U.S.NRC
United States Nuclear Regulatory Commission
*Protecting People and the Environment*

RIL 2022-11

# Human Performance Test Facility (HPTF) Volume 6

## Effects of Automation Level and Task Demands on Workload, Performance, and Trust in Simulated Nuclear Power Plant Operations


Prepared by:
Heather Watkins[1]
John Grasso[1]
Jinchao Lin[2]
Gerry Matthews[3]
Kelly Dickerson[1]
Niav Hughes Green[1]

United States Nuclear Regulatory Commission[1]
11555 Rockville Pike
Rockville, MD 20852

University of Central Florida[2]
3100 Technology Pkwy
Orlando, FL 32833

George Mason University[3]
4400 University Dr
Fairfax, VA 22030

NRC Project Managers:
John Grasso
Heather Watkins
Kelly Dickerson


**Research Information Letter**
**Office of Nuclear Regulatory Research**

# DISCLAIMER

Legally binding regulatory requirements are stated only in laws, NRC regulations, licenses, including technical specifications, or orders; not in Research Information Letters (RILs). A RIL is not regulatory guidance, although NRC's regulatory offices may consider the information in a RIL to determine whether any regulatory actions are warranted.

# PREFACE

## HPTF RIL Series (RIL 2022-11) Preface

Much of the basis for current NRC human factors engineering (HFE) guidance comes from data from research conducted in other domains (e.g., aviation, defense), qualitative data from operational experience in nuclear power plants (NPPs), and a limited amount from empirical studies in a nuclear environment. The Commission, in staff requirements memorandum (SRM) SECY-08-0195, approved the staff's recommendation and directed the staff to consider using generic simulator platforms for addressing human performance issues, as simulators provide a tool to gather more empirical nuclear-specific human performance data. These data would enhance the current information gathering process, thus providing stronger technical bases and guidance to support regulatory decision making.

The former Office of New Reactors (NRO) issued a user need for the Office of Nuclear Regulatory Research (RES) to update its human factors (HF) review guidance with regards to emerging technologies (User Need NRO-2012-007) and more recently the Office of Nuclear Reactor Regulation (NRR) issued a follow-on user need with the same purpose (User Need NRR-2019-008). In the spring of 2012, the NRC sponsored a project to procure a low-cost simulator to empirically measure and study human performance aspects of control room operations to address the human performance concerns related to current as well as new and advanced control room designs and operations. Using this simulator, the Human Factors and Reliability Branch (HFRB) in the RES Division of Risk Assessment (DRA) began a program of research known as the NRC Human Performance Test Facility (HPTF) to collect empirical human performance data with the purpose of measuring and ultimately better understanding the various cognitive and physical elements that support safe control room operation. Additionally, the baseline methodology documented in these volumes will enable HRA data research that will address key gaps in available data for topics such as dependency and errors of commission, improving the state-of-the-art of Human Reliability Analysis (HRA) and thus dual HF and HRA data missions.

Recognizing the essential role of data to our HF and HRA programs, the NRC historically approached data collection through multiple avenues – all with their inherent strengths and weaknesses:

1. Licensed Operators – controlled experiments at the Halden Reactor Project
2. Licensed Operators – the Scenario Authoring, Characterization, and Debriefing Application (SACADA) database capturing training scenarios
3. Novice populations – scientific literature, laboratory settings – non-nuclear

The HPTF program captures data from both novice and operational populations and the work is specifically targeted to the nuclear domain. In addition, the HPTF methodology expands upon these data collection methods by also including formerly licensed operators and other individuals with nuclear domain expertise (e.g., former plant designers, PRA experts, NRC staff). The HPTF methodology (described in detail in RIL 2022-11 Volume 1) enables the NRC to fill in the gaps from the other 3 data collection activities and conduct responsive research to support the informational needs of our users (e.g., NRR HFE technical reviewers and HRA analysts).

The intent of the HPTF was to design experiments that balanced domain realism and laboratory control sufficiently to collect systematic, meaningful, human performance data related to 3-2 execution of common nuclear main control room (MCR) tasks. Three large-scale experiments were conducted to address challenges associated with developing a research methodology for using novices in a highly complex, expert driven domain. These three experiments are reported as Studies 1 and 2 in RIL 2022-11 Volume 1, which describes the approach and methodology underlying this research effort and the resulting findings for the series of studies. In RIL 2022-11 Volume 2, the Volume 1 findings were further validated via a fourth data collection by testing a formerly licensed operator population using a full-scale, full-scope simulator. Cross-experiment comparisons were enabled by leveraging a formerly licensed operator as a member of the research team to serve as senior reactor operator (SRO) and ensure participants received experience as similar and structured as possible to the studies in Volume 1[1].

The HPTF team works with the technical staff in the user offices to ensure pertinent research questions can be addressed within the constraints of the HPTF methodology. HPTF research questions are formulated collaboratively between NRC staff and a contractor with an identical simulator and performance assessment capabilities. Three experimental design workshops have been held to date. The first workshop was held on March 5 and 6, 2018, upon completion of the first three HPTF experiments. The direction resulting from this first workshop was to validate the methodology and generalize the findings from the baseline HPTF experiments by using formerly licensed operators as participants to complete an experimental scenario using an analog, full-scope, full-scale simulator and a digital, part-task simulator. RIL 2022-11 Volume 2 describes the research approach and findings for the fourth experiment in the series.

The second workshop was held on August 20 and 21, 2019. The direction resulting from this second workshop was to perform a reanalysis of all HPTF experiments thus far to investigate: 1) Workload Measure Sensitivities (RIL 2022-11, Volume 3) 2) Task Order Effects and 3) Touchscreen Ergonomics (forthcoming RIL 2022-11 Volume 5). Due to the COVID-19 health crisis, the third workshop was held as a virtual series consisting of six 2-hour blocks between October 29 to November 20, 2020. The future direction topics discussed during the most recent workshop are described in RIL 2022-11 Volume 6 (this volume).

These volumes of research illustrate the NRC's ongoing effort to perform systematic human performance data collection using a simulator to better inform NRC guidance and technical bases in response to SRM SECY-08-0195 and SRM-M061020. The HF and HRA data are essential to ensure that our HFE guidance documents and HRA methods support the review and evaluation of "state-of-the-art" HF programs (as required by 10 Code of Federal Regulations (CFR) 50.34(f)(2)(iii)).

1 Systematic experimentation is challenging in the nuclear domain using real operators and full, dynamic scenarios because operators can take many paths to achieve a successful outcome. This variability represents a condition that is not conducive to controlled laboratory study. By including a confederate SRO in the study using a dynamic scenario, this hard to control variability is managed, thereby, enabling stable observations. See RIL 2022-11 Volumes 1 and 2 for examples of these methodological benefits.

# ABSTRACT

As nuclear power plants modernize their control rooms with advanced digital systems and higher levels of automation, understanding how these changes affect operator workload, performance, and trust is essential for maintaining safe and effective operations. This Research Information Letter describes two simulator-based studies conducted at the Human Performance Test Facility to examine the effects of level of automation (LOA) and task demands on operators performing simulated emergency operating procedures in a pressurized water reactor control room environment. Participants, functioning as reactor operators within a three-person crew, completed tasks under two intermediate LOA conditions (i.e., management-by-consent and management-by-exception) while researchers measured subjective workload, trust, performance, and a range of physiological indicators.

The results show that higher LOA does not consistently reduce workload. In some scenarios, management-by-exception increased temporal demand and subjective workload, particularly for less complex tasks. Detection tasks remained the most cognitively demanding at lower LOA settings. Performance benefits from higher LOA were observed in certain contexts, including faster response times on primary and secondary tasks, but these benefits were not uniform across all task types. Trust ratings did not differ significantly between LOAs overall. However, dispositional trust, which was measured prior to task performance, predicted how participants interacted with and evaluated automation. Physiological measures, including electroencephalogram, electrocardiogram, and other workload-sensitive measures, captured effects of LOA and task type that were not always detected through subjective ratings, underscoring the value of multivariate workload assessment. The findings also suggest that individual trust profiles, such as high “perfect automation” expectations, may influence how workload and performance are affected by automation, providing a potential avenue for targeted training or interface design adjustments.

These studies contribute to a more comprehensive understanding of how intermediate LOAs function in complex operational settings. The results highlight that increasing automation is not inherently beneficial and that its effects on workload, vigilance, and operator engagement depend on both task demands and individual operator characteristics. Insights from this work can inform our understanding of the varied impact of control room design, review practices for human factors engineering reviewers, and potential for targeted operator training programs to ensure that automation supports, rather than undermines, human performance in safety-critical nuclear power operations.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# EXECUTIVE SUMMARY

This Research Information Letter (RIL) summarizes two experimental studies conducted at the Human Performance Test Facility (HPTF) to evaluate how different levels of automation (LOA) and task demands affect workload, performance, and trust during simulated nuclear power plant emergency operations. This work supports license reviews by advancing our understanding of validated methods for workload assessment under a range of automation-supported concepts of operations.

Across both studies, findings demonstrated that higher LOA (management-by-exception) did not consistently lower workload. In some cases, higher LOA increased workload, specifically temporal demand and subjective measures of workload. These workload effects were particularly noticeable in the less complex tasks. HPTF studies use several common rule- and skill-based tasks, including detection, parameter confirmation (referred to as checking), and control use to execute a response (response implementation). Among the three task types, detection produced the greatest amount of cognitive demand in the lower LOA conditions. Management-by-exception produced faster primary and secondary task responses in certain contexts, although these benefits were not observed across all task types.

Trust in automation did not significantly differ between LOA conditions. However, individual differences in dispositional trust influenced how participants interacted with automation. Participants with expectations of “perfect automation” experienced higher workload and greater performance variability. These findings suggest opportunities for targeted training or interface design to better calibrate operator trust in automation.

Physiological measures such as electroencephalography and electrocardiography were sensitive to LOA and task effects not always captured by subjective assessments. This finding underscores the value of multivariate workload evaluation. These objective indicators may complement traditional tools like the NASA Task Load Index in providing a more comprehensive understanding of operator workload.

Overall, the findings underscore the need for careful human factors engineering review of proposed automation configurations. Intermediate LOAs can alter workload and vigilance patterns in ways that are not always beneficial. It may be beneficial to consider physiological and secondary task measures in reviews of complex digital control systems to provide a more accurate and comprehensive understanding of operator workload. Additionally, efforts to characterize individual trust profiles may support focused operator training strategies aimed at preventing complacency or under-reliance on automation.

# SECTION 1: INTRODUCTION

## 1.1 Automation and Digitalization in Nuclear Power Operations

As nuclear power plant (NPP) reactor technology modernizes and evolves, main control rooms (MCRs) must follow suit. However, not all technology changes are upgrades, as some changes may introduce new human factors challenges (Joe, Boring, & Persensky, 2012). The MCR is where all the Instrumentation and controls (I&C) are housed that control the reactor and associated safety systems. MCR controls are defined in 10 CFR 50.54 as any apparatus that directly affects the reactivity or power output levels of the nuclear reactor (NRC, 1998). Licensed personnel are required to continually staff the MCR when the reactor is in any operational mode other than refueling or shutdown. Per-shift, on-site staffing of a MCR crew is dependent on the number of reactor units and control rooms at the NPP site and is defined in 10 CFR 55.54(2)(i). As such, there are multiple configurations and levels of capability that warrant description. In this study, we addressed the impacts of digitalization and automation on MCR crew member workload.

The task of operators in the MCR of an NPP is complex and potentially demanding. Depending on plant design, a typical MCR may have over 8000 displays and controls, including over 1000 annunciator tiles, 700 analogue and digital indicators and trend recorders, and over 2000 switches (Anokhin et al., 2010). Operators are thus vulnerable to elevated cognitive workload, which, in turn, increases risk of human error. Adverse impacts of workload may also be exacerbated by the quantitative and qualitative variability of sources of demand as operator activities shift; for example, vigilant monitoring of displays tends to be especially taxing (Reinerman-Jones, Matthews & Mercado, 2016). The complexity of NPP operations can impose a cognitive burden on operators during emergencies, when issues need to be diagnosed very rapidly (Lin, Yenn, Jou, Hsieh, & Yang, 2013). Operator perception of being overloaded also drives stress responses, which may further elevate the likelihood of human error (Matthews & Reinerman-Jones, 2017).

Workload plays a central role in human reliability and, therefore, the assessment approach is an important element of the safety review performed by the NRC's Human Factors Engineering (HFE) staff. The HFE methodology for reviewing license applications is defined in NUREG-0711 (O'Hara, Higgins, Fleger & Pieringer, 2012). One of the high-level goals for the review is to ensure that "the design will support personnel in maintaining vigilance over plant operations and provide acceptable workload levels, i.e., minimize periods of under- and over-load" (NUREG-0711, p. 11). This goal is accomplished through identifying the alarms, displays, controls and task support needs required for performing personnel tasks and evaluating the acceptability of workload for each task as assigned to staff positions or roles within the control room (e.g., reactor operator, balance of plant). Workload assessment contributes to prospective Human Reliability Analysis (HRA) (NUREG/CR-1278, NRC, 1983) through identifying aspects of operator tasking that potentially impose excessive cognitive load and raise error probabilities and threats to safety.

Digital technologies have, for some years, been incorporated into existing plants across the world, typically to enhance I&C (International Atomic Energy Agency, 2009). Guidance on the potential impact of computerized input devices and displays on operator performance is provided to NRC reviewers in NUREG-0700, Rev. 3 (O'Hara & Fleger, 2020). As existing plants look to extend their operational life, they also must consider shifting some analog systems to digital, creating hybrid control rooms. New builds (either large light water, small modular reactors, or advanced reactors) will be either primarily or entirely digital (Porthin, Liinasuo & Kling, 2020).

The Department of Energy's Light Water Reactor Sustainability program, Joe and Kovesdi (2018) list several enabling digital technologies for modernized or newly built control rooms.

- Desktop-based operator workstations
- Operator-at-the-boards controls
- Soft controls
- Large overview displays
- Intelligent alarm systems
- Computer-based procedures (CBPs)
- Wireless technologies
- Advanced diagnostic tools to aid in detection, diagnosis, and troubleshooting by integrating additional sensors and field information with existing control room indications
- Automation

These enabling digital technologies present both opportunities and challenges for workload management. Digital systems are typically designed to lower operator workload by reducing the perceptual, cognitive and response demands of tasks (Lee & See, 2004). According to Porthin et al. (2020), digital systems have the capability to include a greater depth of information than analog displays, which can help support operators. Additional functionality of digital systems includes trend analyses, place-keeping functions when several paths are in process simultaneously, and constraints to prevent certain operator errors. Workload mitigation is typically a primary goal for system automation. Some research in the nuclear domain has demonstrated expected workload benefits. For example, Jou, Yenn, Lin, Yang and Lin (2009) investigated a reactor shutdown task and found that the introduction of automation reduced operator workload and human errors.

However, digitalization can also have unanticipated adverse effects. For example, operators may find it challenging to understand some forms of additional information such as the temporal sequence of events and causes of alarm states (Lin, Yenn & Yang, 2010; Liu & Li, 2016; Porthin et al., 2020), and they can be difficult to represent digitally. When events are unanticipated, the operator may have to search within the interface for additional information on plant system status (Medema, Savchenko, Boring & Ulrich, 2019). While helpful, the wealth of information available from digital displays can impair operator awareness and the ability to predict future plant states (Liu & Li, 2016). Digital interfaces may also impose additional navigation and interface management tasks that burden and distract the operator (Bye, 2023; O'Hara et al., 2002). For example, plant system status information may be buried within a hierarchical menu system instead of persistently visible on the hard panels.

In addition to the potential adverse effects of digitization, system automation can introduce comparable challenges for operators. One such challenge is the potential degradation of awareness, as automation often necessitates increased monitoring, thereby imposing additional cognitive demands (Medema et al., 2019). Rather than reducing workload, the need to manage and supervise automation may increase it, particularly in unanticipated operational states that require operators to interpret or troubleshoot automated system behavior (Skjerve & Skraaning, 2004). Furthermore, there is a risk that operators may adopt inaccurate mental models, or "frames," for understanding automation. Nystad, Kaarstad, and McDonald (2019) found that under high workload conditions, operators may be prone to diagnostic errors due to reliance on inappropriate frames, particularly when failing to consider multiple hypotheses for anomalous instrument readings.

## 1.2 Assessment of Workload in HPTF Research

The Human Performance Test Facility (HPTF) has two facilities, one at the NRC Headquarters in Rockville MD and the other at the Institute of Simulation and Training, University of Central Florida (UCF). The objective of the HPTF is to conduct research to assess the impact of new and existing MCR designs on human performance using human-in-the-loop experimental techniques. The intent of the HPTF was to design experiments that balanced domain realism and laboratory control sufficiently to collect systematic and meaningful human performance data related to execution of common MCR tasks. This includes supporting the NRC's mission by advancing, validating, and documenting workload assessment methodology for NPP MCR operations using a generic plant simulator (Hughes, D'Agostino & Reinerman-Jones, 2017). Full details have been documented previously (see RIL-2022-11, Volume 1 and 2 for full methodological details). It utilizes a GSE Generic Pressurized Water Reactor (GPWR) simulator that can be configured to provide experimental control over the task elements performed by operators. The simulator is also configurable, such that it can be used with novice populations or trained operators. Early HPTF work demonstrated that novices can perform and show workload levels similar to operators for skill- and rule-based tasks (RIL 2022-11, Volumes 1 and 2). Key steps in the methodology design process included specification of checking, detection, and monitoring task elements that could be performed effectively by novices and the scaled (43%) reduction in the number of controls on the panels, which equalized complexity between novices and operators.

Previous HPTF studies have focused on multivariate assessment of workload and performance under different task configurations (Hughes et al., 2023a, 2023b). This approach is consistent with the evidence that different workload measures can either converge or diverge, providing a means to measure neurocognitive change with variation in task demands (Matthews et al., 2015; Matthews & Reinerman-Jones, 2017). Studies have assessed subjective workload dimensions and stress together with objective central nervous system, autonomic, and hemodynamic measures and performance. These studies investigated issues such as workload differences between task types, finding that detection imposes the highest workload and vulnerability to vigilance decrement (Reinerman-Jones, Matthews & Mercado, 2016). They also tested for differences between desktop monitor and touchscreen displays, contrasting two approaches to digitalization of I&C (Reinerman-Jones, Harris, Hughes & D'Agostino, 2017). A further study (Reinerman-Jones et al., 2019) utilized a similar method with a sample of former NPP operators, finding workload responses comparable to those seen in novice participants, with some differences in the details (Lin et al., 2022).

To assess workload, HPTF studies have employed both subjective and objective measures. The NASA Task Load Index (NASA-TLX: Hart & Staveland, 1988) is used by many licensees as a workload assessment tool, however, it has several limitations (de Winter, 2014). It relies on subjective responses that may be subject to various biases, and it must be administered post-task. It also assumes a unitary view of attentional capacity or resources, whereas performance research suggests that humans access multiple resource pools for cognitive task performance (Wickens, 2008). Licensees must balance many practical considerations during HFE validation events and therefore may rely on only one well-established workload measure, like the NASA-TLX. The HPTF does not have to balance these same practical considerations and thus has an opportunity to use methods that complement the NASA-TLX. The Instantaneous Self-Assessment (ISA: Tattersall & Foord, 2008) probes workload during the task itself, providing a concurrent assessment, though it cannot track workload change over short time durations. The Multiple Resources Questionnaire (MRQ: Boles, Bursk, Phillips & Perdelwitz, 2007) assesses demands on 17 different resources, based on an extension of Wickens' (2008) resource theory.

