Ecological Interface Design in the Nuclear Domain: An Empirical Evaluation of Ecological Displays for the Secondary Subsystems of a Boiling Water Reactor Plant Simulator

Nathan Lau, Greg A. Jamieson, *Member, IEEE*, Gyrd Skraaning, Jr., and Catherine M. Burns

***Abstract*—Laboratory studies have shown that ecological interfaces can enhance operator performance in process control. However, limited verification and validation studies in representative settings are impeding the adoption of the Ecological Interface Design (EID) framework in the nuclear domain. A companion article presents an application of EID to the secondary side of a Boiling Water Reactor plant simulator, demonstrating that the framework can lead to display features and verification criteria relevant to supporting operators in both anticipated and unanticipated situations. This article presents an empirical study as a first step towards the validation of EID in the nuclear domain. The results suggest that ecological displays have a marked advantage in supporting operator performance during monitoring for unanticipated events as compared to mimic-based displays. The ecological displays did not support operator performance differently for other types of tasks. This study provides supporting or validation evidence that EID is effective at a scale and level of complexity that is representative of nuclear power plant operations. The implications for introducing ecological displays into NPP control rooms are discussed.**

***Index Terms*—Control room, ecological interface design, nuclear power plant.**

I. Introduction

ECOLOGICAL INTERFACE DESIGN (EID) is a theoretical framework for designing human-computer interfaces for complex socio-technical systems [1], [2]. The framework explicitly aims to support worker adaptation, especially during unanticipated events, thereby facilitating robust designs of user interfaces. Research on the EID framework has progressed significantly since its introduction over fifteen years ago [3]. Proof-of-concept ecological interfaces have been reported in many domains (see [2]) and performance benefits have been demonstrated in many empirical studies (see [4]). Despite its theoretical strength and accumulating research evidence, however, EID has yet to be widely adopted by industry.

One factor precluding industry from gaining the knowledge and confidence to adopt EID is a shortage of representative studies that address the verification and validation of design products in specific industrial settings. Verifying examples are critical to demonstrate applicability of the design framework while validating studies are crucial to confirm performance benefits. To address this research issue, the University of Toronto, University of Waterloo and the OECD Halden Reactor Project established a research program to provide design, verification and validation evidence for EID in the nuclear domain.

In a companion article [5], we present the applications of EID for both design and verification. This includes applying EID to the secondary side of a Boiling Water Reactor (BWR) plant simulator and analytically demonstrating that the framework specifies information requirement and design principles that could complement current verification practices. The companion article describes the EID framework and reviews previous EID applications in the open literature, outlining the need for verification research in the nuclear domain.

In this article, we turn to the validation or evaluation component of our research program that specifically examines operator task performance supported by ecological displays. We have also collected other data pertinent to the evaluation of human-system interfaces. Refer to [6]–[8] for our empirical evaluation on the support for Situation Awareness [9] provided by EID in this study. Refer to [10] for a qualitative analysis of interview and video data collected in this study. We first review the empirical foundations upon which our research program is built. Then, we present an empirical study evaluating the ecological displays for the secondary side of the BWR simulator described in the companion article, and discuss the benefits that EID could bring to the nuclear industry.

A. Review of EID Research—Empirical Studies

The empirical foundation of EID is built upon studies of process control 'microworld' simulations [4], [11]–[14], which have served as experimental platforms for a large number of studies covering many aspects of the interface design problem. Vicente [4] provides a comprehensive review of these results, concluding that EID can lead to robust and usable interfaces. Participants are frequently more effective and efficient in completing laboratory tasks using ecological

Manuscript received January 10, 2008; revised May 16, 2008. Current version published December 31, 2008. This work was supported by the Natural Science and Engineering Research Council (Canada) and the OECD Halden Reactor Project.

N. Lau and G. A. Jamieson are with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S3G8, Canada (e-mail: nathan.lau@utoronto.ca; jamieson@mie.utoronto.ca).

G. Skranning, Jr. is with the Division of Industrial Psychology, Institute for Energy Technology, Halden NO-1751, Norway (e-mail: gyrds@hrp.no).

C. M. Burns is with the Systems Designs Engineering, University of Waterloo, Waterloo ON N2L 3G1, Canada (e-mail: c4burns@uwaterloo.ca).

Digital Object Identifier 10.1109/TNS.2008.2005725

interfaces than with traditional (i.e., mimic-based) interfaces under unanticipated or uncommon conditions. Unanticipated events typically force operators to engage knowledge-based work, involving reasoning about safety and operating goals, and the sometimes conflicting means of achieving them (see [15], [16]; also cf., knowledge-based behaviors in [16]). The advantages with knowledge-based tasks are usually obtained without performance decrements under anticipated conditions when operators typically engaged in procedure-guided tasks, primarily involving rule-based-decision-making (see [15], [16]; also cf., rule-based behaviors in [16]).

These empirical studies, however, lack the scale and complexity needed to evaluate the performance benefits of EID in the process industries. Nevertheless, they motivate continued research, including proof-of-concept applications and a selection of empirical studies.

Of the ecological interfaces for industrial process systems mentioned in the companion article [5], three were empirically evaluated. Jamieson [17], [18] developed an ecological interface for a simulated petrochemical process and evaluated it in a full-scope simulator with licensed operators. The results corroborated many of the findings of the foundational studies, supporting the generalization that ecological interfaces improve monitoring and control performance in comparison to conventional computer interfaces.

The two other EID empirical studies are situated in the power generation domain. Ham and Woon [19], [20] presented an empirical evaluation of ecological interface content in a nuclear plant simulator. They found that student participants presented with the full suite of information identified by a Work Domain Analysis were more effective at diagnosing unfamiliar faults than those presented with a subset that was similar to the information content identified by task-based approaches. Because the evaluation was limited to the content (i.e., there was no manipulation of graphical representation), the study did not completely assess the EID framework, which typically employs configural graphics to communicate system information.

Burns [21], [22] developed and implemented three ecological displays based on a single work domain analysis for a simulated prototype fossil fuel power plant. Each of the displays used a different information integration and navigation technique (as opposed to the two previous investigations where alternative interfaces were developed through different design approaches). The displays afforded an empirical comparison of these techniques employing university students as participants. While the findings provide valuable guidance for the design of interfaces for the process industries, they are not intended to serve as validation evidence for performance benefits of ecological interfaces over conventional interfaces.

B. *Implications for the Nuclear Industry*

Empirical findings in representative settings to date support the conclusion that the benefits of EID observed in microworlds can generalize to real world applications for process control. However, the weight of empirical evidence collected under conditions representative of the industrial environment and user population is insufficient to validate the claims of EID and facilitate broad adoption in industry.

This shortage of representative empirical evaluations may impede the adoption of EID in the nuclear domain. To acquire support from management and regulators, demonstrating the ability of ecological displays to meet operational and safety goals or to obtain benefits over existing technology is crucial. Given the validation evidence currently available, it is unrealistic to expect widespread adoption of EID in the nuclear industry. In effect, the nuclear industry would not be able to capitalize on the potential value of EID providing support for operators to cope with unanticipated events [15], [23].

C. *Overview of the Current Study*

To support the nuclear industry in managing unanticipated events, EID researchers should begin the validation process by gathering empirical evidence in representative settings. This article presents the first empirical evaluation of an ecological interface that was developed for the secondary side of a high fidelity simulator [5]. This study provides empirical evidence on the relative levels of operator performance for ecological displays in comparison with mimic-based displays during realistic nuclear power plant events. The findings are the first validation evidence for EID in the nuclear domain.

II. Method

A. *Participants*

Six licensed operator crews[1] $(\mathrm{n} = 6)$ were recruited from a BWR power plant identical to the simulated process. Each crew consisted of one reactor operator (RO) and one turbine operator (TO), responsible for the primary and secondary side of the simulated process, respectively. In two cases, participants currently working as ROs operated the secondary side. This substitution should not affect generalization of the results given that all ROs must previously or currently hold TO licenses. Because the Ecological displays are only developed for the secondary side, the results and discussion in this paper only pertain to the performance data collected on the TOs. (The performance data collected on the ROs was not analyzed here.)

B. *Experimental Environment*

We used the HAlden Man-machine laboratory BOiling water reactor (HAMBO) [24], [25] as the experimental platform for this study. HAMBO, a high-fidelity simulator of a 1200 MW boiling water reactor plant (in operation), offers a realistic environment of industrial nuclear processes and features for sophisticated graphics [5]. The access to licensed operators from the operating plant corresponding to the simulation was also a crucial factor for our decision to use HAMBO, as this increases the representativeness of the study. However, we confined our

[1] A "crew" as organized by the nuclear plant where operators were recruited is composed of: (a) two turbine operators, (b) one reactor operator, and (c) one shift supervisor. In this study, we were able to recruit three four-operator crews and reorganized them into six two-operator crews. In general, we reserved the role of the turbine operators in the re-organized crews for the turbine operators (leaving the role of reactor operators in the study to either reactor operators or shift supervisors). This re-organization was a compromise for the limited access to licensed operators but should not significantly affect the outcome of the experiment, which focused on the performance of turbine operators.

