistic work environments. The normality assumption was not violated for this effect.



Cell No.	Confidence intervals					
	Interface type	Scenario period	Mean	Std.Err.	-95.00%	+95.00%
1	Traditional	Detection	2.97	0.31	2.16	3.77
2	Traditional	Mitigation	4.67	0.31	3.88	5.46
3	Ecological	Detection	3.12	0.46	1.95	4.29
4	Ecological	Mitigation	4.25	0.33	3.40	5.10
5	Advanced	Detection	3.20	0.32	2.38	4.02
6	Advanced	Mitigation	3.75	0.30	2.97	4.53

*Figure 32. Two-way Interaction effect for Workload (interface type*scenario period)*

The systematic interaction between scenario type and scenario period ($F(1,5)=8.59$, $p=.03$) is illustrated in Figure 33. It can be seen from the graph that the operators experienced less Workload during Detection phase, and more Workload during the Mitigation phase of Beyond design basis scenarios when compared to Within design basis scenarios – in which the operators experience more Workload during the Detection phase and less Workload in the Mitigation phase. This interaction effect explains 17 % of the total variance ($\omega^2=.17$), which is a relatively large effect for experiments in realistic operating environments. The normality assumption was not violated for this effect.



Cell No.	Confidence intervals Scenario type	Scenario period	Mean	Std.Err.	-95.00%	+95.00%
1	Within	Detection	3.37	0.41	2.32	4.41
2	Within	Mitigation	4.08	0.17	3.65	4.51
3	Beyond	Detection	2.82	0.33	1.98	3.66
4	Beyond	Mitigation	4.37	0.18	3.91	4.82

*Figure 33. Two-way Interaction effect for Workload (scenario type*scenario period)*

7. DISCUSSION

As mentioned in section 1.3, this report does not include all findings from the experiment. That is, user feedback on the display types is reported in Welch et al. (2007), while the results of the Contextual Assessment of Systems Usability are available in Norros & Salo (2007). A full discussion of the results is therefore beyond the scope of the text. In this chapter, we will still present some initial interpretations of the statistical results, and discuss the methodological limitations of the study.

7.1 Overall performance level

According to the process expert who was present during the data collection, the overall human performance level in the experiment was satisfactory. This means that the performance was close to what we should expect from licensed NPP operators working on a familiar plant with unfamiliar computerized user interfaces and a reduced crew size. There was no shift supervisor to manage the work and this may have slowed down the operators' decision making compared to normal operation at the home plant. It is also likely that the relatively short operator training and the rather difficult Beyond design basis scenarios influenced performance negatively. However, there were no situations where the participants completely lost control and the simulation became meaningless.

7.2 Support for hypotheses

As stated by the research approach (section 2.1), the hypotheses were written in natural language before the experiment (chapter 4) and concerned the anticipated effect of the manipulated variables. Statistical analysis was used to evaluate whether the hypotheses were supported by the data. The analysis was not a formal test, but a goal-driven statistical exploration to aid an intelligent interpretation of the data. In this exploration, we decided that graphical representation of confidence limits (section 5.5) and comparisons of means (section 5.8) should be avoided. This may be well justified methodologically, but our "minimalist approach" to statistical analysis (Howell, 2006) may have increased the risk for overinterpretation of the data. It is therefore especially important to compare the experimental results to the hypotheses. We may also search for "reasonable findings", such as hypothesised effects that are reproduced by statistically unrelated dimensions of one theoretical construct, or effects that go in opposite directions for measures that reflect opposite phenomena (see section 5.8). Findings that are not backed up by the hypotheses, or do not fit into a logical pattern should probably be ignored, or at least interpreted with great care.

Table 20 compares the hypotheses and the experimental results from the study. For the Situation Awareness measures, we predicted the three-way interaction effect between display type, scenario type and scenario period. The results showed that the Ecological displays supported both Process overview (section 6.3.1) and Scenario understanding (section 6.3.2) in the Detection phase of the Beyond design basis scenarios. This finding seems trustworthy, since the two SA measures were uncorrelated (section 6.2). The improvement in Scenario understanding brought the operator's Situation Awareness up to the level of their understanding during the Mitigation phase of the scenario, which was a very promising result.

Table 20. Comparison of hypotheses and experimental results

Hypothesis	Anticipated effect	Finding	Supported
Main effect of display type on SA	Ecological displays will produce highest SA, while Traditional displays will produce lowest SA	No systematic effect for Process overview and Scenario understanding. Systematic effect for Metacognitive accuracy: Better SA with Advanced displays.	No
Main effect of scenario type on SA	Higher SA in Within design basis scenarios	No systematic effect for any SA measure.	No
Main effect of scenario period on SA	No prediction	Effects of scenario period in opposite directions for Process overview and Scenario understanding.	Partly
Interaction effect between display type and scenario type on SA	Ecological displays provide better SA in (both phases of) Beyond design basis scenarios	Systematic interaction effects for Process overview and Scenario understanding. Possible advantages of Ecological displays for Scenario understanding.	Partly
Interaction effect between display type and scenario period on SA	Ecological displays provide better SA in the Detection phase (for both scenario types)	No systematic effect for any SA measure.	No
Interaction effect between display type, scenario type and scenario period on SA	Ecological displays provide better SA in the Detection phase of Beyond design basis scenarios	Systematic effect for Process overview and Scenario understanding – both in the hypothesized direction	Yes
Main effect of display type on Workload	Ecological displays produce lower Workload	No systematic effect.	No
Main effect of scenario type on Workload	There is more Workload during Beyond design basis scenarios	No systematic effect.	No
Main effect of scenario period on Workload	There is more Workload during the Mitigation phase	Effect of scenario period in the hypothesized direction	Yes
Interaction effect between display type and scenario type on Workload	Ecological displays produces less Workload in Beyond design basis scenarios	No systematic effect.	No
Interaction effect between display type and scenario period on Workload	Ecological and Advanced displays produce less Workload in the Detection phase of the scenarios	Systematic interaction effect in hypothesized direction.	Yes
Interaction effect between scenario type and scenario period on Workload	Larger increase in Workload between detection and Mitigation phase for Beyond design basis scenarios	Systematic interaction effect in hypothesized direction.	Yes

We originally expected a general beneficial effect of the Ecological displays on SA, but the hypotheses that involved interface type were <u>not</u> strongly supported by the data (Table 20). In other words, the Ecological displays did not improve SA for all scenario types and/or scenario periods as anticipated. This may be due to insufficient operator training in the ecological display condition. Ecological displays are radical and innovative, and it was probably unrealistic to expect that the participants would be able to transform their mental model of the nuclear process after only a few hours of pre-experimental training (see section 2.5). It therefore seems reasonable to assume that the participants were unable to utilize the full potential of EID.

The beneficial effect of the Ecological displays did not extend to the Mitigation phase of the scenarios. This could be explained by the hybrid nature of the EID prototype that was developed for the experiment (Welch et al., 2007). That is, the operators used integrated, compressed and traditional process mimic diagrams to intervene and control the system, while the ecological design elements only provided process information. Intervention and control was an important operator activity in the Mitigation phase, but not in the Detection phase of the scenarios. Therefore, the presumed positive effect of the ecological design elements may have been diluted by the use of traditional design elements for intervention and control during the Mitigation phase. It is even possible that the hybrid interface forced the operators to shift frequently between incompatible mental models of the process while they mitigated the disturbances – which may have masked the beneficial effects of the ecological display elements. Thus, future testing of EID in nuclear process control should employ design prototypes that fully support mitigation through ecological display elements.

The hypothesized main effect of display type on Situation Awareness probably expressed a theoretical ideal rather than a realistic expectation. After all, EID was specifically conceived as a design framework that identifies information needed to support anomaly detections, decision making and action under beyond-design basis events (as noted in section 4.1.4). It was therefore not surprising that the Ecological displays failed to support effective Situation Awareness in Within design basis scenarios. This highlights the concern that EID is based solely on Work Domain Analysis (see section 2.2.1), which while motivation useful information visualizations, does not particularly help operators to perform procedural tasks. When using the EID approach, designers should take care to integrate a task based approach (such as GDTA, or Control Task Analysis) in order to support procedural tasks properly (see Jamieson, in press). This deficiency in the EID approach may also explain the absence of the hypothesized main effect of display type on Workload, where Ecological displays were supposed to alleviate Workload under all operating conditions.

In general, the interpretability of main effects is not guaranteed when interaction effects are present (Howell, 1997). Given the complex patterns of the interactions observed in this study, the meaningfulness of the hypothesized main effects of display type can therefore be questioned (in retrospect).

The anticipated main effect of scenario type on Workload and Situation Awareness was not observed. This may suggest that the manipulation of scenario type was not strong and clear, but rather subtle as it showed up in several complex interaction effects (consistently and in accordance with the hypotheses). We were able to predict that Ecological and Advanced displays would produce less Workload in the Detection phase of the scenarios, as compared to the Traditional displays. The other Workload hypotheses were trivial and generally supported by the data.

7.3 Other findings

The findings described in this section were not supported by hypotheses. This type of analysis is accounted for by the research approach (section 2.1), which stated that statistical techniques may be used to search for unanticipated effects that can contribute to the development of new theories and the formation of hypotheses for later experiments.

We observed a logical pattern of results for the main effect of scenario period on Situation Awareness. That is, the operators' Process overview was better in the Detection phase than in the Mitigation phase (section 6.3.1); while the Scenario understanding was better in the Mitigation phase than in the Detection phase (section 6.3.2). This seems reasonable since the operators gradually learn to handle the disturbance throughout the scenario, and because monitoring of the process state is especially important in the early phases of the scenario when the upcoming process deviations are more or less unknown. Findings like this may help to validate the NPPSA model outlined in section 3.3.

The operators assessed their own performance efficiency more accurately during Within design basis scenarios than Beyond design basis scenarios (see section 6.3.3). This finding seems psychologically reasonable, and suggests that the metacognitive accuracy indicator is sensitive to variations in the task conditions.

Traditional displays seem to improve SA in the Mitigation phase of Within design basis scenarios. This is the only non-hypothesized effect that showed a consistent pattern for both Process overview and Scenario understanding. The finding can be explained by the operators' familiarity with the Piping and Instrumentation Diagrams of the plant, which is the basis for the Traditional displays (see section 2.2.1). In the Mitigation phase of proceduralized scenarios, the operators probably fell back to their standard mode of operation and were able to concentrate more on the task – as opposed to new interface design features.

The Process overview in Beyond-design basis scenarios dropped dramatically during the Mitigation phase with both the ecological and Traditional displays (section 6.3.1). While it can be argued that this is not important, since operators were then focused on solving the problem and not maintaining a large view of the process, the Advanced display showed an improvement in Process overview. This display should be examined further to determine what features might be providing this benefit.

Another possible advantage of the Advanced displays was the improved Metacognitive accuracy under all operating conditions (section 6.3.3). It is hard to imagine why traditional process mimic with integrated mini trends and added configural design features would improve the operators' ability to assess their own performance. The positive effects of the Advanced displays can not be corroborated by independent facts and should therefore be interpreted with great care (see section 5.8). However, Advanced display type may gradually become the new industry standard for NPP control rooms. These findings are therefore interesting and could motivate further research.

7.4 Limitations of the study

7.4.1 Operator training

Prior to the trials, the participants were familiarised with the HAMMLAB environment and trained to use the various interfaces. In total, the operators were given ~7 hours of instruction. Although more time was devoted to the Ecological displays, it was not possible for participants to fully grasp and understand certain graphical elements designed for longer term use, e.g., patterns produced by trend graphs and balance displays. In addition, participants were provided with reference sheets as memory aids because of the overwhelming number of features found in each interface and in particular the Ecological displays. Many of the visual elements intended to improve Situation Awareness, e.g., mass balance bar connections to mimic graphics, were perhaps used less effectively than anticipated due to limited experience with the interface. Given that some of the participants had prior experience with the Traditional and Advanced displays, as well as possible positive learning transfer from the existing control room interface, novice Ecological display users can be said to develop Situation Awareness on par with slightly more experienced Traditional and Advanced display users.

When examined alone, we may speculate if the Ecological and Advanced displays resulted in riskier projections and strategy selections, whereas the Traditional displays motivated more conservative decision making. It is also possible that visual elements such as emergent features and historical trend plots in the non-Traditional displays lead to looser projections. However, a more likely explanation is that training was limited, particularly in the use of advanced graphical features, some of which require more experience than time allowed. For example, the individual graphics in the Ecological displays, such as the mass balances, were intended to support skill-based behaviour including the perception of process deviations. The graphics were organised in a manner to support knowledge-based behaviour that develops primarily over time through experience. The effect was perhaps not seen in the Traditional displays due to similarities with the existing control room interfaces, which are generally approached in a conservative manner.

7.4.2 Operator Experience

Though unlikely to have significantly influenced the SA results, it should be noted that some of the current Reactor Operators (RO) and shift supervisors (SS), i.e., former Turbine Operators (TO), were asked to perform the tasks of a TO. Although these operators still hold TO licences, a greater number of licensed TO participants would have been preferable. Participation was open to all control room operators with no experience restrictions.

It is possible to confound the results by employing operators already familiar with the control condition, in this case the Traditional displays. Since the interfaces are currently still under evaluation, i.e., not implemented in the control room of the operators' home plant, it was hoped that any past experience with HAMMLAB and the Traditional displays was negated through training. A little over half of the operators had participated in past HAMMLAB studies. It is possible that proper rule-based corrective actions were carried out as a result of experience in spite of minimal interface training, which could have led to non-ideal Situation Awareness.

7.4.3 Crew composition

As noted previously, a typical control room crew consists of one TO, one RO, and one SS. Due to the limited number of participants, the SS role was excluded from the study. Furthermore, the study focused on the effects of different interfaces on TO-Situation Awareness alone, as team Situation Awareness was beyond the scope of the study. However, it was observed that the lack of a SS was detrimental to TO-performance on several occasions, in which certain cross-system information such as that from the reactor side would have been relayed to the TO (as explained by the process expert).

7.4.4 Scenario design

The high complexity of both the Within and Beyond design basis scenarios is likely to have affected some of the scenario type results. Within design basis scenarios were designed to involve situations that are intended to be resolved through standard procedures. Beyond design basis scenarios included a series of events that could not be diagnosed through traditional rule-based strategies. It was observed that the sequence of events in both types of scenarios was at times rendered more complex than expected due to operator interaction with the system. Actions performed by the operators resulted in new events affecting the development of the scenario and thus the corrective actions required for a resolution, e.g., an anticipated scenario becoming more difficult to handle through standard procedures.

7.4.5 Simulator constraints

Certain elements in the EID information displays were re-designed as a result of technical limitations in the simulator and interface design platform (Welch et al., 2007). In particular, the polar star graphics representing valve positioning were replaced due to implementation difficulties involving the scaling of the graphics. As well, the lack of mass sensors in the turbine subsystems required that the values had to be calculated in the simulator, which was not always accurate. Some variables traditionally not included in control room interfaces were deemed essential in Ecological displays, but could not be fully implemented. A few unexpected technical glitches in the simulator were also encountered during the experiment, one of which required a scenario restart. Any learning effects from the incident are not expected to have altered the results significantly as only one trial was affected.

7.4.6 Ecological display prototype

As mentioned several times, the Ecological displays included traditional process mimic for intervention and control of the system, while ecological design elements only provided process information to the operators. Another weakness is that the ecological display prototype did not cover all of the process information presented on the Turbine Operator work station (see Welch et al., 2007). Furthermore, the Turbine Operators had a traditional large screen display and alarm system available in all experimental conditions. One could therefore argue that the experiment evaluated an inconsistent hybrid interface with integrated ecological design elements, which could <u>not</u> be expected to generally outperform traditional computerized displays.

It may be methodologically problematic that the Ecological displays were spatially dedicated to the Turbine Operator screens, while the traditional and advanced process formats could be displayed freely to the same screens depending on the Turbine Operator's preferences (see section 2.6). This is a systematic difference between the display type prototypes that may be difficult to justify theoretically with reference to the HSI design frameworks. An experiment is internally valid when systematic variation in the dependent variable can only be attributed to the experimental manipulation, i.e., there is an unambiguous causal relationship between treatment and outcome variables, and no third-variable can explain the experimental effect (Pedhazur & Pedhazur Schmelkin, 1991). This difficulty may therefore be a threat to the internal validity of the experiment.

A similar problem is that the EID prototype utilized sensor information that was unavailable in the traditional and advanced interfaces. However, this is probably easier to defend with reference to the HSI design frameworks. See Welch et al. (2007) for a full discussion of the *sensor problem*.

7.4.7 Generalizability

The generalizability of the results is limited to situations where the detection phase of the scenario comes prior to the mitigation phase (the presentation order of the scenario periods were not counterbalanced). This is probably an unproblematic restriction since most mitigations are preceded by periods where the operators monitor and detect process deviations.

The experimental findings are based upon the performance of six participants. Even though NPP operators are well trained and a relative homogeneous group, we know that there can be strong variations in operator performance. It can therefore be questioned if a sample of six participants can represent the operator population. Given that the performance variation induced by the task is known to be substantial in realistic simulator research, one may also question whether it is possible to generalize from a sample of three scenarios to a broad category of operating conditions (such as within design basis scenarios).

Finally, the experimental results are representative of Situation Awareness acquired and developed mainly through the turbine side interfaces.

8. CONCLUSION

Our study produced evidence that Ecological Interface Design improves the Situation Awareness of NPP operators in the Detection phase of Beyond design basis scenarios. Developing Ecological displays that fully support operator intervention with the nuclear process could possibly extend these benefits to the Mitigation phase of Beyond design basis scenarios as well.

Effective utilization of Ecological displays implies that the operators change their mental model of the nuclear process, e.g., by forming abstract representations of energy balances. The learnability of Ecological displays may therefore represent a challenge. We were, however, able to demonstrate some benefits of EID even with a limited amount of operator training, which suggests that the true potential of the design framework may be large. The potential of EID to dramatically improve Situation Awareness for unanticipated events is quite important since serious accidents are most likely to occur in this type of situation.

9. REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation Awareness and the cognitive management of complex systems. *Human Factors, 37*(1), 85-94.

Anderson, J. R. (1985). *Cognitive psychology and its implications.* New York, NY: Freeman.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2004). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods,* **10** (4)

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors*, **34**(5), 513-533.

Bennett, K. B., Nagy, A. L., & Flach, J. M. (2006). *Visual displays. In G. Salvendy (Ed.), Handbook of Human Factors and ergonomics* (3rd ed., pp. 1191-1219). Hoboken, NJ: Wiley.

Braarud, P. Ø. (2000). *Subjective task complexity in the control room* (HWR-621). Halden, Norway: OECD Halden Reactor Project.

Braarud, P. Ø. and Brendryen, H. (2001). *Task demand, task management, and teamwork* (HWR-657). Halden, Norway: OECD Halden Reactor Project.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and psychological measurement*, **33**, 107-112

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin.* **112** (1), 155-159.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist,* **49** (12), 997-1003.

Cook, T.D. and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Chicago: Rand McNally.

Cumming, G. & Finch, S. (2005). Inference by eye. *American psychologist,* **60** (2), 170-180

Drøivoldsmo, A. (2003). *New tools and technology for the study of human performance in simulator experiments.* Dr.polit. dissertation NTNU, Trondheim, Norway

Durso, F.T., Hackworth, C.A., Truitt, T.R., Crutchfield, J.M., Nikolić D. and Manning, C.A. (1998). Situation Awareness as a Predictor of Performance for En Route Air Traffic Controllers. *Air Traffic Control Quarterly*, **6**, 1-20

Durso, F. T., Bleckley, M. K., & Dattel, A. R. (2006). Does Situation Awareness add to the validity of cognitive tests? *Human Factors, 48*(4), 721-733.

Edwards, E., & Lees, F. P. (1974). *The human operator in process control.* New York: Halsted Press.

Endsley, M. R. (1988). Design and evaluation for Situation Awareness enhancement. *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting,* Santa Monica, CA. 97-101.

Endsley, M. R. (1995a). Toward a theory of Situation Awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R. (1995b). Measurement of Situation Awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R. (2000a). Theoretical underpinnings of Situation Awareness: A critical review. In M. R. Endsley, & D. J. Garland (Eds.), *Situation Awareness: Analysis and measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2000b). Direct measurement of Situation Awareness: Validity and use of SAGAT. In M. R. Endsley, & D. J. Garland (Eds.), *Situation Awareness: Analysis and measurement* (pp. 147-174). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2004). Situation Awareness: Progress and directions. In S. Banbury, & S. Tremblay (Eds.), *A cognitive approach to Situation Awareness: Theory and application* (pp. 342-351). Aldershot, Hampshire; Burlington, VT: Ashgate Pub.

Fern, E. F., & Monroe K. B. (1996). Effect size estimates: Issues and problems in interpretation. *Journal of Consumer Research,* **23**, 89-105

Field, A. (2000). *Discovering Statistics Using SPSS*, 2nd ed. London, UK: Sage Publications.

Fiore, S. M., Johnston, J. H. & Smith, C. A. P (2005). Evaluating computerized decision support systems for teams: Using cognitive load and metacognition theory to develop team cognition measures, *Proceedings from the 11th international conference on human-computer interaction* (HCII 2005), Las Vegas, Nevada US.

Flach, J. M. (1995). Situation Awareness: Proceed with caution. *Human Factors, 37*(1), 149-157.