HTPF research has also utilized physiological measures for workload (Hughes et al., 2023c; Matthews et al., 2015). These measures provide continuous assessment of workload throughout the duration of task performance. Sensors included the electrocardiogram (ECG) from which heart rate variability (HRV) was calculated as a workload measure. Central nervous system activity was recorded as the electroencephalogram (EEG). Higher cognitive activity tends to be associated with greater spectral power in frontal theta, as well as increased beta and decreased alpha power. Two hemodynamic indices of workload reflected brain metabolic activity were obtained bilaterally from left and right hemisphere sensors. Functional near infrared spectroscopy (fNIRS) indexes oxygen saturation in pre-frontal areas.

To date, this line of research has shown a reasonable degree of convergence between the physiological measures and NASA-TLX in their sensitivity to the differing task demands imposed by checking, detection, and response implementation tasks, with some variation across measures and studies (Hughes et al., 2023c). Along with the NASA-TLX, brain oxygenation measured with fNIRS provided the most consistent discrimination between tasks. The physiological data also provided sensitivity to task differences beyond that provided by NASA-TLX. EEG and ECG data registered higher workload on response implementation relative to checking in novices using a touchscreen. This task effect was not evident in the subjective data.

Another potential area of interest is using secondary task methodology for evaluating automation impacts. Performance-based workload assessment uses a secondary task to assess "spare capacity" not allocated to primary tasks (Eggemeier, Wilson, Kramer & Damos, 1991). The assumption is that increasing primary task workload will lead to deterioration in secondary task performance, such as slowing responses to probe stimuli. This approach is popular in some human factor domains, especially vehicle driving (De Waard & Brookhuis, 1996). Kaber and Endsley (2004) used secondary task performance as a metric to demonstrate that automation reduced workload. However, this method has not been widely applied in NPP research. Gao et al. (2013) state that it may be difficult to design secondary tasks where demands match those of operators' primary tasking. One relevant study was conducted by Hwang et al. (2008), who had operators perform a simulated reactor shut down task. The secondary task involved mental calculation. Findings showed that errors on the secondary task were correlated with workload measured by NASA-TLX. Similarly, researchers argued for the validity of performance in an arithmetic secondary task as an index of workload on alarm monitoring and identification in NPP operations (Wu & Li, 2013). While secondary task methodology has not previously been applied to HPTF research, it appears promising for evaluating automation impacts.

## 1.3 Levels of Automation and Computer Based Procedures

Design of Human Systems Interfaces (HSIs) that support automation of plant operations is critical for technological advancement and full nuclear plant modernization (Joe & Kovesdi, 2018; O'Hara, Higgins & Brown, 2009). Automation is also critical for novel plant design concepts including Small Modular Reactors (SMRs) and microreactors (Hildebrandt et al., 2023). By automating some or all of the tasks that could be performed by automation, additional operator capacity becomes available enabling multi-unit concepts of operation. According to O'Hara et al. (2009, p. 231), "…alarm systems automate monitoring and detection, computerized operator support systems automate aspects of situation assessment, computer-based procedures automate aspects of response planning, and high-level control and process automation provide automatic response implementation." NUREG 0700, Revision 3 (O'Hara & Fleger, 2020) includes sections on (1) evaluating displays for monitoring automation, (2) alerts, notifications, and status indications, and (3) interacting with and controlling automation.

Automated systems vary in ways that are critical to human-systems integration (HSI). For example, as outlined in NUREG-0700, Revision 3 (O'Hara & Fleger, 2020), systems may operate in distinct modes, such as shutdown, refueling, startup/hot standby, and run, each of which is associated with specific operational characteristics in boiling-water reactors. Distinct from modes are levels of automation (LOAs), which define different allocations of function and decision authority between the human and the machine. In human factors, LOAs are typically described using a numerical scale, with scale points ranging from full human control to full machine control. For example, Parasuraman, Sheridan and Wickens (2000) defined a 10-point scale and described its application to understanding automation of four elements of information-processing: information acquisition, information analysis, decision and action selection, and action implementation. NUREG 0700, Revision 3, defines five LOAs for NPP applications, illustrated in Table 1 (O'Hara & Fleger, 2020).

**Table 1** *Levels of Automation for NPP Applications (O'Hara & Fleger, 2020)*

| Level | Automation Tasks | Human Tasks |
|---|---|---|
| (1) Manual Operation | No automation | Operators manually perform all tasks |
| (2) Shared Operation | Automatic performance of some tasks | Operators perform some tasks manually |
| (3) Operation by Consent | Automatic performance when directed by operators to do so, under close monitoring and supervision | Operators monitor closely, approve actions, and may intervene to provide supervisory commands that automation follows |
| (4) Operation by Exception | Essentially autonomous operation unless specific situations or circumstances are encountered | Operators must approve of critical decisions and may intervene |
| (5) Autonomous Operation | Fully autonomous operation. System cannot normally be disabled but may be started manually | Operators monitor performance and perform back up, if necessary, feasible, and permitted |

Various additional LOA models have been proposed, and debate over the utility of models in system design is ongoing (Kaber, 2018). However, defining and choosing LOAs remains essential in determining how to distribute or allocate tasks between operators and automation, setting workload levels to avoid under- or over-load, promoting appropriate levels of trust in automation, and avoiding incorrect use and human errors caused by complacency or human bias (Janssen, Donker, Brumby, & Kun, 2019). LOA models may also have to accommodate for adaptive automation, where LOA varies according to operator demand (Kaber, Riley, Tan & Endsley, 2001).

There is extensive research literature on automation and workload management, in a variety of human factors domains (Parasuraman & Manzey, 2010; Hancock 2023; Thomson et al. 2024). The design intent is that introducing automation reduces operator workload, which in turn reduces probabilities of operator error and enhances situation awareness (SA: Endsley, 1996). These benefits have indeed been demonstrated in a variety of contexts (Balfe, Sharples & Wilson, 2015; Endsley, 1996).

However, re-allocating functions from humans to machines is far from a panacea for workload issues, and human performance challenges at higher LOAs have emerged. A basic issue is that at high LOAs the operator's role changes from active control to supervisory monitoring of the automation (Sheridan, 1988). For instance, at Level 4 automation, the operator is placed in the role of monitoring the automation's decisions in specific situations, as described in Table 1 above. This role as overseer requires continuous sustained attention, leading to the issues associated with vigilance: performance decrement, boredom and stress, and paradoxically elevated workload (Reinerman et al., 2016; Warm, Parasuraman & Matthews, 2008). Thus, intermediate LOAs may be preferable to high LOAs because they maintain operator engagement with the task and mitigate vigilance and “out of the loop” effects. Using an intermediate level of workload likely also has situational assessment (SA) benefits. Kaber, Onal and Endsley (2000) found that there was a greater loss of SA when automation failed in higher automation level conditions than lower levels of automation. Closely related to these concerns is the issue of automation complacency, in which operators may become overly reliant on automated systems and fail to adequately monitor for potential failures (Parasuraman & Manzey, 2010), particularly under multitasking conditions. The introduction of automation can also lead to performance issues such as skill degradation, resulting from increased reliance on automated systems and reduced opportunities for practice. Additionally, it may give rise to novel and less understood error types, including mode errors (Kim & Park, 2018).

Studies in the nuclear domain have shown that automation can have mixed impacts on workload and performance, depending on the context and implementation. In a study using a simulated reactor microworld, Hall et al. (2023) showed that use of level 2 and 3 CBPs enhanced performance and reduced subjective workload in some cases when compared to level 1 CBPs. However, CBPs and similar forms of automation can increase workload due to additional interface management demands, system complexity, need to track operational modes, and challenge of correctly diagnosing automation errors (Kim & Park, 2018). These findings underscore the complex and sometimes contradictory nature of automation’s impact. For instance, Huang et al. (2006) found that automatic alarm reset was faster and required less effort than manual reset, yet experienced operators still preferred manual control. Similarly, Jou et al. (2009) reported that automation reduced workload during a reactor shutdown task but not during an alarm reset task. Qing et al. (2021) found that while greater automation of procedures reduced workload compared to paper-based procedures, it also diminished SA, consistent with findings in other domains (Kaber et al., 2000).

Adverse impacts of automation can potentially be mitigated through increasing information transparency to give operators insight into how the automation is functioning. Skraaning, Eitrheim and Lau (2010) compared impacts of transparent and non-transparent interfaces for interaction with automation in an NPP control room. The transparent interface enhanced operator SA and trust in the automation but had no overall performance impact. In fact, transparency impaired operator’s ability to detect disturbance, possibly because the interface overloaded their capacity to simultaneously monitor automation and detect deviations in system states. Overall, these various studies provide only a limited picture of automation impacts, but they do highlight the need for further research to evaluate benefits and costs of automation across different automation configurations and operational scenarios.

### 1.4 Measurement of Trust in Automation

There is a large human factors literature on trust in automation (Lee & See, 2004). Operators should match their trust in the system to its actual reliability, or there are risks of misuse, disuse, and abuse of automation (Parasuraman & Riley, 1997). Trust calibration and optimization is a major issue for design and evaluation of automated systems; both over- and under-trust are potentially harmful in operational settings (Lee & See, 2004; Parasuraman & Manzey, 2010). Much of the empirical work in this area addresses trust in systems of moderate or even low reliability, which are unlikely to be found in the nuclear domain (however, see Bye, 2003; Skjerve & Skraaning, 2004 for limited nuclear examples). Skjerve and Skraaning documented concerns about operators’ reliability in evaluating the trustworthiness of automated systems. Additionally, Tasset, Charron, Miberg and Hollnagel (1999) found that operator trust was inversely related to the efficiency of the automation. Given the limited number of studies conducted thus far, there is a need for more research on automation impacts.

There are both trait and state measures of trust reported in the literature (Kohn et al., 2021). Trait measures assess the person's general willingness or disposition to trust automated systems, whereas state measures are administered following task performance to assess level of trust during a single interaction with the automation. Operators with high trust in automation might be over-reliant on the automation and at risk of complacency and failure to monitor and check the automation. Conversely, low trust operators would be inclined to neglect the automation and perform tasks manually, leading to unnecessary (and unplanned) workload.

Trust is a multifaceted construct, and various scales have been developed that show only limited convergence (Matthews et al., 2021). Two trait trust scales that appear to be well-suited to the nuclear domain are the Human Interaction with Technology (HIT; Lyons & Guznov, 2019) and Perfect Automation Schema (PAS: Merritt, Unnerstall, Lee & Huber, 2015) scales. The HIT is based on a leading theory of trust. The HIT can distinguish between several key components of trust, including performance, benevolence, and integrity (Mayer, Schoorman & Davis, 1995). It assesses intentions to rely on the automation and so it is expected to be predictive of reliance behavior. The perfect automation schema assesses the presence of an all-or-nothing attitude that automation either works or it does not (Merritt et al., 2015). Automated systems in NPPs are anticipated to be highly reliable. Individuals with high PAS scores may then be at risk of assuming that the system is perfectly reliable, leading to possible complacency. THE PAS questionnaire (Merritt et al., 2015) also has a scale for high expectations and a scale that measures overall trust. In addition to these trait measures, various scales have been used to measure immediate states of trust (Kohn et al., 2021). The Checklist of Trust between People and Automation (CTPA; Jian, Bisantz, & Drury 2000) has been validated for assessment of trust in process control systems and appears to be well-suited to the nuclear domain (Sauer, Chavaillaz & Wastell, 2016).

## 1.5 Current Studies

This Research Information Letter summarizes two studies that examined how varying LOAs and task demands affect workload, performance, and trust in nuclear power plant control room operations. HPTF researchers at the University of Central Florida collected and analyzed the data for these studies. Both studies used simulated emergency operating procedures (EOPs) within a generic pressurized water reactor environment, enabling controlled manipulation of automation and task variables while capturing performance, subjective, and physiological measures.

For both studies (i.e., Experiment 6 and Experiment 7), each experimental session involved a three-person crew consisting of one experimental participant and two trained confederates. The participant served in the Reactor Operator 1 (RO1) role, while the Senior Reactor Operator (SRO) and Reactor Operator 2 (RO2) roles were performed by members of the research team following scripted procedures. Within this crew context, the participant was responsible for executing procedure steps and interacting with automation, while the overall crew worked through an EOP designed to bring the plant to a safe state during an emergency. Performance measures focused on the participant's activities, including communication exchanges, navigation and identification of controls, and accuracy in completing action steps.

Experiment 6 focused on the effects of LOA and task type in a relatively constrained setting. Participants performed three types of tasks that are representative of steps in EOPs: checking, which required locating and verifying the state of a parameter or control; detection, which required monitoring instruments to identify abnormal values; and response implementation, which required manipulating a control to achieve a prescribed change in system status. To probe workload more fully, a secondary monitoring task was included, requiring participants to acknowledge gauge changes displayed on the interface throughout the scenario. These primary and secondary tasks were embedded within a modified “Loss of All AC Power” scenario, allowing systematic comparison across automation conditions. Two intermediate LOAs were tested (management-by-consent, management-by-exception) to assess how the degree of operator involvement in procedure execution influenced task performance. Experiment 6 provided an initial test of how LOA and task type interact to shape workload, vigilance, and trust, with particular emphasis on comparing subjective ratings and physiological indicators of workload.

Experiment 7 expanded on this approach by introducing more complex operational demands, specifically a power correction procedure representative of load-following tasks. This scenario required participants to balance turbine output with grid demand while responding to time-sensitive changes in plant conditions. To further stress operator capacity, a workload manipulation was introduced, allowing assessment of how increasing cognitive demands affect operator performance and reliance on automation. Measures were largely consistent with those in Experiment 6 with some variances noted in Section 2: General Methodology. The design allowed researchers to examine how individual differences in dispositional trust moderated workload and performance outcomes, providing insight into why operators may vary in their responses to automation in complex environments.

Together, these studies provide a basis for evaluating how automation level and task complexity influence operator workload, performance, and trust. The following section outlines the general methodology used to support these investigations.

# SECTION 2: GENERAL METHODOLOGY

The methodologies for these two studies were developed to systematically examine the relationship between automation level, task demands, workload, performance, and trust in simulated nuclear power plant operations. Both studies were conducted using a full scope, generic pressurized water reactor simulator, configured to support experimental control of emergency operating procedures. Each study incorporated behavioral performance data, subjective workload and trust assessments, and physiological measures. Building on previous HPTF research, the studies employed a multivariate approach to workload assessment. This approach aligns with evidence that different workload measures can either converge or diverge, offering a more comprehensive view of neurocognitive changes that occur in response to varying task demands (Matthews et al., 2015; Matthews & Reinerman-Jones, 2017).

In Experiment 6, the primary focus was on intermediate LOAs and task type in a controlled environment. Participants completed EOP tasks involving checking, detection, and response implementation while operating under management-by-consent or management-by-exception conditions. Workload was assessed through subjective ratings (NASA-TLX, MRQ, ISA), physiological indices (EEG, ECG, fNIRS), and a secondary monitoring task. Trust in automation was measured using validated scales administered before and after the scenarios.

Experiment 7 extended this methodology by introducing a more complex operational scenario: a power correction procedure representative of the actions a balance of plant operator might perform during load-following. In addition to manipulating LOA, this study incorporated a workload manipulation to evaluate how increasing cognitive demands influence operator performance and trust in automation. The measures used were largely consistent with those in Experiment 6, allowing direct comparison across studies.

Both studies shared a common methodological framework, including use of the HPTF simulator, standardized training protocols, and multivariate workload assessment. Key differences included the complexity of the task environment, the inclusion of explicit workload manipulations in Experiment 7, and the scope of physiological instrumentation. This shared framework ensured consistency while providing a robust foundation for evaluating the impact of automation and task demands on human performance in safety-critical nuclear operations.

The following subsections describe the core methodological components that were common to both studies, beginning with the simulation environment used to replicate nuclear power plant operations.

## 2.1 Simulator

The GSE Generic Pressurized Water Reactor (GPWR) NPP simulator (Figure 1) was modified to provide the interface of the I&Cs in a main control room and support performance data recording. The GSE GPWR is a digital full scope simulator that has the capability to simulate all the physical and underlying thermodynamics occurring in the would-be plant. The Experimental Platform for Instrumentation and Controls (EPIC) was developed to modify the GPWR to a part-task simulator dedicated to the experimental scenario. EPIC removes the physics models and underlying thermodynamics so that the simulation provides a repeatable experience between participants on the two panels with digital representation of emulating analog I&C. Each panel consisted of four 27-inch touchscreen monitors (2560 × 1440 pixels in resolution) arranged two high by two wide. In addition, an in-house developed computerized procedure system was displayed on a separate monitor providing participants with instructions and prompts regarding the ongoing tasks.

**Figure 1 GSE GPWR NPP Simulator, Set up for Two Operators - Two, 2 x 4 Screen Arrays**

## 2.2 Physiological Measures

A suite of three physiological instruments (Advanced Brain Monitoring's B-Alert X10, Spencer Technologies' ST3 Digital Transcranial Doppler, and Somantics' Invos Cerebral/Somatic Oximeter) was used to monitor workload states during experimental sessions. Details about the specific signals these systems monitored are explained below.

### 2.2.1 Electroencephalogram (EEG)

The Advanced Brain Monitoring B-Alert X10 system was used to assess nine-channels of EEG and one channel of ECG (Figure 2). Following the international standard 10-20 System, the sampling rate of 256 Hz captures signals from Fz, F3, F4, Cz, C3, C4, Pz, P3, and P4. Reference electrodes were placed on each participant's mastoid bone. The Power Spectral Density (PSD) analysis techniques were used to analyze activity in three standard bandwidths: theta (4-8 Hz), alpha (9-13 Hz), and beta (14-30 Hz; Wilson, 2002). Each bandwidth was collected at the nine electrode sites. They are combined to compare left and right hemispheres and the front, temporal, and parietal lobes.


**Figure 2 ABM's X10 EEG/ECG System**

### 2.2.2 Electrocardiogram (ECG)

The Advanced Brain Monitoring System B-Alert X10 system was used to monitor the ECG, sampling at 256 Hz. Single-lead electrodes were placed on the center of the right clavicle and one on the lowest left rib (Figure 3). Heart Rate (HR) was computed using peak cardiac activity to measure the interval from each beat per second. "So and Chan" QRS detection methods were used to calculate Interbeat Interval (IBI) and Heart Rate Variability (HRV: Taylor, Reinerman-Jones, Cosenzo, & Nicholson, 2010). This approach maximizes the amplitude of the R-wave (Henelius, Hirvonen, Holm, Korpela, & Muller, 2009).

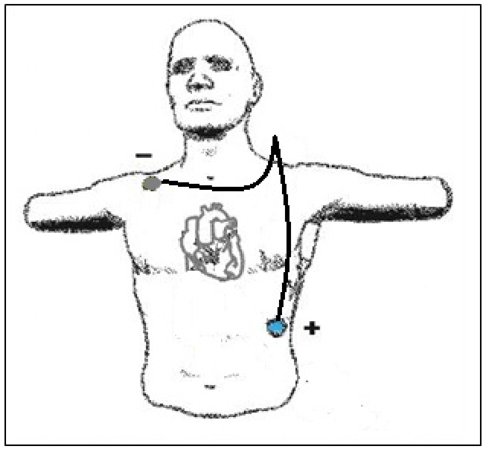


**Figure 3 Electrode Locations for the ECG System**

### 2.2.3 Functional Near Infrared Imaging (fNIRS)

The Somantics' Invos Cerebral/Somatic Oximeter, Model 5100C (Figure 4), was used to monitor (hemodynamic) changes in oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-HB) in the left and right hemisphere prefrontal cortex (Ayaz et al., 2011; Chance, Zhuang, UnAh, Alter, & Lipton, 1993).