Fig. 1. HAMMLAB: reactor operator workstation, left; large screen display, center; turbine operator workstation, right.

Fig. 2. LCD panel layout for the turbine operator workstation in the Ecological display condition.

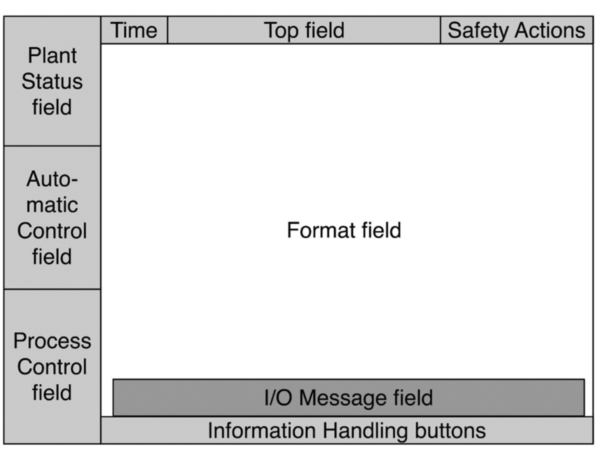

Fig. 3. Display layout.

scope to the secondary side of the plant because of the higher availability of TOs.

The HAMBO simulator resides within the HAlden Man-Machine Laboratory (HAMMLAB) facility, where the experiment was conducted. Fig. 1 shows the reactor and turbine operator workstations, and the large screen display in HAMMLAB[2]. Each operator station consists of twelve 19" LCD panel displays. Fig. 2 schematically shows the layout of the displays for the turbine operator workstation. All navigation and interaction is carried out via keyboard and mouse. The large screen display in the centre is shared between both reactor and turbine operators and provides a condensed mimic diagram of both the primary and secondary side with key instrumentation outputs.

C. Experimental Manipulations

1) Display Types: Three display types—Traditional, Advanced, and Ecological—were selected for comparison. Prior to illustrating their distinct characteristics, we first describe the shared features of the interfaces.

As mentioned in the companion article [5] and illustrated in Fig. 2, the design scope is limited to the secondary side. Given this scope, alarm information was communicated in the same manner across the display conditions (i.e., they have the same alarm displays). Furthermore, all three types of displays share the general layout shown in Fig. 3, in which the gray areas are

[2]In HAMMLAB, a supervisor workstation is also available but not shown as it is not part of this study.

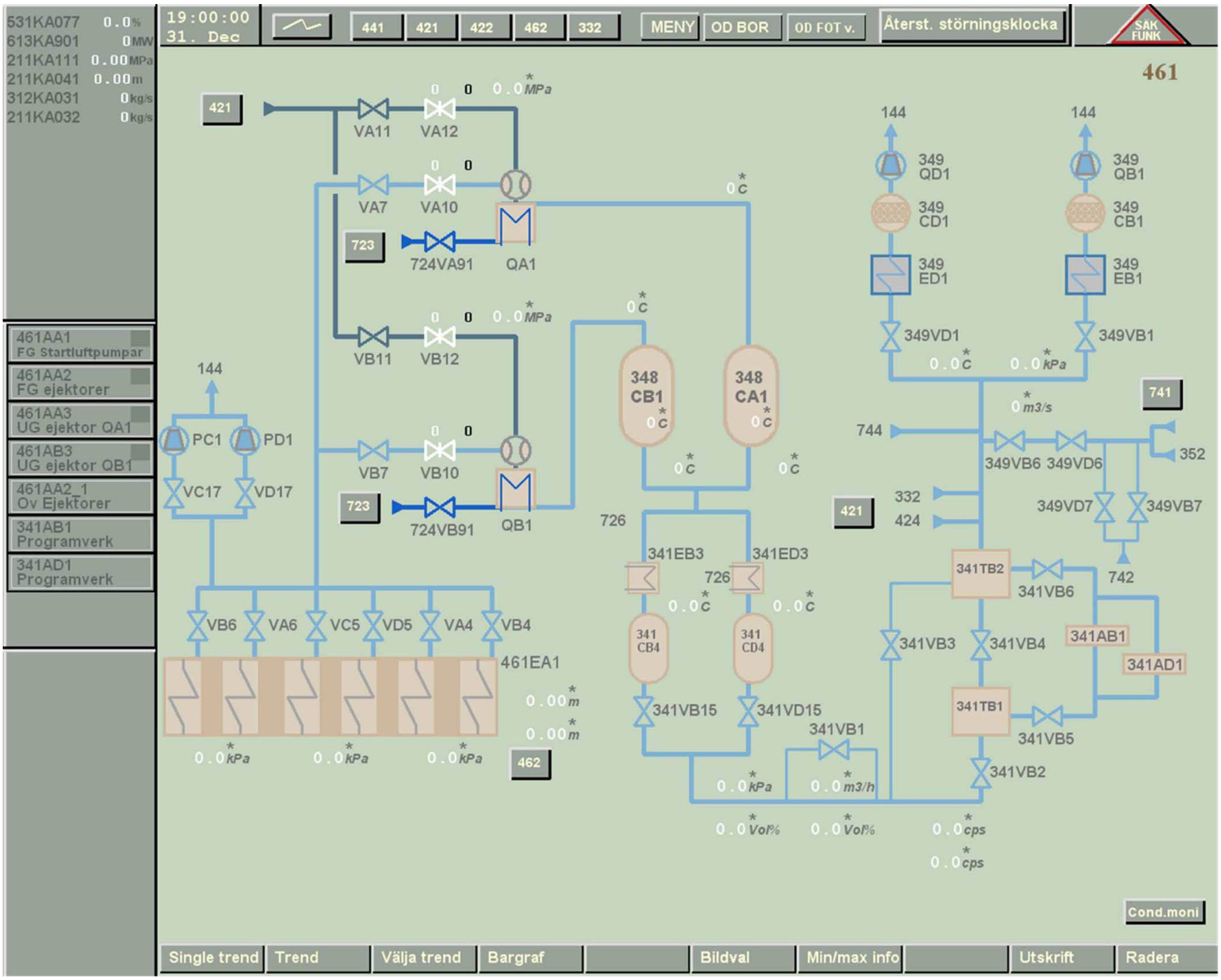

Fig. 4. An example of a Traditional display.

the same across the three display types. The interaction style was also consistent across display type. Specifically, operators could click on a plant component icon in the format field of any display to access/view the equipment status and related control variables that appeared in the automatic and/or process control fields. To execute control actions, operators could key in desired values in corresponding variable entry fields. In contrast, for displays that were within the design scope (i.e., the secondary side), the format fields (i.e., the white area in Fig. 3) varied according to the description as follows.

a) Traditional: The Traditional displays are the computerized version of the hard-wired wall panels originally installed in the operating nuclear plant (Fig. 4). The Traditional displays roughly represent the "state-of-practice" design, characterized by mimic diagrams of the facilities with numeric outputs of instrumentation (i.e., mimic-based displays). Although the design process might not adhere strictly to a particular framework, the design was largely informed by task analyses and possibly some user input as mandated by regulators (see [26]).

b) Advanced: The Advanced displays are an improved version of the Traditional displays (Fig. 5). The Advanced displays retain the mimic-diagrams of the Traditional displays; however, they also contain some configural graphics (e.g., [27], [28]) and "mini-trends" strategically developed or inserted by process experts. For this display condition, the large screen display contained some advanced visualization features that were absent from both Traditional and Ecological display conditions in addition to the condensed mimic diagrams with key instrumentation outputs. The Advanced displays generally represent the latest implemented design, characterized by some novel visualization or leading edge features on top of mimic-diagrams with numerical outputs of instrumentation. The Advanced displays resemble the displays developed in many control room modernization projects. Though potentially similar to configural graphics typifying ecological displays, the new visualizations in the Advanced displays are based on expert opinions that are often considered as products of a user-centered approach and evolutionary design strategy.

c) Ecological: The Ecological displays were designed according to the EID framework as described in the companion article [5]. Because the design scope was limited, the participants had access to the Traditional displays for plant processes