Fracker, M. L. (1991). *Measures of Situation Awareness: Review and future directions*. Wright-Patterson AFB OH: Armstrong Laboratory, Crew Systems Directorate.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, N.J.: Lawrence Erlbaum.

Goodstein, L. P. (1981), *Discriminative display support for process operators, in J. Rasmussen and W. B. Rouse (eds), Human Detection and Diagnosis of System Failures* (Plenum, New York), 433-439.

Karlsson, T., Jokstad, H., Meyer, B. D., Nihlwing, C., Norrman, S., Karita E., Raussi, P. & Tiihonen O. (2001). *The HAMBO BWR Simulator of HAMMLAB.* (HWR-633). Halden, Norway: OECD Halden Reactor Project.

Hart & Staveland (1988). *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.* In Hancock, P.A. & Meshkati, N. (1988). Human Mental Workload . Elsevier Science Publishers.

Hart, S. G. and Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Booher (Ed.) Manprint: *An Approach to Systems Integration*. (pp. 257-296). New York: Van Nostrand Reinhold.

Hintze, J. (2004). *NCSS and PASS. Number Cruncher Statistical Systems.* Kaysville, Utah. www.ncss.com

Howell, D. C. (1997). *Statistical methods for psychology (4th ed.)*, CA: Wadsworth.

Howell, D. C. (2006). *Multiple comparisons with repeated measures.* Material prepared for the next version of 'Statistical methods for psychology'. http://www.uvm.edu/~dhowell/StatPages/

Huet, N. (1999). Assessment device of metacognition involved in a working situation. *International Journal of Industrial Ergonomics* **24**, 243-251

Jamieson, G. A. (in press). *Ecological Interface Design for Petrochemical Process Control: An Empirical Assessment.* To appear in IEEE Transactions on Systems, Man & Cybernetics.

Jones, G. D. and Endsley M. R. (2004). Use of real-time probes for measuring Situation Awareness. *International Journal of Aviation Psychology*. **14**(4), 343-367.

Jones, J. L. (1995). *Understanding psychological science.* New York, NY: Harper Collins.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374-419.

Kèohler, W. (1971). *The selected papers of Wolfgang Kèohler*. New York, USA: Liveright.

Kershaw, T. C., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 3-13.

Kimmel, H. D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin*, **54** (4).

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences, 3rd ed.* Pacific Grove, CA: Brooks Cole.

Lewis, J. R. (1989). Pairs of Latin squares to counterbalance sequential effects and paring of conditions and stimuli. *Proceedings of the Human Factors 33rd annual meeting*. Denver, CO.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review,* **1**, 476-490.

Logan, G. D., & Zbrodoff, N. J. (1999). Selection for cognition: Cognitive constraints on visual spatial attention. *Visual Cognition, 6*(1), 55-81.

Luce, R. D. (1963). Detection and recognition of human observers. In R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology. vol. 1*. New York, USA: Wiley.

Maki, R. H. (1998). Metacomprehension of text. Influence of absolute confidence level on bias and accuracy. *The psychology of learning and motivation*. Vol. 38.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective.* Mahwah, NJ: Lawrence Erlbaum Associates

McGuinness, B. (2004). Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self-ratings. Paper presented at the *Proceedings of the Ninth International Command and Control Research and Technology Symposium,* Copenhagen, Denmark. Retrieved February 6, 2007, from http://www.dodccrp.org/events/9th_ICCRTS/CD/papers/159.pdf

McNicol, D. (2005). *A primer of signal detection theory*. Mahwah, N.J: Lawrence Erlbaum Associates, Inc.

Mendoza, J. L., Toothaker, L. E. & Nicewander, W. A.(1974). A Monte Carlo comparison of the univariate and multivariate methods for the groups by trials repeated-measures design. *Multivariate Behavioral Research*, **9**, 165-177.

Mill, J. S. (1843). *A System of Logic*. London: Longmans

Montgomery, D. C. (1997). *Design and analysis of experiments*. New York: John Wiley.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors, 42*(1), 36.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Norros, L. & Salo, L. (2007). *Communicating Functional Meaning of Process State Information – Comparing Operator Practices when using three Forms of Information Presentation.* Proceedings from the Enlarged Halden Programme Group meeting. Storefjell, Norway.

O'Grady, K. E (1982). Measures of explained variance: cautions and limitations. *Psychological Bulletin,* **92** (3), 766-777

O'Hara, J. M., Higgins, J. C., Persensky, J. J., Lewis, P. M. and Bongarra, J. P. (2002) *Human Factors Engineering Program Review Model.* (Tech. Rep. NUREG-0711) Washington, D.C., USA: US Nuclear Regulatory Commission

Payton, M. E., Greenstone, M. H. & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance. *Journal of Insect Science,* 3:34.

Pew, R. W. (1995). Experimental analysis and measurement of Situation Awareness. *Proceeding of the International Conference on Experimental Analysis and Measurement of Situation Awareness,* Daytona Beach, FL. 7-15.

Pedhazur, E. J. & Pedhazur Schmelkin, L. (1991). *Measurement, design, and analysis. An integrated approach.* NJ: Hillsdale.

Rasmussen, J. (1969). "Man-machine communication in the light of accident records," S- 1-69. *Danish Atomic Energy Commission.* Research Establishment Risoe, Roskilde. Denmark. 1969.

Reising, D. V., & Sanderson, P. M. (2002). Ecological interface design for pasteurizer II: A process description of semantic mapping. *Human Factors, 44*(2), 222-247.

Rosenthal, R., Rosnow, R. L. & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research.* UK: Cambridge University Press.

Schenker, N. & Gentleman, J. F. (2001). On judging the significance of differences by examining overlap between confidence intervals. *The American Statistician*, **55** (3)

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, **26**, 113-125

Seagull, F. J., Xiao, Y., Mackenzie, C. F., & Wickens, C. D. (2000). Auditory alarms: From alerting to informing. *Proceedings of the IEA 2000/HFES 2000 Congress,* Santa Monica, CA. I-223-I-226.

Skraaning, G. (1998). *The Operator Performance Assessment System (OPAS).* Halden Work Report 538: OECD Halden Reactor Project.

Skraaning, G. (2003). *Experimental Control versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments.* OECD Halden Reactor Project. HPR-361, dr.philos. dissertation NTNU, Trondheim, Norway

Smith, K., & Hancock, P. A. (1995). Situation Awareness is adaptive, externally directed consciousness. *Human Factors, 37*(1), 137-148.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

StatSoft, Inc. (2004). *STATISTICA for Windows* [Computer program manual]. Tulsa. OK.

St-Cyr, O., & Vicente, J. K. (2005). Sensor noise and ecological interface design: Effects of increasing noise magnitude on operators' performance. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting,* Orlando, FL. 417-421.

Sørenssen, A., de Carvalho, V., Karlsson, T., Meyer, B., Nihlwing, C., Nystad, B., Saarni, R. (1999). *The Operator Interface of the BWR Simulator in HAMMLAB* (HWR-592). Halden, Norway: OECD Halden Reactor Project.

Tzelgov, J. (1997). NOTE. *Consciousness and Cognition, 6*(2-3), 441-451.

Underwood, B. J. (1957). *Psychological research.* New York: Appleton-Century-Crofts.

Veland, 2004. *State-of-the art report on Ecological Interface Design.* OECD Halden Reactor Project, HWR-750, Halden, Norway

Vicente, K. J., and Rasmussen, J. (1992) 'Ecological interface design: Theoretical foundations', *IEEE Transactions on Systems, Man and Cybernetics* 22: 1-18.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Vicente K. J. (2002), "Ecological interface design: Progress and challenges," *Human Factors*, vol. 44, pp. 62-78.

Weisberg, R. W., & Alba, J. W. (1981a). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General, 110*, 169-192.

Weisberg, R. W., & Alba, J. W. (1981). Gestalt theory, insight and past experience: Reply to dominowski. *Journal of Experimental Psychology: General, 110*, 193-198.

Weisberg, R. W., & Alba, J. W. (1982). Problem solving is not like perception: More on gestalt theory. *Journal of Experimental Psychology: General, 111*, 326-330.

Welch, R., Friberg, M., Nystad, E., Teigen, A. & Veland, Ø. (2005). *EID – Prototype design & user test 2004.* OECD Halden Reactor Project, HWR-790, Halden, Norway

Welch, R., Braseth, A. O., Nihlwing, C., Skraaning, G.., Teigen A., Veland, Ø., Burns, C. M., Kwok, J., Jamieson, G., & Lau, N. (2007). *The 2005 Ecological Interface Design Process and the Resulting Displays.* OECD Halden Reactor Project, HWR-847, Halden, Norway

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wilkinson, L. (1999). Statistical methods in psychology journals. Guidelines and explanations. *American Psychologist*, **54** (8), 594-604.

Woods, D. D. (1994). Cognitive demands and activities in dynamic fault management: Abductive reasoning and disturbance management. In N. Stanton (Ed.), *The human factors in alarm design* (pp. 63-92). London, UK: Taylor & Francis.

Xiao, Y., & Seagull, F. J. (1999). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting,* Santa Monica, CA. 256-260.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70*(1), 145-154.

10. APPENDIX 1: ACTUAL TASK PERFORMANCE ITEMS

The numbers in brackets behind the response alternatives denote the points that could be earned by the Turbine Operators if they performed the associated problem solving behaviour. The response alternatives per item are mutually exclusive, meaning that an observed operator activity can be associated with only one of the alternatives. The web-based scoring form used by the process expert during the data collection therefore used a forced choice response format for each performance item.

Within-1

Detection phase

1. Detects the temperature difference after each steam super-heater 422KA504 and KB504
    a. Yes [3]
    b. No [0]
2. Detects that the position of the water level regulating valve for the feedwater tank 462VA5 is changing
    a. Yes [2]
    b. No [0]
3. Detects the increasing flow to the feedwater tank 462KB301
    a. Yes [3]
    b. No [0]
4. Makes inference about leakage in the steam super heater 422EA1
    a. No inference or wrong inference [0]
    b. Partly correct inference [1]
    c. Mostly correct inference [2]
    d. Fully correct inference [3]
5. Selects the correct procedure
    a. Yes [3]
    b. No [0]
6. Informs RO about the upcoming power reduction
    a. Yes [3]
    b. No [0]

Mitigation phase

1. Correctly takes care of core coolant pump 313PB2 trip
    a. Yes [2]
    b. No [0]
2. Correctly takes care of steam super heater 422EA1 leakage
    a. No action or wrong action [0]
    b. Detects the disturbance after the alarm, and intervenes according to the procedure [2]
    c. Correct diagnosis before the alarm [3]
3. Detect and take care of the problem in the turbine plant main steam system 421
    a. Does not take care of the 421 problem correctly [0]
    b. Detects and takes care of the 421 problem correctly within 7 minutes [1]
    c. Detects and takes care of the 421 problem correctly within 5 minutes [2]
    d. Detects and takes care of the 421 problem correctly within 3 minutes [3]

Within-2

Detection phase

1. Detect the deviating state of drainage valve 463VA20 and the following increasing level in the high pressure pre-heater drain tank 463EB1
    a. Detects nothing [0]
    b. Detects only that the level in 463EA1 is different from the level in 463EB1 [1]
    c. Detects that the level in 463EB1 is increasing, understands that the reason for this is 463VA20, and tries to open the valve by hand [2]
    d. Immediately detects that 463VA0 does not regulate correctly and tries to open the valve by hand [3]

Mitigation phase

1. Detects the increased level, and that 463VA20 is completely closed, and that H1 in 463EB1 will lead to HPPH BP
   a. Detects nothing [0]
   b. Detects the alarm 463KA402H1, but not that 463VA20 has closed, and informs RO when HPPH BP begins [1]
   c. Detects that 463VA20 has closed, and informs RO about HPPH BP after the alarm 463KA402H1 [2]
   d. Detects that 463VA20 has closed, and informs RO about HPPH BP before the alarm 463KA402H1 [3]
2. Detects that the condenser pressure is starting to increase and that this is caused by increasing sea water temperature
   a. Detects nothing [0]
   b. Detects that the condenser pressure is starting to increase and believes that this is caused by an air leakage (wrong but plausible conclusion) [1]
   c. Detects that the condenser pressure is starting to increase [2]
   d. Detects that the condenser pressure is starting to increase and concludes that this is caused by increasing sea water temperature [3]
3. Tries to open the water level regulating valve for the feedwater tank 462VA5 or VB5 manually
   a. Unable to manage the disturbance without evoking safety systems [0]
   b. Manages the disturbance without evoking safety systems [1]
   c. 462VB5 is opened to produce the same flow to the feedwater tank as before the valve closed, and correct flow balance is maintained during the mitigation time, and there is no H1/L1 alarm from the feedwater tank [3]
4. Detect that main flow valve 463VA4 for feedwater pump 463PA1 has opened, change feedwater pump, and report deviating behaviour of VA4
   a. Detects nothing and/or carries out wrong actions [0]
   b. Changes feedwater pump without knowing why [1]
   c. Detects after more than 5 minutes that VA4 has opened, changes feedwater pump, and reports deviating behaviour of VA4 [2]
   d. Detects before 5 minutes that VA4 has opened, changes feedwater pump, and reports deviating behaviour of VA4 [2]

Within-3

Detection phase

1. Detects the first problem with respect to the drainage change in the steam reheat system 422 from the condenser to the HPPHs
   a. Detects nothing
   b. Detects that the level in 463EA2 is increasing [1]
   c. Detects the level in 463EA2, and concludes that the reason is 463VA21, and tries to open this valve by hand [2]
   d. Immediately detects that 463VB20 is stuck at 30% and tries to open the valve manually, and tries to open 422VA10 by hand and balance the flow [3]
2. Detects the second problem with respect to the drainage change in 422 from the condenser to the HPPHs
   a. Detects nothing [0]
   b. Only detects that the level in 422TA2 is decreasing [1]
   c. Detects that the level in 422TA2 is decreasing and regulates 422VA10 manually by hand [2]
   d. Detects that the level in 422TA2 is decreasing, and regulates 422VA10 manually by hand, and reports the deviation on 422VA10 [3]

Mitigation phase

1. Analysis of the system state should lead to a deviation report on 463VA21
   a. Detects and/or reports nothing [0]
   b. Detects the alarm 463KA402H1, but not that 463VA21 is stuck, and informs RO when HPPH BP starts [1]
   c. Detects the alarm 463KA402H1, and that 463VA21 is stuck, and informs RO when HPPH BP starts [2]
   d. Detects that 463VA21 is stuck, and informs RO when HPPH BP starts, but before the alarm 463KB404H1 [3]
2. Detect that 422KA404 has drifted and report the deviation
   a. Detects and/or reports nothing [0]
   b. Detects the low level in 422TA1 [1]
   c. Reports that 422VA10 has opened too much [2]
   d. Concludes that 422KA404 has drifted, observes on the measurement point display that two are similar and two different, and reports the deviation [3]

Beyond-1

Detection phase

1. Detects the increasing level in 422TB1, and understands that this is caused by the stuck valve 422VB6
   a. Detects nothing [0]
   b. Detects the increasing level in 422TB1 [2]
   c. Detects the increasing level in 422TB1, and understands that 422VB6 is stuck, and reports this deviation [3]
2. Detect the increasing level in 422TB1 and concludes that this is caused by the stuck valve 422VB6
   a. Detects nothing [0]
   b. Observes that the H1 alarm at 2.8m is missing [3]
3. Inform RO about TS
   a. Does not inform RO [0]
   b. Informs RO about upcoming TS at 3.3m [2]

Mitigation phase

1. Observes that 422VB6 closes
   a. Does not detect that 422VB6 closes [0]
   b. Detects that 422VB6 closes completely, but after the H2 alarm in 422TB1 and after the TS [1]
   c. Detects that 422VB6 closes completely after the H2 alarm in 422TB1, but before the TS [2]
   d. Detects that 422VB6 closes completely before the H2 alarm in 422TB1 [3]
2. Detect that the generator is not separated from the grid
   a. Fails to detect that the generator is not separated from the grid
   b. Detects that the generator is not separated from the grid within 60 sec. after TS [1]
   c. Detects that the generator is not separated from the grid within 45 sec. after TS [2]
   d. Detects that the generator is not separated from the grid within 30 sec. after TS [3]

Beyond-2

Detection phase

1. Detects leakage in the condensate cleaning building
   a. Detects nothing [0]
   b. Detects that the position of the water level regulating valve for the feedwater tank 462VA5 increases [2]
   c. Detects increased flow from the condenser and that the water level regulating valve for the feedwater tank 462VA5 increases [3]

Mitigation phase

1. Diagnose the leakage in the condensate cleaning building

   a. No diagnosis made [0]
   b. Diagnose the leakage in the condensate cleaning building after alarm indicating high level in room Z02.10 [1]
   c. Diagnose the leakage in the condensate cleaning building before alarm indicating high level in room Z02.10 [3]
2. Try to bypass condensate cleaning building
   a. No action [0]
   b. Tries to bypass condensate cleaning building [2]
3. Close water level regulating valve for the feedwater tank 462VA5
   a. No action [0]
   b. Closes 462VA5 in order to manually close 332VB2 [1]
   c. Closes 462VA5 in order to manually close 332VB2, but before the low level alarm in the feedwater tank [3]

Beyond-3

Detection phase

1. Detect that 441PB1 reduces rotation speed
   a. Detects nothing [0]
   b. Detects increasing level in the condenser and/or trip of 441PB1 [1]
   c. Detects increasing level in the condenser, and then the reduced rotation speed of 441PB1 [2]
   d. Detects the reduced rotation speed of 441PB1, and then the level increasing in the condenser [3]

Mitigation phase

2. Detect the increasing sea water temperature
   a. Detects nothing [0]
   b. Detects the increasing sea water temperature within 5 minutes after the H1 alarm indicating high pressure in the condenser [1]
   c. Detects the increasing sea water temperature within 1 minutes after the H1 alarm indicating high pressure in the condenser [1]
   d. Detects the increasing sea water temperature before the H1 alarm indicating high pressure in the condenser [1]
3. Suspect additional problems with the condenser that are not related to the sea water temperature
   a. Does not suspect additional problems with the condenser [0]
   b. Suspects additional problems with the condenser [3]
4. Detect that bypassing of steam to the condenser occurs without open bypass steam cooling spray valves
   a. Observes nothing and fails to perform correct actions [0]
   b. Detects that bypassing of steam occurs and decreases the power output to close the bypass steam valves [2]
   c. Detects that bypassing of steam occurs and that none of the bypass steam valves are open, and opens the bypass steam valve and decreases the power output - leading to closed bypass steam valves [3]
5. Observe that the sea water temperature is increasing further
   a. Does not detect that the sea water temperature is increasing further [0]
   b. Detects that the sea water temperature is increasing further [2]
6. Detect that the 421VA3v1 valve has opened and that no bypass steam valves are open
   a. Detects nothing [0]
   b. Detects that the 421VA3v1 valve is stuck at 30% within 1 minute [2]
   c. Detects that the 421VA3v1 valve is stuck at 30% and that no bypass steam valves are open within 1 minute [3]

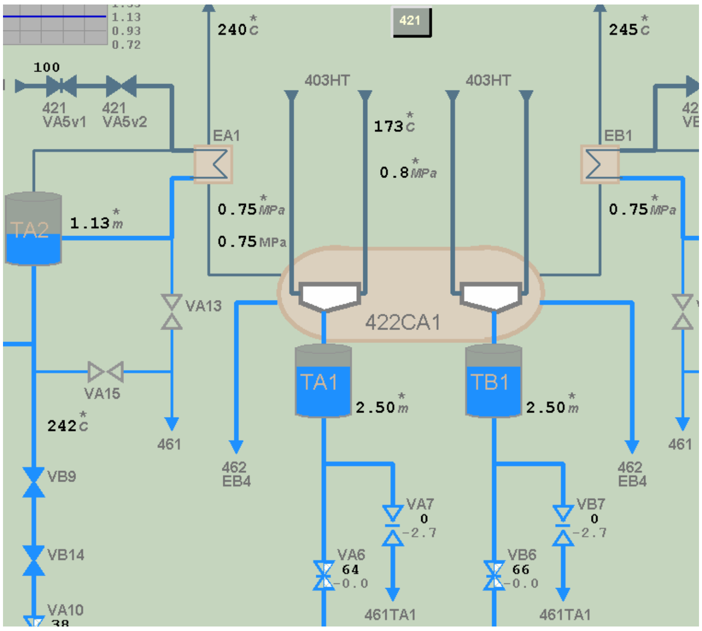
11. APPENDIX 2: SCENARIO DESCRIPTIONS

WITHIN-1

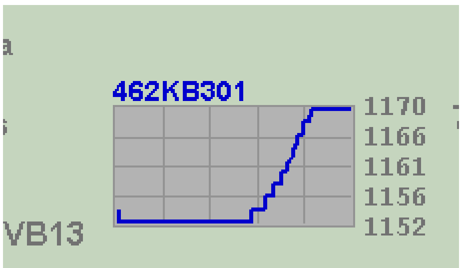
Start conditions
109.3 % reactor power
19.5 C in 316 cooling from SUB C.
Sea water temperature 12 C.
Pressure in containment -3.6 kPa

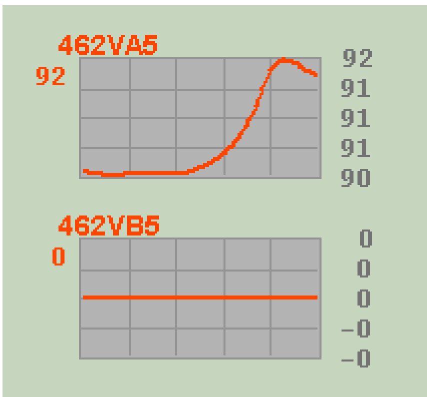
Leak in the intermediate super heater
After 5 minutes, a leak from tube side to shell side starts in the intermediate super heater 422EA1. This makes the high-pressure steam leak over to the super heated steam that is on its way to the low-pressure turbine.