**Figure 4 fNIRS Screen**

## 2.3 <u>Subjective Measures</u>

Subjective measures were used to assess perceptions of workload.

### 2.3.1 NASA-Task Load Index (NASA-TLX)

The NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988) is a widely used multi-dimensional measurement of subjective workload. In this study, it was used to measure the perceived workload at the end of each condition (management by exception, management by consent) or after the entire experimental scenario, depending on the simulator condition. It consists of six rating scales for workload-relevant factors, including mental demand, physical demand, temporal demand, performance, effort, and frustration. All factors, except performance, are rated on a 0 - 100 scale from "Low" to "High". Performance is rated on a 0 - 100 scale from "Good" to "Poor".

### 2.3.2 Multiple Resource Questionnaire (MRQ)

The MRQ was used to characterize the nature of the mental processes used during a task (Boles & Adair, 2001). The items on the questionnaire were derived from factor analytic studies of lateralized processes (Boles, 1991, 1992, 1996, 2002). Participants received a copy of the scale with definitions and completed the MRQ at the end of each task type or the scenario depending on the simulator condition using a computerized version of the questionnaire. The following 14 of 17 scales were included for the present study:

- auditory emotional process
- auditory linguistic process
- manual process
- short-term memory process
- spatial attentive process
- spatial categorical process
- spatial concentrative process
- spatial emergent process
- spatial positional process
- spatial quantitative process
- visual lexical process
- visual phonetic process
- visual temporal process
- vocal process

### 2.3.3 Instantaneous Self-Assessment (ISA)

The Instantaneous Self-Assessment (ISA: Tattersall & Foord, 1996) is a subjective unidimensional workload rating method that provides a continuous and concurrent assessment of task demand on perceived workload. In this study, it was used to measure the perceived workload for each task type. Participants were asked to verbally rate their workload for completing each of the three task types (checking, detection, and response implementation) using a 5-point Likert scale ranging from “1 = Very Low” to “5 = Very High”.

## 2.4 Individual Difference Measures

Individual differences measures were used to assess subjective perceptions of automation.

### 2.4.1 Perfect Automation Schema (PAS)

The PAS (Merritt et al., 2015) consists of 8 items relating to people’s trust in automated systems, with two subscales: High expectations and All-or-none belief. Each item is scored on a 5-point Likert scale from “strongly disagree” to “strongly agree”.

### 2.4.2 Human Interaction with Technology (HIT)

The HIT trust scale (modified from Lyons & Guznov, 2019) comprises 10 items asking about intentions to trust automation. Items are answered on a 7-point scale, 1 being “not at all true” and 7 being “very true”.

### 2.4.3 Checklist of Trust between People and Automation (CTPA)

The CTPA (Jian et al., 2000) is composed of 12 items which are rated on a 7-point Likert scale from “not at all” to “totally agree”. An example statement is “The system was reliable.” Participants rated their trust in the automation following performance.

The physiological, subjective, and individual difference measures described above were applied consistently across both experiments. The following sections detail the specific methodologies and hypotheses for each experiment, beginning with Experiment 6, which examined the effects of automation level and task type on operator workload, performance, and trust.

# SECTION 3: EXPERIMENT 6 METHODOLOGY

## 3.1 Overview of Experiment 6: Effects of Automation Level and Task Type on Workload, Performance, and Trust

Experiment 6 aimed to explore the effects of varying LOA and task type on workload, performance, and trust in a simulated EOP. The study focused on the use of CBPs, which are hypothesized to influence workload due to their novel demands (Kim et al., 2014; O'Hara et al., 2000). Participants performed two CBPs, switching between them as directed by the procedure. Performance was measured primarily by response time, with secondary tasks requiring monitoring of 12 gauges to measure workload.

The study compared two intermediate LOAs, which were management-by-consent and management-by-exception. These LOAs are defined in NUREG 0700, Revision 3. The management-by-consent LOA required participants to approve or override automation decisions, while management-by-exception allowed automation to advance the procedure without direct participant approval unless an error was detected. The study tested the following hypotheses:

1. **Workload**

    a. It was hypothesized that detection tasks would elicit higher workload than checking and response implementation tasks.

    b. Additionally, it was anticipated that the higher LOA (management-by-exception) would result in lower workload, particularly in the more demanding detection task.

2. **Performance**

    Shorter response times were expected for the management-by-exception LOA for both:

    a. primary operational tasks and

    b. secondary gauge monitoring.

3. **Trust**

Trust was measured as an exploratory variable, with the expectation that higher LOA might lead to complacency and increased trust, potentially influencing automation use.

The results of this study aim to provide insights into how intermediate automation levels affect workload and performance, particularly in high-stakes tasks, while also exploring how individual differences in trust may influence the overall effectiveness of automation.

## 3.2 Participants

A total of 45 participants (*M* = 20.93, *SD* = 5.78) were recruited from the University of Central Florida. Participants received course credits (1 credit/hour) in compensation for their time. There were 26 men, 15 women, one that chose not to disclose gender, and three that did not respond. None of the participants had prior experience operating an NPP control system of any kind. All participants reported normal or corrected-to-normal vision and no color vision deficiency or history of neurological disorders.

The number of participants varies between measures and scenarios in this study. This is noted as a limitation, as it introduces interpretability biases. In some cases, participants failed to complete both scenarios, therefore the post-surveys for those cases were inevitably missing. There were also a reduced number of cases for some physiological measures due to sensor-related issues (e.g., missing raw data, connection failure, data parsing errors, etc.).

## 3.3 Training

Training involved an overview using PowerPoint slides to go through the task environment and each component of the system, a demonstration with the live training scenario on the simulator, as well as a hands-on training of the simulator under necessary guidance and supervision to make sure that the participants were familiar with the simulator. All participants completed the practice tasks successfully and exhibited proficiency in the simulator interface and controls.

## 3.4 Experimental Design

A two (LOA: management-by-consent and management-by-exception) × three (Task type: checking, detection, and response implementation) within-subjects design was adopted in this study. Participants completed two experimental scenarios, one in management-by-consent and one in management-by-exception LOA settings respectively in a balanced order. Each experimental scenario consisted of twenty-five task steps, including 17 checking tasks, four detection tasks, and four response implementation tasks. The number and order of the task steps were not balanced due to the nature of the original EOP, which requires steps to be taken in a prescribed sequence.

## 3.5 Independent Variables

The independent variables in this experiment were level of automation (management-by-consent and management-by-exception) and task type (checking, detection, and response implementation).

### 3.5.1 Level of Automation

Automation was offered at two levels, management-by-consent and management-by-exception respectively. Management-by-consent is the lower LOA (see Table 1). In this condition, explicit endorsement or over-ride of the automation is needed. Whereas the higher LOA (management-by-exception) implemented the automation unless the operator chose to over-ride within a set time window. All twenty-five task steps were supported by automation. The automation reliability was set to be approximately 90% and was described as highly reliable but not perfect to participants during training. The imperfect automation was implemented by creating incorrect automation outcomes for two of the checking steps. There were no automaton failures for detection and response implementation task types due to the small number of steps available for those task types.

### 3.5.2 Task Type

Participants performed three types of tasks, checking, detection, and the implementation of a response. The checking task was rule-based and required a one-time inspection of an I&C to verify that it was in the state called for by the EOP. Participants were required to locate various I&Cs and indicate identification by clicking on the correct I&C. The detection task was also rule-based and required participants to correctly locate an instrument and then to continuously monitor that instrument parameter for identification of change. The response implementation was a skill-based task and required participants to locate a control on a given panel and subsequently manipulate the control in the required direction (i.e., open or shut (Vicente & Rasmussen 1988).

In addition to the general tasks in the three above types, participants were also instructed to monitor a group of twelve gauges on the panel throughout the scenario as the secondary task. When the gauges fell below 20%, participants were supposed to react by tapping the label of the gauge to acknowledge. This happened 10 times during each scenario.

## 3.6 Scenario Setup

The experimental scenario was developed based on a generic version for a "Loss of All Alternating Current (AC) Power (ECA-0.0)" EOP but modified for experimental use. There was a total of twenty-five task steps, including 17 checking tasks, four detection tasks, and four response implementation tasks. A computerized procedure tool was provided on a separate monitor to prompt participants to follow the instructions to complete the tasks. Not only did the computerized procedure tool provide task instructions, but it offered automation aid at different levels based on LOA condition. The system was presented in a two-column layout. Instructions associated with each workstation were shown in its dedicated column.

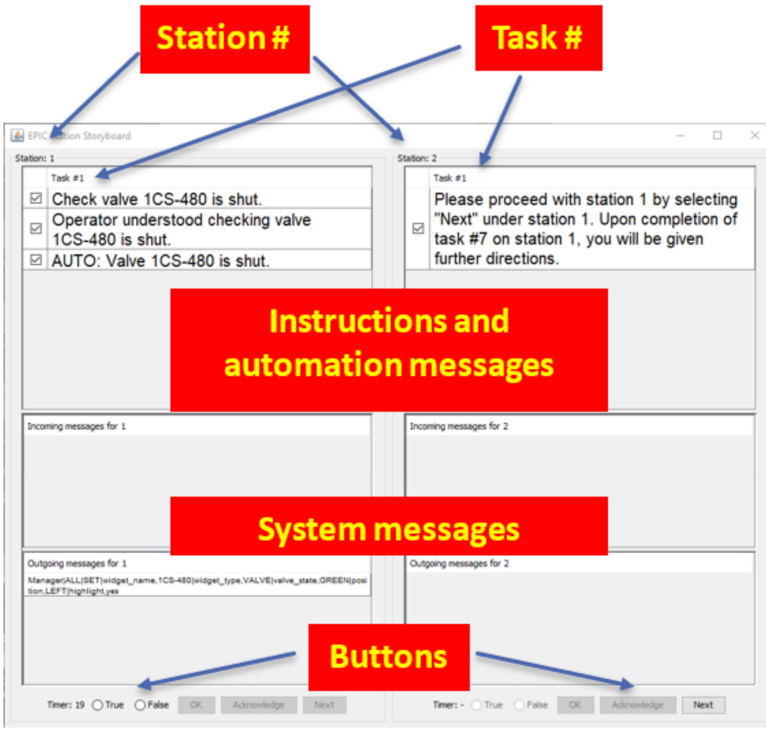


**Figure 5 Computerized Procedure Tool**

## 3.7 Measures

### 3.7.1 Workload

Workload was measured using the NASA TLX, Multiple Resource Questionnaire, and Instantaneous Self-Assessment.

### 3.7.2 Performance

Performance was measured as reaction time. Reaction time was measured for each task type including checking, detection, and response implementation tasks, as well as the secondary task.

### 3.7.3 Trust

Individual difference measures were used to assess trust. These included the Perfect Automation Schema (PAS; Merritt et al., 2015), Human Interaction with Technology (HIT; modified from Lyons & Guznov, 2019) and the Checklist of Trust between People and Automation (CTPA; Jian et al., 2000).

## 3.8 Procedures

Participants were asked to read a consent form approved by the UCF Internal Review Board (IRB) and were then briefed on the experimental procedure. Following consent, they were equipped with the noninvasive physiological sensors, including EEG, ECG, and fNIRS, and then sat through a 5-minute baseline for each measure. They took a pre-task survey set which included a demographics questionnaire, the PAS, and the HIT trust scale. Once these were complete, participants were given a short training with a PowerPoint presentation and then a hands-on practice session, both of which were done with the researcher present. Participants interacted with two scenarios, each with a different level of automation. During the task, participants were asked to rate their workload level on a scale of 1 to 5 (ISA self-report scale). After each scenario was complete, participants filled out a series of post-task questionnaires including the NASA-TLX, the MRQ, and the CTPA. After questionnaires were completed, the sensors were removed, and the participant was debriefed then dismissed.

# SECTION 4: EXPERIMENT 6 RESULTS

## 4.1 Subjective Measures

### 4.1.1 NASA-TLX

Paired-samples *t*-tests with Bonferroni corrections were conducted to determine whether there was a statistically significant mean difference between the subjective workload measured by NASA-TLX in different LOA manipulations. It is acknowledged here that the use of Bonferroni corrections for the NASA-TLX subscale comparisons may have been overly conservative. This may have obscured meaningful differences (e.g., physical demand, effort, and global workload), limiting the interpretability of workload effects across LOA manipulations. Participants reported higher temporal demand in the management-by-exception condition ($M$ = 28.06, $SD$ = 25.24) than in the management-by-consent condition ($M$ = 19.14, $SD$ = 21.13), $t(34) = -3.04$, $p < .01$. No significant difference was revealed from other subscales or the global workload score (Table 2).

**Table 2**
*Summary of NASA-TLX between LOA Manipulations*

| | Management-by-Consent | | | Management-by-Exception | | | 95% CI for Mean Difference | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | | *t* | *df* |
| Global Workload | 22.21 | 14.95 | 35 | 25.43 | 16.00 | 35 | [-7.16, 0.73] | -1.66 | 34 |
| Mental Demand | 37.00 | 28.88 | 35 | 38.63 | 32.32 | 35 | [-10.20, 6.95] | -.39 | 34 |
| Physical Demand | 17.29 | 14.52 | 35 | 21.89 | 21.01 | 35 | [-11.62, 2.42] | -1.33 | 34 |
| Temporal Demand | 19.14 | 21.13 | 35 | 28.06 | 25.24 | 35 | [-14.87, -2.95] | -3.04** | 34 |
| Effort | 20.71 | 18.03 | 35 | 25.77 | 24.56 | 35 | [-10.94, 0.82] | -1.75 | 34 |
| Frustration | 18.49 | 27.23 | 35 | 17.97 | 20.39 | 35 | [-7.87, 8.90] | .13 | 34 |
| Performance | 21.19 | 28.74 | 36 | 22.47 | 29.88 | 36 | [-13.37, 10.81] | -.22 | 35 |

**$p < .01$, *$p < .05$

### 4.1.2 MRQ

Paired-samples *t*-tests with Bonferroni corrections were conducted to determine whether there was a statistically significant mean difference between the mental processes measured by MRQ in different LOA manipulations. Participants reported higher visual phonetic process in the management-by-exception condition ($M$ = 37.25, $SD$ = 39.01) than in the management-by-consent condition ($M$ = 27.82, $SD$ = 34.06), $t(43) = 2.31$, $p < .05$. No significant difference was revealed from other subscales (Table 3).

**Table 3**
*Summary of MRQ between LOA Manipulations*

| | Management-by-Consent | | | Management-by-Exception | | | 95% CI for Mean Difference | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | | *t* | *df* |
| Auditory Emotional | 9.30 | 22.41 | 44 | 6.91 | 16.89 | 44 | [-4.44, 9.21] | .71 | 43 |
| Auditory Linguistic | 10.16 | 21.85 | 44 | 13.70 | 36.19 | 44 | [-12.32, 5.23] | -.82 | 43 |
| Manual Process | 54.98 | 36.19 | 44 | 55.02 | 37.29 | 44 | [-8.01, 7.92] | -.01 | 43 |
| Short Term Memory | 51.48 | 35.66 | 44 | 57.23 | 36.44 | 44 | [-12.36, .86] | -1.75 | 43 |
| Spatial Attentive | 65.45 | 38.68 | 44 | 61.80 | 36.95 | 44 | [-3.76, 11.08] | 1.00 | 43 |
| Spatial Concentrative | 50.80 | 35.67 | 44 | 50.05 | 36.36 | 44 | [-6.98, 8.48] | .20 | 43 |
| Spatial Categorical | 44.80 | 35.53 | 44 | 40.91 | 36.50 | 44 | [-4.28, 12.05] | .96 | 43 |
| Spatial Emergent | 57.14 | 37.63 | 44 | 56.80 | 39.55 | 44 | [-8.15, 8.83] | .08 | 43 |
| Spatial Positional | 56.25 | 36.82 | 44 | 60.34 | 37.25 | 44 | [-13.23, 5.05] | -.90 | 43 |
| Spatial Quantitative | 49.50 | 38.12 | 44 | 48.39 | 39.37 | 44 | [-7.93, 10.16] | .25 | 43 |
| Visual Lexical | 59.55 | 37.78 | 44 | 58.93 | 37.92 | 44 | [-10.47, 11.69] | .11 | 43 |
| Visual Phonetic | 27.82 | 34.06 | 44 | 37.25 | 39.01 | 44 | [-17.74, -1.12] | -2.29* | 43 |
| Visual Temporal | 53.55 | 39.66 | 44 | 53.45 | 38.31 | 44 | [-7.63, 7.82] | .02 | 43 |
| Vocal Process | 15.18 | 27.86 | 44 | 12.61 | 24.34 | 44 | [-2.01, 7.15] | 1.13 | 43 |

**$p < .01$, *$p < .05$

### 4.1.3 ISA

A two (LOA: management-by-consent and management-by-exception) × three (task type: checking, detection, and response implementation) repeated ANOVA was run to test the effects of experimental manipulations on the subjective workload measured by ISA. A significant main effect of LOA was revealed, $F(1,36) = 5.81$, $p < .05$, $\eta_p^2 = .14$. Overall, participants reported higher workload in the management-by-exception condition. Additionally, the interaction of LOA and task type was significant, $F(2,72) = 8.14$, $p < .001$, $\eta_p^2 = .18$. In the management-by-consent condition, the detection task ($M = 1.59$, $SD = .69$) was rated as the task type with the highest workload. However, in the management-by-exception condition, the detection task received the lowest workload rating ($M = 1.51$, $SD = .80$), and the checking task ($M = 1.76$, $SD = .90$) was rated as the highest workload (Table 4). It is noted that there were more checking trials compared to detection trials, which potentially could have led to the difference in ISA ratings.

**Table 4**
*Two-way ANOVA Statistics for ISA Ratings*

| | Management-by-Consent | | | Management-by-Exception | | | ANOVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | Effect | *F* | *df* | $\eta_p^2$ |
| Task type | | | | | | | | | | |
| Checking | 1.41 | .64 | 37 | 1.76 | .90 | 37 | LOA | 5.81* | 1, 36 | .14 |
| Response | 1.24 | .44 | 37 | 1.62 | .72 | 37 | Task type | 1.59 | 2, 72 | .04 |
| Detection | 1.59 | .69 | 37 | 1.51 | .80 | 37 | L × T | 1.23** | 2, 72 | .18 |

**$p < .01$, *$p < .05$, L × T: LOA × Task type

### 4.1.4 CTPA

Paired-samples *t*-tests were conducted to determine whether participants trusted the two LOAs differently. There were no significant findings from this analysis.

## 4.2 Physiological Measures

All physiological dependent variables entered into the ANOVAs were calculated as percentage differences in percentage five-minute resting baseline. This method helps account for individual differences when comparing group means as is the case when running ANOVAs.

### 4.2.1 Physiological Measures Electroencephalogram (EEG)

Brain activity was recorded at 9 EEG sensor sites, the EEG data were analyzed by grouping sensor sites by hemispheres (i.e., compare brain activity between the left and right hemispheres) and lobes (i.e., compare brain activity among the frontal, parietal and occipital lobes).