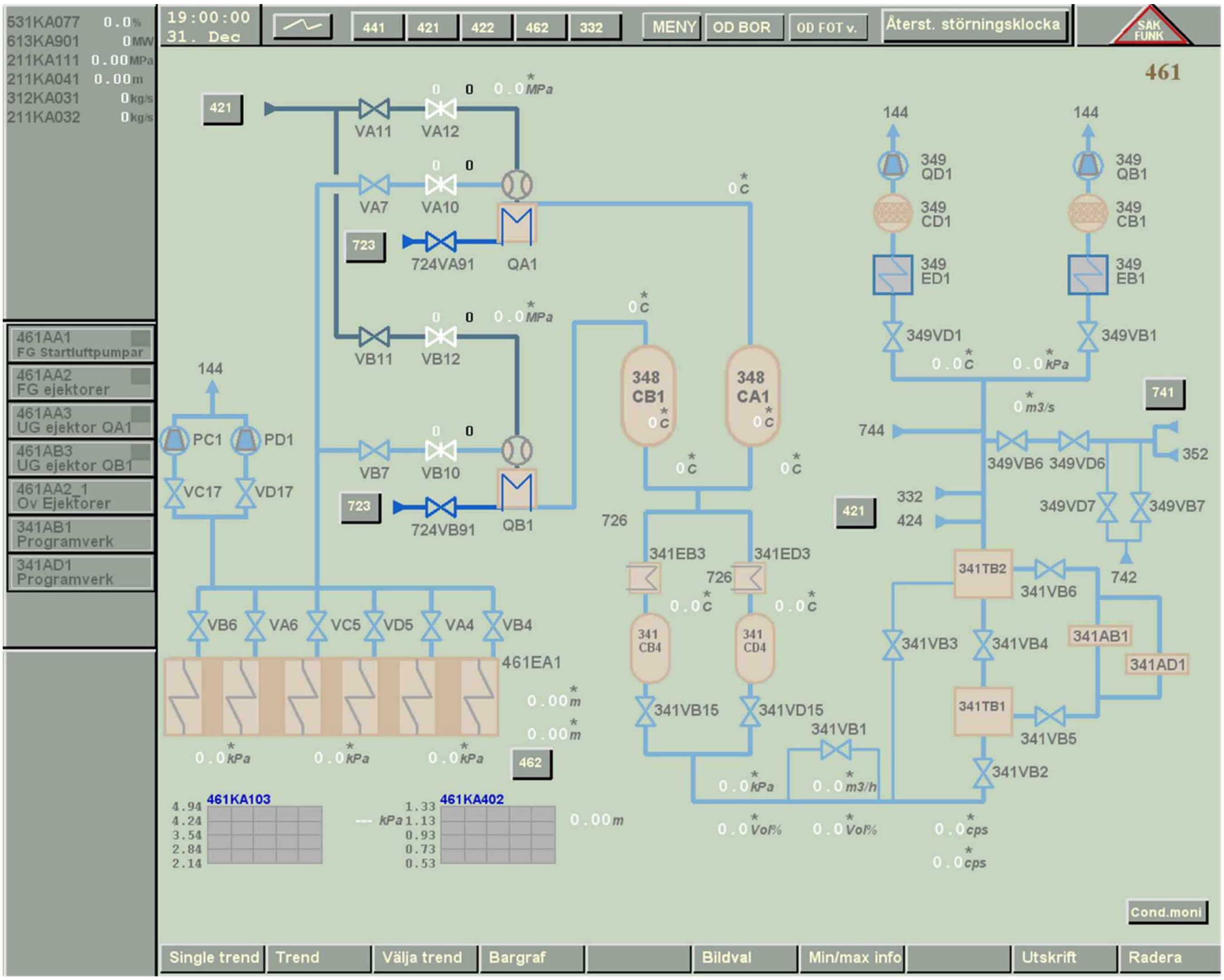

Fig. 5. An example of an Advanced display.

that were not represented by the Ecological displays. Furthermore, the operators had access to the large screen display in the Traditional display condition.

2) Scenario Type: This study contained three Procedure-guided and three Knowledge-based scenarios (Table I). In general, equipment failures anticipated by the utilities and job responsibilities familiar to operators characterized the Procedure-guided scenarios, while unanticipated failures and unfamiliar responsibilities characterized the Knowledge-based scenarios.

For the purpose of this study, Procedure-guided scenarios were defined by a set of disturbances that could be resolved by referencing plant procedures. For instance, one Procedure-guided scenario involved a leak in the intermediate superheater at full power. An alarm (based on the temperature difference between two instruments) could identify this fault, which could be resolved by following a procedure for mitigating the alarm(s). Operators who did not follow the procedure might encounter more errors or pursue a less efficient solution path.

Scenarios in which disturbances could not be resolved by procedures were classified as Knowledge-based. For instance, one Knowledge-based scenario challenged the operator to respond to a sudden temperature increase of seawater, the primary cooling source for the entire power plant. Although alarms would sound as equipment became overheated, prescribed solutions did not exist in any procedure. In this scenario, the best solution was to reduce power manually, even though maintaining full power was one of the operational objectives.

3) Scenario Phase: Each scenario started with a "Detection" phase, a time period just before the first alarm sounded, and ended with a "Mitigation" phase that consisted of all subsequent events. (Fig. 6 illustrates the detailed structure of the scenarios.) The two phases afforded separate assessments of the effectiveness of the displays in supporting both monitoring and intervention.

D. Experimental Design

A $3 \times 2 \times 2$ within-subjects design was employed with treatments of display type (Traditional, Advanced and Ecological), scenario type (Procedure-guided and Knowledge-based), and scenario phase (Detection and Mitigation). The treatments were completely crossed and counterbalanced using a Latin-square technique. Table II presents the final experimental design and

TABLE I
DEFINING CHARACTERISTICS OF THE PROCEDURAL-GUIDED AND KNOWLEDGE-BASED SCENARIOS IN THE EXPERIMENT

Procedural-guided scenarios	Defining Characteristics
1. Leak in the intermediate super heater	A leak from the tube side of the intermediate super heater led to an alarm that would direct operators to various procedures to resolve the leak and stabilize the plant.
2. Drain switching and high-pressure pre-heater bypass	Malfunctioned valves and cooling water pumps led to difficulties in executing the part of a procedure where steam should be released to the high-pressure preheaters. Alarms should direct the operators to various procedures for corrective responses.
3. Drain switching problem	Malfunctioned valves and a drifting instrument led to difficulties in executing the part of a procedure where drain flows from intermediate superheater should be switched to the high-pressure pre-heater. Alarms and redundant instruments should direct the operators to various procedures for corrective actions.

Knowledge-based scenarios	Defining Characteristics
1. Turbine trip with generator connected to the grid	Some malfunctioned valves and alarm system errors led to a sudden turbine trip without a generator breaker trip. There were no specific procedures to support the operators in a turbine situation while the generator was still connected to grid. The best course of action is to scram the plant to avoid emergency and equipment stress.
2. Leak in condensate cleaning system	A leak in the condensate cleaning equipment and a set of malfunctioned valves caused unanticipated fluctuation of condensate levels in the condenser and feedwater tank. The procedures and automation were not designed to manage such a situation. The operators should identify the appropriate standby valve to manually regulate the water levels and stabilize the situation.
3. High temperature sea water	A raise in the seawater temperature led to reduced cooling capability across the entire plant. There were no procedures anticipating such situation. The best course of action is to lower power output manually.

TABLE II
THE FINAL EXPERIMENTAL DESIGN: DISPLAY TYPE $\times$ SCENARIO TYPE $\times$ SCENARIO PHASE. THE INDEPENDENT VARIABLES WERE COMPLETELY CROSSED. THE RUN ORDERS WERE COUNTER-BALANCED USING A LATIN SQUARE. (NOTE THAT SCENARIO 1, 2 AND 3 ARE PROCEDURE-GUIDED SCENARIOS; WHEREAS, SCENARIO 4, 5 AND 6 ARE KNOWLEDGE-BASED SCENARIOS.)

Crew	Procedure-guided Scenarios						Knowledge-based Scenarios					
	Traditional Displays		Advanced Displays		EID Displays		Traditional Displays		Advanced Displays		EID Displays	
	Det. Phase	Mit. Phase	Det. Phase	Mit. Phase	Det. Phase	Mit. Phase	Det. Phase	Mit. Phase	Det. Phase	Mit. Phase	Det. Phase	Mit. Phase
1	Scenario 1		Scenario 2		Scenario 3		Scenario 5		Scenario 6		Scenario 4	
2	Scenario 3		Scenario 1		Scenario 2		Scenario 5		Scenario 4		Scenario 6	
3	Scenario 3		Scenario 1		Scenario 2		Scenario 4		Scenario 5		Scenario 6	
4	Scenario 3		Scenario 2		Scenario 1		Scenario 5		Scenario 4		Scenario 6	
5	Scenario 3		Scenario 2		Scenario 1		Scenario 4		Scenario 5		Scenario 6	
6	Scenario 1		Scenario 3		Scenario 2		Scenario 4		Scenario 6		Scenario 5	

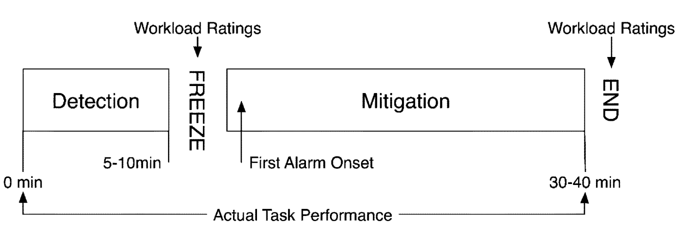

Fig. 6. Basic structure of the scenarios.

the assignment of the six crews to the six scenarios that were divided into two phases (N = 72).