The steam flow to the condenser will increase and also the drain steam from the LP-turbine because the steam that normally goes through the pipes in the super heater and after it has condensed to water back in to the feed water tank will now go through the low pressure turbine and to the condenser. This will generate an increasing drain flow from LP-pre heater 462EB2 and 462EB3 and this is possible to see for the operator if he is looking at the position of drain valve 462VB29.



In this picture we see the situation after 3 minutes. The leak flow is 39 kg/s now. The leak flow is on EA1 and the temperature to LP turbine is now 240 C on that side and 245 on the EB1 side. The super heating is not working as it should because of the leak flow On the drain side the VA6 valve is more closed than VB6 because of the decreased drain flow on the side were the leakage is.





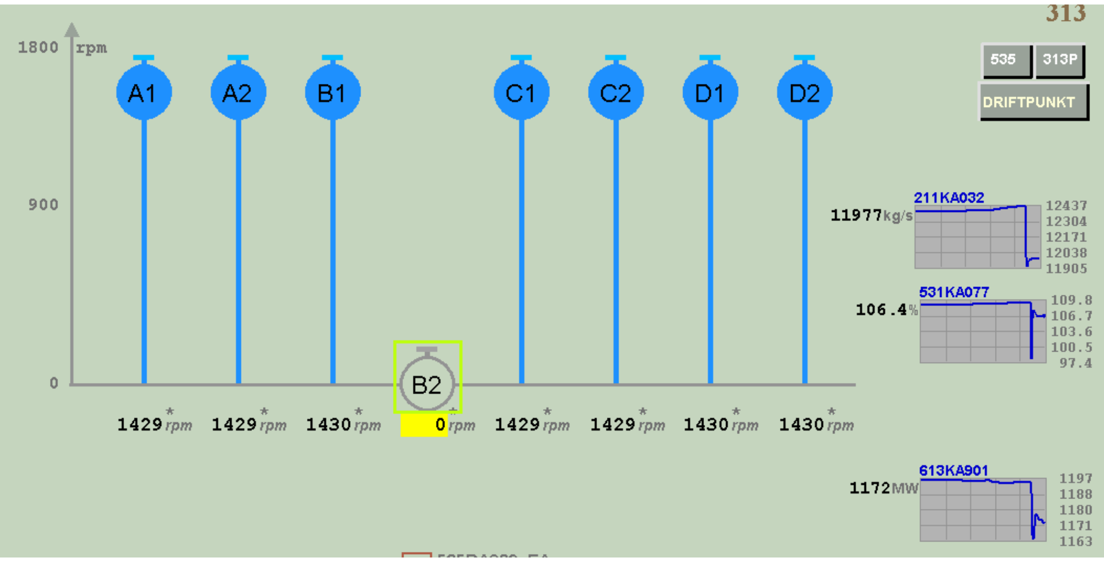
The condense flow 462KB301 is increasing and the level regulating valves for the feed water tank is opening more This can be seen in the trends. No alarms yet.

HWR-833

The turbine efficiency will decrease and because of that the 313 internal core collant pumps will increase automaticly to try to keep the electrical power out constant. If the operator diagnose the problem here, it is knowlidge based.

Scenario break
First break will be after 9.5 minutes just before the 313 pump stop.

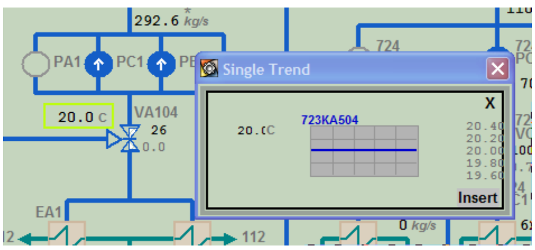
The leak will now stop increasing and after 60 second one internal core colent pump 313PB2 trip on a relay protection. This will generate alarm “313KB732L1 PB2 not running”. The Reactor Operator has to report it to the maintenance group and they will fix the pump, and 5 minutes after it is reported the Reactor Operator will be asked to start it. This problem on the reactor side can make it harder for the Turbine Operator to detect the leak in the super heater because of some power oscilations.



This is how it looks like after the 313PB2 pump has tripped. Before the trip we can see that the reactor power was increasing from 109.3 to 109.8 percent. Maybe the Reactor Operator will notice this, but probably not. After the trip the power decreases to 106.4 percent. When the 313PB2 is reported ready to use again the Reactor Operator has to reduce the power to 90 percent and then start the pump, and after that he can increase the power to 109.3 percent again.

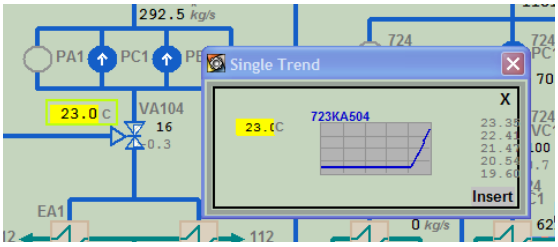
Five minutes after the trip of 313PB2 the leak in 422EB1 is starting to increase again and this time to alarm level. The alarm on diff. temp from the super heaters outlet “422KW504L1 Temp diff. EA1 – EB1” the leak flow will then be round 100 kg/s, when this alarm is coming the scenario changes from knowledge base to procedure based. In the procedure for this alarm, the operator can see that it is depending on a leak in the 422 super heaters and refers to a plant disturbing procedure. From that procedure, the operator can see that he shall decrease the power of the plant to 90 percent and then close the super heat steam valves 421VA/B5v1 until the temperature difference is below 10 C. After that the plant can run at 109 percent again.

Next malfunction is on the reactor side and a regulation valve in cooling system 723 is starting to drifting and slowly closing. This will lead to increasing temperature in cooling system 723 and the Reactor Operator will get an alarm.

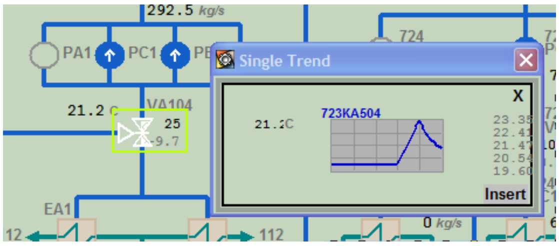


This is how it looks like before the malfunction.

Alarm on high temperature in system 723 is coming at 23 C and it looks like this:

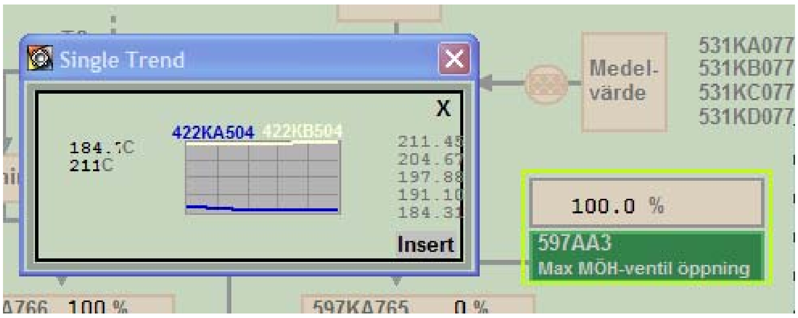


The operator has to put valve 723VA104 in hand and open it to get increasing cooling, like this:



He has then to report the problem to maintenance.

It takes about 7 minutes to run down to 90% and after that the Reactor Operator start 313PB2, all this can be done from display 313 or if that not is available, he has to use display 313P and 535. The Turbine Operator decreases the opening of 421VA/B5 to reach a diff temperature of 10 degrees. To do this he need displays 421 and 597 or if he wants he can use the popup trends of the temperatures 422KA504 and 422KB504 like this:



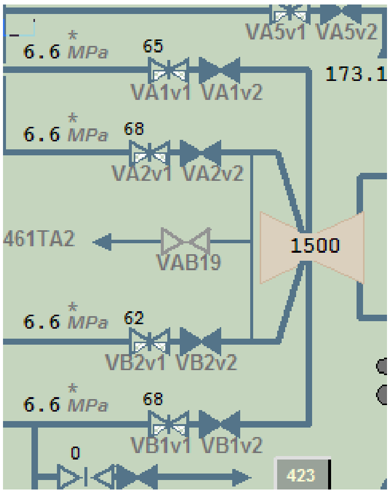
But the operator needs to see the position of the 421VA/B5 valves in display 421.

To do this will take about 10 minutes and after that the power shall increase to 109.3% again.

At 95% power (531KW077) a malfunction makes the valve 421VB2v1 stuck. It is one out of four main steam valves to the HP turbine. It is important that the Turbine Operator detect this, before reaching 100% and the running up stops. After the detection (or before maybe?) the last break will be.

HWR-833

Here we can se the diff of 421VB2v1 compared to the other three valves:



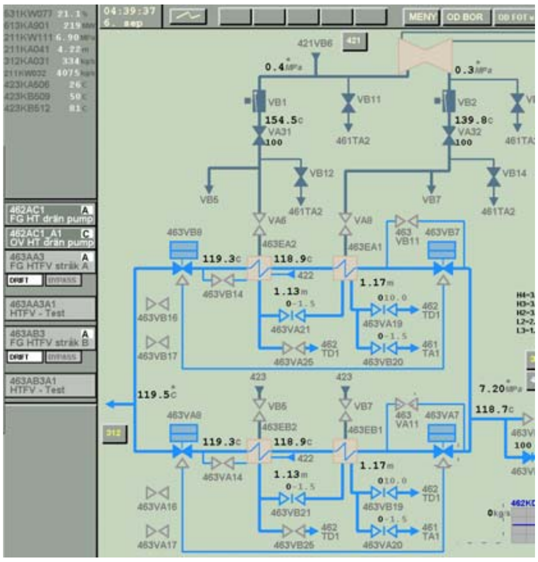
Scenario time/action scale:

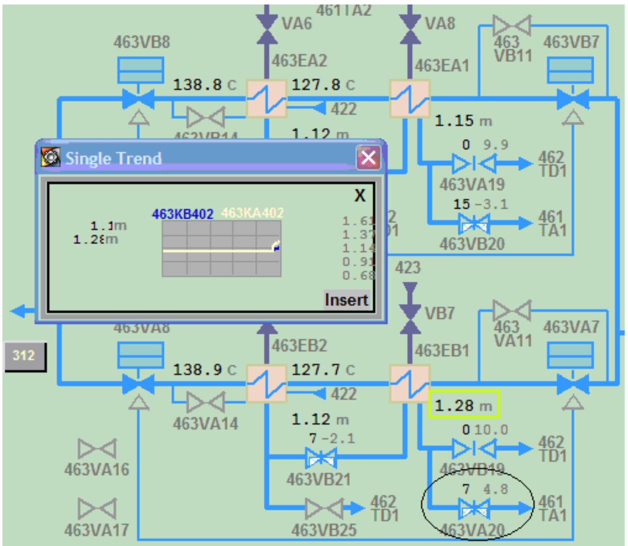
Time min.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Leak 422EA1 kg/s						0	20	30																																
Trip of 313PB2 313KB723L1										X																														
Leak 422EA1 continue															30	31	32																							
Running down power to 90 %																O	O	O	O	O	O	O																		
Malfunction on 723VA104																		x																						
Larm 422KW504L1																	X																							
Larm 422KW504L2																		X																						
Larm 723KA504H1																			X																					
Start pump 313PB2																									O	O	O	O												
Closing valves 421VA5/VB5																									O	O	O	O	O	O	O	O	O	O						
Running up power to 109 %																																			O	O	O	O	O	O
421VB2v1 stuck at > 95% power																																				x				
Detect stuck 421VB2v1																																						X	X	X
Break for questions									x																														x	
Calling control room																		x																						

WITHIN-2

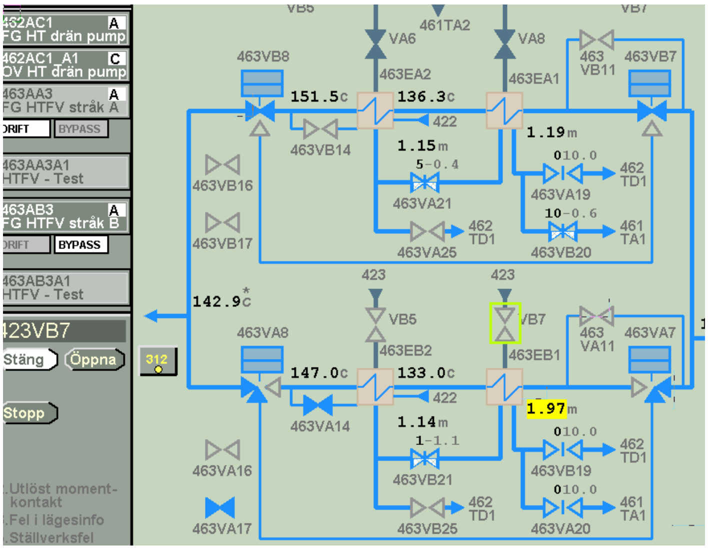
Start conditions
20 % reactor power
15 C in 316 no cooling.
Sea water temperature 10 C.
Pressure in containment -0.51 kPa and alarm. This means that the Reactor Operator can decrease the pressure if noticed.

Problem with drain switching and high-pressure pre heater bypass
The starting point for this scenario is at 20 % power just in time for opening steam to the high-pressure pre heaters. Procedure DI-3 007 page 10 Sequence 43+. The Reactor Operator start running out rods and the electrical power increase over 220 MW the limit for automatically opening of valves 423VB5, VA6, VB7, VA8 that provides the heater with steam.





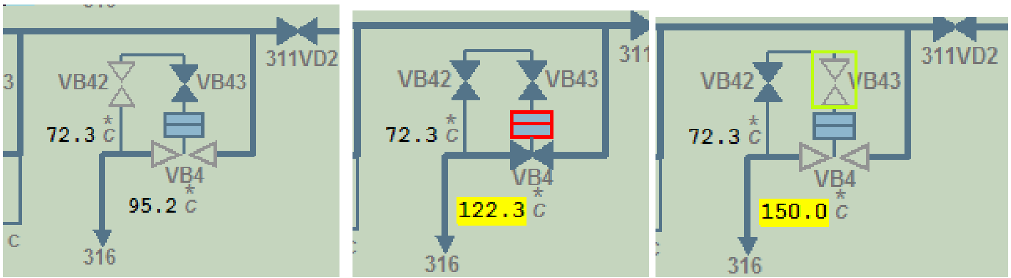
There is a malfunction that prevent the drain valve 463VA20 from open as it should to drain the water to the condenser and after one minute it closes completely. This will lead to high level in the pre heater and a bypass. The other drain valve from that pre heater 463VB19 has an auto close in this power area and cannot be opened. There is no way to prevent a bypass, because if they try to close 423VB5 and VB7, will VB7 not close depending on a malfunction.



The operators should detect that the valve does not open as it should and report it to maintenance. The valve will be fixed and works as it should in a short time. After that the operators should reset the bypass when the water level in the pre heater is OK. When they reset, the bypass 423VB7 will not open depending on the power that now can be below 220MW el and also depending on logic that in this area wants the power over 220MW el and valve 463VA20 closed for open. We want to force the crew to detect this and the way to solve it is to increase power over 220MW el and then close 463VA20 for a short time to get the auto open pulse that is necessary for valve 423VB7 to open. If the crew tries to open 423VB7 manually, it will not work due to a malfunction.

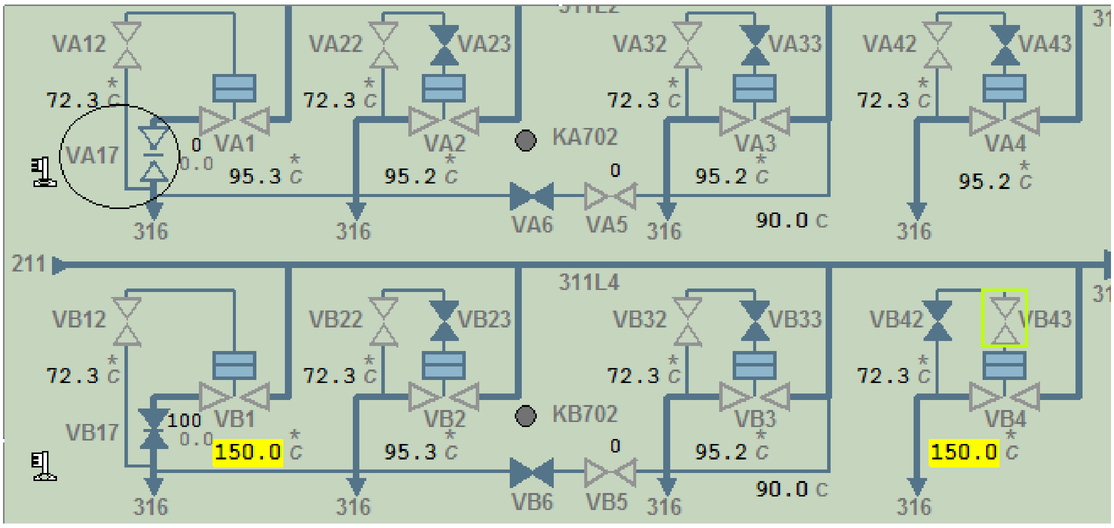
Scenario break
First break is coming after 5 minutes just before the high level alarm in 463EB1.

A few minutes after the bypass, the sea water temperature is increasing from 10 degrees to 15 degrees and this will affect the efficiency of the plant. Three minutes after that a relief valve 314VB4 is opening because of a malfunction. The Reactor Operator will get an alarm indication this and he/she can force-close it quickly. When the relief valve open, steam from reactor tank will blow down to the containment pool and it will affect the whole plant.



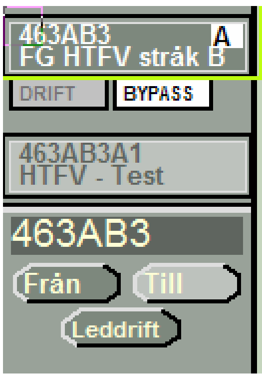
Before VB4 is open Forced closed

If the Reactor Operator is observant and power is less than 25% he/she will notice that the valve 314VA17 is closed and it should be open. The valve should be opened it is noticed.

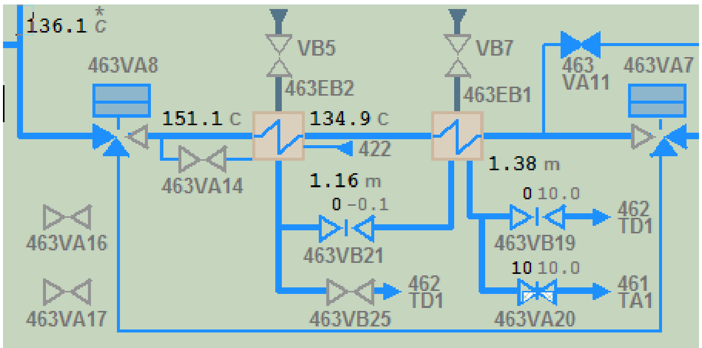


There is 314VA17 comparing to where 314VB4 is.

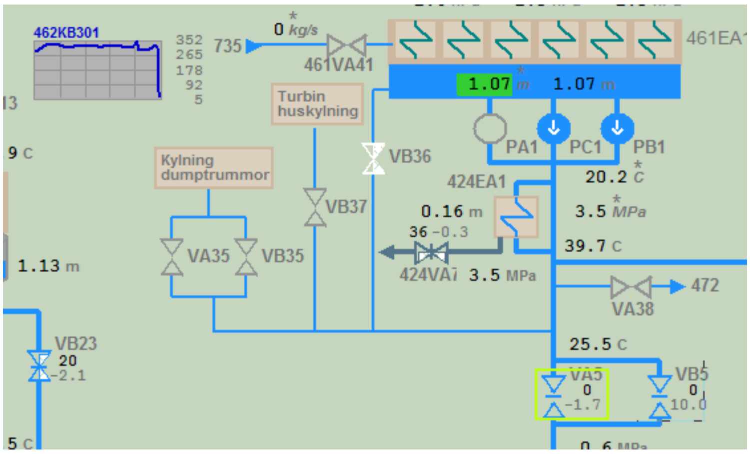
When the maintenance report 463VA20 is OK, it is time for the Turbine Operator to reset the bypass. He will do that from the function group 463AB3 by pushing the “Från” button.