#### *4.2.1.1 Hemispheres*

Two (LOA: management-by-consent and management-by-exception) × three (task type: checking, detection, and response implementation) × two (Hemisphere: left and right) repeated ANOVAs were run for theta, alpha, and beta frequency bands separately. These ANOVAs provided insight into the overall effects of LOA and task type manipulations on the left and right hemispheres.

A follow-up two (LOA: management-by-consent and management-by-exception) × three (task type: checking, detection, and response implementation) repeated ANOVA on EEG theta band revealed a significant main effect of LOA, $F(1,64) = 4.93$, $p < .05$, $\eta_p^2 = .13$, such that management-by-exception ($M = 705.33$ $SE = 234.11$) elicited greater theta increase than management-by-consent ($M = 321.34$ $SE = 76.40$) (Figure 6).

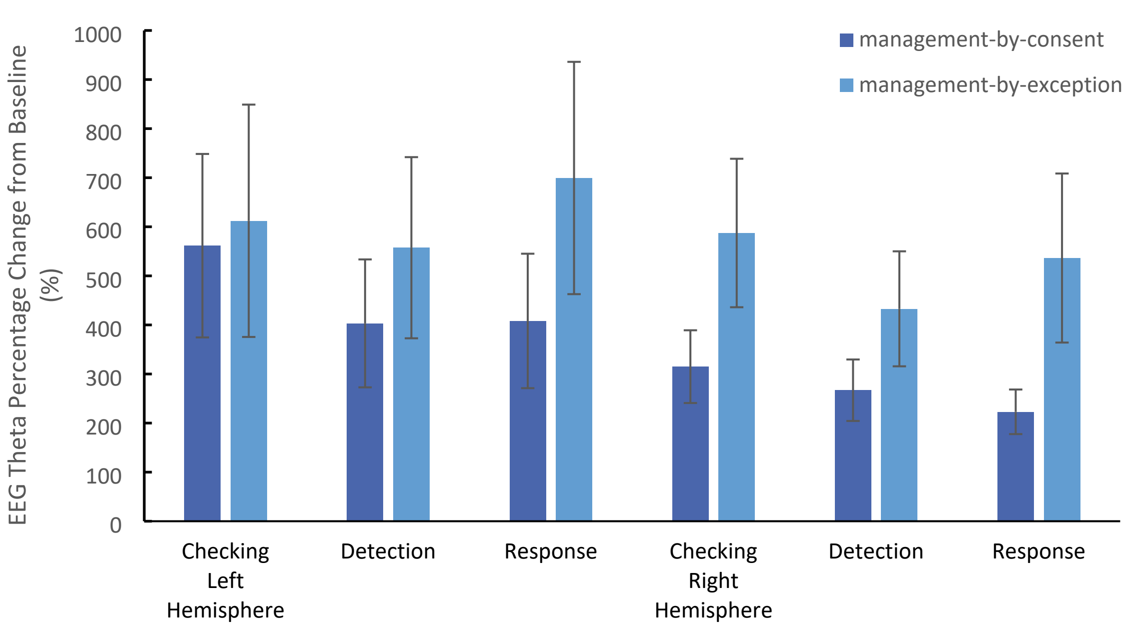


**Figure 6 EEG Theta Percentage Change from Baseline by Task Type and Hemisphere (error bars denote standard errors)**

In terms of alpha band, a significant main effect was found for task type, $F(1.68,52.19) = 4.96$, $p < .05$, $\eta_p^2 = .14$. Overall, the detection task type ($M$ = 172.61 $SE$ = 50.27) showed a smaller increase from baseline than the checking task type ($M$ = 254.68 $SE$ = 77.31). In the right hemisphere, management-by-exception elicited greater alpha increase in all three task types; however, such trend was not observed in the left hemisphere (Figure 7).

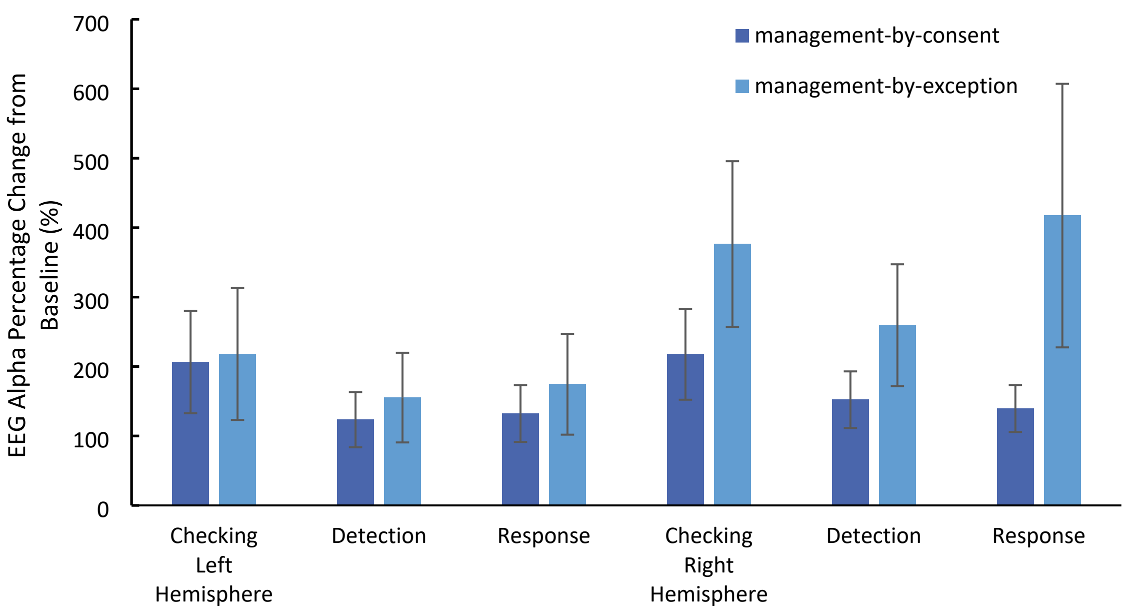


**Figure 7 EEG Alpha Percentage Change from Baseline by Task Type and Hemisphere (error bars denote standard errors)**

For EEG beta band, a significant main effect was found for task type, $F(1.59,49.36) = 11.54$, $p < .01$, $\eta_p^2 = .27$. Compared to the detection task type ($M = 123.51$ $SE = 17.76$) and the response implementation task type ($M = 135.03$ $SE = 18.94$), the checking task type ($M = 169.62$ $SE = 25.87$) showed a greater increase from baseline.

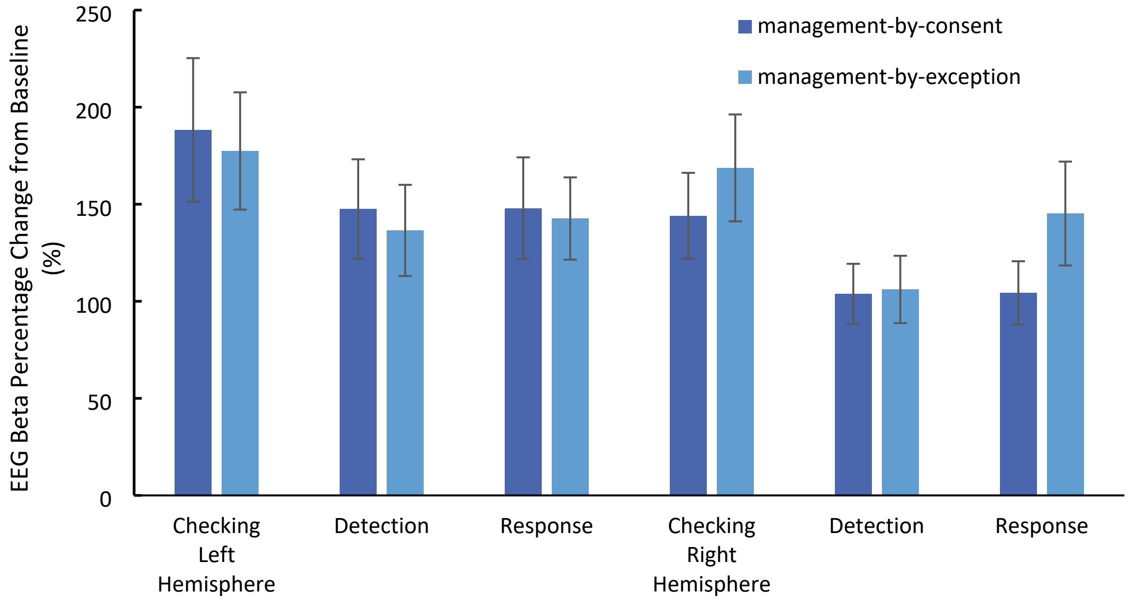


**Figure 8 EEG Beta Percentage Change from Baseline by Task Type and Hemisphere (error bars denote standard errors)**

#### ***4.2.1.2*** *Lobes*

Two (LOA: management-by-consent and management-by-exception) × three (task type: checking, detection, and response implementation) × three (Lobe: front, parietal, and occipital) repeated ANOVAs were run for theta, alpha, and beta frequency bands separately. These ANOVAs provided insight into the overall effects of LOA and task type manipulations on the frontal, parietal, and occipital lobes.

For theta frequency band, a significant main effect of LOA was found, $F(1,31) = 4.56$, $p < .05$, $\eta_p^2 = .13$, such that the management-by-exception condition ($M = 580.09$ $SE = 153.44$) elicited a greater increase in theta band than the management-by-consent condition ($M = 418.81$ $SE = 118.09$). In addition, the main effect of task type was also significant, $F(2,62) = 3.53$, $p < .05$, $\eta_p^2 = .10$. Compared to the detection task type ($M = 452.88$ $SE = 125.93$), participants showed greater increase in theta during checking tasks ($M = 558.86$ $SE = 143.16$).

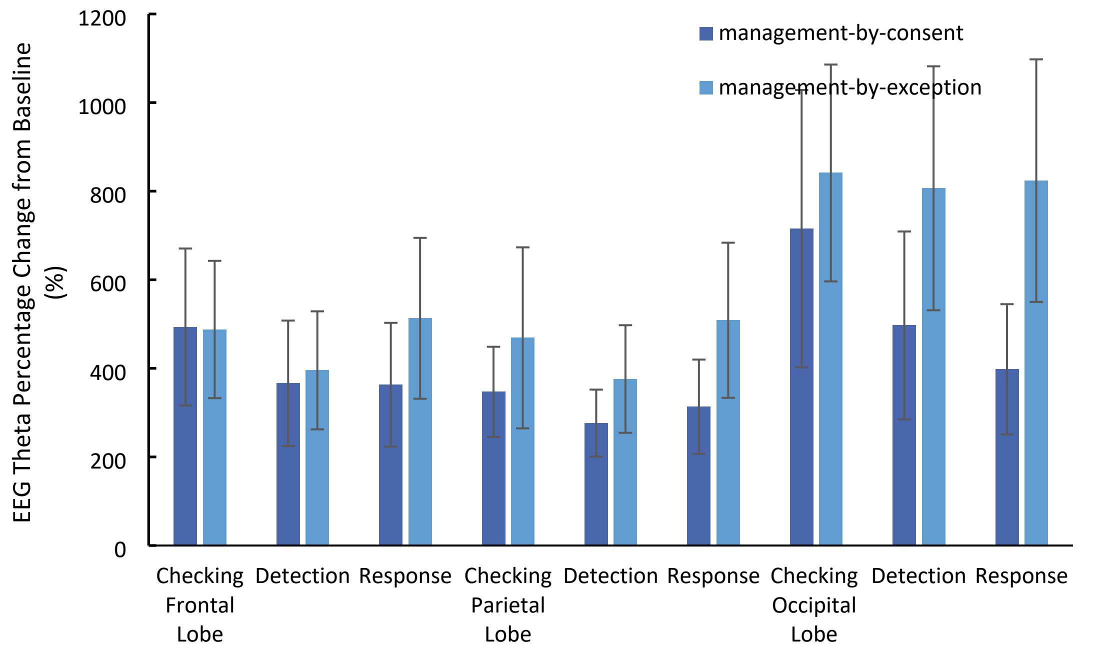


**Figure 9 EEG Theta Percentage Change from Baseline by Task Type and Lobe (error bars denote standard errors)**

For alpha frequency band, a significant main effect was found for task type, $F(1.46,45.17) = 5.84$, $p < .05$, $\eta_p^2 = .16$. Overall, the detection task type ($M = 180.70$, $SE = 65.11$) showed a smaller increase from baseline than the checking task type ($M = 257.75$ $SE = 91.49$).

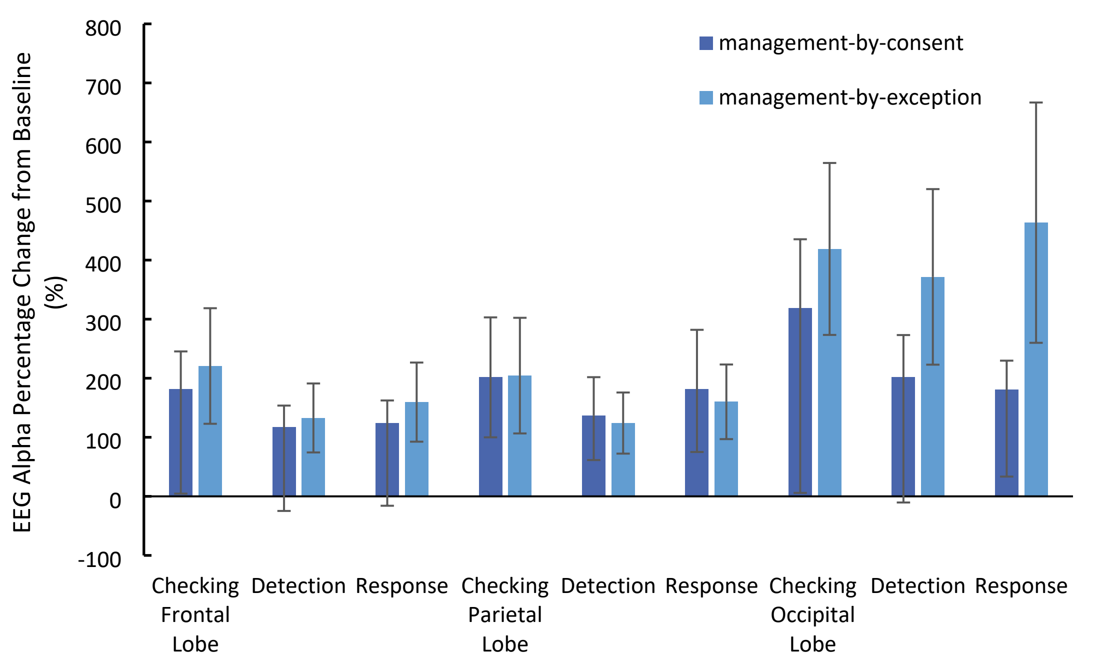


**Figure 10 EEG Alpha Percentage Change from Baseline by Task Type and Lobe (error bars denote standard errors)**

For EEG beta band, a significant main effect was found for task type, $F(1.50,46.52) = 11.71$, $p < .01$, $\eta_p^2 = .27$. The checking task type ($M$ = 170.69 $SE$ = 24.00) showed a greater increase from baseline than both the detection task type ($M$ = 127.11 $SE$ = 17.05) and the response implementation task type ($M$ = 140.96 $SE$ = 18.98).


**Figure 11 EEG Beta Percentage Change from Baseline by Task Type and Lobe (error bars denote standard errors)**

### 4.2.2 Functional Near-Infrared Spectroscopy (fNIRS)

A two (LOA: management-by-consent and management-by-exception) × three (task type: checking, detection, and response implementation) × two (Hemisphere: left and right) repeated ANOVA was run to determine the effect of LOA and task type manipulations on regional oxygen saturation ($rSO_2$) for the left and right prefrontal cortex. The ANOVA results revealed no significant main effect for LOA or task type and no significant interaction.

### 4.2.3 Electrocardiogram (ECG)

Two (LOA: management-by-consent and management-by-exception) × three (task type: checking, detection, and response implementation) × two (Hemisphere: left and right) repeated ANOVAs were conducted to determine the effect of LOA and task type manipulations on HR, HRV, and IBI. The ECG measures, HR, IBI, and HRV, were derived from R-Peak detections using the So-Chan QRS algorithm from the raw ECG signal. The only significant finding regarding the ECG measures was from HRV (Table 5). The main effect of LOA was significant, $F(1,32) = 4.25$, $p < .05$, $\eta_p^2 = .12$. Participants had greater HRV change from the pre-task baseline in the management-by-consent condition ($M$ = 64.03, $SE$ = 19.51) than in the management-by-exception condition ($M$ = 35.44, $SE$ = 11.08).

**Table 5**

*Two-way ANOVA Statistics for HRV*

| | Management-by-Consent | | | Management-by-Exception | | | ANOVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | Effect | *F* | *df* | $\eta_p^2$ |
| HRV | | | | | | | | | | |
| Checking | 76.64 | 143.09 | 33 | 36.23 | 71.31 | 33 | LOA | 4.28* | 1, 32 | .12 |
| Response | 49.75 | 116.10 | 33 | 29.93 | 77.49 | 33 | Task type | 1.16 | 2, 64 | .04 |
| Detection | 65.70 | 129.10 | 33 | 40.17 | 62.94 | 33 | L × T | .52 | 2, 64 | .02 |

**$p < .01$, *$p < .05$, L × T: LOA × Task type

## 4.3 Individual Difference Measures

Correlational analyses were computed to assess the relationship between individual difference measures (e.g., HIT and PAS) and subjective outcomes as well as physiological responses.

### 4.3.1 Individual Difference Measures and Trust

HIT score was positively associated with the trust in automation measured by CTPA in both LOA conditions (management-by-consent, $r = .374$, $p < .05$; management-by-exception, $r = .429$, $p < .01$). Neither of the PAS subscales predicts trust in automation in this study (Table 6).

**Table 6**

*Correlations between Individual Difference Measures and Trust*

| | HIT | PAS_High Expectation | PAS_All or None Thinking |
|---|---|---|---|
| CTPA_ Management-by-Consent | .374* | .295 | -.003 |
| CTPA_ Management-by-Exception | .429** | .261 | .011 |

**$p < .01$, *$p < .05$

### 4.3.2 Individual Difference Measures and Subjective Workload

Table 7 shows the correlations between dispositional trust measures and subjective workload measures measured by NASA-TLX and MRQ in two LOA conditions. Generally, the All or None thinking score derived from PAS was predictive of, in terms of weak to moderate correlation, some of the mental processes measured by MRQ. The high expectation score derived from PAS was negatively associated with subjective workload induced by detection tasks (measured by ISA). Such negative association showed a similar pattern in both LOA conditions. The correlation was significant in the management-by-consent condition, $r = .\text{-}369$, $p < .05$, but showed only a nonsignificant trend in the management-by-exception condition, $r = .\text{-}221$, $p > .05$. In addition, HIT was negatively associated with subjective workload induced by response implementation tasks (measured by ISA). Similarly, such negative association was only statistically significant in the management-by-consent condition, $r = .\text{-}358$, $p < .05$, but not in the management-by-exception condition, $r = .\text{-}272$, $p > .05$. No significant correlation was found between dispositional trust and subjective workload measured with the NASA-TLX.