E. Hypotheses

The theoretical foundations of EID [1] and previous empirical results [4], [18] suggest that Ecological displays would support operators better than both Traditional and Advanced displays. In particular, the performance advantage of the Ecological displays was anticipated to be most pronounced in Knowledge-based scenarios, in which problem solving would be the primary means to resolving process disturbances.

F. Measures

1) Actual Task Performance: Actual task performance was captured and quantified using the Operator Performance Assessment System (OPAS) [6], [29], [30]. OPAS provides a structure for the assessment of whether operators carry out their task work in accordance with scenario solutions prescribed a priori by experts in control room operation.

Prior to data collection, process experts analyzed the scenarios and developed optimal solutions by identifying items that expressed the desired performance. In principal, any discrete performance criterion that can be verified against actual operator data (i.e., observable operator behaviors) may serve as a performance item. Typical items that could differentiate between levels of task performance across experimental conditions include omissions, commissions, response time, and strategies (Table III; also see e.g., [31]). Performance items may also be perceived in the light of process operation activities in terms of

TABLE III
GENERIC EXAMPLES OF OPAS PERFORMANCE ITEMS.
(SCORES VALUES IN SQUARE BRACKETS.)

Omission Example	Detect a specific event (e.g., an alarm):
	No [0] Yes [2]
Commission Example	**Performing specific actions (e.g., starting a specific pump):**
	Do not start or start wrong pump [0] Start the correct pump without informing field operator [2] Start correct pump and inform field operator [3]
Response Time Example	**Perform a specific actions (e.g., close a valve) after a specific event (e.g., an alarm):**
	Do not close or close valve after 8 min [0] Close valve between 5 to 8 min [1] Close valve between 3 to 5 min [2] Close valve within 3 min [3]
Strategy Example	**Performing specific actions in a certain sequence (e.g., managing faults in appropriate priorities):**
	Fail to stop or notice leakage [0] Stop leakage after managing sub-disturbances [1] Stop leakage before managing sub-disturbances [3]

TABLE IV
EXAMPLES OF OPAS PERFORMANCE ITEMS FOR THIS EXPERIMENT.
(SCORES VALUES IN SQUARE BRACKETS.)

Detection Example	Detect and take care of the problem in the turbine plant main stream systems 421
	a) Does not take care of the 421 problem correctly [0] b) Detects and takes care of the 421 problems correctly with in 7 minutes [1] c) Detects and takes care of the 421 problems correctly with in 5 minutes [2] d) Detects and takes care of the 421 problems correctly with in 3 minutes [3]
Inference Example	**Make inference about leakage in the steam super heater 422EA1**
	a) No or wrong inference [0] b) Partly correct inference [1] c) Mostly correct inference [2] d) Fully correct inference [3]
Action Example	**Close water level regulating valve for the feedwater tank 462VA5**
	a) No action [0] b) Close 462VA5 in order to manually close 332VB2 [1] c) Close 462VA5 in order to manually close 332VB2, but before the low level alarm in the feedwater tank sounded [3]
Teamwork Example	**Inform reactor operator about the upcoming power reduction**
	a) No [0] b) Yes [3]

safety (i.e., preventing/mitigating transients or accidents), production (i.e., maintaining power production level), and preservation activities (i.e., minimizing physical stress on equipment).

Process experts specify different performance items (i.e., content of the task performance construct) according to the defining characteristics of the scenarios. In effect, performance items vary with scenarios according to the judgments of the experts. OPAS is designed to rely on experts extensively because past research [32] indicated that expertise is a critical to understanding human task performance in complex domains.

For this experiment, one process expert analyzed the scenarios and specified the performance items. A simple scoring system was used, where the operators earned points for completing performance items. Each item depicted alternative operator activities that were rewarded by 0, 1, 2 or 3 points. In this experiment, the performance items may be classified as one of the four categories: detection, inference, action and teamwork/communication[3] (see examples in Table IV).

During the experiment, a process expert registered the points earned by operators in completing the predefined activities within each performance item based on observations of operator verbalization, physical behaviors, problem solving, and system states. Studies have shown that real time expert rating is comparable to objective data logs (e.g., simulator logs and video recordings) and that a single expert rater is adequate given the high inter-rater reliability [29] of the OPAS instrument. The employed performance index is the unweighted average of all performance items defined for a scenario.

The OPAS index reflects the discrepancy between operator performance and predefined optimal solutions to scenarios. Due to its relativistic nature, the OPAS index cannot establish any general acceptance criteria, as it is only meaningful for comparisons between indices across situations. Nevertheless, operator performance relative to the optimal level can be psychologically meaningful. OPAS assesses the degree of conformance with performance expectations that remain constant across task conditions; thus, the raw scores originating from different scenarios can be compiled into one performance index. In addition, OPAS is similar to training and licensing assessment situations in the nuclear domain, for which human performance constructs are often ill-defined and may be difficult for non-experts to understand due to domain complexity.

The ill-defined nature of human performance in complex domains is partially attributable to the fact that measures of task performance often include multiple and interacting aspects of human performance. Some aspects of human performance, such as expertise or past experience, are not generally considered part of task performance but can significantly affect task performance. In other words, task performance measures often cannot distinguish between interacting aspects of human performance, even though these aspects may be psychologically or conceptually distinct. For this reason, task performance may be interpreted from multiple perspectives depending on the combinations of aspects of interest.

In this study, we are particularly concerned with distinguishing between task performance and workload. Workload is largely driven by the nature of the scenarios, which also determines the OPAS performance items. For some scenarios, operators may experience high workload from completing many relatively simple performance items in a short amount of time. Other operators may experience high workload from completing only a few complex performance items. In either case, operators need to overcome workload demand to achieve high OPAS indices. From one perspective, workload is an integral part of task performance as both are always present when performing work. From another, workload remains distinct from task performance, as workload could mediate but not determine task performance. To consider task performance

[3]Performance items were classified during the analysis phase.

TABLE V
WORKLOAD ITEMS OF THE SUBJECTIVE TASK-COMPLEXITY SCALE

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	May 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

TABLE VI
TRAINING PROGRAM

	Activity	Day	Time (min)
1	Familiarizing with the facility, interface navigation, alarm systems, large screen displays	1	60
2	Training on the Traditional and Advanced interfaces	1	60
3	Training on the differences between physical/home plant and simulated plant	1	30
4	Training on the data collection procedures	1	30
5	Training on the Ecological interface	1	180
6	Re-training on Day 1 materials prior to data collection	2	60

from both perspectives, the results include the effects of experimental manipulations on (a) Workload and (b) Actual task performance controlling for the mediating effect of Workload (see IV Discussion).

2) Workload: Workload is generally accepted to have a significant impact on performance. Improved task performance at the expense of higher workload is usually not desirable; thus, we collected workload data using a subjective task-complexity scale developed by the OECD Halden Reactor Project [33]. The scale is a self-rating instrument focusing on task-related difficulties that control room operators experience while they work. Participants rate five items (Table V) in a seven-point Likert scale anchored by ‘very difficult’ (1) and ‘very easy’ (7). Several psychometric evaluations and experimental studies indicate that the scale is more reliable and predictive of task performance in representative nuclear process control settings [6], [7] than the NASA-TLX [34]. [35] discusses the subjective task-complexity scale in detail.

G. Procedure

The participation of each crew was divided over three consecutive days. The first day was dedicated to the training program after obtaining informed consent and demographic information. Six hours of training occurred on the first day as outlined in Table VI. The second day started with a one-hour training session to refresh the materials presented on the first day, followed by three scenarios with fifteen-minute breaks in between. The third day started with three scenarios also with fifteen-minute breaks in between, followed by a debriefing/closing session.

For all scenarios, crews were asked to maintain the original power level and safe operation. The process expert (who analyzed the scenarios) registered OPAS scores to corresponding performance items at various points of the scenarios by observing the participants while they monitored system states and resolved disturbances. The participants also responded to the subjective task-complexity questionnaire during a short simulator freeze and at the end of each scenario. The simulation freeze occurred at the end of the Detection phase, which took up the first five to ten minutes of the scenario as depicted in Fig. 6. The scenario then continued with the Mitigation phase, which was marked by the onset of the first alarm within the first minute. The Mitigation phase usually lasted for 30 to 40 minutes, followed by another administration of the subjective task-complexity questionnaire at the end of the scenario.

TABLE VII
ANOVA RESULTS FOR WORKLOAD

	SS	df	MS	F	p
Display	1.423	2	0.7117	0.59	0.5736
Error	12.10	10	1.210		
Scenario	0.2939	1	0.2939	0.92	0.3814
Error	1.596	5	0.3192		
Phase	22.89	1	22.89	23.03	0.0049
Error	4.969	5	0.9939		
Display*Scenario	0.7078	2	0.3539	0.32	0.7304
Error	10.91	10	1.091		
Display*Phase	3.968	2	1.984	5.39	0.0258
Error	3.679	10	0.3679		
Scenario*Phase	3.125	1	3.125	8.59	0.0326
Error	1.818	5	0.3637		
Display*Scenario *Phase	1.270	2	0.6350	0.75	0.4983
Error	8.497	10	0.8497		

III. RESULTS

Two statistical models were built to analyze the dependent variables. The first is a three-way analysis of variance (ANOVA) on Workload, and the second is an analysis of covariance (ANCOVA) on Actual task performance controlling for Workload.