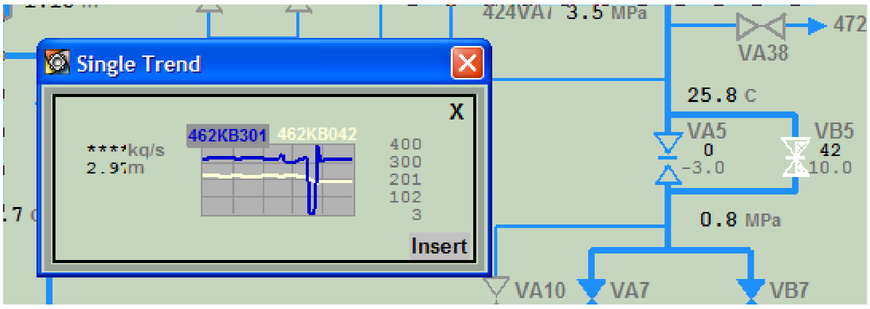
When he does that a series of actions will be performed from the function group like pressurising the pipe between the bypass valves and then heating up the pipe and after that open the bypass valves.



Under the reset of the pre heater, the regulating valve 462VA5 is closing and the standing by valve 462VB5 is not opening. This means that there will now not go any flow to the feed water tank and the Turbine Operator will get some alarms that telling him there is some problem with 462VA5.



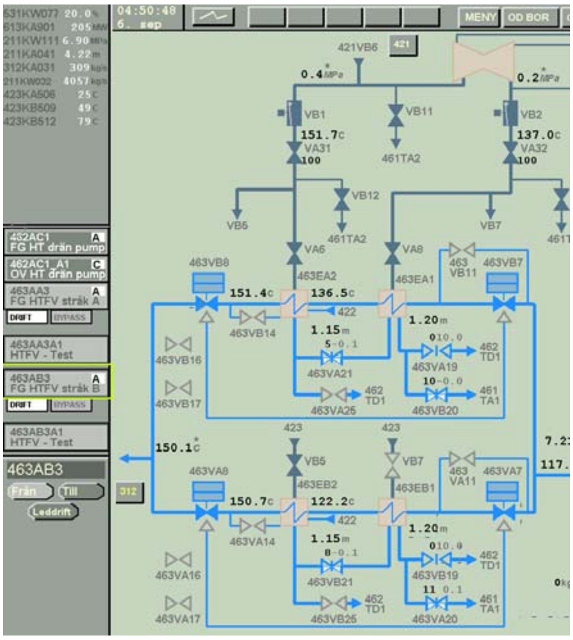
The Turbine Operator has now to test-open one of VA/B5 manually and only VB5 is possible to open. The operator has to manually regulate the water level in the feed water tank. It is easy to see on the mini trend that the flow has a drop to zero and to make it easier to get back to, and maintain constant flow the operator can use a pop up trend with the flow 462KB301. If the operator not opens any of 462VA/B5, the plat will trip after some time depending on the water level in the feed water tank. The Turbine Operator shall report the problem to maintenance and they will fix 462VA5.



Here is a pop up with the flow to and the water level in the feed water tank.

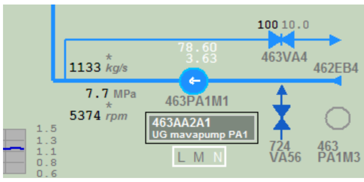
Three minutes after the problem with the 462VA/B5 a cooling water pump is tripping on the reactor side, 713PC1 and no automatic start of stand by pump is coming. The Reactor Operator has to manually start the pump 713PA1 and report the deviation to maintenance.

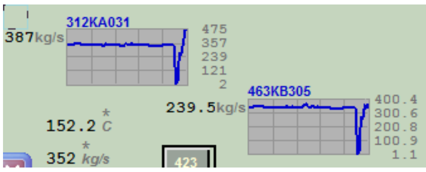
HWR-833

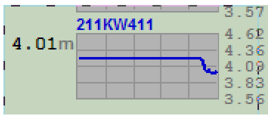


After reset bypass: Depending on how much the Reactor Operator has increased the power can the situation after reset of bypass look like this (above) and it is important that the Turbine Operator knows that the valve 423VB7 will open after the power is over 220MW electric. If the power is over 220MW when reset is taking place, the valve 423VB7 will open.

The next problem is coming after the break and that is the min flow valve for the feed water pump PA1 is open fully, 463VA4.







This will not generate any alarms, but can be seen as a flow dip in the feed water flow and decreasing water level in the reactor tank. The Turbine Operator has to shift feed water pump and stop PA1. And report the valve 463VA4 to maintenance.

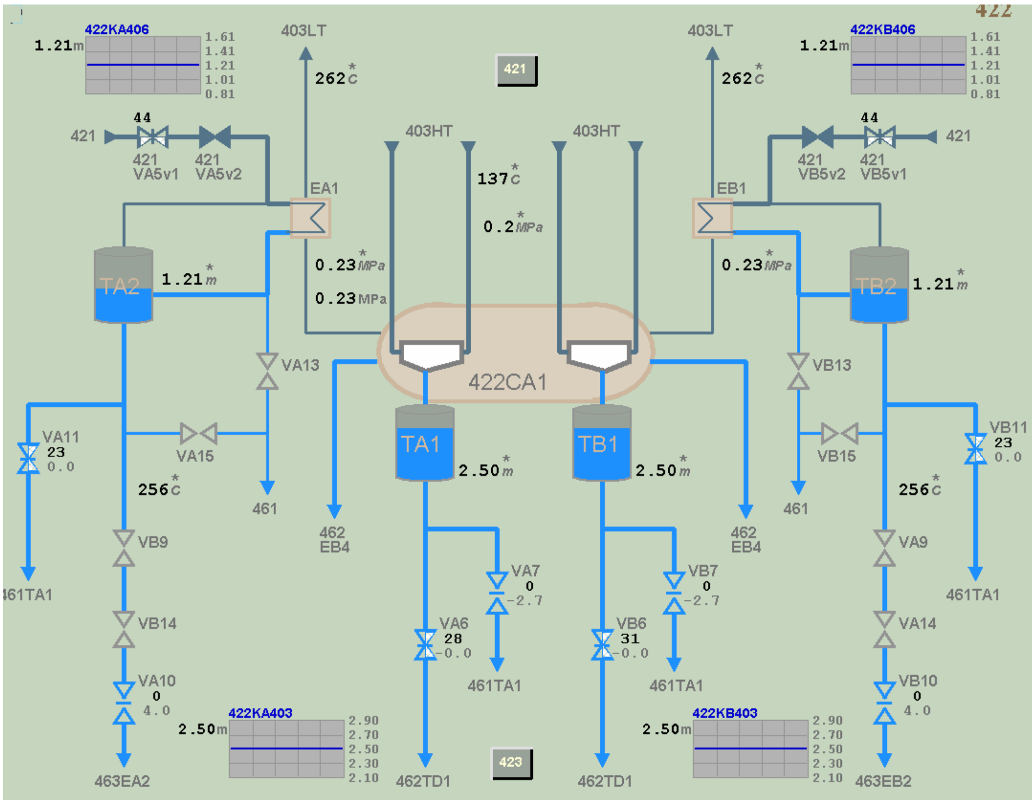
Scenario time/action scale:

Time min.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Malfunct. 423VB7	x																													
Wait for 463VA20 to open				x	x	x																								
Malfunct. 463VA20 stuck				x	x	x																								
Wait for 423VB7 pos open					x	x	x																							
Malfunct. 463VA20 closing					x	x	x																							
Larm 463KA403H1						x	x	x																						
HPPH bypass						x	x	x																						
Sea water temp inc. To 15 C									x	x	x																			
463VA20 ok reset HPPH BP												O	O	O	O															
Malfunct. 314VB4 is opening													x																	
Clos 314VB43													O																	
Malfunct. 462VB5 closing														x																
Open 462VA5 manually															O															
Malfunct. 713PC stopping																		x												
Manually start 713PA1																			O											
Malfunct. 463VA4 opening																					x									
Manually shift feed water pump																						O								
Break for questions									x											x										
Calling control room															x															

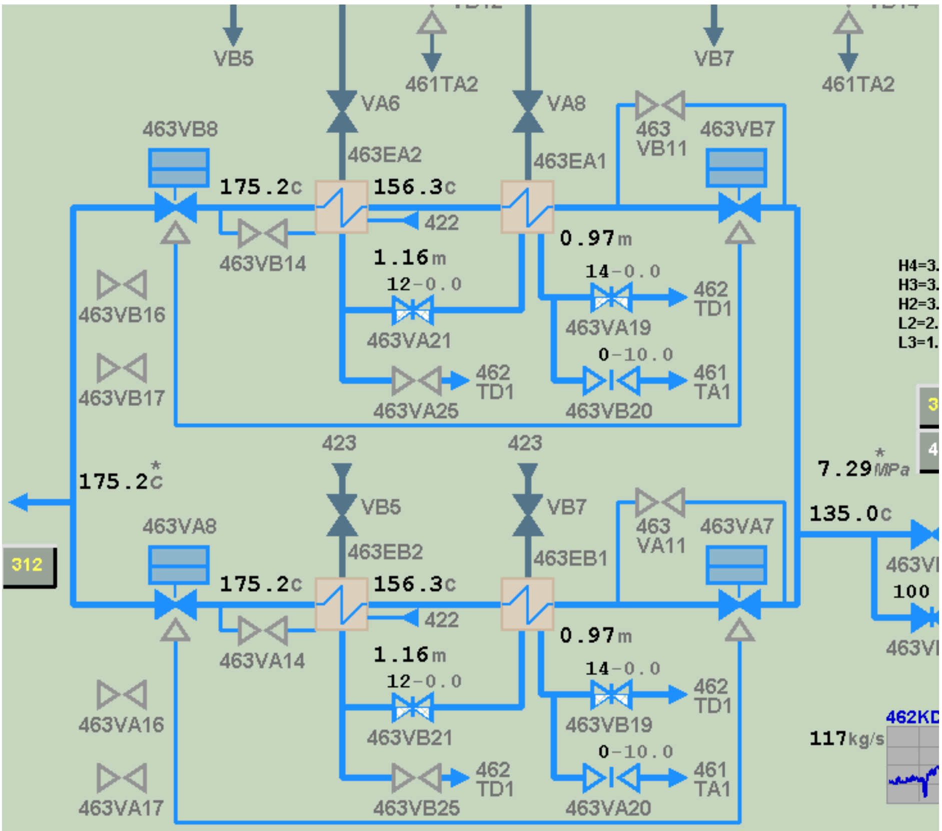
WITHIN-3

Start conditions
44 % reactor power
17 C in 316 D SUB cooling.
Sea water temperature 12 C.
Pressure in containment -0.431 kPa and alarm. This means that the Reactor Operator can decrease the pressure if he notices it.

Problem with drain switching
The starting point for this scenario is at 44 % power just in time for opening drain flow from intermediate super heater to the high-pressure pre heater. Procedure DI-3007 page 11 sequence 45+ 440MW.

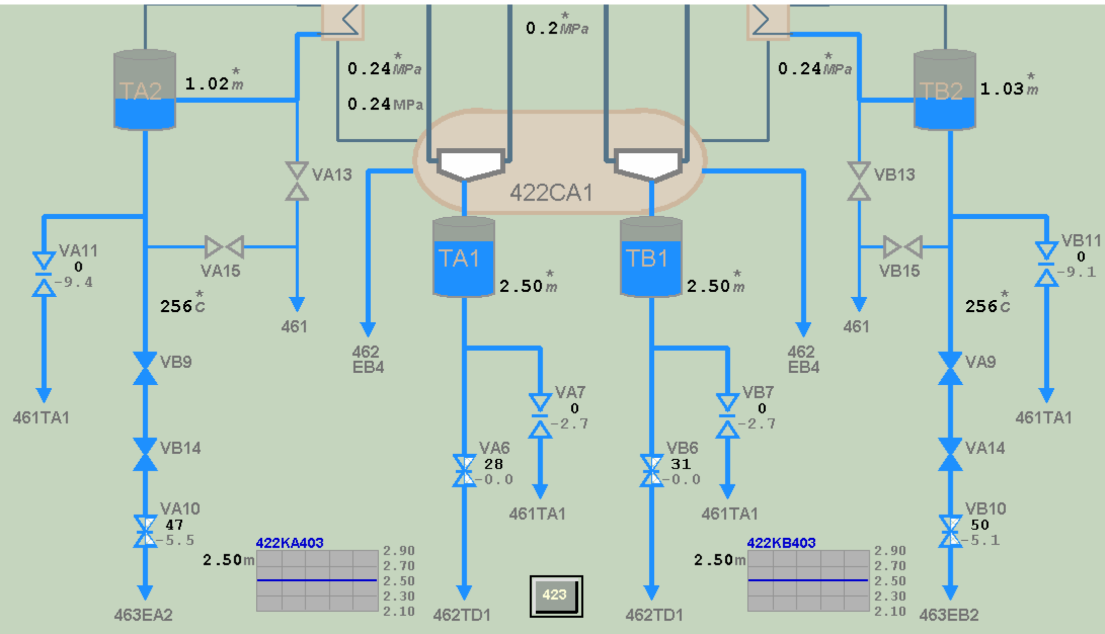


The first thing happening is that valves 422VB9, VA9, VB14, VA14 are opening. After that, the regulating valves 422VA10, VB10 start to regulate the water level in drain tanks TA2 and TB2. The regulating valves 422VA11, VB11 is then closing because they use a higher water level as setpoint than 422VA10, VB10. When this happens, the drain flow to the pre heater is increasing and level regulating valves 463VA21 and VB21 has to increase position.

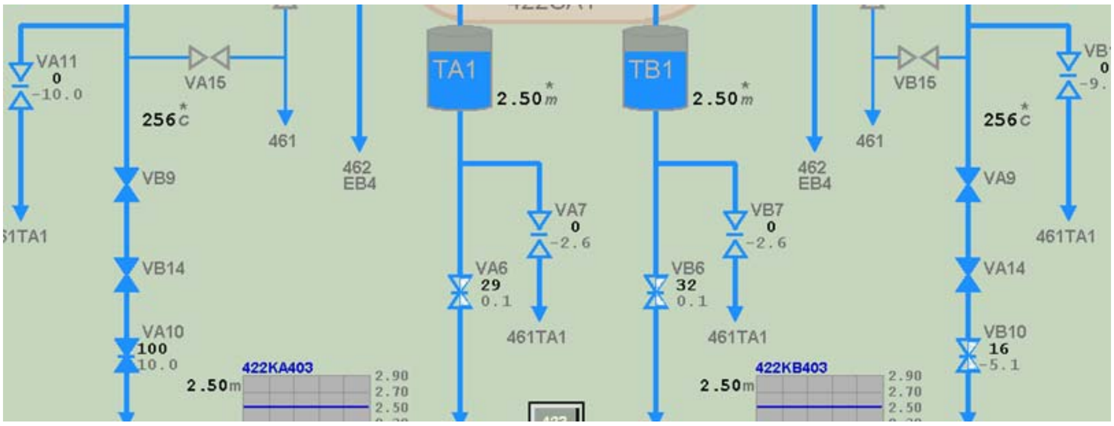


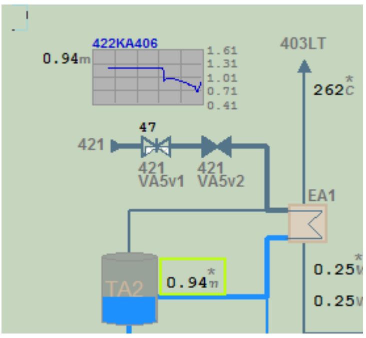
Above is the first malfunction, and valve 463VB21 is stuck. Next malfunction is that the set point that 422VA10 is using for regulating the level in tank 422TA2 will slowly decrease. The water level in 422TA2 will slowly decrease.

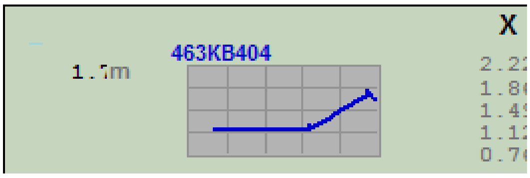
Scenario break
The break will be before low level 1 in 422TA2 and high level 1 in 463EA2 is reached.



After the break, measurement signal 422KA404, the one that control valve 422VA10 is regulating from, goes to its max value and generates a high level alarm and 422VA10 is opening to 100%. This will also generate a protection signal for the turbine. But only one out of three and two out of three is needed for the safety action.

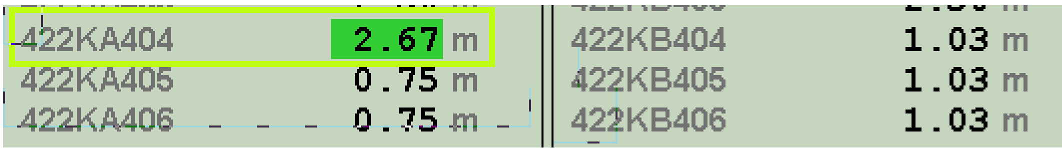




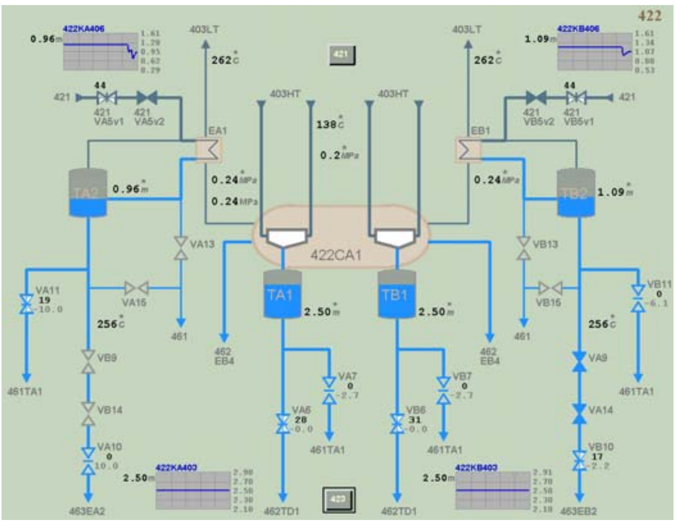


Water level in 422TA2 Water level in 463KB404

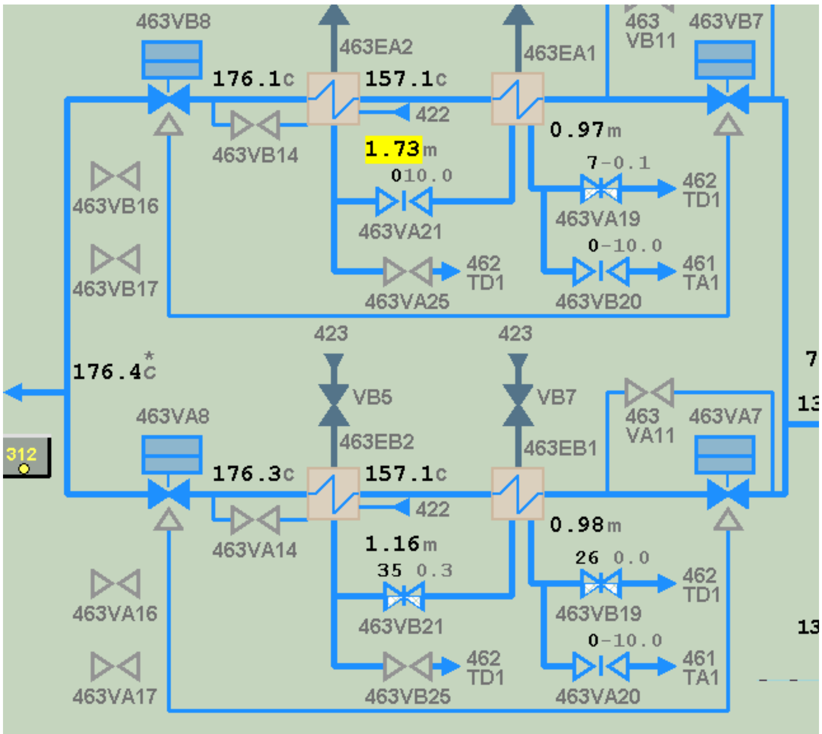
It is just one out of three measurements that indicate high water level in TA2. KA404, 405, 406 is in 422TA2.



The water level will now decrease in 422TA2 and at low level 1 from measurement 422KA406 valves 422VB9 and VB14 will close. The water level will then increase again and valve 422VA11 will take care of the water level. Five seconds after this malfunction, another error is forcing 463VA21 to close completely and the water level in high-pressure pre heater 463EA2 is increasing to bypass level. There is nothing the operator can do to prevent this.



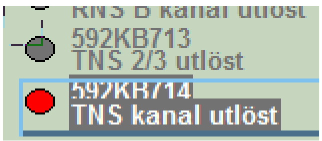
HWR-833



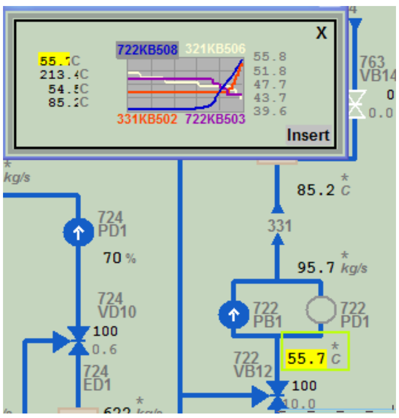
The crew has to identify these two problems and report them correctly to maintenance. Both problems will be solved, first the pre heater and then the drifting measurement. Then the crew can reset the bypass and open valves 422VB9 and VB14 to get valve 422VA10 regulating the water level in 422TA2.