**Table 7**
*Correlations between Individual Difference Measures and Subjective Workload*

| | Management-by-Consent | | | Management-by-Exception | | |
|---|---|---|---|---|---|---|
| | HIT | PAS High Expectation | PAS All or None Thinking | HIT | PAS High Expectation | PAS All or None Thinking |
| **NASA-TLX** | | | | | | |
| Global Workload | .032 | .118 | .012 | -.165 | -.054 | .136 |
| Mental Demand | -.035 | .236 | .089 | -.208 | -.032 | .225 |
| Physical Demand | -.005 | .250 | .148 | -.164 | -.028 | .200 |
| Temporal Demand | .109 | -.008 | .144 | -.095 | -.021 | .119 |
| Effort | -.073 | -.063 | .102 | -.134 | -.066 | .064 |
| Frustration | .194 | .043 | -.145 | .041 | -.071 | -.013 |
| Performance | -.076 | .017 | -.159 | -.212 | -.142 | -.168 |
| **MRQ** | | | | | | |
| Auditory Emotional | -.064 | .216 | .044 | .092 | .210 | -.073 |
| Auditory Linguistic | -.135 | .199 | .051 | .110 | .208 | -.105 |
| Manual Process | .127 | .110 | .339* | .248 | .234 | .320* |
| Short Term Memory | .029 | .093 | .325* | .286 | .233 | .270 |
| Spatial Attentive | .050 | .096 | .253 | .116 | .077 | .191 |
| Spatial Concentrative | .124 | .089 | .331* | .121 | .037 | .259 |
| Spatial Categorical | -.004 | -.005 | .371* | .035 | .007 | .383* |
| Spatial Emergent | .077 | -.045 | .315* | .068 | .058 | .136 |
| Spatial Positional | .098 | .100 | .213 | .112 | .078 | .035 |
| Spatial Quantitative | -.043 | .009 | .247 | .144 | .171 | .027 |
| Visual Lexical | .054 | .076 | .365* | .168 | .120 | .063 |
| Visual Phonetic | -.030 | -.052 | .131 | .116 | .017 | .143 |
| Visual Temporal | .085 | .040 | .198 | .129 | .187 | .110 |
| Vocal Process | -.129 | -.045 | .382* | -.066 | -.029 | .468** |
| **ISA** | | | | | | |
| ISA Checking | -.056 | .129 | -.191 | -.210 | -.181 | -.017 |
| ISA Detection | -.277 | -.369* | -.018 | -.119 | -.221 | .001 |
| ISA Response | -.358* | -.158 | .009 | -.272 | -.254 | .115 |

**$p < .01$, *$p < .05$

### 4.3.3 Individual Difference Measures and Physiological Measures

Table 8 shows the correlations between dispositional trust measures and physiological responses measured by EEG, ECG, and fNIRS in two LOA conditions. Both dispositional trust measures showed some correlations with participants' cardiac responses to both LOA and task type manipulations in terms of HRV. For example, HRV during detection tasks was associated with high expectation (management-by-consent, $r = -.527$, $p > .01$; management-by-exception, $r = -.437$, $p < .01$) and HIT (management-by-exception, $r = -.384$, $p < .05$). HRV during response implementation tasks was associated with PAS (management-by-consent, $r = -.448$, $p > .01$; management-by-exception, $r = .390$, $p < .05$). HRV in checking tasks was only associated with PAS All or None thinking in the management-by-exception condition, $r = .377$, $p < .05$. Additionally, regional oxygen saturation in the left prefrontal cortex during detection tasks was negatively associated with high expectation in the management-by-consent condition, $r = -.239$, $p < .05$.

**Table 8**
*Correlations between Individual Difference Measures and Physiological Measures*

| | Management-by-Consent | | | Management-by-Exception | | |
|---|---|---|---|---|---|---|
| | HIT | PAS High Expectation | PAS All or None Thinking | HIT | PAS High Expectation | PAS All or None Thinking |
| **EEG** | | | | | | |
| Alpha F Checking | -.030 | -.174 | -.210 | -.105 | -.228 | -.062 |
| Alpha F Detection | -.029 | -.226 | -.252 | -.161 | -.267 | -.085 |
| Alpha F Response | .008 | -.023 | -.111 | -.151 | -.221 | -.101 |
| Beta F Checking | .081 | .060 | -.080 | -.038 | -.123 | .137 |
| Beta F Detection | .045 | -.040 | -.099 | -.193 | -.241 | .056 |
| Beta F Response | .136 | .136 | .049 | -.047 | .012 | .180 |
| Theta F Checking | -.010 | .030 | -.210 | -.151 | -.090 | -.075 |
| Theta F Detection | -.087 | -.034 | -.247 | -.212 | -.224 | -.120 |
| Theta F Response | .026 | .113 | -.059 | -.170 | -.123 | -.168 |
| **ECG** | | | | | | |
| Heart Rate Checking | -.204 | .029 | .032 | -.117 | .062 | .023 |
| Heart Rate Detection | -.150 | -.087 | -.006 | -.130 | .116 | .025 |
| Heart Rate Response | -.184 | .065 | -.001 | -.166 | .068 | -.084 |
| HRV Checking | -.086 | -.225 | .084 | -.176 | -.295 | .377* |
| HRV Detection | -.201 | -.527** | .270 | -.384* | -.473** | .230 |
| HRV Response | -.203 | -.448** | .015 | -.291 | -.205 | .390* |
| IBI Checking | .086 | -.053 | -.173 | .005 | -.067 | -.130 |
| IBI Detection | .034 | -.090 | -.146 | .011 | -.116 | -.146 |
| IBI Response | .062 | -.069 | -.164 | .062 | -.052 | -.030 |
| **Oximeter** | | | | | | |
| $rSO_2$ Left Checking | -.149 | -.299 | -.249 | .103 | .042 | -.113 |
| $rSO_2$ Left Detection | -.166 | -.329* | -.242 | .087 | .021 | -.162 |
| $rSO_2$ Left Response | -.154 | -.286 | -.240 | .053 | -.024 | -.182 |
| $rSO_2$ Right Checking | -.050 | .061 | .006 | -.023 | .088 | .145 |
| $rSO_2$ Right Detection | -.082 | -.066 | -.018 | -.003 | .103 | .183 |
| $rSO_2$ Right Response | -.058 | .010 | -.007 | -.002 | .115 | .211 |

**$p < .01$, *$p < .05$

## 4.4 Performance Measures

### 4.4.1 Reaction Time

Reaction time was measured for each task type including checking, detection, and response implementation tasks, as well as the secondary task.

A two (LOA: management-by-consent and management-by-exception) × four (task type: checking, detection, response implementation, and secondary) repeated ANOVA was run to test the effects of experimental manipulations on performance in terms of reaction time. The analysis revealed a significant main effect of LOA, $F(1,34) = 18.00$, $p < .01$, $\eta_p^2 = .35$., and a main effect of task type, $F(1.51, 51.47) = 28.08$, $p < .01$, $\eta_p^2 = .45$. Overall, participants had faster reaction times in management-by-exception condition (*M* = 7.31, *SE* = .33) than in management-by-consent condition (*M* = 8.76, *SE* = .33). Participants had the slowest reaction time for detection tasks (*M* = 8.76, *SE* = .39). The reaction time was significantly slower than the reaction time for both response implementation tasks (*M* = 7.74, *SE* = .39) and checking tasks (*M* = 5.40, *SE* = .39).

Paired sample *t*-tests were used to compare the secondary task reaction times between LOA manipulations as well as with other task types. In the management-by-exception condition, the secondary task reaction time was significantly slower than the reaction time for checking tasks (*M* = 4.66, *SE* = .25), $t(30) = 6.89$, $p < .01$ and the reaction time for response implementation tasks (*M* = 6.87, *SE* = .49), $t(30) = 3.10$ $p < .01$. In the management-by-consent condition, the secondary task reaction time was also significantly slower than the reaction time for checking tasks (*M* = 6.07, *SE* = .30), $t(32) = 8.54$, $p < .01$ and for response implementation tasks (*M* = 8.63, *SE* = .57), $t(32) = 2.78$ $p < .01$ (Figure 12).

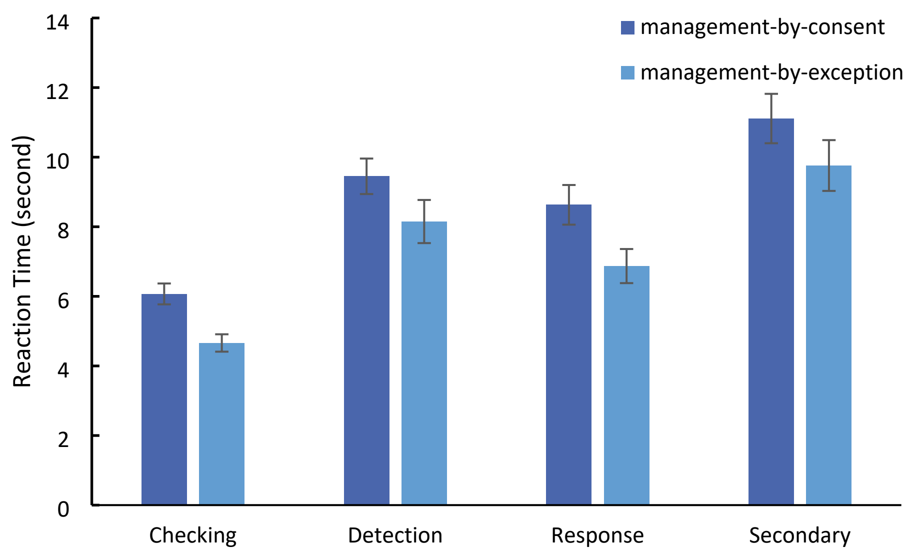


**Figure 12 Reaction Time by Task Type (error bars denote standard errors)**

# SECTION 5: EXPERIMENT 6 DISCUSSION

Plant modernization and digitalization in the nuclear industry has multiple implications for HFE. Technological advancements may enhance safety in various respects, including mitigating operator workload, but they can also have unintended consequences that contribute to human error. There are HRA needs to assess novel potential error modes and failure mechanisms as well as performance shaping factors (Bye, 2023; Porthin et al., 2020). The present study aimed to contribute to understanding the impacts of digitalization by examining workload during execution of CBPs supported by automation. In this section, we discuss how task type and LOA influenced subjective and objective outcomes during the simulated EOP. We conclude by reviewing implications of the study for introducing CBPs supported by automation, choosing the optimal LOA, and understanding the role of operator trust.

Hypothesis 1a proposed that detection tasks would elicit higher workload than checking and response implementation tasks. This hypothesis was not supported by the EEG alpha, beta, and theta data, which showed lower activity during the detection task compared to the others. According to the ISA data, in the management-by-consent condition, the detection task had the highest reported workload, but in the management-by-exception condition, detection had the lowest reported workload. All other measures revealed no significant findings for this hypothesis. These contradictory findings suggest that detection tasks impose unique demands related to sustained attention rather than simply higher workload.

Hypothesis 1b anticipated that the higher LOA (management-by-exception) would result in lower workload. This hypothesis was not supported. Contrary to expectation, participants reported higher temporal demand (NASA-TLX) and greater visual phonetic process (MRQ) in the management-by-exception condition. EEG theta data and ECG HRV measures also indicated higher workload in the management-by-exception condition. All other measures revealed no significant findings related to this hypothesis. This suggests that management-by-exception may introduce additional monitoring demands that offset workload benefits of higher automation.

Hypothesis 2 predicted shorter response times for the management-by-exception LOA for both primary and secondary tasks. This hypothesis was supported for the primary tasks, but there were no significant findings for secondary tasks. Response times were significantly faster in the management-by-exception condition across task types, aligning with expectations that higher automation would facilitate more rapid responses.

Hypothesis 3 suggested that higher LOA might lead to increased trust and potential complacency. This hypothesis was not supported by our data. While CTPA ratings were slightly higher for management-by-exception, this difference was not statistically significant. Individual differences in trust disposition (measured by the HIT scale) predicted state trust, but we found little evidence that the higher LOA systematically affected trust levels or induced complacency.

## 5.1 Effects of Task Type and LOA on Subjective Workload

The analyses of subjective measures revealed several significant effects of the experimental manipulations, including task type and LOA. In the current study, the main effect of task type tended to be less significant than in previous studies with no automation support. The only significant task type effects came from the ISA data in the management-by-consent condition, in which the detection task was rated most demanding, and in the management-by-exception condition, in which the detection task was rated the least demanding. Such workload difference regarding the detection task may imply that the automation, especially the higher level of automation, is helpful in reducing the workload burden of detection tasks. As operators tend to be more prone to errors during extended detection tasks in conventional NPPs, the computerized procedure system and automation functionality may mitigate such vulnerability by alleviating workload and reducing the possibility of human errors.

Utilizing a computerized procedure system and automation could potentially bring several benefits to NPP operations, including improving procedure execution, enhancing human performance, supporting future technology, and reducing maintenance cost (Lew et al. 2018; Lipner & Kerch 1994). Two intermediate levels of automation were employed in the experimental manipulations. Some impact of the two LOAs on subjective workload was found. Both NASA-TLX and MRQ revealed some workload differences between two LOAs on some dimensions. For example, participants reported greater temporal demand and visual phonetic process in the management-by-exception condition. Results from ISA ratings also supported the trend revealed by NASA-TLX and MRQ that the workload in the management-by-exception condition tended to be higher. A possible explanation is that in the management-by-exception condition, which is the higher LOA, participants may have experienced additional attentional demands because of the greater need to visually monitor both the automation and their secondary task.

**Regulatory significance:** These findings have implications for workload assessment. The current results suggest that reviews may benefit from considering that task type can affect workload differently depending on the measurement approach used. Detection tasks may appear low in workload by some measures but may impose significant monitoring demands that could affect reliability. Specifically, detection tasks showed the lowest performance accuracy among primary tasks, slower reaction times, and were associated with reduced EEG beta and theta activity. Additionally, higher levels of automation may increase certain aspects of workload (e.g., temporal demand) rather than globally reducing it. This may be important to consider during evaluations of proposed systems.

## 5.2 Effects of Task Type and LOA on Physiological Measures

Experimental manipulations of task type and LOA may have different impacts on subjective workload, psychophysiological indicators of brain response, and objective performance measures. Thus, the subjective measures, although they are simple to administer, do not provide a comprehensive workload assessment. Therefore, to provide a comprehensive picture of operator response to experimental manipulations, workload was also assessed with an integrated suite of psychophysiological sensors, including EEG, ECG, and fNIRS.

The effects of task type and LOA revealed by subjective measures were partially supported by physiological measures. Based on the subjective data, workload was perceived differently depending on the LOA condition. However, no strong task type by LOA interaction effect was revealed by physiological measures. In terms of the effect of task type, EEG theta and beta bands suggested that workload associated with detection tasks was lower, which was consistent with the trend in the management-by-exception condition revealed by the ISA data; whereas the EEG alpha band data indicated that workload associated with detection tasks was highest, which was consistent with trend for management-by-consent condition revealed by the ISA data.

In terms of the effect of LOA, the physiological measures showed a more consistent trend than the subjective measures. Both EEG theta and ECG HRV data suggested that management-by-exception was associated with greater workload than management-by-consent, which was consistent with the trend observed by NASA-TLX, MRQ, and ISA.

Typically, as demands on attention and workload increase, activity in frontal theta and beta increases while alpha activity decreases (Gevins & Smith, 2003). As EEG alpha did not show a convergent trend as expected, it was possible that EEG, especially alpha band, was picking up something else. Divergence of physiological data and other measures is not uncommon. Literature shows that speech, respiration, muscle activity, body position, physical fitness, and many other factors can affect cardiac and other physiological responses to workload manipulations in an experiment (Wilson, 1992; Jorna, 1992). In this experiment, posture change, physical movement, and verbal communication were inevitable given operator tasks and may have influenced EEG response. Physiological measures can be not only sensitive to study manipulations but also to extraneous factors and in some cases, such complexity can make the physiological data hard to interpret and explain.

No significant finding was revealed on measures from fNIRS. fNIRS has been found to have good sensitivity and diagnosticity in experimental scenarios requiring sustained attention. It is possible that the automation introduced to this study mitigated the effort needed for sustained attention and, in turn, reduced the sensitivity of fNIRS.

**Regulatory significance:** The divergent physiological responses observed across measures highlight the importance of comprehensive measurement approaches in evaluations. For human factors engineering reviews, these findings suggest that validation testing may benefit from including multiple complementary physiological measures to effectively capture different aspects of operator cognitive states. An overreliance on single metric, physiological or otherwise could lead to incomplete or misleading conclusions about the adequacy of human-system interface designs in safety-critical environments.

## 5.3 Effects of Task Type and LOA on Performance Measures

In regard to LOA, reaction times were significantly faster in the management-by exception condition than management-by-consent, which indicates a lower workload in the management-by-exception condition. These findings were also supported by reaction times for the secondary task which tended to be faster in the management-by-exception condition, however these findings were not significant.

In regard to task type, reaction times for the detection tasks were significantly slower than both checking and response implementation tasks, which indicates a higher workload for detection tasks. The reaction time for the secondary task was also significantly slower than both the checking task and the response implementation task in both LOAs, however no significant findings were found between the secondary task and the detection task in regards to reaction time.

**Regulatory significance:** The findings have implications for understanding human performance in relation to the LOA implemented in a given system. For example, while higher automation levels may yield faster response times, they may also introduce trade-offs, such as reduced accuracy and diminished situation awareness. Regulatory reviews may benefit from looking for validation testing that includes extended operational scenarios to more effectively capture potential vigilance failures, particularly for detection tasks. Additionally, the findings support the value of incorporating task-specific automation effects into human reliability analysis, rather than relying solely on generalized assumptions about automation.

## 5.4 Individual Differences and Correlations

The data showed that individuals varied in both trait and state trust scale scores. The CPTA (Jian et al., 2000) measured the participant's trust in the automation to perform the steps in the CPB assigned to it. The possible range of scores on the CPTA is 12-84, with a scale midpoint of 48. The observed CPTA means for the two LOA conditions were in the mid-50s, above the midpoint but considerably below the top end of the scale. It appears that participants were generally under-trusting of the automation given that it was objectively highly reliable. Lack of familiarity and knowledge of the system may have generated uncertainty in participants that lowered trust. However, data also showed substantial variability in trust scores and there was no significant difference in CPTA scores across LOAs.

Trust in the CBP automation might reflect both the person's general disposition to trust automation as well as their direct experience of interacting with it during the simulated EOP. The correlational analyses showed that the HIT trait scale (Lyons & Guznov, 2019) predicted CPTA scores at both LOAs, consistent with expectation. The PAS (Merritt et al., 2015) was not significantly correlated with the CPTA, implying that this instrument is less effective than the HIT in capturing pre-existing tendencies towards trusting automation. The High Expectation subscale showed non-significant positive correlations in both conditions; these might have reached significance with a larger sample size. The All or None Thinking subscale – the direct measure of the perfect automation schema – showed near-zero correlations with the CPTA. No evidence was found for higher scores on the PAS leading to complacency, possibly because participants were generally somewhat under-trusting of the automation.