A. Assumptions

The validity of the two statistical models rest on several assumptions. The normality assumption for the Workload and Actual task performance measurements is not satisfied according to the Shapiro-Wilks’ W tests on the distributions for every combination of treatments. However, both ANOVA and ANCOVA are generally robust against the violation of normality, except for some specific characteristics of population distribution. (See [36] and [37] for a discussion of normality violations on ANOVA and ANCOVA, respectively). For the normality assumption, we thus examined the distributions for every treatment level using histograms and normal probability plots for these specific violations and did not find any major threats to the validity of the statistical results. The spherity assumption also applies to the statistical models, which are both repeated-measures. The Mauchley’s tests indicate that all effects of both models satisfy the sphericity assumption. ANCOVA requires an additional assumption—homogeneity of slopes/within group regression or parallelism. The homogeneity of slopes assumption is satisfied according to visual examinations of the scatter plots, and the interaction terms between independent variables and the continuous predictor in a general linear model.

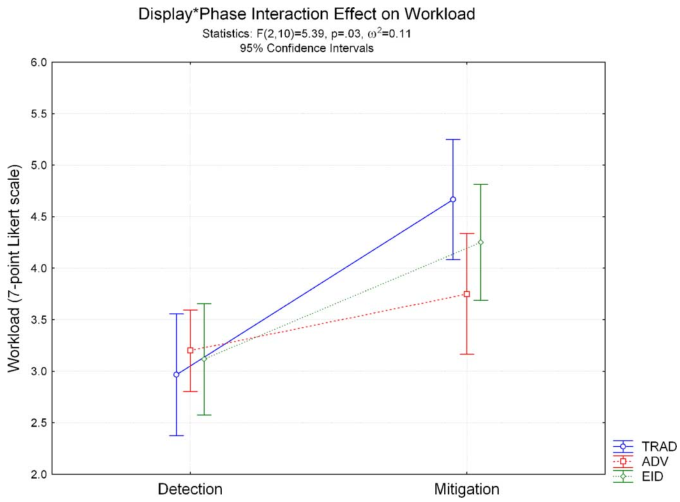

Fig. 7. Interaction plot of Display and Phase for Workload. The plot is drawn according to the method proposed by [44] to remove within-subject variance. (Note that overlaps between confidence intervals do not necessarily indicate that the means are not significantly different. See [45], [46] for a discussion.)

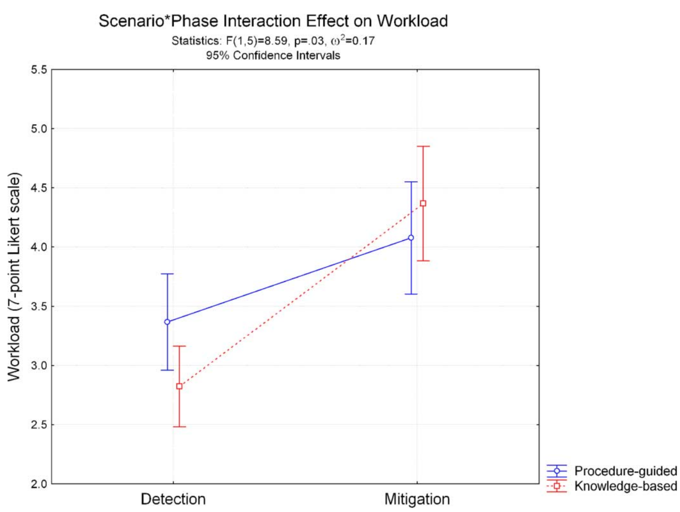

Fig. 8. Interaction plot of Scenario and Phase for Workload. The plot is drawn according to the method proposed by [44] to remove within-subject variance. (Note that overlaps between confidence intervals do not necessarily indicate that the means are not significantly different. See [45], [46] for a discussion.)

B. Workload—The Covariate

Workload was measured by a subjective self-rating scale, which was confirmed to have a sufficiently high inter-item reliability ($\alpha = 0.89$) for further analysis. Workload was analyzed in an ANOVA with fixed factors of display type (Traditional, Advanced and Ecological), scenario type (Procedure-guided and Knowledge-based), and scenario phase (Detection and Mitigation), and with a random factor of crew. The ANOVA is an over-parameterized model built on Type II sums of squares.

This analysis explores the fixed effects on Workload that are helpful for interpreting the next model—an ANCOVA on Actual task performance. Workload and Actual task performance have a low correlation, $\mathrm{r}(72) = .31, \mathrm{p} < 0.01$. Table VII presents the results of the ANOVA for all effects. A significant main effect of phase $(\mathrm{F}(1,5) = 23.03, \mathrm{p} < 0.01)$, and significant two-way interaction effects for display and phase $(\mathrm{F}(2,10) = 5.39, \mathrm{p} < 0.05)$, and scenario and phase $(\mathrm{F}(1,5) = 8.59, \mathrm{p} < 0.05)$ were observed.

We omit the plot of the phase main effect as it provides limited and redundant information as compared to the significant two-way interaction effects. The display and phase interaction plot (Fig. 7) illustrates no practical difference in Workload between the three display types in the Detection phase, but the Workload increase to the Mitigation phase is highest with the Traditional displays and lowest with the Advanced displays. The display and phase interaction effect accounts for 11% of the total variance $(\omega^2 = 0.11)$[4].

The scenario and phase interaction plot (Fig. 8) illustrates that the participants experienced less Workload during the Detection phase and more Workload during the Mitigation phase of Knowledge-based when compared to Procedure-guided scenarios (in which the participants experienced more Workload during the Detection phase and less Workload in the Mitigation phase). The scenario and phase interaction effect accounts for 17% of the total variance $(\omega^2 = 0.17)$.

C. Actual Task Performance Controlled for Workload

Actual task performance (i.e., the OPAS indices) was analyzed in an ANCOVA with fixed factors of display type (Traditional, Advanced and Ecological), scenario type (Procedure-guided and Knowledge-based) and scenario phase (Detection and Mitigation), a random factor of crew, and a covariate of Workload. The ANCOVA was an over-parameterized model built on Type II sums of squares.

This analysis explores the fixed effects on Actual task performance controlled for Workload, assessing the support for problem solving provided by each display type while limiting the mediating effect of task demand. The results provide empirical evidence on whether EID could introduce performance benefits according to its theoretical foundation. Table VIII presents the results of the ANCOVA for all effects. The significant effects on Actual task performance after controlling for Workload are the two-way interaction of display and phase $(\mathrm{F}(2,10) = 8.09, \mathrm{p} < 0.01)$, and the three-way interaction of display, scenario and phase $(\mathrm{F}(2,9) = 6.08, \mathrm{p} < 0.05)$.

Because the two-way interaction only provides limited and redundant information, we present the three-way interaction plot. Fig. 9 suggests that the Ecological displays enhanced Actual task performance in the Detection phase of Knowledge-based scenarios. The performance difference between interfaces in other experimental conditions appeared negligible. The three-way interaction effect accounts for 12% of the total variance $(\omega^2 = 0.12)$.

A post-hoc analysis using Tukey's Honestly Significant Difference (HSD) criterion for significance was conducted to confirm the performance advantage of Ecological displays in the Detection phase of Knowledge-based scenarios. As

[4][36] provides the formulae for calculating the effect sizes based on F-value, number of levels for each treatment (p, q, r), and number of blocks/operators (n) for main and two-way interaction effects. We extended the formulas provided by [36] for the three-way interaction effects. The formulae applied to calculated the effect sizes in this article are: $\omega_p^2 = ((\mathrm{p}-1)(\mathrm{F_p}-1)/((\mathrm{p}-1)(\mathrm{F_p}-1)+\mathrm{npqr})$; $\omega_{\mathrm{pq}}^2 = ((\mathrm{p}-1)(\mathrm{q}-1)(\mathrm{F_{pq}}-1)/((\mathrm{p}-1)(\mathrm{q}-1)(\mathrm{F_{pq}}-1)+\mathrm{npqr})$; $\omega_{\mathrm{pqr}}^2 = ((\mathrm{p}-1)(\mathrm{q}-1)(\mathrm{r}-1)(\mathrm{F_{pqr}}-1)/((\mathrm{p}-1)(\mathrm{q}-1)(\mathrm{r}-1)(\mathrm{F_{pqr}}-1)+\mathrm{npqr})$.