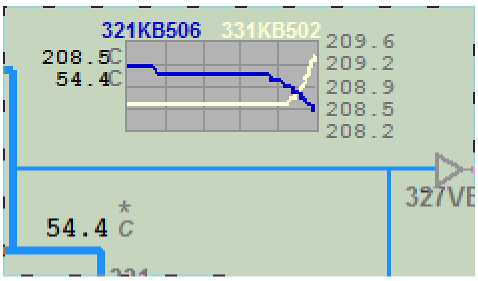
1	6. sep	18:52:47.7	592KB714	TNS kanal utlöst		0.27
2	6. sep	18:52:47.7	463KB404H1	Nivå EA2	1.72	1.88
3	6. sep	18:52:47.7	422KA404H2	Nivå i TA2	2.10	2.70
2	6. sep	18:52:48.9	422KA406L1	Nivå i TA2	0.74	1.14
2	6. sep	18:53:34.6	463KB404H1	Nivå EA2	1.72	1.88

The alarm list looks like this (above) and can be a little confusing because it indicates a remaining high level alarm in 422TA2 422KA404H2, that is a malfunction, and then it indicates a low level alarm in disappearing status 422KA406L1, and that one has been real. The high level alarm is setting off 592KB714 TNS channel.

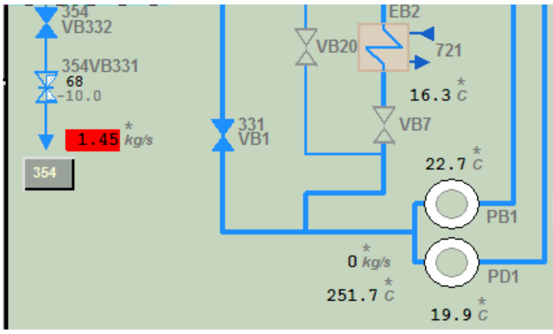


Above we can see it from display 592, one TNS channel but not TNS 2/3.



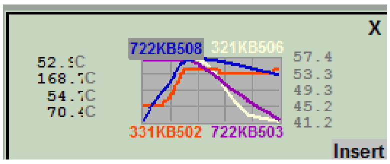


On the reactor side the heat exchanger 722EB1 and 722ED1 is starting to lose efficiency and after a few minutes it will lead to high temperature 722KB508H1 (55 C) and trip of 321 pumps. There is a possibility to see it from trends in the 321 display if the operator has it open before the alarm. Now the Reactor Operator can not run out rods depending on no cooling flow (354KB301) to the rods.

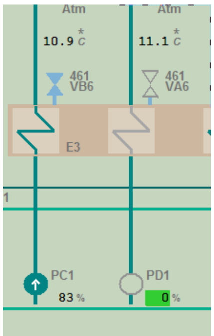


Display 321 is shown above.

The temperature in 722 will then decrease again and the operator may try to start 321 again.



Five minutes after this a main cooling seawater pump 441PD1 is stopping because a malfunction closes valve 461VA6. The crew should be able to detect why the pump was tripping by inspecting the event list.



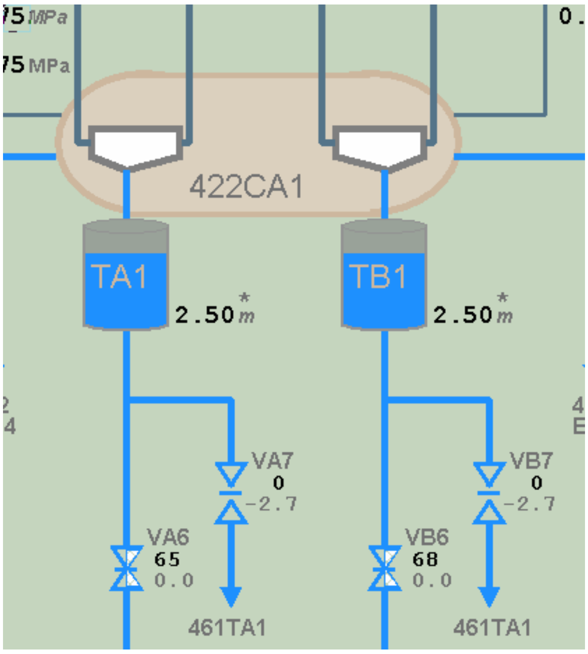
Scenario time/action scale:

Time min.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Wait for 422VB9 to open			x	x	x																									
Malfunct. 463VA21 stuck 30%			x	x	x																									
Malfunct. 422KA404 drifting				x	x	x																								
Malfunct. 422KA404 max						x																								
Alarm 422KA404H2 Nivå 422TA2						x																								
Alarm 422KA406L1 Nivå 422TA2						x																								
Malfunct. 463VA21 closing							x																							
Alarm 463KB404H1 Nivå 463EA2							x																							
HTFVBP								x																						
463VA21 ok reset HPPH BP											O																			
422KA404 ok												O																		
Reset HPPHBP												O	O	O	O	O	O													
Malfunct. 722 cooling eff. decr.								x	x	x	x	x	x	x	x															
TRIP 321 pumps																x														
Malfunct. 461VA6 closing																						x								
TRIP 441PD1																						x								
Break for questions						x																	x							
Calling control room															x															

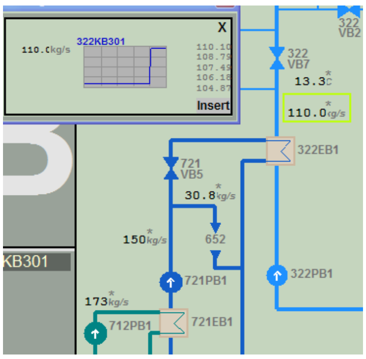
BEYOND-1

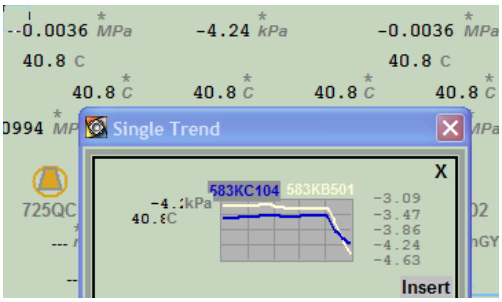
Start conditions
109.3 % reactor power
19.5 C in 316 cooling from SUB C.
Sea water temperature 12 C.
Pressure in containment -3.6 kPa

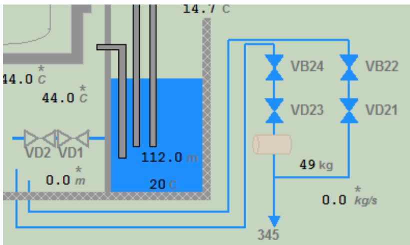
Turbine trip with the generator still connected to the grid
After 2 minutes a malfunction on the 422VB6 makes it close very little and stuck. This makes the water level in drain tank 422TB1 increase very slow from normal level 2.5m. When the water level reach 2.8m a high level alarm 422KB403H1 should set off but that fails and the stand by valve 422VB7 is stuck. At a water level at 3.3m a high level alarm 422KB401(2, 3) H2 is coming and trip the turbine after 20 seconds. Before this alarm, the break should be and the operators should know what's coming.

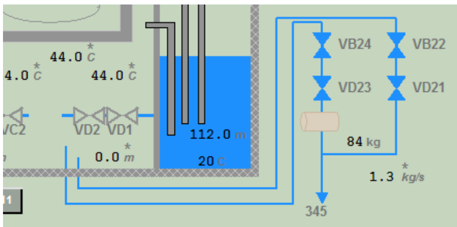


On the reactor side, a valve that uses in case of sprinkling of dry well (422VB7) starting to leak and a flow of 8 kg/s is going in to drywell.







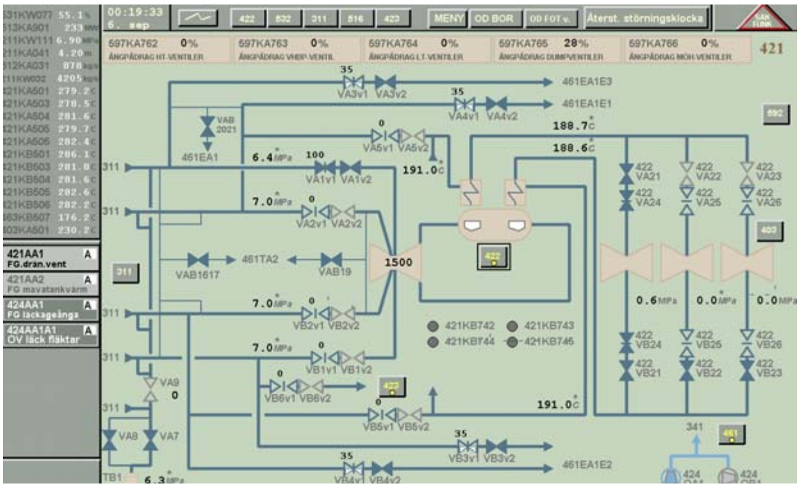


The leak flow from containment increases.

Consequences of this are that the flow from 322PB1 is increasing and pressure and temperature in dry well will decrease abnormally in upper drywell. The right thing to do here is to stop 322PB1 and start cooling of wet well from another SUB, but this is hard to find.

Scenario break
The break comes after 7 minutes just before 422VB6 will close.

After the turbine trip, one steam line with valves 421VA1, VA2, 422VA21, VA24, VB24, VB21 is still open and the generator is producing 230 MW electric. If everything is working as it should the generator breaker will trip on low power (11 MW).



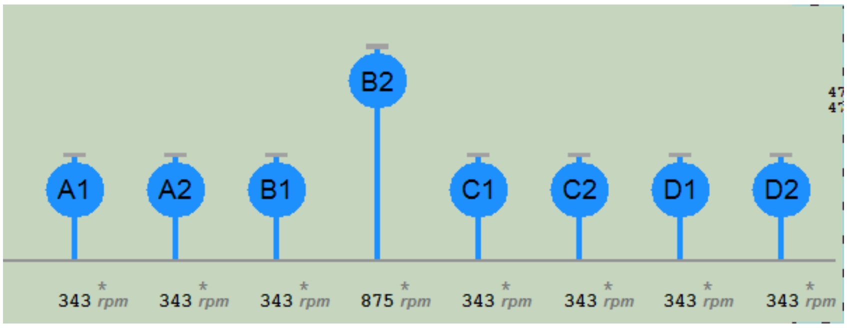
This is how it looks with the open valves from one steam line to the HP turbine and through one LP turbine. The rest of the steam is going throe the bypass valves to the condenser.

These are some actions that the crew can do:

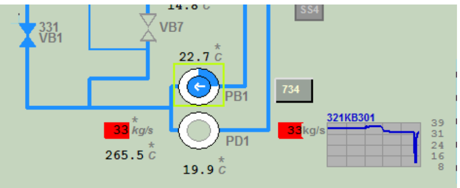
1. Try to close the main steam valves in the steam line that is feeding the HP turbine. This is a good action but it will not work due to malfunctions, the 311 valves in that steam line is not possible to close.
2. Open the generator breaker. This will make the turbine rush up in speed and is not a good action.
3. Contact the maintenance group. If the operators do this, they can make the valve 421VA1v1 close from 100% to 22% and no more.
4. Scram the plant. This is the only way to solve the problem, but they need to know that it will make the pressure in the reactor tank decreasing fast and this will lead to that the feed water pumps top fill the reactor tank if not the Turbine Operator run them manually. If the reactor will be top filled, a feed water isolation chain set off and stops the feed water pumps and close all 311 steam valves except the two with malfunction. No steam to the ejectors means that the pressure in the condenser will increase and the leaking steam from the turbine will leak out in the turbine plant, which is not good.

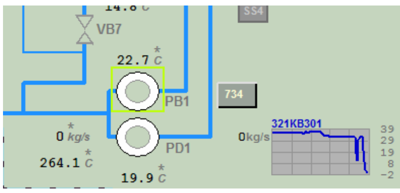
A few minutes after TS a 313 pump on the reactor side increase its speed and has to be stopped by the operator and reported to maintenance.

HWR-833



Three minutes after the 313 pump stops, a malfunction on the low flow setpoint for 321KB301 makes the 321 pump PB1 trip. The operator shall detect that low flow never disappears when he try to start the pump and then report it to maintenance.





The low flow alarm disappears when the pump stops. This scenario is tricky and the crew needs to think before taking action.

Scenario time/action scale:

Time min.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Malfunct. 422KB401				x																										
Malfunct. 422VB6				x																										
Malfunct. 322VB4							x																							
Malfunct. 422VB6 close									x																					
Larm 422KB403H2										X																				
TS										X																				
Malfunct. 313PB2												x																		
Stop pump 313PB2													O																	
Malfunct. 321KB301															x															
Trip of 321PB1																X														
Break for questions							x																							
Calling control room												x																		

BEYOND-2

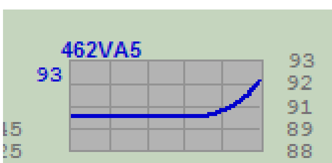
Start conditions

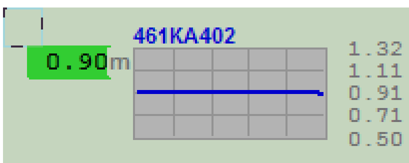
109.3 % reactor power
19.5 C in 316 cooling from SUB C.
Sea water temperature 12 C.
Pressure in containment -3.6 kPa

Leak in condense cleaning building KRA 332

After 2 minutes a small leak in the KRA starts growing up to 42 kg/s in 3 minutes. After these 3 minutes, the first alarms are coming. It is a low level alarm in the condenser and a high level alarm in the KRA building.

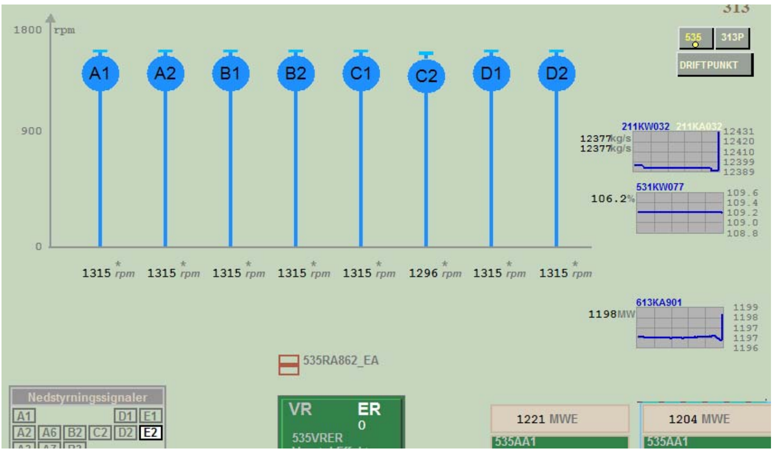
In the process, the only parameters that have changed are the position of level regulating valve in the feed water tank 462VA5 and the water level in the condenser. The flow from the condenser to the condense cleaning system (KRA) 332 has increased.



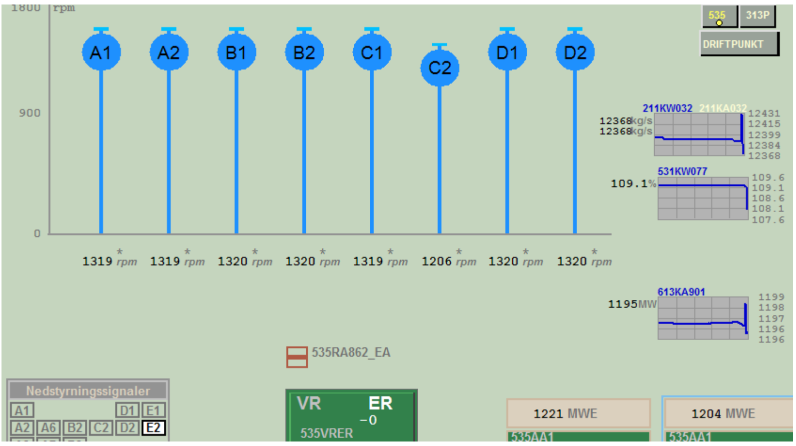


Scenario break
Before alarm 345KD437H2 high water level in room Z02.10 is coming.

After the break, the operators will understand that they have a leak in the KRA building when the alarm 345KD437H2 high water level in room Z02.10 is coming. After one minute, core cooling pump 313PC2 is increasing to max speed and the leak starts to increase up to 170 kg/s in 3 minutes.



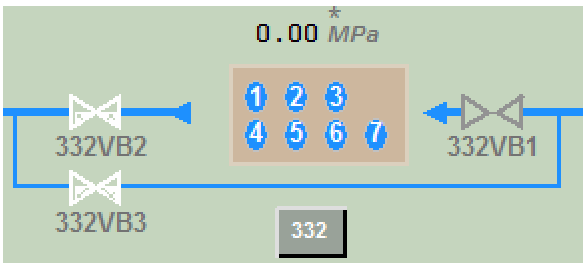
After 10 seconds the same pump is decreasing speed to 1206 rpm.



After 313PC2 has decreased.

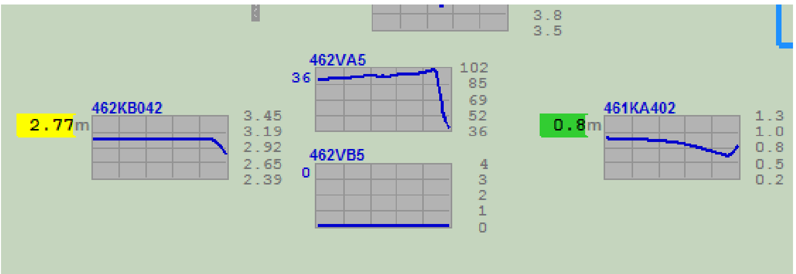
The Reactor Operator shall stop this pump and report it to maintenance.

During these 3 minutes while the leak flow to KRA is increasing the crew will probably decide to bypass KRA to see if they can stop the leak. If they do so, the leak will decrease to 48 kg/s because one valve 332VB2 is stuck at 4% open. They can report this or send out a field operator and in that case the valve will be closed after some time, but no flow from 462, 462VA/B5 shall be closed a short time.



HWR-833

The operators will now get a low level 2 and 3 in the condenser. At the same time, the level-regulating valve for the feed water tank 462VA5 is starting to close slowly and the water level in the condenser is increasing at the same time as the water level in the feed water tank starts to decrease.



Here we can see the position of 462VA5 is decreasing, and the water level in feed water tank 462KB042 is decreasing - will increase in condenser 461KA402.

The standing by valve for 462VA5 that is 462VB5 is not opening but booth valves are possible to operate manually. The interesting part of this scenario is to see if the operators identify the leak before the alarms and what they decide to do after understanding the situation. The problem with 462VA5 can take some time to detect and one operator has to regulate the water level in the feed water tank by hand until maintenance has fixed the valve.

Scenario time/action scale:

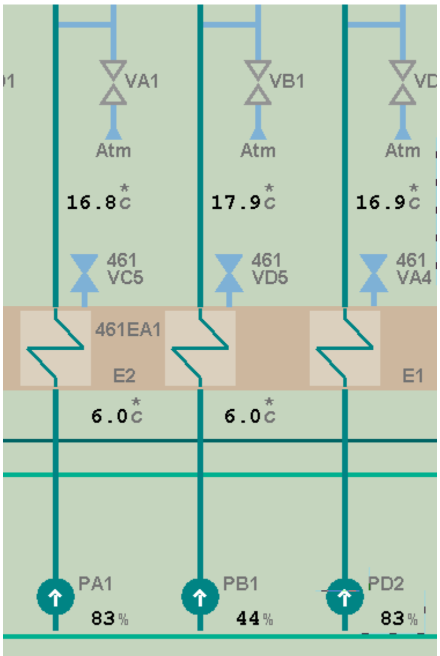
Time min.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	3
Malfunct. Leak KRA			x	x	x																												
Larm 461KA401L1						X																											
Malfunct. 313PC2						x																											
Larm 461KA401L2										X																							
Larm 461KA401L2												X																					
Malfunct. 462VA5 closing													x	x	x																		
Larm 462KB042L1 fedw.tank																X																	
Leak in room B01.56																		x															
H-isolation																															X		
Break for questions						x																										x	
Calling control room															x																		

BEYOND-3

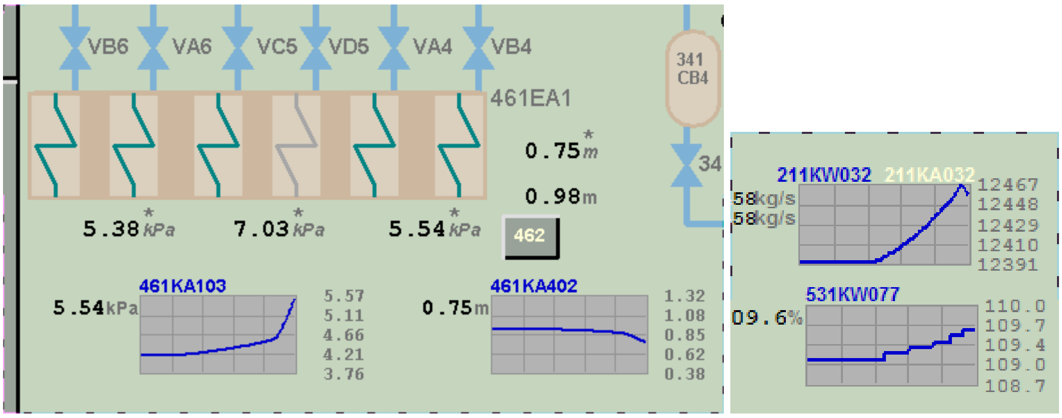
Start conditions
109.1 % reactor power
19.5 C in 316 cooling from SUB C.
Sea water temperature 12 C.
Pressure in containment -4.0 kPa

High temperature in the sea
The scenario begins with a seawater-cooling pump in the condenser starting to decrease in speed until it trips after 5 minutes. The pressure in the condenser will increase in chamber 461EA1.E2. The flow will decrease from that pump and the temperature after 461EA1E1 will increase.