Our initial expectation was that trust might be inversely related to workload, because participants low in trust would be inclined to assign greater mental effort to monitoring and verifying the decisions of the automation. The level of correlation shown in the analyses was fairly low: trust and workload were not strongly inter-related. It is possible that regardless of the level of trust, workload is derived primarily from actual task performance and interactions with the automation and is only a minor factor. The correlation analysis showed significant positive associations between the All or None Thinking subscale of the PAS and some of the scales of the MRQ, especially manual and vocal process, short term memory, several of the spatial demand scales, and visual lexical demands. More of those correlations were significant in the management-by-consent LOA, but correlational magnitudes were similar across the two LOA conditions. Participants may have been uncertain about system reliability, and uncertainty may be more impactful on all-or-none-thinkers causing an increase in their perceptions of demands.

Overall, the correlational analyses draw attention to individual variation of trust in the automation utilized in this study and the role of pre-existing dispositions. It is unclear whether results would generalize to more experienced operator samples, but evaluations of trust may need to accommodate impacts of inter-individual variability. For example, a system could be reviewed to ensure its safety when operated by over- and under-trusting individuals.

**Regulatory Significance:** These findings emphasize the potential impact of individual differences in trust on operator outcomes. Specifically, inter-individual variability in trust can affect operator performance, workload perception, and safe use of automation. Reviewers may consider how this variability has been accounted for regarding HSI design, task analysis, training, and V&V. It may be beneficial to consider how a system can support safe operation across a diverse operator population, including those with extreme trust tendencies.

## 5.5 General Implications and Conclusion

These findings have implications for the NUREG-0711 workload assessment guidelines. The current results suggest that regulatory review should consider that task type affects workload differently depending on the measurement approach used. Detection tasks may appear low in workload by some measures but actually impose significant vigilance demands that could affect reliability. Additionally, higher levels of automation may increase certain aspects of workload (e.g., temporal demand) rather than universally reducing it, which has implications for how automation is evaluated during licensing reviews.

The experiment reported here is the first in which the HPTF workload methodology was applied to the HSI issues raised by the introduction of system automation, in the context of CBP usage. Experimental manipulations influenced the profile of workload response across multiple subjective and objective workload measures. Both task type and LOA influenced workload. The experiment also introduced trust assessment into the methodology for the first time and suggests further research into this topic.

Several findings from this study have implications for future research. First, the performance data for detection tasks suggest that even with automation support, detection remains vulnerable to vigilance-related failures. This has safety implications for scenarios where operators must detect abnormal conditions over extended periods. Such vigilance vulnerabilities should be specifically addressed in safety analyses. Second, the paradoxical workload findings (higher subjective workload but faster response times with management-by-exception) raise concerns about optimal attention allocation. Faster responses may come at the cost of reduced situation awareness or decreased ability to detect anomalies outside the primary task focus. Safety analyses should consider these potential tradeoffs rather than assuming faster response times necessarily indicate improved safety. Third, the low trust scores despite high automation reliability suggest operators may be undertrusting the system. While undertrust can prevent complacency, it may also lead to unnecessary manual interventions and increased workload during critical operations. This potential source of operational instability warrants safety consideration.

In this concluding section, we address three focal questions. First, what are the implications of introducing CBPs supported by automation for operator workload? The current task design appeared to lower workload and reduce inter-task differences compared to previous studies using more traditional configurations. Second, how does LOA affect workload response? We found workload differences between management-by-consent and management-by-exception conditions which tended to favor the lower LOA, although the operational significance of these LOA effects remains to be determined. Third, how does operator trust affect safety of automated systems? Analyses of trust in this experiment highlight individual variation in trust but further work is needed to explore safety implications.

### 5.5.1 Level of Automation has Mixed Impacts

Choice of LOA is critical for implementation of automation in safety critical industries (Janssen et al., 2019; Parasuraman et al., 2000). In general, intermediate LOAs tend to be preferred. Low LOAs may not offer much benefit over manual control and high LOAs can elicit out-of-the loop issues, loss of situation awareness, and vigilance decrement (Kaber et al., 2000; Qing et al., 2021). The present study showed that, in some cases, the lower intermediate LOA (management-by-consent), tended to elicit lower workload than a higher intermediate LOA, (management-by-exception). This effect of LOA was most evident in ISA and EEG measures for workload. NASA-TLX data implied that the paradoxically higher workload of management-by-consent might relate to temporal demands. The operator must keep track of the time window allowed for over-riding the automation which may raise cognitive demands. On the other hand, response times were faster in the management-by-exception condition, consistent with previous findings on automation impacts (e.g., Huang et al., 2006). Further research is necessary to determine whether the workload costs of the higher LOA justify the speeding of operations. Overall levels of workload remained low with management-by-exception which may support the use of this LOA.

### 5.5.2 Trust Impacts Require More Research

Operator trust in automation and its performance and safety impacts is a major and well-research issue in human factors (Lee & See, 2004; Parasuraman & Riley, 1997). However, empirical research on trust impacts in NPP operations is lacking. The investigation of trust was a subsidiary issue in the current research. However, results showed considerable individual variation in trust at both trait and state levels. Individual variation may well be reduced by training and experience. HPTF research has examined the comparability of workload data in novice and experienced samples (Hughes et al., 2023a), but it has not investigated trust data in experienced operators, so degree of generalization is unclear. Nevertheless, the current data point to the risk that some operators are over-trusting whereas others are under-trusting leading to the performance failings associated with over- and under-reliance respectively (Parasuraman & Riley, 1997). Further attention to trust in future research is warranted.

# SECTION 6: EXPERIMENT 7 METHODOLOGY

## 6.1 Overview of Experiment 7: Impact of Workload Manipulation and Task Type on Performance, Workload, and Trust

Experiment 7 built on Experiment 6 by introducing additional task types and manipulating workload during a simulated EOP, incorporating automation failures as an unexpected event. This study focused on the integration of a task similar to what would be performed during load-following operations, a potentially relevant challenge for future power plants, especially SMRs and other advanced designs that may operate flexibly with variable energy sources (Clayton & Wood, 2010; O'Hara & Higgins, 2012). Load-following requires operators to adjust power output in response to changes in grid demand, which may add complexity and strain on human resources during operations.

This study also introduced a novel workload manipulation where the relevant I&Cs were highlighted with varied specificity. This manipulation was designed to investigate whether sensitivity to workload measures varies depending on workload level. Participants also performed a secondary task of monitoring gauges, and their response times were recorded as a behavioral workload measure.

The study used management-by-consent as the LOA, which was the same level used in Experiment 6, to explore how task complexity and varying workload impact operator performance and trust. The specific hypotheses tested included:

1. **Workload:**
    a. Task types, including a power level change, detection, and checking, were expected to show differences in subjective and objective workload measures. The power level change task required multiple steps and a period of monitoring. It is expected that the power change task would impose higher workload due to its complexity, though this might diminish with practice.

    b. The less specific control highlighting will induce a higher workload and exacerbate task-type differences, particularly for detection tasks.

2. **Performance:**

a. Detection tasks, requiring vigilance, were expected to have the lowest performance accuracy, particularly when the workload manipulation was introduced.

b. The power level change task, due to their complexity, were also expected to show reduced accuracy.

3. **Trust:**

   Trust was measured using the CPTA scale, with the hypothesis that automation failure would lead to reduced trust, especially at higher workload levels. Individual differences in trust, previously observed by Lin et al. (2024), were also explored to examine how dispositional trust impacts automation use, workload, and performance.

## 6.2 Participants

A total of 31 participants (*M* = 21.65, *SD* = 6.63) were recruited. Participants received course credits (1 credit/hour) in compensation for their time. There were 23 men, 8 women. They had no prior experience operating an NPP control system of any kind. All participants reported normal or corrected-to-normal vision and no color vision deficiency or history of neurological disorders.

The number of participants varied between measures and scenarios for multiple reasons. In some cases, participants failed to complete both scenarios, therefore the post-surveys for those cases were inevitably missing. There were also a reduced number of cases for some physiological measures due to sensor-related issues (e.g., missing raw data, connection failure, data parsing errors, etc.).

## 6.3 Training

Training followed the same structure as the previous study: an overview using PowerPoint slides to go through the task environment and each component of the system, a demonstration with the live training scenario on the simulator, as well as a hands-on training of the simulator under necessary guidance and supervision to make sure that the participants were familiar with the simulator. All participants completed the practice tasks successfully and exhibited proficiency in the simulator interface and controls.

## 6.4 Experimental Design

A two (Workload level: low and high) × four (Task type: checking, detection, response implementation, and power change) within-subjects design was adopted in this study. Participants completed two experimental scenarios in low and high workload configurations respectively in a balanced order. Each experimental scenario consisted of 45 task steps, including 17 checking tasks, four detection tasks, four response implementation tasks, and 20 additional power change tasks. The number and order of the task steps were not balanced due to the nature of the original procedures, which requires steps to be taken in a prescribed sequence. ECA-00 and OP-131.01 were used in this study.

Automation was available in the management-by-consent mode to provide navigational and operational assistance for checking, detection, and response implementation tasks. Specifically, the navigational assistance varied between the two experimental scenarios to induce the workload manipulations (see more in 2.4.1). In terms of operational assistance, the automation would assess the system and provide its evaluation for each task. Explicit endorsement or over-ride of the automation is needed from the human operator. The automation reliability was set to be close to 90% and was described as highly reliable but not perfect to participants during training. In the total of forty-five task steps, two checking steps, one detection step, one response implementation step, and two additional steps in the power change tasks were given incorrect automation outcomes.

## 6.5 Independent Variables

The independent variables in this experiment were workload level (low and high) and task type (checking, detection, response implementation, and power change).

### 6.5.1 Workload Level

Workload level was manipulated through the navigational assistance provided by the automation. Specifically, the navigation cue provided by the automation varied between the two scenarios. In the low workload scenario, the automation would highlight the specific target I&C for each task by adding a blue box around the I&C (Figure 13a). By contrast, in the high workload scenario, the automation would highlight a larger area in which the target I&C is located (Figure 13b). In this case, the target I&C may be any one of the cluster of I&Cs inside the larger blue box. This navigation process was expected to demand more effort from participants and, in turn, induce higher workload.

a b

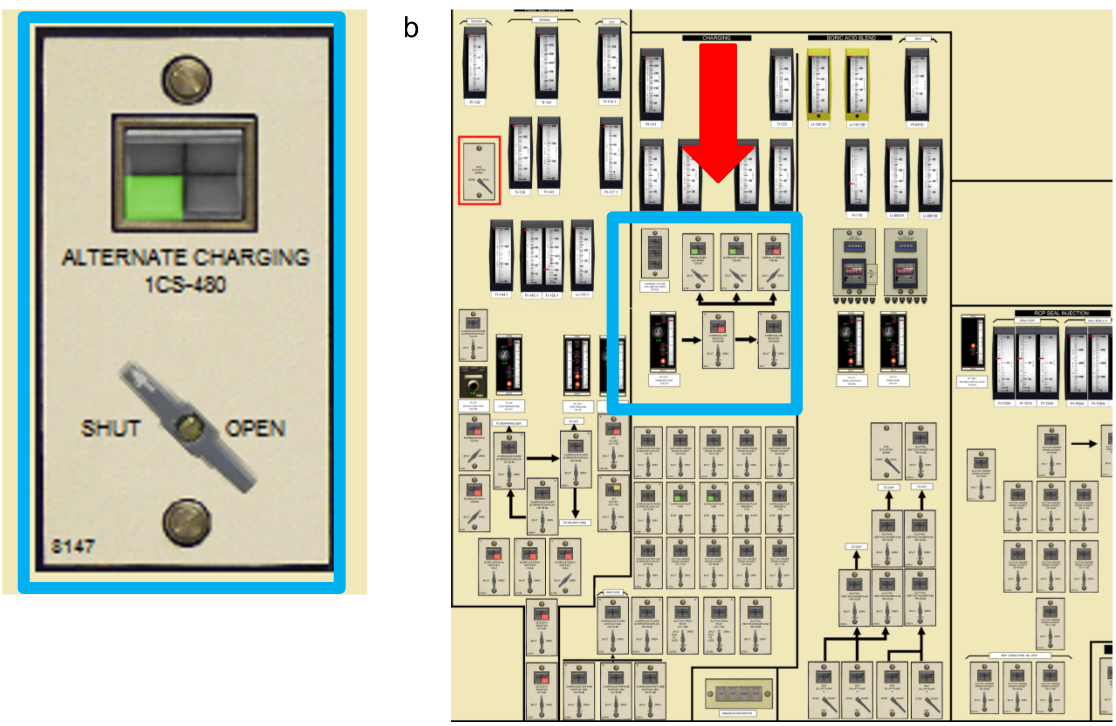


**Figure 13 Navigation Cues**

### 6.5.2 Task Type

The task type consisted of four conditions. The checking task type required a one-time inspection of an I&C to verify that it was in the state called for by the EOP. Participants were required to locate various I&Cs and indicate identification by clicking on the correct I&C. The detection task type required participants to correctly locate an instrument and then to continuously monitor that instrument parameter for identification of change. The response implementation task type required participants to locate a control and subsequently manipulate the control in the required direction (i.e., open or shut). The power change task type required participants to follow the OP-131.01, minor power correction procedure to authorize the automation-suggested action when the load rate is within the safe range.

The power change task type was a novel addition to task type manipulations. The demand for electricity fluctuates throughout the day and power plants may adjust their power output (or load rate) accordingly. Participants were required to follow the power change procedure to authorize the automation-suggested action when load rate is within the safe range using the load following panel (Figure 14). Specifically, after acknowledging the task in the computerized procedure tool, participants needed to tap OPER AUTO button on the load following panel to start the authorization process. Then participants needed to determine whether the load rate provided by automation exceeded 50 MW/min (the safe range set for this study) or not. If the rate was less than 50 MW/min, participants were asked to click true, OK (in the computerized procedure tool), tap GO (on the panel); otherwise, if the rate was greater than 50 MW/min, they were instructed to click false, OK (in the computerized procedure tool), tap RESET (on the panel), and finally click next (in the computerized procedure tool) to proceed.

**Figure 14 Load Following Panel**

In addition to the general tasks in the three above types, participants were also instructed to monitor a group of twelve gauges on the panel throughout the scenario as the secondary task. When the gauges fell below 20%, participants were instructed to react by tapping the label of the gauge to acknowledge. This happened 10 times for each scenario.

## 6.6 Scenario Setup

The experimental scenario was developed based on a generic version for a “Loss of All Alternating Current (AC) Power (ECA-0.0)” emergency operating procedure (EOP) but modified for experimental use. There was a total of 45 task steps, including 17 checking tasks, four detection tasks, four response implementation tasks, and 20 additional power change tasks. A computerized procedure tool was provided on a separate monitor to prompt participants to follow the instructions to complete the tasks (Figure 15). Not only did the computerized procedure tool provide task instructions but offered automation aid to provide navigational and operational assistance. The system was in a two-column layout. Instructions associated with each workstation were shown in its dedicated column.

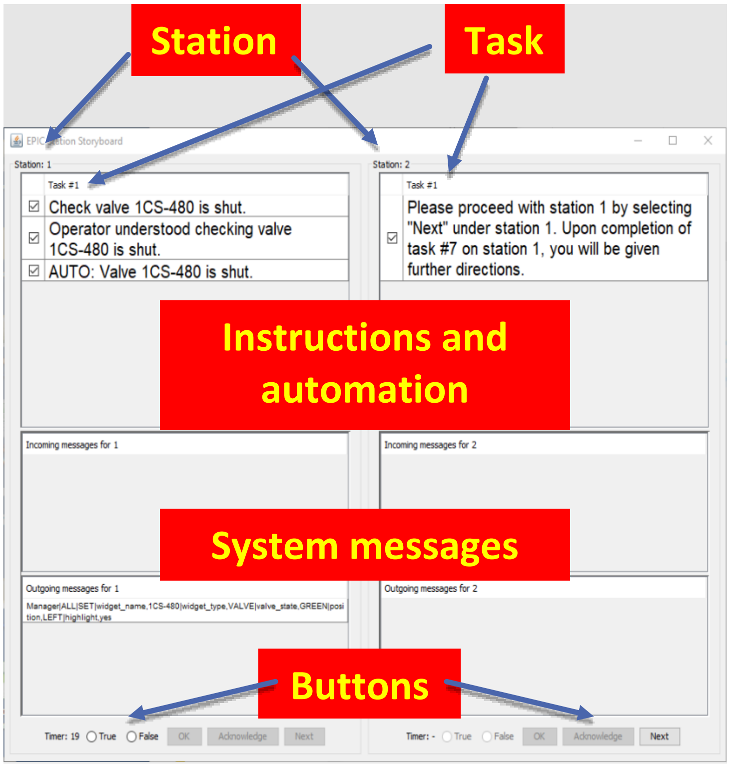


**Figure 15 Computerized Procedure Tool**

Automation was configured at an intermediate level (management-by-consent), which requires participants to approve evaluations made by the automation. The reliability of the automation was approximately 90%. In other words, participants would encounter six automation failures (automation providing incorrect evaluations) during each experimental scenario. In checking tasks, automation would provide its evaluation in 10 seconds after the task was acknowledged. In response implementation tasks, automation would not only provide its evaluation but complete the desired action in case participants did not or chose to not perform the manual action in 10 seconds after the task was acknowledged. In detection tasks, which required five-minute continuous monitoring on a specific gauge for each step, automation was monitoring the status of the specific gauge and would provide its evaluation after the five-minute period with a 10 second delay. In load following tasks, automation provided a suggested load rate instead of a status evaluation. Participants were told that the automation was not always correct, and it would provide incorrect information occasionally. In case that automation failed, it would provide an incorrect evaluation of the system status for checking, detection, and response implementation tasks (e.g., indicate a valve is shut when it is actually open etc.) or provide a load rate that exceeded the safe range for load following tasks (e.g., suggest a rate that is greater than 50 MW/min). They were asked to approve or reject the evaluations or suggested load rate made by the automation based on their strategy and judgement and use the true or false buttons in the computerized procedure tool to complete each task.

## 6.7 Performance Measures

Performance measures included multiple measures to characterize accuracy (e.g., hits, correct rejections, misses, false alarms) and reaction times for primary and secondary tasks. Not all measures were applicable for every task type due to the nature of the task type.

### 6.7.1 Task Accuracy Measures

The definitions of each accuracy metric for primary and secondary tasks are provided in Table 9.