TABLE VIII
ANCOVA RESULTS FOR ACTUAL TASK PERFORMANCE WITH WORKLOAD AS THE COVARIATE

	SS	df	MS	F	p
Display	0.9986	2	0.4993	0.99	0.4064
Error	5.0524	10	0.5052		
Scenario	0.8219	1	0.8219	1.88	0.2294
Error	2.190	5	0.4381		
Phase	0.0300	1	0.0300	0.14	0.7156
Error	0.8501	5	0.1700		
Display*Scenario	3.813	2	1.907	1.97	0.1912
Error	9.290	10	0.9290		
Display*Phase	4.400	2	2.199	8.09	0.0074
Error	2.799	10	0.2799		
Scenario*Phase	1.113	1	1.113	1.95	0.2176
Error	3.237	5	0.6475		
Display*Scenario *Phase	1.637	2	0.8187	6.08	0.0213
Error	1.211	9	0.1346		

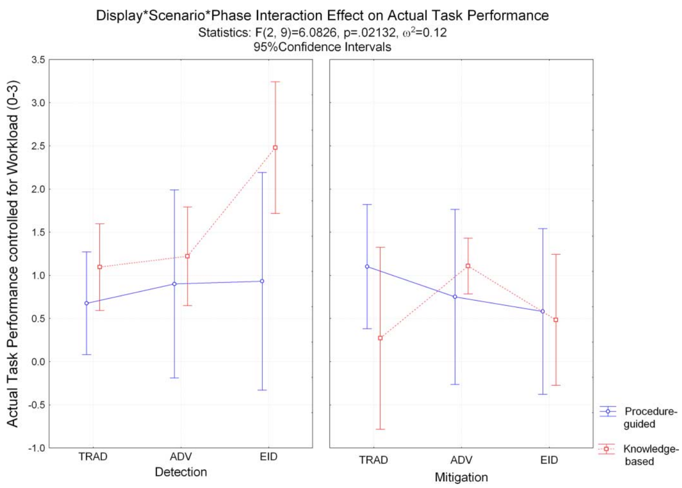

Fig. 9. Interaction plot of Display, Scenario and Phase for Actual Task Performance controlled for Workload. The plot is drawn according to the method proposed by [44] to remove within-subject variance. (Note that overlaps between confidence intervals do not necessarily indicate that the means are not significantly different. See [45], [46] for a discussion.)

common statistical software applications do not support the post-hoc test for the above ANCOVA model, we applied a technique discussed in [38], [39]. The technique (1) formulates a new set of scores by subtracting the standardized scores between the dependent variable and the covariate, and (2) applies ANOVA on the new scores, which exclude the variance contributed by the covariate. In our case, we built an ANOVA model on the differences between the standardized scores of Actual task performance and Workload. The final Tukey's HSD post-hoc analysis indicated that the performance of Ecological displays $(\mathrm{M} = 1.81, \mathrm{SD} = 0.89)$ was significantly higher than Traditional $(\mathrm{M} = 0.09, \mathrm{SD} = 0.72, \mathrm{p} < .01)$ and Advanced $(\mathrm{M} = 0.28, \mathrm{SD} = 0.87, \mathrm{p} < .01)$ in the Detection phase of Knowledge-based scenarios.

IV. DISCUSSION

The results have direct implications for interface design in the nuclear industry. In this section, we will first discuss effects of the experimental manipulations on Workload to seek a greater understanding of the covariate. Then, we examine the analysis on the Actual task performance controlling for Workload that provides the first empirical and validation evidence on EID in the nuclear domain. This section concludes with limitations, contributions and suggestions for future work.

A. *Workload—The Covariate*

The analysis with Workload as the dependent measure confirms the effectiveness of the experimental manipulations. A main effect of phase was expected given that the Detection phase only required monitoring while the Mitigation phase required intervention in addition to monitoring. The scenario and phase interaction effect was also expected. As a result of low familiarity with or poor anticipation of process events, participants would be more likely to miss the early indications of system disturbances in the Detection phase of Knowledge-based scenarios than of Procedure-guided scenarios, thereby assuming normal operating states and experiencing less Workload. However, in the Mitigation phase of Knowledge-based scenarios, participants must compensate for the unanticipated disturbances and late intervention leading to a substantial increase in Workload. On the other hand, when participants could detect the early indication of disturbances in Procedure-based scenarios, the increase in Workload from the Detection to Mitigation phase was less pronounced relative to Knowledge-based scenarios.

The analysis on Workload also illustrates some differences between the display types. The results indicate that both Advanced and Ecological displays induced lower increases of Workload from the Detection to Mitigation phase than the Traditional displays. This finding is encouraging in that the new visualization techniques do not result in higher Workload. The reduced workload supported by the Ecological displays in comparison to the Traditional displays is consistent with NASA-TLX results in [14]. The minimum Workload increase was unexpectedly observed with the Advanced displays. We postulate that the evolutionary improvements based on the user-centered approach for the Advanced displays may be concentrated on improving efficiency and thereby reducing Workload. In contrast, the ecological approach typically emphasizes interface effectiveness against unanticipated events that does not necessarily correlate with Workload.

B. *Validation Evidence: Actual Task Performance Controlling for Workload*

The theoretical foundations and accumulated empirical results pertaining to EID indicate that the primary contribution of introducing ecological displays would be superior support for knowledge-based or problem solving tasks relative to displays based on conventional approaches. More specifically, ecological displays improve task performance beyond the benefits from the mediating effect of workload reduction. From this perspective, task performance independent of workload variation could highlight the precise effect of ecological displays intended by the EID framework.

Actual task performance, however, as defined by OPAS is likely to include the influence of workload (see II Method). To limit the influence of workload while assessing the level of support for problem solving provided by the displays, we conducted an ANCOVA on Actual task performance with Workload as a covariate. The ANCOVA removed the variance associated with Workload in each scenario from the Actual task performance. (The Situation Present Assessment Method [40] applies a similar approach to separate workload effects from situation awareness measurements.) In effect, the ANCOVA results provided an indication of performance more confined to problem solving in comparison to an ANOVA on Actual task performance.

The decision to limit the mediating effect of Workload on Actual task performance in our analysis does not imply the independence between workload and task performance in general. The integral perspective (that task performance is a synthesis of many interacting aspects of human performance) is important, particularly for a summative evaluation such as those in integrated system validation [15]. Readers should not interpret the ANCOVA Actual task performance results independent of the ANOVA Workload results. Removing variance associated with Workload from Actual task performance through ANCOVA is a technique to clarify and emphasize of the intended effects of the EID framework rather than a proposal for altering interpretation of performance constructs within the nuclear domain. In consideration that both the Advanced and Ecological displays appear superior to the Traditional displays in terms of Workload according to the ANOVA two-way interaction effect (see section above and Fig. 7), this analysis approach appeared appropriate and meaningful as new visualization techniques do not seem to induce excessive Workload.

The ANCOVA results extend the available confirming evidence on the theoretical claim [1], [41] that EID could improve operator support for knowledge-based or problem solving tasks beyond alleviation of workload, corroborating the general findings of the previous EID studies [4], [18]. The three-way interaction plot (Fig. 9) and post-hoc analysis illustrate a marked advantage for the Ecological displays in the Detection phase of Knowledge-based scenarios over both Traditional and Advanced displays, whereas other performance differences were relatively negligible. This unique performance advantage also directly translates to the significant two-way interaction effect between display type and scenario phase.

The marked advantage for the Ecological displays in the Detection phase of Knowledge-based scenarios indicates that EID could lead to displays which better support operators in monitoring for unanticipated events or early phases of problem solving (i.e., problem identification and formulation) than mimic-based displays. Monitoring for critical events evolving from ‘normal’ operating states is a key part of supervisory control. Effective monitoring facilitates early intervention that can prevent process deviations developing into major disturbances or even accidents (see [42]). Furthermore, investigations have repeatedly indicated that major accidents are often preceded by unanticipated events [43], [44]. Thus, the unique advantage of the Ecological displays demonstrated in this study is encouraging in that EID could be a design solution for coping with unanticipated events, which have largely been neglected by conventional approaches.

The theoretical foundations of EID [1], [2], [41] support the argument that the framework could contribute to this benefit in two ways. First, the information content and structure identified by the Work Domain Analysis are explicitly selected to support operators in coping with all events, including unanticipated ones. In contrast, conventional approaches only explicitly identify information requirements of anticipated events. While all of these design approaches could effectively support monitoring for anticipated events (as suggested by the negligible performance difference between display types in the Detection phase of Procedure-guided scenarios (see Fig. 9)), the information content and structure in Ecological displays should better support operators in coping with unanticipated events. Second, the graphical forms in Ecological displays followed the SRK taxonomy, which served as an overarching framework to guide design towards high compatibility with human information processing (for all levels of cognitive control). On the other hand, conventional approaches usually contain specific, rather than overall, design heuristics and principles (e.g., [26]) to ensure compatibility with information processing. Thus, Ecological displays could communicate process information more effectively to operators than displays based on conventional approaches. This advantage would also be most prominent for information related to knowledge-based rather than rule-based decision making when common monitoring strategies do not apply.