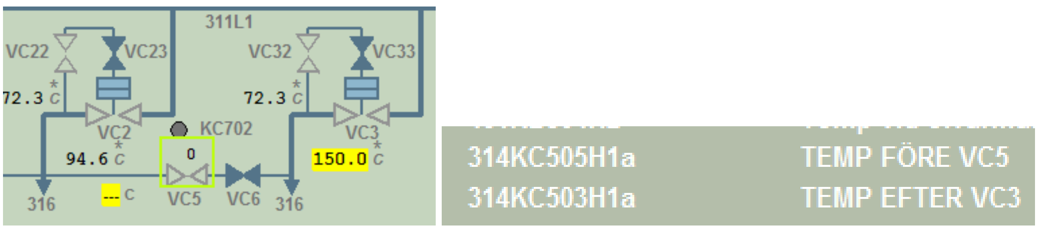
Scenario break
After the seawater-cooling pump trips (5 minutes).



After that the seawater temperature starts to increase from 12 C to 18 C in five minutes. This will lead to increasing pressure in the condenser and higher temperature. The plant efficiency will decrease and the Reactor Operator has to reduce power.

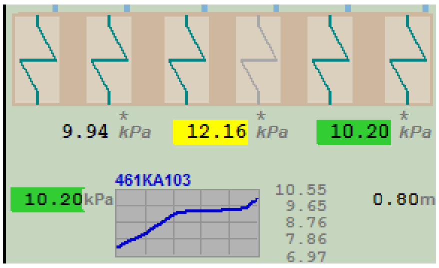


During the next five minutes, a relief valve 314VC5 starts to leak and after some time, an alarm is telling the operator about the problem. The operator shall close 314VC6 and open 314VC5.



Then the heat transfer in the condenser starts to decrease and the pressure in the condenser increases more.

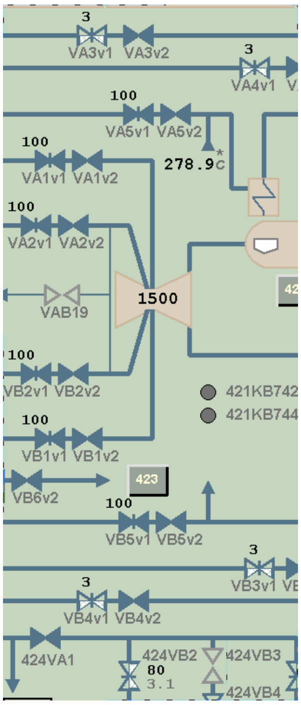
HWR-833



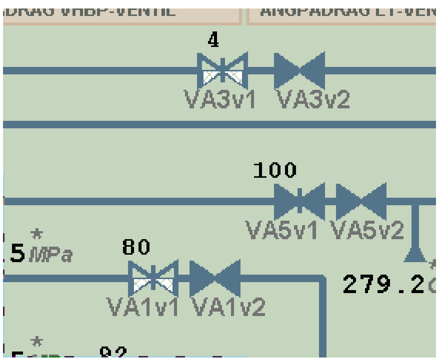
Alarms from the cooling system will now come, as it gets hotter.

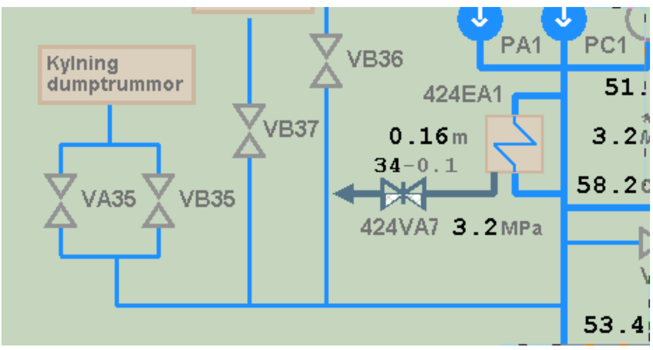
6. sep	00:49:29.0	461KB108H1	Tryck i kylare QB1.E3	0.05	0.05
6. sep	00:49:29.0	461KA108H1	Tryck i kylare QA1.E3	0.05	0.05
6. sep	00:52:18.1	461KB107L1	Tryck e.reglervent VA12	0.70	0.97
6. sep	00:52:18.1	461KA104H1	Ej I tr. i kond	10.00	11.34
6. sep	00:53:46.0	724KB505H1	Temp efter 724-KYLARE	25.00	27.09

Later, the load equipment malfunction and the bypass valves open a few percent; no alarms are coming on this because of malfunctions.

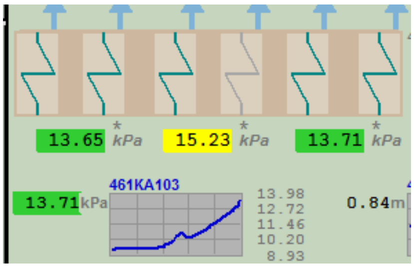


The crew will now decrease power at list to the bypass valves are closed. After that, the heat transfer in the condenser decreases even more, and the seawater temperature increases to 25 C in 5 minutes.
Now a new malfunction makes bypass valve 421VA3v1 open a few percent.

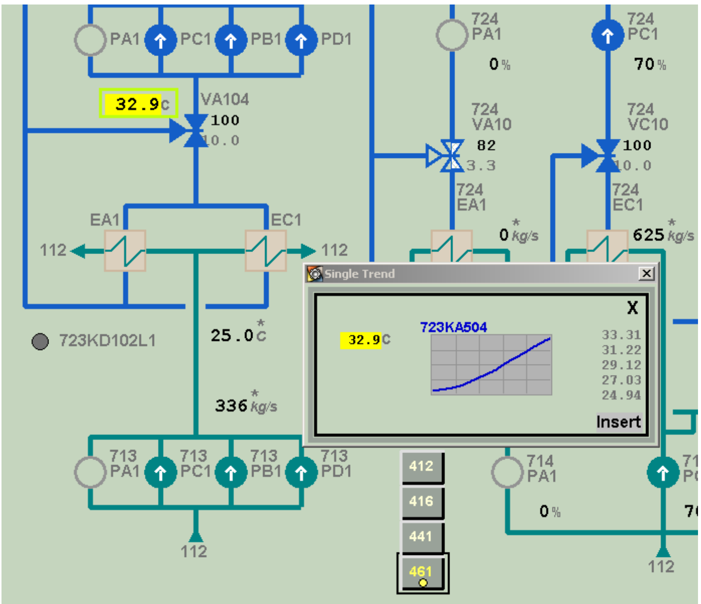




No cooling flows in the bypass inlet to the condenser. The operator can open it manually.



The pressure in the condenser is now over high level 2 in one chamber and after a short time, the reactor is automatically decreasing power to 73% if not the crew has done this already.



On the reactor side, the efficiency of the heat exchanger 723EA1 and EB1 is decreasing and the temperature increases in the cooling system and leads to problem in cooling objects and the containment atmosphere.

The crew needs to stop the station, but no manually protection chains are possible to set off except containment isolation, but it is better to start power reduction manually.

Scenario time/action scale:

Time min.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Malfunct. 441PD1 decr. Speed		x	x	x	x	x																											
Sea water temp 12 - 18 C						x	x	x	x	x																							
Larm 461KA104H1							X																										
Malfunct. 314VC5										x	x	x	x	x																			
Larm 314KC505H1												X																					
Larm 314KC503H1												X																					
Closing 314VC6												O																					
Eff.Loss in condenser														x	x																		
Malfunct. Bypass valves open																x																	
Decrease power to close BPV																	O																
Sea water temp 18 - 25 C																x	x	x	x	x													
Eff.Loss in condenser																					x												
BPV 421VA3v1 open 3%																					x												
Eff.Loss in 723 reaktor cooling																						x	x	x	x	x							
Break for questions						x																										x	
Calling control room																		x															

12. APPENDIX 3: PROCESS OVERVIEW ITEMS

Instructions to participants

During the experimental training, the Process overview questionnaire was demonstrated to the operators. It was explained that the term ‘recently’ refers to:

- The period of time towards the end of the scenario phase where a question is meaningful. It does not make sense, for example, to ask the operators whether the level of an empty tank has decreased. This question would therefore refer to process states that occurred after the tank was filled with water.
- The period of time towards the end of the scenario phase where the addressed process state behaved consistently, i.e. the process state did not change between an increasing, stable, and decreasing behaviour.

The participants were instructed to indicate for each questionnaire item whether the process state had recently increased, remained stable or decreased.

Within-1

Detection phase

1. Recently, the position of 422VA6 has:
   decreased - remained stable - increased
2. Recently, the number of 441 main coolant pumps in operation has:
   decreased - remained stable - increased
3. Recently, the electrical output power of 613KA901 has:
   decreased - remained stable - increased
4. Recently, the flow from high pressure drain tank 462TD1 has:
   decreased - remained stable - increased
5. Recently, the level in condenser 461KA40x has:
   decreased - remained stable - increased
6. Recently, the temperature after the condensate pumps 462KA503 has:
   decreased - remained stable - increased
7. Recently, the state of the high pressure drain tank regulating valve 462VC19 has:
   decreased - remained stable - increased
8. Recently, the level in 422TB1 422KB403 has:
   decreased - remained stable - increased
9. Recently, the temperature in the combined tube before HPPH 463KB501 has:
   decreased - remained stable - increased
10. Recently, the pressure in condenser 461KA1xx has:
    decreased - remained stable - increased
11. Recently, the position of the level regulating valve 462VA5 for the feedwater tank has:
    decreased - remained stable - increased
12. Recently, the feedwater temperature after HPPH 463KB507 has:
    decreased - remained stable - increased

Local items: 1, 3, 4, 7, 11, 12
Global items: 2, 5, 6, 8, 9, 10

Mitigation phase

1. Recently, the condense flow to the feedwater tank 462KB301 has:
   decreased - remained stable - increased
2. Recently, the pressure in the condenser has:
   decreased - remained stable - increased
3. Recently, the turbine rotation speed has:
   decreased - remained stable - increased

4. Recently, the position of 421VB1v1 has:
   decreased - remained stable - increased
5. Recently, the position of 421VB2v1 has:
   decreased - remained stable - increased
6. Recently, the temperature after 422EA1 422KA504 has:
   decreased - remained stable - increased
7. Recently, the number of closed 421 valves has:
   decreased - remained stable - increased
8. Recently, the electrical output power of 613KA901 has:
   decreased - remained stable - increased
9. Recently, the temperature difference between 422KA504 (after 422EA1) and 422KB504 (after EB1) has:
   decreased - remained stable - increased
10. Recently, the position of valve 462VA/B35 has:
    decreased - remained stable - increased
11. Recently, the number of 463PX1 feedwater pumps in operation has:
    decreased - remained stable - increased
12. Recently, the level in condenser 461KA40x has:
    decreased - remained stable - increased

Local items: 1, 2, 4, 5, 9, 10
Global items: 3, 6, 7, 8, 11, 12

Within-2

Detection phase

1. Recently, the position of valve 462VC19 has:
   decreased - remained stable - increased
2. Recently, the flow from 462TD1 (462KD302) has:
   decreased - remained stable - increased
3. Recently, the position of 422VA6 has:
   decreased - remained stable - increased
4. Recently, the level of 422TB1 422KB403 has:
   decreased - remained stable - increased
5. Recently, the number of 441 main coolant water pumps in operation has:
   decreased - remained stable - increased
6. Recently, the flow from the condenser to feedwater tank 462KB301 has:
   decreased - remained stable - increased
7. Recently, the electrical output power of 613KA901 has:
   decreased - remained stable - increased
8. Recently, the 403KA7xx turbine bearing vibrations have:
   decreased - remained stable - increased
9. Recently, the temperature in the combined tube before HPPH 463KB501 has:
   decreased - remained stable - increased
10. Recently, the position of valve 462VA5 has:
    decreased - remained stable - increased
11. Recently, the feedwater temperature after HPPH 463KB507 has:
    decreased - remained stable - increased
12. Recently, the number of 462Px1 condensate pumps in operation has:
    decreased - remained stable - increased

Local items: 1, 2, 3, 7, 10, 11
Global items: 4, 5, 6, 8, 9, 12

Mitigation phase

1. Recently, the level in 462TD1 has:
   decreased - remained stable - increased
2. Recently, the pressure in the condenser has:
   decreased - remained stable - increased
3. Recently, the temperature in the combined tube before HPPH 463KB501 has:
   decreased - remained stable - increased
4. Recently, the number of 462Px1 condensate pumps in operation has:
   decreased - remained stable - increased
5. Recently, the temperature after the 462KA503 condensate pumps has:
   decreased - remained stable - increased
6. Recently, the flow from the condenser to feedwater tank 462KB301 has:
   decreased - remained stable - increased
7. Recently, the output power (613KC901) has:
   decreased - remained stable - increased
8. Recently, the number of 462Px1 condensate pumps in operation has:
   decreased - remained stable - increased
9. Recently, the position of the main flow valve 463VA4 to feedwater pump A has:
   decreased - remained stable - increased
10. Recently, the flow to offgas system 341 (341KB301) has:
    decreased - remained stable - increased
11. Recently, the condenser flow from low pressure drain pump 462KB305 has:
    decreased - remained stable - increased
12. Recently, the pressure in condenser 461KA1xx has:
    decreased - remained stable - increased

Local items: 1, 2, 7, 9, 10, 11
Global items: 3, 4, 5, 6, 8, 12

Within-3

Detection phase

1. Recently, the flow from 462TD1 (462KD302) has:
   decreased - remained stable - increased
2. Recently, the level in the steam super heater drainage tank 422TA2 has:
   decreased - remained stable - increased
3. Recently, the level in condenser 461KA40x has:
   decreased - remained stable - increased
4. Recently, the feedwater flow 312KA031 has:
   decreased - remained stable - increased
5. Recently, the level in HPPH drainage tank 463EA2 has:
   decreased - remained stable - increased
6. Recently, the flow to offgas system 341, 341KB301 has:
   decreased - remained stable - increased
7. Recently, the position of 463VA21 has:
   decreased - remained stable - increased
8. Recently, the number of closed 421 valves has:
   decreased - remained stable - increased
9. Recently, the position of the backup drainage regulating valve from HPPH EA1 463VB20 has:
   decreased - remained stable - increased
10. Recently, the number of 462Px1 condensate pumps in operation has:
    decreased - remained stable - increased
11. Recently, the position of the auxiliary drainage valve from 422TA2, 422VA11 has:
    decreased - remained stable - increased
12. Recently, the feedwater temperature after HPPH 463KB507 has:
    decreased - remained stable - increased

Local items: 1, 2, 5, 7, 9, 11
Global items: 3, 4, 6, 8, 10, 12

Mitigation phase

1. Recently, the position of the vacuum valve from 461EA1E1 441VA6 has:
   decreased - remained stable - increased
2. Recently, the sea water temperature (441KA509) has:
   decreased - remained stable - increased
3. Recently, the flow from 462TD1 (462KD302) has:
   decreased - remained stable - increased
4. Recently, the level in 422TB1 422KB403 has:
   decreased - remained stable - increased
5. Recently, the level in condenser 461KA40x has:
   decreased - remained stable - increased
6. Recently, the number of 441 main coolant pumps in operation has:
   decreased - remained stable - increased
7. Recently, the temperature on 441KX511/512 has:
   decreased - remained stable - increased
8. Recently, the temperature after HPPH (463KB507) has:
   decreased - remained stable - increased
9. Recently, the flow to offgas system 341, 341KB301 has:
   decreased - remained stable - increased
10. Recently, the total feedwater flow 312KA031 has:
    decreased - remained stable - increased
11. Recently, the condensate flow to the feedwater tank 462KB301 has:
    decreased - remained stable - increased
12. Recently, the pressure in the condenser (461KA105) has:
    decreased - remained stable - increased

Local items: 1, 2, 3, 7, 8, 12
Global items: 4, 5, 6, 9, 10, 11

Beyond-1

Detection phase

1. Recently, the level in 422TB1 has:
   decreased - remained stable - increased
2. Recently, the feedwater temperature after HPPH 463KB507 has:
   decreased - remained stable - increased
3. Recently, the level in feedwater tank 462KB402 has:
   decreased - remained stable - increased
4. Recently, the flow to 462TD1 has:
   decreased - remained stable - increased
5. Recently, the number of 462Px1 condensate pumps in operation has:
   decreased - remained stable - increased
6. Recently, the number of closed 421 valves has:
   decreased - remained stable - increased
7. Recently, the condenser pressure 461KA1xx has:
   decreased - remained stable - increased
8. Recently, the flow to offgas system 341, 341KB301 has:
   decreased - remained stable - increased
9. Recently, the number of 441 main coolant pumps in operation has:
   decreased - remained stable - increased

Local items: 1, 4, 8 (it was impossible to find more than 3 local items)
Global items: 2, 3, 5, 6, 7, 9

HWR-833

Mitigation phase

1. Recently, the turbine rotation speed 452KA812 has:
   decreased - remained stable - increased
2. Recently, the feedwater flow 463KB305 has:
   decreased - remained stable - increased
3. Recently, the electrical output power 613KC901 has:
   decreased - remained stable - increased
4. Recently, the pressure in the condenser has:
   decreased - remained stable - increased
5. Recently, the number of 441 main coolant pumps in operation has:
   decreased - remained stable - increased
6. Recently, the electrical output power 613KA901 has:
   decreased - remained stable - increased
7. Recently, the position of 463VA4 has:
   decreased - remained stable - increased
8. Recently, the turbine bearing vibrations 403KA7xx has:
   decreased - remained stable - increased
9. Recently, the turbine rotation speed 452KA812 has:
   decreased - remained stable - increased
10. Recently, the temperature in the combined tube before HPPH 463KB501 has:
   decreased - remained stable - increased
11. Missing from web-questionnaire
12. Missing from web-questionnaire

Local items: 1, 2, 3, 4, 7 (one item missing)
Global items: 5, 6, 8, 9, 10 (one item missing)

Beyond-2

Detection phase

1. Recently, the flow from the condenser has:
   decreased - remained stable - increased
2. Recently, the flow to the feedwater tank has:
   decreased - remained stable - increased
3. Recently, the flow to offgas system 341, 341KB301 has:
   decreased - remained stable - increased
4. Recently, the position of 462VA5 has:
   decreased - remained stable - increased
5. Recently, the turbine rotation speed 452KA812 has:
   decreased - remained stable - increased
6. Recently, the number of 462Px1 condensate pumps in operation has:
   decreased - remained stable - increased
7. Recently, the number of closed 421 valves has:
   decreased - remained stable - increased
8. Recently, the temperature after the condensate pumps 462KA503 has:
   decreased - remained stable - increased
9. Recently, the condensate flow to feedwater tank 462KB301 has:
   decreased - remained stable - increased

Local items: 1, 2, 4 (it was impossible to find more than 3 local items)
Global items: 3, 5, 6, 7, 8, 9

Mitigation phase

1. Recently, the temperature after the condensate pumps 462KA503 has:
   decreased - remained stable - increased

2. Recently, the level in the condenser has:
   decreased - remained stable - increased
3. Recently, the number of 462Px1 condensate pumps in operation has:
   decreased - remained stable - increased
4. Recently, the number of 441 main coolant pumps in operation has:
   decreased - remained stable - increased
5. Recently, the pressure in the condenser has:
   decreased - remained stable - increased
6. Recently, the flow to offgas system 341, 341KB301 has:
   decreased - remained stable - increased
7. Recently, the position of 462VA5 has:
   decreased - remained stable - increased
8. Recently, the condensate flow 462KB301 has:
   decreased - remained stable - increased
9. Recently, the temperature in the combined tube before HPPH 463KB501 has:
   decreased - remained stable - increased
10. Recently, the electrical output power 613KA901 has:
    decreased - remained stable - increased
11. Recently, the level in the feedwater tank has:
    decreased - remained stable - increased
12. Recently, the feedwater temperature after HPPH 463KB507 has:
    decreased - remained stable - increased

Local items: 2, 5, 7, 8, 10, 11
Global items: 1, 3, 4, 6, 9, 12

Beyond-3

Detection phase

1. Recently, the turbine oil temperature after bearing 403KA5xx has:
   decreased - remained stable - increased
2. Recently, the temperature after condensate pump 462KA503 has:
   decreased - remained stable - increased
3. Recently, the temperature after the condenser pumps has:
   decreased - remained stable - increased
4. Recently, the temperature 462KA503, after the condenser pumps has:
   decreased - remained stable - increased
5. Recently, the level in condenser 461KA40x has:
   decreased - remained stable - increased
6. Recently, the pressure 341KB101 before TB1 has:
   decreased - remained stable - increased
7. Recently, the number of 441 pumps in operation has:
   decreased - remained stable - increased
8. Recently, the electrical output power 613KA901 has:
   decreased - remained stable - increased
9. Recently, the temperature in the combined tube before HPPH 463KB501 has:
   decreased - remained stable - increased
10. Recently, the pressure in the condenser has:
    decreased - remained stable - increased
11. Recently, the flow 341KB301 has:
    decreased - remained stable - increased
12. Recently, the number of 463PX1 feedwater pumps in operation has:
    decreased - remained stable - increased

Local items: 4, 6, 7, 8, 10, 11
Global items: 1, 2, 3, 5, 9, 12

Mitigation phase

1. Recently, the temperature after the condensate pumps 462KA503 has:
   decreased - remained stable - increased
2. Recently, the temperature after the condensate pumps has:
   decreased - remained stable - increased
3. Recently, the turbine rotation speed 452KA812 has:
   decreased - remained stable - increased
4. Recently, the level in the condenser has:
   decreased - remained stable - increased
5. Recently, the temperature after the 724 pumps 724KB505 has:
   decreased - remained stable - increased
6. Recently, the pressure in the condenser has:
   decreased - remained stable - increased
7. Recently, the number of 441 main coolant pumps in operation has:
   decreased - remained stable - increased
8. Recently, the flow to offgas system 341, 341KB301 has:
   decreased - remained stable - increased
9. Recently, the electrical output power 613KA901 has:
   decreased - remained stable - increased
10. Recently, the level in HPPH 463EB1 463KA402 has:
    decreased - remained stable - increased
11. Recently, the condensate flow 462KB301 has:
    decreased - remained stable - increased
12. Recently, the turbine bearing vibrations 403KA7xx has:
    decreased - remained stable - increased

Local items: 4, 5, 6, 9, 10, 11
Global items: 1, 2, 3, 7, 8, 12

13. APPENDIX 4: THEORETICAL SHORTCOMINGS OF SACRI

Introduction

Situation Awareness Control Room Inventory (SACRI) was developed as part of the Halden Reactor Project (HRP) to assess Situation Awareness (SA) of operators in reactor control rooms (Hogg, Folessø, Torralba, and Volden, 1994). SACRI was unique at its time[6] because it capitalized on the use of the probe technique (e.g. SAGAT; 1995b) and Signal Detection Theory (SDT) to measure SA. However, SACRI has not matured as there is an absence of further review and/or application beyond the original studies for its initial development.