**Table 9**
*Definitions of Task Accuracy Measures*

| Metric | Definition |
|---|---|
| **Checking Task** | |
| Correct Actions | Correct hits and correct rejects |
| • Correct Hits | The number of actions that participants agree with the correct automation evaluations |
| • Correct Rejects | The number of actions that participants reject the incorrect automation evaluation successfully |
| Misses | The number of actions that participants fail to agree with the correct automation evaluations |
| False Alarms | The number of actions that participants fail to reject the incorrect automation evaluation |
| **Detection Task** | |
| Correct Hits | Correct hits auto and correct hits manual |
| • Correct Hits Auto | The number of actions that participants notice a threshold is reached (responded in computerized procedure tool) |
| • Correct Hits Manual | The number of actions that participants notice a threshold is reached (responded on the panel manually) |
| Misses | The number of actions that participants fail to notice a threshold is reached (fail to agree with the correct automation evaluations) |
| False Alarms | The number of erroneous actions (tap when the threshold is not reached) |
| Additional Actions | The number of duplicate (unnecessary) clicks (e.g., unnecessary actions on the target I&Cs) |
| Invalid Actions | The number of incorrect clicks a participant takes (e.g., actions on other I&Cs) |
| **Response Task** | |
| Correct Hits | Correct hits auto and correct hits manual |
| • Correct Hits Auto | The number of actions that participants agree with the correct automation evaluations (responded in computerized procedure tool) |
| • Correct Hits Manual | The number of actions that participants take correct manual actions (responded on the panel manually) |
| Misses | The number of actions that participants fail to agree with the correct automation evaluations |

| | |
|---|---|
| Additional Actions | The number of duplicate (unnecessary) clicks (e.g., unnecessary actions on the target I&Cs) |
| Invalid Actions | The number of incorrect clicks a participant takes (e.g., actions on other I&Cs) |
| **Power Change Task** | |
| Correct Actions | Correct hits and correct rejects |
| • Correct Hits | The number of actions that participants select TRUE when the rate is less than 50 |
| • Correct Rejects | The number of actions that participants select FALSE when the rate is greater than 50 |
| Misses | The number of actions that participants select FALSE when the rate is less than 50 |
| False Alarms | The number of actions that participants select TRUE when the rate is greater than 50 |
| Additional Actions | The number of duplicate (unnecessary) clicks (e.g., unnecessary actions on the target I&Cs) |
| Invalid Actions | The number of incorrect clicks a participant takes (e.g., actions on other I&Cs) |
| Correct OperAuto | The number of tasks with clicks on OperAuto button |
| Correct Go | The number of actions that participants tap GO when the rate is less than 50 |
| Correct Reset | The number of actions that participants select RESET when the rate is greater than 50 |
| Miss OperAuto | The number of tasks with no click on OperAuto button |
| **Secondary Task** | |
| Correct Hits | The number of actions that participants notice a gauge is below 20% |

### 6.7.2 Reaction Time Measures

Table 10 Table 10
*Definitions of Reaction Time* summarizes the definitions of each reaction measure for each task type manipulation.

**Table 10**
*Definitions of Reaction Time Measures*

| Metric | Definition |
|---|---|
| **Checking Task** | |
| Hit RT | The time a participant takes to check and take a correct action (hit) |
| Reject RT | The time a participant takes to check and correctly withhold action (correct rejection) |
| Miss RT | The time a participant takes to check but take an incorrect action (miss) |
| **Detection Task** | |
| Hit RT | The time a participant takes to notice a change and take a correct action |
| **Response Task** | |
| Hit RT | The time a participant takes to respond and take a correct action (hit) |
| Miss RT | The time a participant takes to respond and take an incorrect action (miss) |
| **Power Change Task** | |
| Hit RT | The time a participant takes to respond and take a correct action in CPT (hit) |
| Reject RT | The time a participant takes to respond and take a correct action in CPT (correct rejection) |
| Correct Go RT | The time a participant takes from acknowledgement to click go on the panel |
| Correct Reset RT | The time a participant takes from acknowledgement to click reset on the panel |
| **Secondary Task** | |
| Hit RT | The time a participant takes to notice a change and take a correct action |

### 6.7.3 Automation Dependence Metric

Automation dependence is defined as the percentage of task steps in which the participant followed the recommendation from the automation. Detailed automation dependence metric formulas for each task type are listed in Table 11.

**Table 11**

*Automation Dependence Measures*

| Metric | Formula |
|---|---|
| Checking | $\frac{Correct\ Hits + Correct\ Rejections}{17} \times 100\%$ |
| Detection | $\frac{Correct\ Hits\ Auto}{4} \times 100\%$ |
| Response | $\frac{Correct\ Hits\ Auto}{4} \times 100\%$ |
| Power Change | $\frac{Correct\ Hits + Correct\ Rejections}{20} \times 100\%$ |

## 6.8 Procedures

Participants were asked to read a consent form approved by the UCF IRB and were then briefed on the experimental procedure. Following consent, they took a pre-task survey set which included a demographics questionnaire, the PAS, and the HIT trust scale. Then participants were given comprehensive training with a PowerPoint presentation, a live scenario demonstration, and a hands-on practice session, all of which were done with the researcher present. Once these were complete, they were equipped with noninvasive physiological sensors, including EEG, ECG, and fNIRS, and a 5-minute resting baseline was collected. Participants interacted with two scenarios, each with a different workload level manipulation. Order of the two scenarios was counterbalanced. During each scenario, participants were asked to rate their workload level on a scale of one to five (ISA self-report scale) after specific steps in each task type. After each scenario was completed, participants filled out a series of post-task questionnaires including the NASA-TLX, the MRQ, and the CTPA. After this, the sensors were removed, and the participant was debriefed then dismissed.

# SECTION 7: EXPERIMENT 7 RESULTS

## 7.1 Subjective Measures

### 7.1.1 NASA-TLX

Paired-samples *t*-tests with Bonferroni corrections were conducted to determine whether there was a statistically significant mean difference between the subjective workload measured by NASA-TLX in different workload manipulations. As with the previous analysis for Experiment 6, the use of Bonferroni corrections may have been overly conservative, possibly masking additional meaningful differences among NASA-TLX subscales in the workload manipulations. This is noted here as a limitation. Participants reported higher mental demand in the high workload condition ($M$ = 31.04, $SD$ = 21.21) than in the low workload condition ($M$ = 23.93, $SD$ = 17.65), $t(27) = -2.38$, $p < .05$.

**Table 12**
*Summary of NASA-TLX between Workload Manipulations*

| | Low Workload | | | High Workload | | | 95% CI for Mean | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | Difference | *t* | *df* |
| Global Workload | 17.11 | 11.12 | 28 | 17.40 | 11.54 | 28 | [-4.87, 4.29] | -.13 | 27 |
| Mental Demand | 23.93 | 17.65 | 28 | 31.04 | 21.21 | 28 | [-13.24, -.98] | -2.38* | 27 |
| Physical Demand | 17.32 | 14.94 | 28 | 16.71 | 16.56 | 28 | [-5.43, 6.65] | .21 | 27 |
| Temporal Demand | 13.54 | 11.18 | 28 | 12.61 | 11.47 | 28 | [-3.8, 5.66] | .40 | 27 |
| Effort | 18.86 | 20.32 | 28 | 19.36 | 17.60 | 28 | [-7.98, 6.98] | -.14 | 27 |
| Frustration | 15.61 | 19.80 | 28 | 13.96 | 22.59 | 28 | [-4.84, 8.13] | .52 | 27 |
| Performance | 13.43 | 20.21 | 28 | 10.71 | 10.49 | 28 | [-5.17, 10.60] | .71 | 27 |

**$p < .01$, *$p < .05$

### 7.1.2 MRQ

Paired-samples *t*-tests with Bonferroni corrections were conducted to determine whether there was a statistically significant mean difference between the mental processes measured by MRQ between different workload manipulations. The differences were not statistically significant (Table 13).

**Table 13**
*Summary of MRQ between Workload Manipulations*

| | Low Workload | | | High Workload | | | 95% CI for Mean | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | Difference | *t* | *df* |
| Auditory Emotional | 6.50 | 19.93 | 28 | 4.18 | 8.59 | 28 | [-5.76, 10.40] | .59 | 27 |
| Auditory Linguistic | 6.14 | 14.40 | 28 | 15.29 | 26.97 | 28 | [-18.92, .64] | -1.92 | 27 |
| Manual Process | 49.46 | 27.16 | 28 | 51.50 | 28.95 | 28 | [-10.35, 6.28] | -.50 | 27 |
| Short Term Memory | 43.39 | 21.90 | 28 | 50.54 | 27.02 | 28 | [-15.68, 1.39] | -1.72 | 27 |
| Spatial Attentive | 61.96 | 26.85 | 28 | 67.86 | 25.66 | 28 | [-14.42, 2.63] | -1.42 | 27 |
| Spatial Concentrative | 43.04 | 28.91 | 28 | 41.43 | 29.87 | 28 | [-10.05, 13.26] | .28 | 27 |
| Spatial Categorical | 52.32 | 32.98 | 28 | 47.68 | 30.44 | 28 | [-6.15, 15.43] | .88 | 27 |
| Spatial Emergent | 52.64 | 30.05 | 28 | 58.71 | 28.30 | 28 | [-15.90, 3.75] | -1.27 | 27 |
| Spatial Positional | 59.29 | 29.15 | 28 | 58.93 | 28.03 | 28 | [-6.30, 7.01] | .11 | 27 |
| Spatial Quantitative | 47.86 | 29.95 | 28 | 48.46 | 28.89 | 28 | [-10.47, 9.25] | -.13 | 27 |
| Visual Lexical | 55.89 | 29.00 | 28 | 60.68 | 29.33 | 28 | [-18.17, 8.61] | -.73 | 27 |
| Visual Phonetic | 29.75 | 30.45 | 28 | 31.00 | 33.46 | 28 | [-11.26, 8.76] | -.26 | 27 |
| Visual Temporal | 40.75 | 31.68 | 28 | 52.14 | 30.47 | 28 | [-23.54, .76] | -1.92 | 27 |
| Vocal Process | 7.04 | 11.92 | 28 | 9.50 | 18.76 | 28 | [-9.18, 4.25] | -.75 | 27 |

**$p < .01$, *$p < .05$

### 7.1.3 ISA

A two (Workload: low and high) × four (task type: checking, detection, response implementation, and load following) repeated ANOVA was run to test the effects of experimental manipulations on the subjective workload measured by ISA. The analysis revealed neither significant main effect nor interaction effect (Table 14)
.

**Table 14**
*Two-way ANOVA Statistics for ISA Ratings*

| | Low Workload | | | High Workload | | | ANOVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | Effect | *F* | *df* | $\eta_p^2$ |
| Task type | | | | | | | | | | |
| Checking | 1.70 | .72 | 27 | 1.89 | .89 | 27 | Workload | 1.91 | 1, 26 | .07 |
| Response | 1.63 | .69 | 27 | 1.89 | .89 | 27 | Task type | 1.20 | 2.5, 63.8 | .04 |
| Detection | 1.70 | .82 | 27 | 1.67 | .78 | 27 | W × T | .80 | 3, 78 | .03 |
| Load | 1.52 | .70 | 27 | 1.63 | .88 | 27 | | | | |

$^{**}p < .01$, $^{*}p < .05$, W × T: Workload × Task type

### 7.1.4 CTPA

Paired-sample *t*-tests were conducted to determine whether participants trusted the configurations in the two scenarios differently. There were no significant findings in the analysis.

## 7.2 Physiological Measures

All physiological dependent variables entered into the ANOVAs were calculated as differences in percentage to five-minute resting baseline. For example, if a participant's heart rate for the five-minute baseline was 60 and the heart rate for the subsequent checking task became 70, the raw difference from baseline would be 10, and the percentage change would be 16.67%. This method helps account for individual differences when comparing group means as is the case when running ANOVAs.

### 7.2.1 Physiological Measures Electroencephalogram (EEG)

Brain activity was recorded at nine EEG sensor sites, the EEG data were analyzed by grouping sensor sites by hemispheres (i.e., compare brain activity between the left and right hemispheres) and lobes (i.e., compare brain activity among the frontal, parietal and occipital lobes). Shapiro-Wilk test of normality suggested that the EEG data were positively skewed which can violate the common assumption of dependent variables being approximately normally distributed. To correct the positively skewed EEG data, logarithmic transformation was applied to the percentage change score from baseline. Specifically, the percentage change score from baseline was transformed by taking $\log_{10}$.

#### *7.2.1.1 Hemispheres*

Two (Workload: low and high) × four (task type: checking, detection, response implementation, power change) × two (Hemisphere: left and right) repeated ANOVAs were run for theta, alpha, and beta frequency bands separately. These ANOVAs provided insight into the overall effects of workload and task type manipulations on the left and right hemispheres.

For theta band, a significant main effect was found for hemisphere, $F(1,27) = 15.10$, $p < .01$, $\eta_p^2 = .36$. Right hemisphere ($M = 1.17$ $SE = .20$) showed greater increase from baseline than left hemisphere ($M = .76$ $SE = .19$). The main effect of task type on EEG theta response was also significant, $F(2.07, 55.91) = 4.61$, $p < .05$, $\eta_p^2 = .15$. Participants showed greater EEG theta increase in checking tasks ($M = 1.07$ $SE = .21$) than in detection tasks ($M = .74$ $SE = .18$).

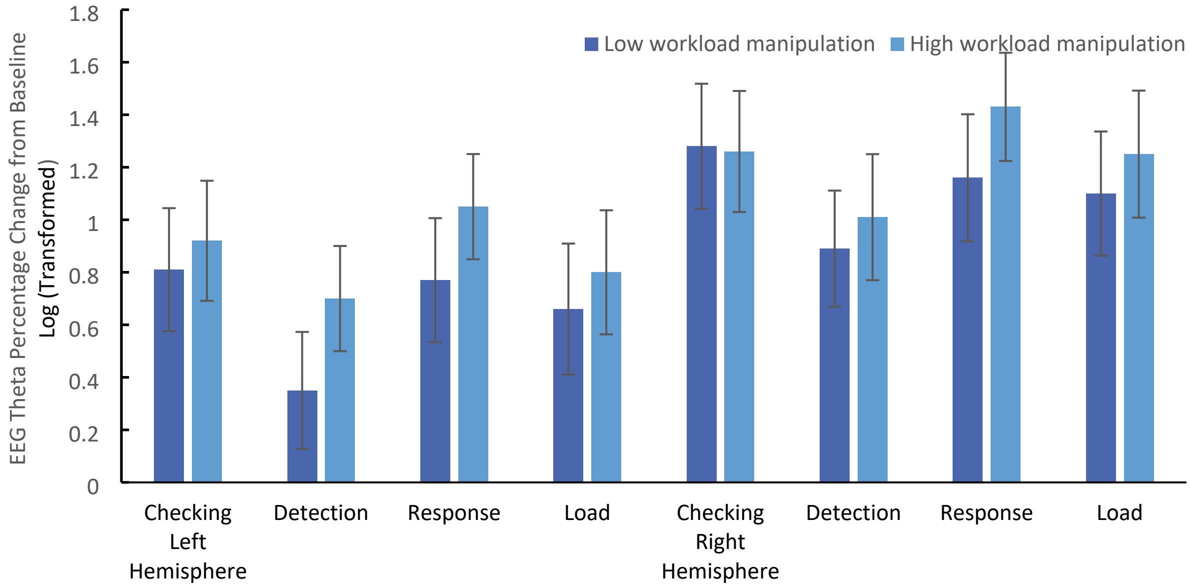


**Figure 16 EEG Theta Log Transformed Percentage Change from Baseline by Task Type and Hemisphere (error bars denote standard errors)**

For alpha band, the main effect of hemisphere was significant, $F(1,27) = 7.19$, $p < .05$, $\eta_p^2 = .21$. Right hemisphere ($M = .85$ $SE = .25$) showed greater increase from baseline than left hemisphere ($M = .41$ $SE = .25$).

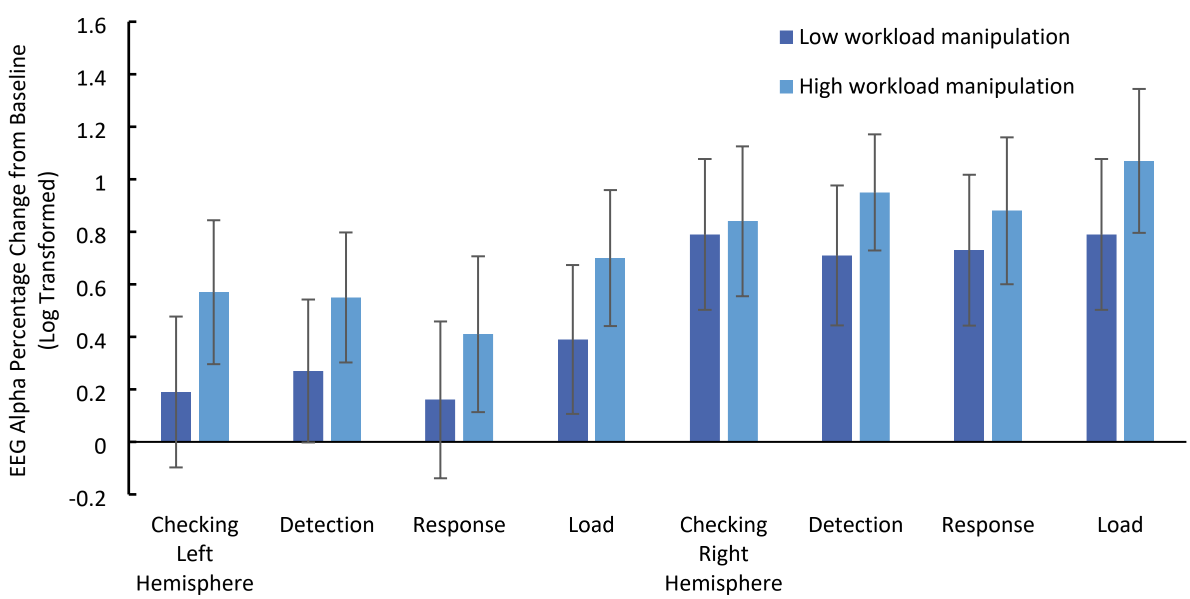


**Figure 17 EEG Alpha Log Transformed Percentage Change from Baseline by Task Type and Hemisphere (error bars denote standard errors)**

In terms of EEG beta band, a significant main effect was found for task type, $F(3, 81) = 7.61$, $p < .01$, $\eta_p^2 = .22$. Overall, the detection task type ($M = 1.27$ $SE = .18$) showed a smaller increase from baseline than the checking task type (*M* = 1.54 *SE* = .17), the response implementation task type (*M* = 1.38 *SE* = .22), and the power change task type (*M* = 1.65 *SE* = .15).

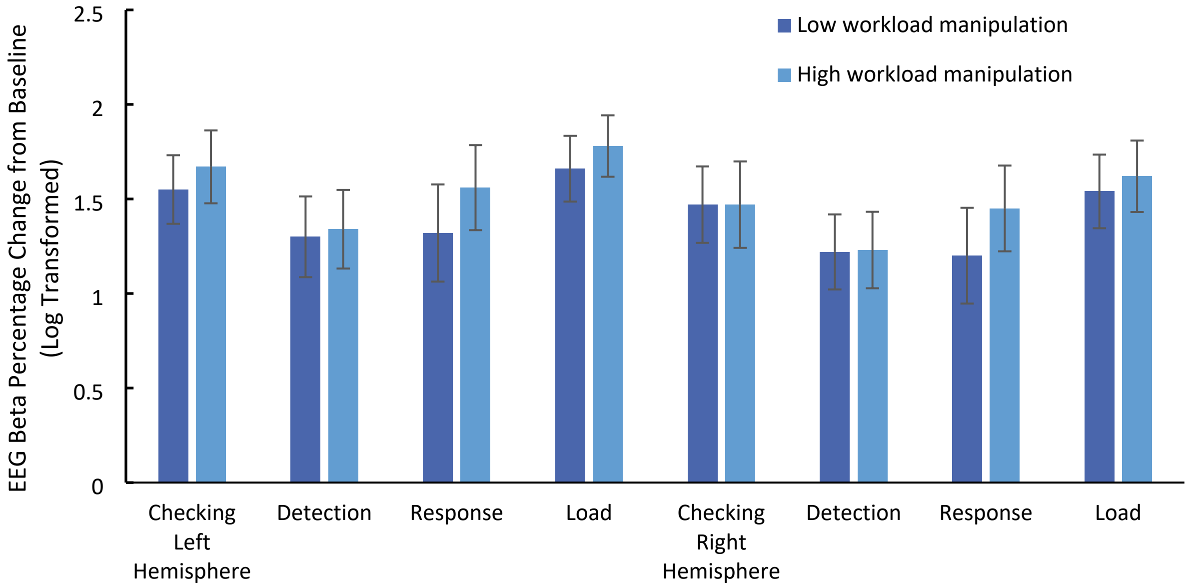


**Figure 18 EEG Beta Log Transformed Percentage Change from Baseline by Task Type and Hemisphere (error bars denote standard errors)**

#### *7.2.1.2 Lobes*

Two (Workload: low and high) × four (task type: checking, detection, response implementation, power change) × three (Lobe: front, parietal, and occipital) repeated ANOVAs were run for theta, alpha, and beta frequency bands separately. These ANOVAs provided insight into the overall effects of workload and task type manipulations on the change of electrical activity in different areas of the brain in terms of frontal, parietal, and occipital lobes.