The performance advantage of the Ecological displays, however, did not persist in the Mitigation phase of Knowledge-based scenarios, as predicted by the framework and observed in previous empirical studies. We postulate four related factors contributing to the diminished performance difference between display types in the Mitigation phase. First, operators were engaged in a greater mixture of tasks during the Mitigation than Detection phases. During the Detection phase, operators monitored process deviations and began problem solving. On the other hand, during the Mitigation phase, operators constructed intervention plans and executed control actions in addition to problem solving tasks. During the Mitigation phase, operators could have engaged some tasks involving rule-based decision-making even in Knowledge-based scenarios, such as executing control actions according to their planned solutions. In other words, the Mitigation phases inherently included tasks besides problem solving even in Knowledge-based scenarios. In effect, the unique support for problem solving provided by the Ecological displays may not be as relevant during Mitigation in comparison to the Detection phase.

The intervention nature of the Mitigation phase also relates to the remaining postulated factors. The second factor is that intervention may be more robust to interface effects as operator responses would rely on multiple skills and resources (e.g., trainings and procedures) as well as representation aides. Therefore, performance advantages induced by any one type of displays would be less discernable during the Mitigation phase due to reduced reliance on representation aides. The third factor is that the Traditional and Advanced displays could contain features particularly effective for intervention. Task-based and user-centered approaches typically emphasize efficiency and precision in executing control actions. Furthermore, the Ecological displays retained the same methods of interaction as the Traditional and Advanced displays, inhibiting a full assessment of the EID framework in supporting intervention. Fourth, enhanced support for other aspects of work, such as intervention as emphasized by the Traditional and Advanced displays, could lead to greater cognitive resources allocated to problem solving, thereby minimizing performance differences across display types. From this perspective, interface designers might regard task-based and user-centered approaches as complementary techniques to be integrated with the ecological approach [2], [17]. While these factors are plausible explanations for the lack of observed performance differences between the display types in the Mitigation phase, the results of this study prompt further empirical examination that explicitly compares visualization features contained in displays based on EID and other approaches.

C. Other Empirical Evidence

Actual task performance and Workload are not the only relevant indicators in human-system interface evaluation. More data were collected and analyzed to evaluate these displays, and these performance indicators should be interpreted in conjunction with one another.

The quantitative results of situation awareness [6]–[8] as a dependent measure converge with the performance indicators in this article. In summary, the analysis of variance results indicate a three-way interaction effect on the Process Overview measure $(\mathrm{F}(2,4) = 9.41, \mathrm{p} < 0.05)$, which is analogical to level 1 situation awareness (i.e., perception as defined in [9]), and the Scenario Understanding measure $(\mathrm{F}(2,4) = 31.07, \mathrm{p} < 0.01)$, which is analogical to level 2 and 3 situation awareness (i.e., comprehension and projection). For both Process Overview and Scenario Understanding, the three-way interaction plots illustrates a performance advantage for the Ecological displays in the Detection phase of Knowledge-based scenarios.

Extensive qualitative data was also collected during the study. A cursory review of the debriefing comments by the designers of the Ecological displays [45] indicates that the participants preferred the conventional simulator (i.e., Traditional and Advanced) displays for their less cluttered information presentation. The participants noted some graphical information in the Ecological displays appeared irrelevant (potentially due to a limited sample of scenarios). The participants did find that the graphical forms on the ecological displays were particularly useful for detecting anomalies but some were often overly complex for efficient diagnosis. All participants thought that the training time was generally insufficient. The formal analysis and report on the qualitative data collected in this study is in progress [10].

D. Limitations

Several limitations to the findings of the study warrant consideration. First, the Ecological displays employed in this study are, in fact, a hybrid Ecological-Traditional interface. The hybrid nature is evident in that: a) the scope of the Ecological displays was limited to the secondary side, b) the overview display in the Ecological display condition was the same as the Traditional one, and c) the interaction was consistent across all display conditions. All of these limitations were a direct consequence of the limited resources available for this study. Still, the reliance on a hybrid implementation raises the question of compatibility between the two display types that has not been investigated in the open literature. Given that the full plant must be in operation during the experiment, crews in the Ecological display condition used different types of displays to control the primary and secondary sides of the plant. In the Ecological display condition, crews used the large screen display in the Traditional display condition because an ecological version was not developed. This choice was influenced by the scope of the research program that excluded design work for the displays of reactor operators (i.e., the primary side), who also used the large screen display for monitoring. Furthermore, the interaction style was devised for the Traditional and Advanced displays, but was also used by participants in the Ecological display condition. This decision was governed by the availability of training time, which was estimated to be sufficient for only one interaction style. For all of these reasons, intervention may have been more be challenging in the Ecological display condition.

It is worth noting that, although a hybrid implementation is not ideal from an experimental perspective, it is actually quite representative of industry practice. In our experience, industry tends to adopt novel interface design techniques in phases, testing new concepts in limited operations while retaining the full suite of traditional controls. Thus, the benefits for the Ecological interface demonstrated in this experiment are likely to be conservative estimates. A comparison of displays that included displays for the primary side and large screen displays, and an interaction scheme based on EID, would provide a more accurate assessment of the merits of the ecological approach.

A final limitation is the limited training provided to operators in using the Ecological displays. Operators were generally more familiar with both the Traditional and Advanced displays, which are very similar to the one in the control room of the nuclear plant being simulated. In an experiment spanning several days, it is not possible to endow operators with the level of familiarity in a novel display that they have attained with displays that are either highly similar (i.e., Traditional) or substantially similar (i.e., Advanced) to those employed in their workplace. Thus, to observe superior performance with the Ecological displays in any condition compared to the others is rather remarkable from a training perspective. Again, it is likely that the benefits for ecological displays shown in this experiment are conservative estimates of the full effect of EID.

E. Contributions

This empirical study marks the beginning of EID validation in the nuclear domain. The findings replicate some of those in the only other EID study that is representative of operations in the process control domain [18]. Taken together these studies demonstrate that the benefits of ecological interfaces observed in laboratory settings can scale up to industry applications. The present study also demonstrated that hybrid implementation of ecological and conventional interfaces did not appear to hinder performance relative to (uniformly implemented) conventional interfaces. This provides reassurance that the common industrial strategy of adopting new visualization techniques in an evolutionary manner is a viable avenue for adopting EID.

F. Future Work

Validation depends on convergent support from a series of empirical studies. Subsequent studies must address several unattended issues. The scope of future assessments must be expanded to include the primary side and other operator support tools (e.g., large screen displays). Studies employing alternative performance measures (e.g., system efficiency) are also needed to obtain both convergent and discriminant validity. A complete set of human performance measures would also illustrate the particular facets of work best supported by EID. A more extensive set of scenarios is also needed to explore the consistency of support provided by ecological displays in other operating modes (e.g., start-up, shut-down, and re-fueling). As recommended in [2] and investigated in [17], [18], our findings suggest that integrating other approaches into the EID framework to explicitly provide procedural supports through ecological interfaces may result in efficient and robust interfaces which may not be achieved with any one design techniques. Thus, future studies also need to assess integrated techniques, probably in both laboratory and industrially representative settings.

V. Conclusion

The objective of our ongoing research program is to collect design, verification and validation evidence to assess the merits of EID in the nuclear domain. A companion article [5] presents our effort in demonstrating that the EID framework can lead to visualization features and verification criteria that are valuable for supporting and ensuring effective monitoring during both anticipated and unanticipated events. This article presents the first empirical evaluation of ecological displays in a setting representative of a nuclear power plant control room with professional operators. The results support the conclusion that ecological displays could provide a marked advantage for monitoring for unanticipated events over other conventional displays while other performance differences between the interfaces are relatively negligible. This conclusion marks a promising beginning of EID validation in the nuclear domain. Furthermore, ecological displays seem to achieve this performance advantage without any workload increments. These results together are particularly encouraging because EID appears to be a design solution for coping with unanticipated events, which have largely been neglected by conventional approaches. This study, therefore, is an important step in the ongoing effort to improve human-system interaction in the nuclear industry.

Acknowledgment

The authors would like to thank C. Nihlwing of IFE for his contributions as the process expert in this study and R. Welch for developing the training program. They would also like to thank J. Kvalem, of IFE, for his effort in making this collaboration possible and A. Teigen, of IFE, for his work in implementing our ecological interface.

References

[1] K. J. Vicente and J. Rasmussen, “Ecological interface design: Theoretical foundations,” *IEEE Trans. Syst., Man, Cybern.*, vol. 22, pp. 589–606, 1992.

[2] C. M. Burns and J. R. Hajdukiewicz*, Ecological Interface Design*. Boca Raton, FL: CRC Press, 2004.

[3] J. Rasmussen and J. K. Vicente, “Coping with human errors through system design: Implications for ecological interface design,” *Int. J. of Human-Computer Studies*, vol. 31, pp. 517–534, 1989.

[4] K. J. Vicente, “Ecological interface design: Progress and challenges,” *Human Factors*, vol. 44, pp. 62–78, Spring 2002.