Theoretical Shortcomings

The review of the SACRI methodology indicates that the original formulation of the measure (i.e., the sensitivity and bias scores) is in disagreement with SDT. First, the probes are formulated as a three alternative forced choice (3AFC; i.e. increase, remain the same, or decrease) task as opposed to a yes/no task. The non-parametric formulas to calculate sensitivity and bias do not account for three-alternative responses. The original SACRI formulation only classified a response as:

> "… a hit [significant parameter shift, detected by subject], miss [significant parameter drift, not detected by subject], correct acceptance [no significant parameter drift, none reported by subject], or false alarms [no significant parameter drift, but one is nevertheless reported as being present]." (Ibid, pp. 7)

It does not include the classification of responses that are opposite in the direction of the actual parameter drift. That is, the operator reports an increase of a parameter which is, in fact, decreasing, or vice versa. Is this a false alarm or a miss? In fact, it is a "miss" in detecting a parameter shift in one direction but a "false alarm" in detecting a parameter shift in opposite direction. This problem might sound trivial in practice because operators, being experts in the field, would not commit such mistakes. Unfortunately, this is not true as indicated by the data collected in this study (see below). In any case, the manipulation in converting 3AFC into yes/no response is never made clear in any SACRI documentation.

Second, in relation to response classification, performance measure for any mAFC task is typically proportion correct, even for 2AFC task that can be converted into yes/no response (see e.g., McNicol, 2005, chap. 2; Stanislaw and Todorov, 1999). Note that d' can be calculated or referenced to d' table if both the noise and signal distributions are Gaussian. However, bias cannot be calculated based on SDT for mAFC tasks, which only use the performance score of proportion correct. Bias is only meaningful for two or more scores (e.g., hit rate and false alarm rate). Given that the probes are in the form of 3AFC tasks, the validity of the bias index is suspicious.

Third, proofs for the proposed non-parametric formulae to calculate sensitivity and bias are absent. The authors have yet to find the equivalent SACRI formulae in the literature (see e.g. Stanislaw and Todorov, 1999; McNicol, 2005, chap. 2; Zhang and Mueller, 2005). The derivation of the non-parametric indices – A' and R:S ratio – are unclear, thus, questionable.

Fourth, even if the non-parametric formulae are correct, the sensitivity formula sometimes yields undefined results, which occur when hit rate and/or correct acceptance rate is 0. Although it is not unusual for sensitivity data to be undefined according to the literature of SDT, the commonly recommended procedure may not be appropriate as the mathematical formula (in the literature) are substantially different. Without any theoretical discussion on the development of the formula for sensitivity, it is difficult to determine the appropriate transformation for the extreme data points.

Fifth, the interpretation of a negative A' of SACRI is ambiguous. Based on all the SDT literature reviewed, negative A' is classified as impossible because A' mathematically denotes the average area of under the ROC curve. Negative areas are not defined for graphs. However, the Hogg et al. (1994) of the SACRI dismissed such concern stating:

[6] McGuinness (2004) developed Quantitative Analysis Situation Awareness (QUASA) for the command and control domain that applies SDT.

> "In fact, it is mathematically possible to derive a negative score: however, due to psychological factors involved in the detection of parameter deviation (on which the underlying principles of Signal Detection Theory are based), an expert operator who is genuinely responding to the best of his/her ability is unlikely to score below the 0.5 mark." (Hogg et al., 1994, pp. 66)

In SDT, it is only mathematically possible to obtain a score below 0.5 but not below zero. The explanation, or dismissal, reflects poor theoretical foundation of SACRI.

Sixth, the interpretation of A' below 0.5 is also ambiguous. Hogg et al. (1994) explained that "a performance score below 0.5 meant the operator performed no better than if he/she had been responding according to chance" (pp. 66). Although this is consistent to the literature of SDT in general, the interpretation of the data becomes an issue for highly representative experiments. A' below 0.5 indicates a *systematic* error either in signal interpretation or probe response. The latter is rare because the queries (e.g. recently, what happened to the level of tank A at the end of the scenario?) and response alternatives (i.e. increase, remain the same, or decrease) should not be difficult to understand and thus should not lead to any systematic effect. On the other hand, if the operator is confusing signal with non-signal/noise, A' below 0.5 cannot be interpreted because the non-signal is undefined. Defining non-signal is impossible due to combinatorial explosion given the nature of complex environment. For highly controlled laboratory settings in which both signal and non-signal can be defined, sensitivity scores below chance level can be interpreted. The concern of A' below 0.5 is dismissed by Hogg et al. (1994) as presented earlier. However, the data collected for this study suggest A' below 0.5 is common while the interpretation or treatment of A' below 0.5 is unclear.

In summary, we have identified six alarming issues of SACRI:

(1) Incomplete classification of responses;
(2) Inadequate explanation for the unconventional treatment of sensitivity and bias scores for a 3AFC task;
(3) Missing proofs for formula of non-parametric measures;
(4) Missing discussion on undefined values of the sensitivity measure for cases when hit rate and/or correct acceptance rate is 0;
(5) Missing interpretation and mathematical foundation for negative A';
(6) Ambiguous interpretation for A' below chance level (i.e. 0.5).

These shortcomings hamper the validity of SACRI and the prevalence for some these concerns are discussed in the next section.

Effects of the Theoretical Shortcomings

Hogg et al. (1994) have dismissed some of the concerns associated with the identified shortcomings of SACRI on a practical basis. However, the data collected for the SACRI measure in this study shows these shortcomings is not merely theoretical. Operators responded in the opposite direction compared to the actual parameter shifts for 42 out of 816 the probes in this study. In other words, 5% of the responses cannot be classified based on the original SACRI methodology. There are 13 out of 72 data points with 0 hit rates, resulting in 18% of undefined values. In addition to undefined cases, twelve (out of 72) A's are below chance level (i.e. 0.5), three of which are negative. The correlation between the sensitivity scores and proportion correct is 0.88 for deleting 13 cases of undefined values and 0.89 for deleting 16 cases of undefined and negative values.

The data collected for the SACRI measure in this study suggests that some of the identified shortcomings deserve attention rather than dismissal. Although the statistics are based on a single study, the presence of undefined or ill-defined treatment of data points is a strong indicator of deficiency in a measure.

Evaluation of SACRI

The analysis reveals that SACRI violates the theoretical foundation of SDT, and the data collected from one study suggests that the violations should not be dismissed on a practical basis. Therefore, the application of SACRI in its original form is **not** recommended. Nevertheless, there are merits in the concept of assessing the sensitivity of operators towards parameter shifts independent of the bias reacting towards those changes. It is economical to obtain the sensitivity as an index of situation understanding and bias as an index of "strategic choices" within a single measure. As such, authors feel that investment on refining SACRI to better couple with SDT should be consider, if not fruitful.

Potential Remedy to the Shortcomings

SACRI can be modified to better couple with SDT. One of the methods is simplifying response alternatives into confidence yes/no ratings similar to QUASA (McGuiness, 2004). The resulting data should generate an ROC curve with three or more data points that more accurate estimates of sensitivity of the operator compared to single point estimation. The non-parametric formula for both sensitivity and bias are readily available in the literature, which also thoroughly discussed the treatment of undefined values (see, e.g., Stanislaw and Todorov, 1999). However, the simplification reduces diagnostic power because the response alternative does not account for directional errors of the parameter shifts (e.g. operator reports an increase of a parameter when it is a decrease). If the original format of the probes – 3 AFC – must be retained for high diagnostic power, SACRI should adopt choice theory (Luce, 1959, 1963; also see McNicol, 2005, chap. 6) as opposed SDT. Choice theory outlines the data collection and analysis methodology to determine response strengths and biases for all alternatives of mAFC tasks. Although it seems to be a viable alterative to SDT, literature on choice theory is limited. The formulation of these remedies is beyond the scope of this report.

14. APPENDIX 5: FOUNDATION OF THE PROCESS OVERVIEW MEASURE

Monitoring and Process overview

Monitoring is generally referred to as a collection of activities performed by operators to gain an overall but imprecise knowledge of the plant operating conditions. This imprecise, overview knowledge of the operating plant conditions is formalized as Process overview. From a practical perspective, Process overview is a view highlighting the plant processes that are deviating from ideal conditions while illustrating others operating under normal conditions in the background. In other words, Process overview enables operators to identify abnormal processes that require further attention or even intervention. From a psychological perspective, Process overview is a high level, or Gestalt, perception of plant processes. Operators perceive NPP processes as a set/whole rather than as individuals/parts, naturally constructing figures (i.e., abnormal processes) and grounds (i.e., normal processes) in their percepts.

The next two sections present the foundation of the Process overview. The first section describes the nature of the nuclear domain and operator work (i.e. monitoring), forming the practical basis for Process overview. The second section connects the practical basis to established psychological theories and findings. Following these sections is a discussion of Process overview in relation to SA literature to date.

The Practical Basis - Nature of Monitoring

Work domain characteristics and operator activities specific to nuclear power plants led to the formulation of Process overview. While most readers may be familiar with the nuclear domain, the key domain characteristics and monitoring activities are described to illustrate the unique contribution of Process overview as a SA concept.

NPPs are large in scale and high in complexity as each plant contains thousands of components and instruments interconnected with various automation agents. At such scale and complexity, even highly reliable equipment can not eliminate component failures on a regular basis. For many of these failures, NPPs could operate safely and efficiently because of redundancy. However, this does not imply that *all* failures would not lead to safety and efficiency implications or even disasters.

Automation has a significant and active role in NPPs. There are many automation agents, such as control loops, installed throughout the plants and they act on the system continuously to maintain or optimize steady state operations. Therefore, plant operation must be assessed according to both values of various parameters and actions of the automation agents. These factors contribute to increased complexity.

NPPs are continuous process control systems, which have unique operating characteristics. The readings of instruments have significantly different meanings in different operating states (e.g., start up, full power, shut down). For instance, normal values for the pressure inside a condenser for start-up may be equivalent to emergency values during full power operating states (see Mumaw et al., 2000; Guerlain and Bullemer, 1996). Control responses are typically slow in comparison to other domains[7] (e.g., aviation). All processes and instruments fluctuate to some degree, generating nuisance (i.e., non-meaningful) variations (see e.g., St-Cyr and Vicente, 2005;

[7] Process control systems usually have slow response time because mass transfer, energy exchange, and chemical reactions are unstable and inefficient in fast conditions.

Reising and Sanderson, 2004). Given that NPPs are intended for continuous operations, many testing and maintenance activities are performed while plants are in full operation. In some cases, operation continues if component failures do not pose safety risks because some repairs only occur during shut down. Furthermore, the operating states (e.g., power level as a result of demand) and other activities (e.g., maintenance) can vary at a regular and irregular basis. In brief, instrument readings and automation actions must be interpreted with respect to frequently changing contexts of plant operations.

Operators engage in monitoring activities that enable them to manage the scale, complexity, and dynamic contexts of NPPs so that they can acquire the necessary situation knowledge to control the systems. Field studies indicate that operators rely on a diverse set of information sources to monitor NPPs (Mumaw et al., 2000). Operators begin their shifts by conducting a shift turnover, a briefing on plant operating conditions between the operators coming on duty and those being relieved. The shift turnover communicates the plant status, activities completed, activities that need to be completed, and other special circumstances. In addition to shift turnovers, logs - chronological records of significant activities (e.g., tests completed, component failures) - are reviewed to gain awareness of recent plant status. Operators also conduct panel walkthroughs to review indicators systematically and field tours to gain a "process feeling". Control room panels and computer displays present the instantaneous values of instruments and control actions of automation; field operators provide other process or equipment information which cannot be accessed in the control rooms.

The domain characteristics and operators activities illustrate that a fundamental challenge faced by the NPP operators is "how to identify and pursue relevant findings against a cognitively noisy background" (Mumaw et al., 2000; also see Woods, 1994). The scale and complexity of NPPs simply prohibit continuous, comprehensive, and reliable sampling of all indicators; thus, operators are always deciding which process areas should be prioritized for close observation. Plant components and instruments are inherently noisy; hence, operators are always deciding what process changes are necessary for further investigation. Operators constantly make these decisions to allocate their cognitive resources to different parts of the plants to manage the complex work environment. To make these decisions, operators rely on an accurate operating context (see Edwards and Lees, 1974; Patterson and Woods, 1997; Mumaw et al., 2000). The operating context is required to interpret values of indicators and actions of automation because normative values or actions satisfying all circumstances rarely exist. NPP operating contexts are also dynamic, further complicating the interpretation of instrument readings and automation actions (c.f., Vicente, 1999). For instance, most, if not all, alarms must be interpreted with respect to the changing context. Some alarms are caused by maintenance activities that vary over time while others are ignored because they are overly sensitive to process fluctuations (i.e., false alarms). Important alarms are those unexpected in a given context, prompting for further investigation (see Mumaw et al., 2000; Xiao and Seagull, 1999; c.f., Seagull et al, 2000). The importance of context is confirmed by the myriad "set up activities" engaged by the operators that enable them to detect, recognize and pursue anomalies (Mumaw et al., 2000). Therefore, operators constantly need to build or update their knowledge of the operating context in parallel with monitoring the instrument readings and automation actions.

The reality that operators need to decide on what processes and changes are important in a dynamic operating context has serious implication in formulating a meaningful and relevant SA component for the nuclear domain. First, for NPP operators, knowing precise indicator values and automation actions is not nearly as important as knowing to which indicators and automation it is necessary to attend (c.f., Level 1 SA proposed by Endsley, 1995). This is because human cognitive resources for attending to such details are only available for parts of, but not the whole NPP because of its scale and complexity. Second, knowing how much indicator values and automation actions have changed on an absolute scale is not nearly as important as knowing which indicator values and automation actions have changed in an abnormal fashion. This is because attending to all changes is not meaningful due to inherent fluctuations in plant processes and components. Furthermore, control engineering principles contend that reacting to all process changes would destabilize the plant processes. It should be emphasized that identifying parameters, automation actions and their changes relevant for attentive observation is highly dependent on the operating context; whereas, knowing the exact values and absolute changes is not dependent on any. To summarize, the first type of situation knowledge acquired by NPP operators is which plant processes are in disturbance or abnormal states in the midst of all other normally fluctuating operations.

Process overview is constructed to reflect the necessary situation knowledge gained through monitoring that allows the operator to function in the complex work environment of the NPP. In practical terms, Process overview is a view of the plant with processes deviating from ideal conditions in focus and others operating under normal conditions in the background. Operators build an operating context to acquire the needed expectation for recognizing the relevance of different indicators, thereby supporting the identification of disturbance areas. The processes that contradict expectation (i.e., the context) are at the focus of attention while those corresponding to expectation are the periphery. Operators functioning under a faulty context would confuse process areas or changes that are normal with those that are abnormal. Process overview reflects the importance of accurate operating context and appropriate strategic distribution of attention that is crucial to address the limitations of cognitive resources in monitoring a complex domain. It is the product of a series of monitoring activities - building and updating context, observing process parameters, deciphering relevance, and ultimately perceiving 'true' process deviations.

Thus far, we have illustrated how domain characteristics and nature of monitoring specify the knowledge requirement at the early stage of Situation Awareness acquisition for NPP operators from a *practical* perspective. The next section presents the psychological foundation of Process overview in connection with the domain specific characteristics and operator work activities.

The Psychological Foundation – Nature of Problem Representation

The psychological foundation of Process overview is shaped by its practical basis – the domain characteristics and monitoring activities. Monitoring involves constantly building and updating operating contexts to identify and pursue anomalies in a noisy and complex environment as opposed to constantly registering parameter values or automation actions to detect changes in a quiescent setting (see above). This illustrates that psychological constructs meaningful to both monitoring and resulting knowledge entail more than just vigilance, memory, and visual perception. According to Mumaw et al. (2000), the nature of monitoring NPPs is "better characterized as active problem solving rather than passive vigilance" (pp. 43).

Operating NPPs bear some key resemblances to solving insight problems. In both cases, the problems are ill defined. In other words, the problems and corresponding solution paths are not evident to the solver. As presented earlier, the challenge for operators is identifying anomalies in parallel with acquiring an operating context, which is highly dynamic. Operating context is an ill defined and holistic concept; thus, anomalies (i.e., the problems) which can only be identified with a context are also ill defined (see Woods, 1994). This points to why automation today cannot fully assume the responsibility of the operators, who are much more adaptive to dynamic environments than computer programs. In addition, the search space for solving insight problems and resolving process anomalies is vast. Thus, exhaustive search for solutions to any problems is not possible or feasible. Operators cannot and do not sample indicators and automation actions in a comprehensive manner to identify problems or consider all possible interventions for each identified anomaly. Rather, they constantly prioritize what parameters to sample and decide on what changes are relevant for further investigation.

Taking the perspective that operating NPPs is analogical to solving insight problems, the psychological foundation for situation assessment and knowledge could be built on cognition studies investigating insight problems. Cognition research indicates that solving insight problems requires effective problem representation which, in turn, reduces the search space or reveals the solution paths (see below). Effective problem representation relies on acquiring “the insight”, or perceiving meaningful constraints or affordances (see below). This finding in cognition research is connected to the psychological aspects of the Situation Awareness construct for the nuclear domain: The first set of situation assessment activities engaged in by operators is similar to representing an insight problem and the first type of situation knowledge acquired by operators resembles a representation of an insight problem, or a reduced search space.

Process overview is constructed to reflect the psychological nature of problem representation in the nuclear domain. It is a high level, or Gestalt, perception of plant processes. Operators perceive constraints and affordances of the systems by building an operating context of the NPP, thereby supporting the process of representing the plant in a meaningful way. This meaningful representation permits a holistic perception of the plant to emerge from observing the indicator values and automation actions, as those in violation of constraints and affordances become the figure while others in conformance become the ground. The concept of figure-ground in Process overview is crucial to address the limitation of cognitive resources. Operators attend to the figure of the percept, which is, in effect, the reduced search space. As mentioned, operators can only afford to devote their cognitive resources to a subset of the system that has a high potential for disturbance. Operating context and observations of parameter values and automation actions can significantly affect problem representation or formulation of figure and ground; thus their accuracy is vital to situation knowledge. For instance, a wrong determination of context could lead to confusion between relevant and irrelevant parameters resulting in incorrect identification of disturbances. To be cognitively efficient, operators must also be judicious of variations to prevent normal fluctuations of the plant processes and instrumentations (generated by noise) in becoming the figure, and thereby, consuming their attention needlessly. In essence, Process overview in monitoring NPPs is the psychological counterpart of problem representation in solving insight problems.

Process overview is classified as a type of perception because problem representation in insight-problem solving is heavily dependent on perceptual factors. Gestalt psychologists argue that effective problem representation is determined by perceptual factors (c.f., Kohler, 1971; Gibson, 1979; Norman, 1988; Kershaw and Ohlsson, 2004). According to the Gestalt tradition, perception is holistic and autonomous (Logan and Zbrodoff, 1999) rather than logical/rational processing of environmental stimuli as in the information processing tradition (see e.g., Wickens & Hollands, 1999). The interpretation of affordances and constraints is often autonomously attached to perception rather than abstract thinking or cognition (e.g., logic or rationale). When affordances and constraints are perceived in the matching contexts of insight problems, the solution paths are obvious as to what should and should not be done. On the other hand, misinterpretation of affordances and constraints leads problem solvers away from the solution paths within the vast search space. In other words, perceptual factors that influence the interpretation of affordances and constraints can significantly alter problem representations, and thereby, affect the success in solving insight problems. In a similar light, Process overview is a holistic and spontaneous plant representation that is highly influenced by the natural interpretation of affordances and constraints. NPP operators build and update context in a spontaneous manner rather than in a detailed, comprehensive fashion because of the demands in the naturalistic monitoring environment. The knowledge of context, though imprecise, provides the basis of affordances and constraints that facilitates the holistic and spontaneous perception of the NPP processes as operators observe process parameters. This perception, Process overview, provides the contrast between parameters in violation of *perceived* constraints (i.e., the figure of perception composed of elements out of expectation) and those being afforded (i.e., the ground of perception composed of elements within expectation) for the situation at hand. The processes and parameters that are in violation of constraints (or cannot be afforded by the current system state) become the reduced search space – the processes identified to be in disturbance that operators must devote resources to diagnose and intervene. The natural interpretation of constraints and affordance is as relevant to Process overview in monitoring as it is to perception in insight-problem solving.

Process overview is not purely influenced by perceptual factors like the perception in insight-problem solving. After all, there are differences between monitoring complex systems and solving insight problems. Process overview is affected by the process factor of physically sensing various indicators of the NPP. Note that the process is *not* the integration of sensed data to generate perception, which is holistic and spontaneous. Unlike monitoring NPPs, sensation is not a significant process factor for solving insight problems, which only contain a "handful of information" (see e.g., Kaplan and Simon, 1990). Process overview is also dependent on knowledge factors (i.e., prior experience/knowledge). Prior experience or knowledge influences interpretation of constraints and affordances (i.e., the context) in an autonomous way, and thus, perception of plant processes. The fact that pattern recognition (c.f., Chase and Simon, 1973; Chabris and Hearst, 2003) and problem solving (c.f., Weisberg and Alba, 1981a, 1981b, 1982) are more automatic with expertise is established in the literature. In most cases, the automaticity in accessing prior knowledge or memory proves highly efficient and effective; however, it could also lead to, and reinforce, a completely wrong perception when mismatches exist between the contexts of the memory and the current situation (e.g., the Stroop effect). Knowledge factors affect problem representation or perception in nuclear power plant monitoring as much as in solving insights problems. Operators capitalize on the automaticity of accessing domain specific memory or knowledge to drive the spontaneous interpretation context and perception of plant processes. Consequently, the accurate identification of significant anomalies through Process overviews is also dependent on knowledge factors.

The psychological foundation of Process overview illustrates some of the parallels between operating NPPs and solving insight problems. Process overview is the counterpart of problem representation that allows operators to manage the ill-defined and vast search space in operating NPPs. This psychological perspective corresponds to the practical perspective that operators do indeed represent plant processes to appropriately allocate their cognitive resources by building/updating contexts in addition to observing parameter values and automation actions. Given that both practical and psychological perspectives of Process overview have been presented, the next section discusses the unique suitability the concept in comparison with other prominent conceptualizations of the SA construct.

Unique Contribution of Process overview

Process overview is derived from the nuclear domain to reflect the characteristics of situation knowledge developed through monitoring. It is a domain specific SA component or concept. Earlier, we argued that domain specific SA models or their components reflect unique aspects of human cognition, domain specific complexities, and their interactions. This section compares the Process overview concept to the SA components in Endsley's model (1995a), illustrates the unique contribution of Process overview to the nuclear domain, and the benefits of developing domain specific SA models.

The SA components in the model proposed by Endsley (1995a) are chosen as the basis of comparison for three reasons. First, the SA model proposed by Endsley (1995a) is established with many empirical studies. Second, Endsley's model emphasizes Situation Awareness as knowledge states that are of interest to formulating measurements[8]. Finally, Endsley's model was derived from the aviation domain; thus, the comparison between the SA components permits inferences on the benefits of domain specific SA models.

Endsley proposed a model containing three levels of SA: Level 1, perception of elements in the environment; Level 2, comprehension of those elements; and Level 3, projection of those elements in the future[9] (1988, 1995a, 2000). The three levels of SA exhibit both similarities and differences to Process overview from a practical and a psychological perspective.

In practice and on a surface level, Process overview appears to include all three levels of SA. In order to acquire a Process overview operators must "perceive" or register indicator values and automation actions in some way as in acquiring the Level 1 SA, perception of the elements in the environment. Since it is a high level view of the plant processes and highlights the abnormal ones, Process overview also includes some form of Level 2 SA, comprehension of the perceived elements. Similarly but to a lesser degree, Process overview includes some primitive form of Level 3 SA, projection of future status for the perceived elements, because operators are anticipating problems from the highlighted processes. Process overview may also contain more precise projection knowledge when Scenario understanding generated from diagnosing faults begins to modify monitoring behaviours and thus Process overview (see section 3.5 on Scenario understanding; c.f., Adams et al., 1995).

Process overview may appear to be an integral of three SA levels. However, alone or in combination, the three levels of SA do not provide the unique perspective of Process overview while Process overview cannot fully represent all three levels of SA. In the nuclear domain, rigid adoption of Level 1 SA, as in registering indicator values by operators, typically does not indicate much of situation knowledge since parameter values are only meaningful in context. As mentioned, registering all indicator values are cognitively unfeasible, and perhaps, unnecessary because operators typically have sufficient time to access indicators of interest before responses are required (see above on the characteristics of NPPs). In fact, attending to and remembering indicator values within normal operating range may suggest inefficient monitoring strategies. In other words, low-level perception without notion of relevance is not meaningful in NPP settings. On the other hand, the amalgamation of Level 1 and 2 SA encompasses cognitive aspects that are

[8] Although a discussion on the process of representing insight problems is presented to support the concept of Process Overview, it is primarily intended to illustrate the characteristics of the specific type of situation knowledge for the nuclear domain.

[9] To the best knowledge of the authors, Endsley has never connected the three levels of SA to any major findings or established theories in psychology. Her description/definition of each level of SA is largely based on common understanding of perception, comprehension, and projection, which are not precisely defined. Thus, it is difficult to relate the three level of SA to other psychological theories in a precise manner.

beyond the scope of Process overview. Process overview is an imprecise view or representation of the plant processes that does not entail a full understanding of why and/or how changes or problems occur. At the same time, a combined Level 1 and 2 SA would bury the distinction between Process overview – perception of significant anomalies – and Scenario understanding– understanding of the scenario disturbances (see section 3.3). While Process overview permits operators to anticipate the process areas that would deviate further from normal operations in the future, it does not include precise projections of process parameters, Level 3 SA unless Scenario understanding modify monitoring behaviours and thus Process overview. In the same lights as Endsley conceptualization Level 3 SA, precise prediction requires detailed reasoning on the disturbances and automation behaviours (i.e. Scenario understanding).

On a psychological level, Process overview and the three levels of SA have similarities and differences as in practice. Process overview is chiefly a perception developed through monitoring that is moderately affected by process and knowledge factors. Ensley's theoretical account of SA captures some aspects of process and knowledge factors affecting Process overview. Endsley's conceptualized SA based on the information processing model that takes process factors well into account. Both Process overview and Level 1 SA are affected by the process of sensing data/information in the environment. The information processing model also provides a good account for knowledge factors. In Endsley's model, the knowledge factors are long-term memory and automaticity that influence all three levels of SA. Long-term memory and automaticity generate expectation, which, in turn, controls operator attention. If the memory accessed is inappropriate[10], the expectation would be incorrect, and thus, attention would be directed at the wrong elements. Process overview is similar in that some expectations are generated in an automatic fashion as a result of prior experience/knowledge.

Endsley's conceptualization of SA, however, does not account for all aspects or factors of Process overview. Endsley has viewed situation assessment as collecting information rather than problem solving; thus, her model does not examine the influences of knowledge factors on problem representation. Specifically, the influence of knowledge factors in affecting the interpretation of affordances and constraints is not reflected in any levels of SA. Furthermore, perceptual factors that are central to Gestalt psychology are almost entirely neglected. None of the SA levels reflects the importance of perceptual factors that are pertinent to problem representation. This marks the most significant difference between Process overview and different levels of SA. These factors are absent from Ensley's account of SA probably because they do not conform to the information processing tradition, which is disconnected from the concept of problem representation.

The differences between Process overview and the three levels of SA are reflections of their domain origins. Endsley account of SA is rooted in aviation (c.f., Pew, 1995; Endsley, 2004). In aviation, the monitoring activities are typically well defined attenuating the significance of problem representation[11]. In other words, what signals are relevant to monitor and what values are considered critical are largely predetermined, permitting a single representation of the domain to be effective in most circumstances. For these domain characteristics, Level 1 SA is meaningful because sensation/registration and perception of indicators are very closely related for static domain/problem representation. For instance, there is practically no difference between sensation and perception of aircrafts on radar. Pilots and air traffic controllers would find every aircraft on radar and distances between them to be always relevant. In most circumstances, pilots

[10] Memory accessed may be inappropriate in two ways: (1) erroneous information is encoded into memory (e.g. the sun is cold); or (2) the memory retrieved is not applicable to the situation (e.g., wrong schema).

[11] Aviation tasks are *not all* well defined. For instance, air traffic controllers have many choices in routing aircrafts in their sectors. Only monitoring tasks in aviation are relatively well defined.

and air traffic controllers represent their domain/problem in the space-time dimension because almost all incidences are the result of loss of separation (between an aircraft and other elements).

The nuclear domain does not permit a static representation of the plant processes or problem because the significance of each indicator varies with the dynamics of the context. In other word, what signals are relevant to monitor and what values are considered critical cannot be predetermined. As a result, sensation and perception of indicators in the nuclear domain are not as closely related as in the aviation domain. Reapplying the same rationale, perception as defined by Level 1 SA (i.e., sensation and registration of indictor values) is more closely related to comprehension (e.g., distance to combat engagement) and projection (e.g., projected time of assaults by enemies) in the aviation domain than in the nuclear domain. The strict adoption of Level 1 SA in the nuclear domain would not provide the meaningful connections to comprehension and projection of elements in the environment as in the aviation domain. The relationships amongst sensation, perception, and comprehension most likely vary across domains.

Based on the field characteristics of monitoring NPP, Process overview exhibits unique properties compared to the three levels of SA proposed by Endsley. Though Endsley's work does not represent all research on SA models, the comparison illustrated that differences between domains in practical terms (e.g. domain characteristics or work activities) do affect psychological conceptualization of SA components. Therefore, field characteristics should not be thoughtlessly dismissed even in discussions at the theoretical/conceptual level[12]. It also demonstrates that meaningful and relevant descriptions of situation knowledge are best developed from a bottom-up approach, based on the domain characteristics and operator activities.

Process overview has a unique foundation, capturing the conceptual and practical characteristics specific to monitoring NPP. For scientific validation and communication (c.f., Underwood, 1959), Process overview requires operationalization, which facilitates the development of performance measures for monitoring.

[12] From the perspective of the authors, SA cannot be developed at a theoretical level because it is a conceptual simplification of theories on awareness/consciousness in psychology. However, the simplification is useful from human factors evaluation perspective.

Cognitive Engineering Laboratory

CEL TECHNICAL REPORT SERIES

CEL 93-01 "Egg-sucking, Mousetraps, and the Tower of Babel: Making Human Factors Guidance More Accessible to Designers"
• Kim J. Vicente, Catherine M. Burns, & William S. Pawlak

CEL 93-02 "Effects of Expertise on Reasoning Trajectories in an Abstraction Hierarchy: Fault Diagnosis in a Process Control System"
• Klaus Christoffersen, Alex Pereklita, & Kim J. Vicente

CEL 94-01 "Cognitive 'Dipsticks': Knowledge Elicitation Techniques for Cognitive Engineering Research"
• Klaus Christoffersen, Christopher N. Hunter, & Kim J. Vicente

CEL 94-02 "Muddling Through Wicked Problems: Exploring the Role of Human Factors Information in Design"
• Catherine M. Burns

CEL 94-03 "Cognitive Work Analysis for the DURESS II System"
• Kim J. Vicente & William S. Pawlak

CEL 94-04 "Inducing Effective Control Strategies Through Ecological Interface Design"
• William S. Pawlak

CEL 94-05 "Research on Factors Influencing Human Cognitive Behaviour (I)"
• Klaus Christoffersen, Christopher N. Hunter, & Kim J. Vicente

CEL 94-06 "Ecological Interfaces for Complex Industrial Plants"
• Nick Dinadis & Kim J. Vicente

CEL 94-07 "Evaluation of a Display Design Space: Transparent Layered User Interfaces"
• Beverly L. Harrison, Hiroshi Ishii, Kim J. Vicente, & Bill Buxton

CEL 94-08 "Designing and Evaluating Semi-Transparent 'Silk' User Interface Objects: Supporting Focused and Divided Attention"
• Beverly L. Harrison, Shumin Zhai, Kim J. Vicente, & Bill Buxton

CEL 95-01 "An Ecological Theory of Expertise Effects in Memory Recall"
• Kim J. Vicente & JoAnne H. Wang

CEL-SP "Strategic Plan"
• Cognitive Engineering Laboratory

CEL-LP "Cognitive Engineering Laboratory Profile"
• Cognitive Engineering Laboratory

CEL 95-04 "A Field Study of Operator Cognitive Monitoring at Pickering Nuclear Generating Station-B"
• Kim J. Vicente & Catherine M. Burns

CEL 95-05 "An Empirical Investigation of the Effects of Training and Interface Design on the Control of Complex Systems"
• Christopher N. Hunter

CEL 95-06 "Applying Human Factors to the Design of Medical Equipment: Patient-Controlled Analgesia"
• Laura Lin, Racquel Isla, Karine Doniz, Heather Harkness, Kim J. Vicente, & D. John Doyle

CEL 95-07 "An Experimental Evaluation of Transparent Menu Usage"
• Beverly L. Harrison & Kim J. Vicente

CEL 95-08 "Research on Factors Influencing Human Cognitive Behaviour (II)"
• Christopher N. Hunter, Michael E. Janzen, & Kim J. Vicente

CEL 95-09 "To the Beat of a Different Drummer: The Role of Individual Differences in Ecological Interface Design"
• Dianne Howie

CEL 95-10 "Emergent Features and Temporal Information: Shall the Twain Ever Meet?"
• JoAnne H. Wang

CEL 95-11 "Physical and Functional Displays in Process Supervision and Control"
• Catherine M. Burns & Kim J. Vicente

CEL 96-01 "Shaping Expertise Through Ecological Interface Design: Strategies, Metacognition, and Individual Differences"
• Dianne E. Howie

CEL 96-02 "Skill, Participation, and Competence: Implications of Ecological Interface Design for Working Life"
• Peter Benda, Giuseppe Cioffi, & Kim J. Vicente

CEL 96-03 "Practical Problem Solving in a Design Microworld: An Exploratory Study'
• Klaus Christoffersen

CEL 96-04 "Review of Alarm Systems for Nuclear Power Plants"
• Kim J. Vicente

CEL 96-05 "DURESS II User's Manual: A Thermal-hydraulic Process Simulator for Research and Teaching"
• Lisa C. Orchanian, Thomas P. Smahel, Dianne E. Howie, & Kim J. Vicente

CEL 96-06 "Research on Factors Influencing Human Cognitive Behaviour (III)"
• Dianne E. Howie, Michael E. Janzen, & Kim J. Vicente

CEL 96-07 "Application of Ecological Interface Design to Aviation"
• Nick Dinadis & Kim J. Vicente

CEL 96-08 "Distributed Cognition Demands a Second Metaphor for Cognitive Science"
• Kim J. Vicente

CEL 96-09 "An Experimental Evaluation of Functional Displays in Process Supervision and Control"
• Catherine M. Burns and Kim J. Vicente

CEL 96-10 "The Design and Evaluation of Transparent User Interfaces: From Theory to Practice"
• Beverly L. Harrison

CEL 97-01 "Cognitive Functioning of Control Room Operators: Final Phase"
• Kim J. Vicente, Randall J. Mumaw, & Emilie M. Roth

CEL 97-02 "Applying Human Factors Engineering to Medical Device Design: An Empirical Evaluation of Two Patient-Controlled Analgesia Machine Interfaces"
• Laura Lin

CEL 97-03 "ADAPT User's Manual: A Data Analysis Tool
for Human Performance Evaluation in Dynamic Systems"
• Xinyao Yu, Farzad S. Khan, Elfreda Lau, Kim J. Vicente, & Michael W. Carter

CEL 97-04 "Research on the Characteristics of Long-Term Adaptation"
• Xinyao Yu, Renée Chow, Greg A. Jamieson,
Rasha Khayat, Elfreda Lau, Gerard Torenvliet,
Kim J. Vicente, & Michael W. Carter

CEL 97-05 "A Comprehensive Experimental Evaluation of Functional Displays in Process Supervision and Control"
• Catherine M. Burns and Kim J. Vicente

CEL 98-01 "Applying Human Factors Engineering to Medical Device Design: An Empirical Evaluation of Patient-Controlled Analgesia Machine Interfaces"
• Laura Lin

CEL 98-02 "Building an Ecological Foundation for Experimental Psychology: Beyond the Lens Model and Direct Perception"
• Kim J. Vicente

CEL 98-03 "Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-based Work"
• Kim J. Vicente

CEL 98-04 "Ecological Interface Design for Petrochemical Processing Applications"
• Greg. A. Jamieson & Kim J. Vicente

CEL 98-05 "The Effects of Spatial and Temporal Proximity of Means-end Information in Ecological Display Design for an Industrial Simulation"
• Catherine M. Burns

CEL 98-06 "Research on Characteristics of Long-term Adaptation (II)"
• Xinyao Yu, Gerard L. Torenvliet, & Kim J. Vicente

CEL 98-07 "Integrated Abstraction Hierarchy and Plan-Goal Graph Model for the DURESS II System: A Test Case for Unified System- and Task-based Modeling and Interface Design"
• Christopher A. Miller & Kim J. Vicente

CEL 98-08 "Comparative Analysis of Display Requirements Generated via Task-Based and Work Domain-based Analyses: A Test Case Using DURESS II"
• Christopher A. Miller & Kim J. Vicente

CEL 98-09 "Abstraction Decomposition Space Analysis for NOVA's E1 Acetylene Hydrogenation Reactor"
• Christopher A. Miller & Kim J. Vicente

CEL 99-01 "Development of an Analytical Framework and Measurement Tools to Assess Adaptive Performance of Individual and Teams in Military Work Domains"
• John R. Hajdukiewicz, Kim J. Vicente, & Robert G. Eggleston

CEL 99-02 "Applying Perceptual Control Theory and Ecological Interface Design to the Control Display Unit"
• Sandra Chéry

CEL 99-03 "Research on the Characteristics of Long-Term Adaptation (III)"
• Gerard L. Torenvliet & Kim J. Vicente

CEL 99-04 "Comparative Analysis of Display Requirements Generated via Task-Based and Work Domain-based Analyses in a Real World Domain: NOVA's Acetylene Hydrogenation Reactor"
• Christopher A. Miller & Kim J. Vicente

CEL 99-05 "A Cognitive Engineering Approach for Measuring Adaptive Behavior"
• John R. Hajdukiewicz & Kim J. Vicente

CEL 00-01 "Differences Between the Eye-fixation Patterns of Novice and Expert Operators of the DURESS II Physical Interface"
• Madhava Enros & Kim J. Vicente

CEL 00-02 “If Technology Doesn’t Work for People, then It Doesn’t Work”
• Kim J. Vicente

CEL 01-01 “A Field Study of Collaborative Work in Network Management: Implications for Interface Design and Evaluation”
• Renée Chow & Kim J. Vicente

CEL 01-02 “A Prototype Ecological Interface for a Simulated Petrochemical Process”
• Greg A. Jamieson & Wayne H. Ho

CEL 01-03 “EID Design Rationale Project: Case Study Report”
• Greg A. Jamieson, Dal Vernon C. Reising & John Hajdukiewicz

CEL 02-01 “Ecological Interface Design for Petrochemical Process Control: Integrating Task-and System-Based Approaches”
• Greg A. Jamieson

CEL 06-01 “Canada Foundation for Innovation (CFI) Emerson DeltaV / MiMiC Industrial Process Control Simulator”
• Antony Hilliard & Laura Thompson

CEL 06-02 “Developing Human-Machine Interfaces to Support Monitoring of UAV Automation”
• Greg A. Jamieson, Lu Wang, Jamy Li

CEL 07-01 "The 2005 Ecological Interface Design Process and the Resulting Displays (HWR-847)"
• Robin Welch, Alf Ove Braseth, Christer Nihlwing, Gyrd Skraaning, Arild Teigen, Øystein Veland, Nathan Lau, Greg A. Jamieson, Jordana Kwok, Catherine M. Burns

CEL 07-02 "The Ecological Interface Design experiment (2005) (HWR-833)"
• Gyrd Skraaning, Nathan Lau, Robin Welch, Christer Nihlwing, Gisle Andresen, Liv Hanne Brevig, Øystein Veland, Greg A. Jamieson, Catherine M. Burns