[5] N. Lau, Ø. Veland, J. Kwok, G. A. Jamieson, C. M. Burns, A. O. Braseth, and R. Welch, “Ecological interface design in the nuclear domain: An application to the secondary subsystems of a boiling water reactor plant simulator,” *IEEE Trans. Nucl. Sci.*, vol. 55, no. 6, pp. XXX–XXX, Dec. 2008.

[6] G. Skraaning, Jr., N. Lau, R. Welch, C. Nihlwing, G. Andresen, L. H. Brevig, Ø. Veland, G. A. Jamieson, C. M. Burns, and J. Kwok, “The ecological interface design experiment (2005),” OECD Halden Reactor Project. Halden, Norway, 2007, HWR-847.

[7] G. Skraaning, Jr., N. Lau, R. Welch, C. Nihlwing, G. Andresen, L. H. Brevig, Ø. Veland, G. A. Jamieson, C. M. Burns, and J. Kwok, “The Ecological Interface Design Experiment (2005),” University of Toronto, Toronto, ON, Canada, Tech. Rep. CEL 07-02, Mar. 2007 [Online]. Available: http://www.mie.utoronto.ca/labs/cel/publications/files/tech_reports/CEL07-02.pdf, [Online]. Available:

[8] C. M. Burns, G. Skraaning, Jr., G. A. Jamieson, N. Lau, J. Kwok, R. Welch, and G. Andresen, “Evaluation of ecological interface design for nuclear process control: Situation awareness effects,” *Human Factors*, vol. 50, pp. 663–679, 2008.

[9] M. R. Endsley, “Toward a theory of situation awareness in dynamic systems,” *Human Factors*, vol. 37, pp. 32–64, 1995.

[10] L. Norros, L. Salo, P. Savioja, and J. Laarni, “The ecological interface design experiment (2005)—The qualitative analysis of operator practices in three interface design conditions,” OECD Halden Reactor Project, Halden, Norway, HWR-889, to be published.

[11] D. V. Reising and P. M. Sanderson, “Minimal instrumentation may compromise failure diagnosis with an ecological interface,” *Human Factors*, vol. 46, pp. 316–333, 2004.

[12] D. V. Reising and P. M. Sanderson, “Ecological interface design for pasteurizer II: A process description of semantic mapping,” *Human Factors*, vol. 44, pp. 222–247, 2002.

[13] O. St-Cyr, “Impact of sensor noise magnitude on emergent features of ecological interface design,” in *Proc. 50th Annual Meeting of the Human Factors and Ergonomics Society*, 2006, pp. 319–323.

[14] A. Garabet and C. M. Burns, “Collaboration with ecological interface design,” in *Proc. 48th Annual Meeting of the Human Factors and Ergonomics Society*, 2004, pp. 543–546.

[15] J. M. O’Hara, J. C. Higgins, J. J. Persensky, P. M. Lewis, and J. P. Bongarra, “Human Factors Engineering Program Review Model,” Washington, D.C., USA, Tech. Rep. NUREG-0711 Rev. 2, Feb. 2004, U.S. Nuclear Regulatory Commission.

[16] J. Rasmussen, “Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models,” *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 257–266, 1983.

[17] G. A. Jamieson, C. A. Miller, W. H. Ho, and K. J. Vicente, “Integrating task- and work domain-based work analyses in ecological interface design: A process control case study,” *IEEE Trans. Syst., Man, Cybern.*, vol. 37, pp. 887–905, 2007.

[18] G. A. Jamieson, “Ecological interface design for petrochemical process control: An empirical assessment,” *IEEE Trans. Syst., Man, Cybern.*, vol. 37, pp. 906–920, 2007.

[19] D. H. Ham and W. C. Yoon, “The effects of presenting functionally abstracted information in fault diagnosis tasks,” *Rel. Eng. Syst. Safety*, vol. 73, pp. 103–119, 2001.

[20] D. H. Ham and W. C. Yoon, “Design of information content and layout for process control based on goal-means domain analysis,” *Cognition, Technology, Work*, vol. 3, pp. 205–223, 2001.

[21] C. M. Burns, “Putting it all together: Improving display integration in ecological displays,” *Human Factors*, vol. 42, pp. 226–241, 2000.

[22] C. M. Burns, “Navigation strategies with ecological displays,” *Int. J. Human-Computer Studies*, vol. 52, pp. 111–129, 2000.

[23] J. M. O’Hara, “A quasi-experimental model of complex human-machine system validation,” *Cognition, Technology, Work*, vol. 1, pp. 37–46, 1999.

[24] T. Karlsson, H. Jokstad, B. D. Meyer, C. Nilhlwing, S. Norrman, E. K. Puska, P. Raussi, and O. Tiihonen, "OECD Halden Reactor Project: The HAMBO BWR Simulator of HAMMLAB," Institutt for Energiteknikk, Halden, Norway, Tech. Rep. HWR-663, Feb. 2001.

[25] F. Øwre, J. Kvalem, T. Karlsson, and C. Nihlwing, "A new integrated BWR supervision and control system," in *Proc. IEEE 7th Conf. Human Factors and Power Plants*, 2002, pp. 4–41–4–47.

[26] J. M. O'Hara and W. S. Brown, "Human-System Interface Design Review Guidelines," Human-System Interface Design Review Guidelines, Washington, D.C., Tech. Rep. NUREG-700, Rev. 2, May 2002.

[27] K. B. Bennett and J. M. Flach, "Graphical displays: Implications for divided attention, focused attention, and problem solving," *Human Factors*, vol. 34, pp. 513–533, 1992.

[28] K. B. Bennett, M. L. Toms, and D. D. Woods, "Emergent features and graphical elements: Designing more effective configural displays," *Human Factors*, vol. 35, pp. 71–97, 1993.

[29] G. Skraaning, Jr., Experimental Control versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments OECD Halden Reactor Project, Halden, Norway, HPR-361, 2003.

[30] G. Skraaning, Jr., "The Operator Performance Assessment System (OPAS)," OECD Halden Reactor Project, Halden, Norway, HWR-538, 1998.

[31] J. Sharit*, Human Error in Handbook of Human Factors and Ergonomics*, G. Salvendy, Ed., 3rd ed. Hoboken, NJ: Wiley, 2006, pp. 708–760.

[32] J. R. Anderson*, Cognitive Psychology and Its Implications*. New York: Freeman, 1985.

[33] P. Ø. Braarud, "Subjective Task Complexity in the Control Room," OECD Halden Reactor Project, Halden, Norway, HWR-621, 2000.

[34] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theorectical research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Amsterdam, The Netherlands: Elsevier, 1988, pp. 139–183.

[35] P. Ø. Braarud and H. Brendryen, "Task Demand, Task Management, and Teamwork," OECD Halden Reactor Project, Halden, Norway, HWR-657, 2001.

[36] R. E. Kirk*, Experimental Design: Procedures for the Behavioral Sciences*, 3rd ed. Pacific Grove, CA: Brooks/Cole, 1995.

[37] B. E. Huitema*, The Analysis of Covariance and Alternatives*. New York: Wiley, 1980.

[38] E. J. Pedhazur and L. P. Schmelkin*, Measurement, Design, and Analysis : An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

[39] D. Howell*, Statistical Methods for Psychology*, 5th ed. Pacific Grove, CA: Duxbury/Thomson Learning, 2002.

[40] F. T. Durso and A. R. Dattel, "SPAM: The Real-Time Assessment of SA," in *A Cognitive Approach to Situation Awareness: Theory and Application*, S. Banbury and S. Tremblay, Eds. Hampshire, U.K.: Ashgate, 2004, pp. 137–154.

[41] K. J. Vicente*, Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Mahwah, NJ: Lawrence Erlbaum Associates, 1999.

[42] R. J. Mumaw, E. M. Roth, K. J. Vicente, and C. M. Burns, "There is more to monitoring a nuclear power plant than meets the eye," *Human Factors*, vol. 42, pp. 36–55, Mar. 22, 2000.

[43] J. Rasmussen, Man-Machine Communication in the Light of Accident Records. Roskilde, Denmark, pp. S-1–S-69, 1969, Danish Atomic Energy Commission, Research Establishment Risø.

[44] J. Reason*, Human Error*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[45] R. Welch, A. O. Braseth, C. Nihlwing, G. Skraaning, Jr., A. Teigen, Ø. Teigen, N. Lau, G. A. Jamieson, C. M. Burns, and J. Kwok, "The 2005 Ecological Interface Design Process and The Resulting Displays," OECD Halden Reactor Project, Halden, Norway, HWR-847, 2007.

[46] D. Cosineau, "Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method," *Tutorials in Quantitative Methods for Psychology*, vol. 1, pp. 42–45, 2007.

[47] G. Cumming and S. Finch, "Inference by eye: Confidence intervals and how to read pictures of data," *American Psychologist*, vol. 60, pp. 170–180, 2005.

[48] G. R. Loftus and M. E. J. Masson, "Using confidence intervals in within-subject designs," *Psychonomic Bulletin and Review*, vol. 1, pp. 476–490, 1994.