For theta band, a significant main effect was found for task type, $F(1.90, 51.41) = 6.01$, $p < .01$, $\eta_p^2 = .18$ (Figure 19). Participants showed significantly less theta increase in detection tasks (*M* = .67 *SE* = .19) than in checking tasks (*M* = 1.05 *SE* = .21).

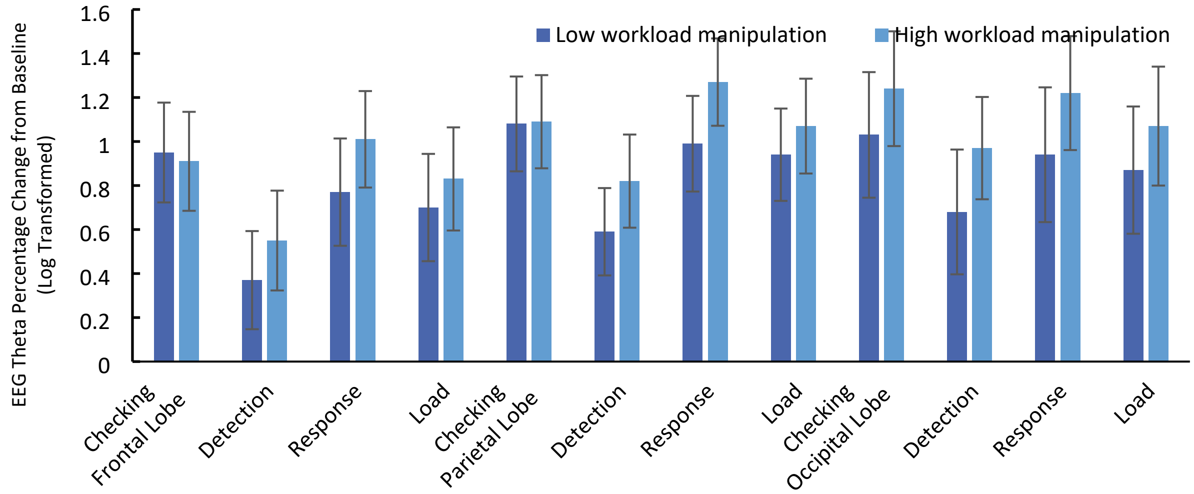


**Figure 19 EEG Theta Log Transformed Percentage Change from Baseline by Task Type and Lobe (error bars denote standard errors)**

For EEG beta band, a significant main effect was found for task type, $F(3, 81) = 8.40$, $p < .01$, $\eta_p^2 = .24$ (Figure 20). Participants showed significantly less beta increase in detection tasks ($M$ = 1.31 $SE$ = .15) than in checking tasks ($M$ = 1.60 $SE$ = .15), response implementation tasks ($M$ = 1.45 $SE$ = .19), and power change tasks ($M$ = 1.67 $SE$ = .14).

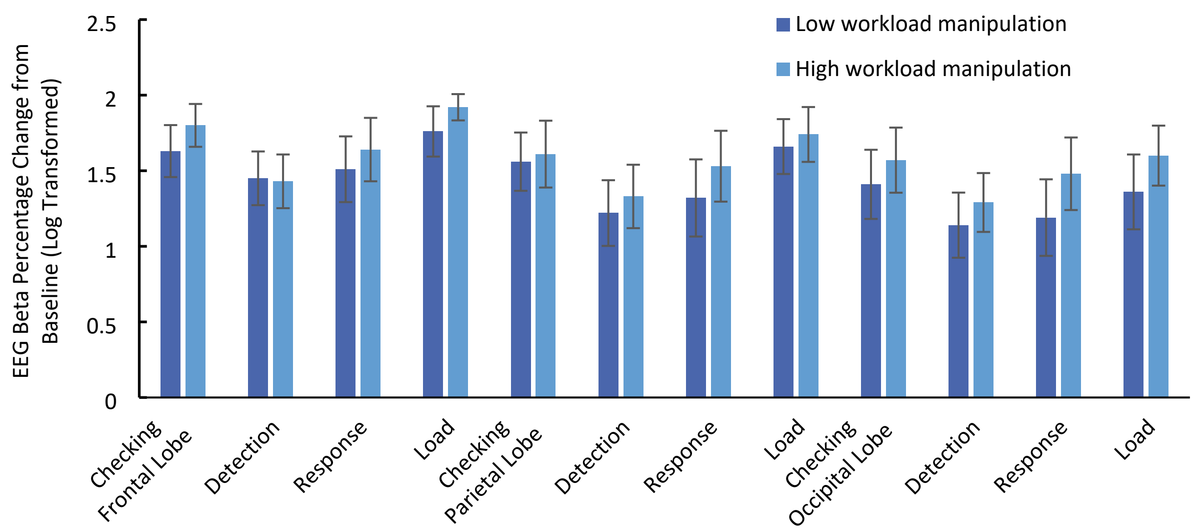


**Figure 20 EEG Beta Log Transformed Percentage Change from Baseline by Task Type and Lobe (error bars denote standard errors)**

### 7.2.2 Functional Near-Infrared Spectroscopy (fNIRS)

A two (Workload: low and high) × four (task type: checking, detection, response implementation, power change) × two (Hemisphere: left and right) repeated ANOVA was run to determine the effect of workload and task type manipulations on regional oxygen saturation (rSO2) for the left and right prefrontal cortex (Figure 21). No significant main effect was found for hemisphere, $F$(1,

26) = 4.24, $p$ = .05, $\eta_p^2$ =.14. A significant main effect was found for task type, $F(2.03, 52.88)$ = 10.17, $p < .01$, $\eta_p^2$ =.28. Specifically, a greater decrease in rSO2 was observed in left prefrontal cortex (M = -2.79 SE = .66) than in right prefrontal cortex (M = -2.15 SE = .63). On a task type basis, the decrease in rSO2 during detection tasks (M = -1.83 SE = .62) was less than the rSO2 decrease during checking tasks (M = -2.74 SE = .69), response implementation tasks (M = -2.64 SE = .62), and power change tasks (M = -2.66 SE = .62). In addition, the interaction between hemisphere and workload was also significant, $F(1, 26)$ = 6.29, $p$ = .05, $\eta_p^2$ =.20. Follow-up four (task type: checking, detection, response implementation, load following) × two (Hemisphere: left and right) repeated ANOVAs were run for two workload conditions separately. It was revealed that the effect of hemisphere was only significant in high workload condition, $F(1, 26)$ = 9.38, $p < .01$, $\eta_p^2$ =.27, not in low workload condition, $F(1, 26)$ = .34, $p$ = .57, $\eta_p^2$ =.01.

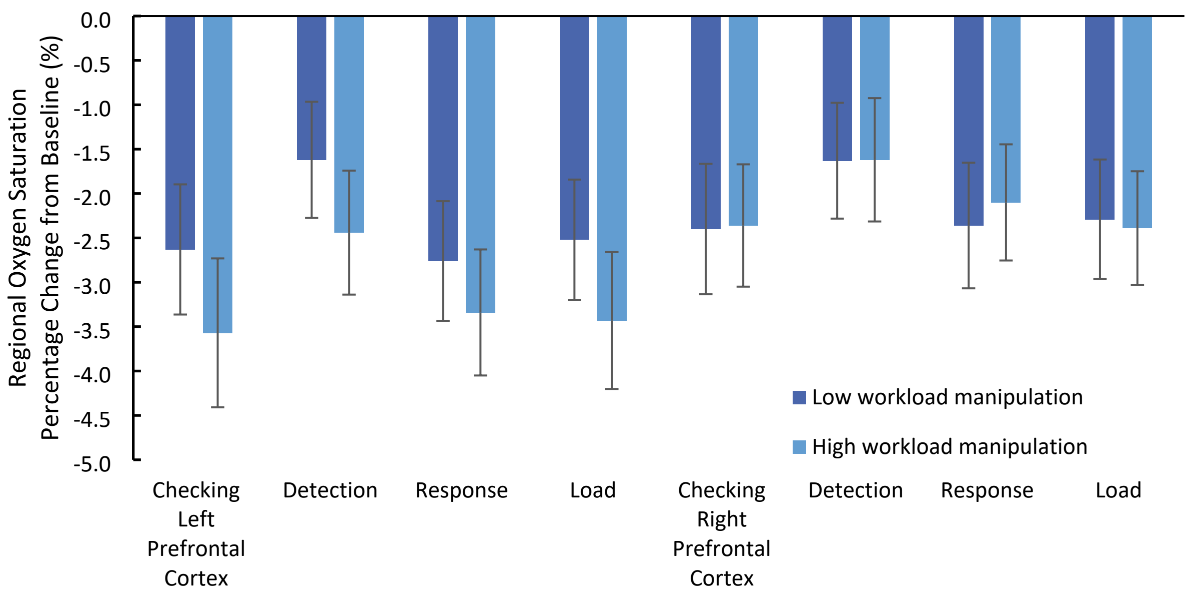


**Figure 21 Regional Oxygen Saturation Percentage Change from Baseline by Task Type and Hemisphere (error bars denote standard errors)**

### 7.2.3 Electrocardiogram (ECG)

A two (Workload: low and high) × four (task type: checking, detection, response implementation, power change) repeated-measures ANOVA was conducted to determine the effect of workload and task type manipulations on HR, HRV, and IBI. The ECG measures, HR, IBI, and HRV, were derived from R-Peak detections using the So-Chan QRS algorithm from the raw ECG signal. No significant main effect or interaction was found for the HR, HRV, or IBI measures.

**Table 15**

*Two-way ANOVA Statistics for Heart Rate*

| | Low Workload | | | High Workload | | | ANOVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | Effect | *F* | *df* | $\eta_p^2$ |
| HR | | | | | | | | | | |
| Checking | 3.28 | 24.14 | 19 | 8.30 | 28.29 | 19 | Workload | .25 | 1, 18 | .01 |
| Response | 14.28 | 23.80 | 19 | 8.25 | 29.91 | 19 | Task type | 2.69 | 3, 54 | .13 |
| Detection | 1.21 | 19.94 | 19 | 2.51 | 23.33 | 19 | W × T | 1.32 | 1.4, 25.2 | .07 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Load | 7.67 | 33.18 | 19 | 12.94 | 32.26 | 19 |

**$p$ < .01, *$p$ < .05, W × T: Workload × Task type

## 7.3 Performance Measures

The accuracy measures were converted to percentages to make comparisons. Analyses in this section will focus on task accuracy (e.g., correct actions) and reaction time for correct actions. The correlational analysis in the next section will include all available measures.

### 7.3.1 Task Accuracy

A two (Workload: low and high) × five (task type: checking, detection, response implementation, power change, secondary) repeated-measures ANOVA was conducted to explore the effect of workload and task type manipulations on task performance in terms of accuracy. The analysis revealed a significant main effect of task type on task accuracy, $F(1.6, 44.4) = 28.37$, $p < .01$, $\eta_p^2 = .50$. Detection tasks received the lowest accuracy ($M$ = 83.19 $SE$ = 3.41) among the four primary task types, and load following tasks received the highest accuracy ($M$ = 97.33 $SE$ = 1.59) among the four primary task types. Compared to primary task accuracy, the secondary task accuracy ($M$ = 58.10 $SE$ = 6.55) was significantly lower.

**Table 16**

*Two-way ANOVA Statistics for Task Accuracy*

| | Low Workload | | | High Workload | | | ANOVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | Effect | *F* | *df* | $\eta_p^2$ |
| Accuracy% | | | | | | | | | | |
| Checking | 95.74 | 6.07 | 29 | 93.10 | 16.42 | 29 | Workload | .00 | 1, 28 | .00 |
| Response | 93.10 | 16.22 | 29 | 91.38 | 21.42 | 29 | Task type | 28.37** | 1.6, 44.4 | .50 |
| Detection | 81.03 | 24.69 | 29 | 85.34 | 21.67 | 29 | W × T | .81 | 2.6, 74.1 | .03 |
| Power Change | 98.45 | 5.02 | 29 | 96.21 | 15.79 | 29 | | | | |
| Secondary | 56.90 | 38.83 | 29 | 59.31 | 34.11 | 29 | | | | |

**$p$ < .01, *$p$ < .05, W × T: Workload × Task type

### 7.3.2 Reaction Time

A two (Workload: low and high) × five (task type: checking, detection, response implementation, power change, secondary) repeated-measures ANOVA was conducted to explore the effect of workload and task type manipulations on reaction time of correct actions. The main effect of task type manipulations on reaction time of correct actions was significant, $F(1.5, 19.0) = 77.16$, $p < .01$, $\eta_p^2 = .86$. Among four primary task types, the average reaction time of response implementation tasks was the slowest ($M$ = 23.79 $SE$ = 1.84), and the average reaction time of detection tasks was the fastest ($M$ = 4.00 $SE$ = .36). The average reaction time of secondary tasks ($M$ = 9.76 $SE$ = .57) was slower than detection tasks, but faster than checking tasks ($M$ = 17.22 $SE$ = .53), response implementation tasks, and load following tasks ($M$ = 14.58 $SE$ = .44).

**Table 17**
*Two-way ANOVA Statistics for Reaction Time*

| | Low Workload | | | High Workload | | | ANOVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *N* | *M* | *SD* | *n* | Effect | *F* | *df* | $\eta_p^2$ |
| RT | | | | | | | | | | |
| Checking | 16.34 | 2.63 | 14 | 18.10 | 2.80 | 14 | Workload | 3.89 | 1, 13 | .23 |
| Response | 22.61 | 8.93 | 14 | 24.95 | 8.10 | 14 | Task type | 77.16** | 1.5, 19.0 | .86 |
| Detection | 3.76 | 1.76 | 14 | 4.23 | 2.62 | 14 | W × T | 1.33 | 1.5, 19.5 | .09 |
| Power Change | 14.62 | 2.38 | 14 | 14.53 | 1.32 | 14 | | | | |
| Secondary | 7.86 | 2.18 | 14 | 11.66 | 3.66 | 14 | | | | |

**$p < .01$, *$p < .05$, W × T: Workload × Task type

### 7.3.3 Automation Dependence

A two (Workload: low and high) × four (task type: checking, detection, response implementation, power change) repeated-measures ANOVA was conducted to explore the effect of workload and task type manipulations on automation dependence in primary tasks. The main effect of task type on automation dependence was significant, $F(1.8, 51.7) = 109.95$, $p < .01$, $\eta_p^2 = .80$. Overall, the dependence on automation was the lowest in detection tasks ($M = 46.55$ $SE = 3.82$), significantly lower than checking tasks ($M = 94.42$ $SE = 1.86$), response implementation tasks ($M = 74.14$ $SE = 2.77$), and load following tasks ($M = 97.33$ $SE = 1.59$). In addition, the analysis also revealed a significant interaction between workload and task type manipulations, $F(1.9, 53.7) = 5.86$, $p < .01$, $\eta_p^2 = .17$. A follow-up comparison on a task type basis revealed that the effect of workload on automation dependence was significant in detection tasks, $F(1, 28) = 6.6$, $p < .05$, $\eta_p^2 = .19$. The dependence on automation in detection tasks under low workload manipulation ($M = 41.38$ $SE = 3.98$) was significantly lower than under high workload manipulation ($M = 51.72$ $SE = 4.63$). Such difference in dependence on automation was not significant in other task types (Table 18).

**Table 18**
*Two-way ANOVA Statistics for Automation Dependence*

| | Low Workload | | | High Workload | | | ANOVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *N* | *M* | *SD* | *n* | Effect | *F* | *df* | $\eta_p^2$ |
| Checking | 95.74 | 6.07 | 29 | 93.10 | 16.42 | 29 | Workload | .10 | 1, 28 | .00 |
| Response | 78.45 | 17.33 | 29 | 69.73 | 22.54 | 29 | Task type | 109.95** | 1.8, 51.7 | .80 |
| Detection | 41.38 | 21.41 | 29 | 51.72 | 24.94 | 29 | W × T | 5.86** | 1.9, 53.7 | .17 |
| Load | 98.45 | 5.02 | 29 | 96.21 | 15.79 | 29 | | | | |

**$p < .01$, *$p < .05$, W × T: Workload × Task type

## 7.4 Correlations

Correlational analyses were computed to assess the relationship between individual difference measures (e.g., HIT and PAS) and subjective outcomes, physiological responses, as well as performance-based measures. In addition, correlations between performance measures in primary and secondary tasks were also calculated to test the potential tradeoff between the tasks. Finally, correlations between CTPA and performance measures were run to verify the association between trust in automation and task performance.

### 7.4.1 Individual Difference Measures and Trust

HIT score was positively associated with the trust in automation measured by CTPA in high workload condition ($r = .411$, $p < .05$). Neither of the PAS subscales predicts trust in automation in this study (Table 19).

**Table 19**

*Correlations between Individual Difference Measures and Trust*

| | HIT | PAS_High Expectation | PAS_All or None Thinking |
|---|---|---|---|
| CTPA_ Low Workload | .242 | .179 | -.261 |
| CTPA_ High Workload | .411* | .150 | -.249 |

**$p < .01$, *$p < .05$

### 7.4.2 Individual Difference Measures and Subjective Workload

Table 20 shows the correlations between dispositional trust measures and subjective workload measures measured by NASA-TLX and MRQ in two workload conditions. The All or None Thinking score derived from PAS was positively associated with subjective workload measured by ISA for checking task in both workload conditions (low workload, $r = .429$, $p < .05$; high workload, $r = .421$, $p < .05$). The All or None thinking score was also positively associated with ISA workload for load following tasks (low workload, $r = .449$, $p < .05$) and for response implementation tasks (high workload, $r = .405$, $p < .05$) in one workload condition. Generally, the All or None score showed weak to moderate correlation with ISA workload in all task types. The high expectation score derived from PAS is positively associated with some spatial mental processes measured by MRQ. For example, positive association was revealed for spatial concentrative process (low workload, $r = .367$, $p < .05$) and spatial quantitative process (high workload, $r = .381$, $p < .05$). The high expectation score was predictive of physical demand measured by NASA-TLX in the high workload scenario, $r = -.380$, $p < .05$. A similar pattern of negative association was also revealed in the low workload condition but not significantly so. HIT was predictive of multiple workload measures from all three subjective measures. Specifically, HIT was negatively associated with physical demand in both workload conditions (low workload, $r = -.515$, $p < .05$; high workload, $r = -.375$, $p < .05$), positively associated with visual phonetic process (low workload, $r = .426$, $p < .05$) and spatial quantitative process (high workload, $r = .517$, $p < .01$). HIT was also negatively associated with ISA workload in detection in both workload conditions, but only significant in the low workload condition, $r = -.400$, $p < .05$.

**Table 20**
*Correlations between Individual Difference Measures and Subjective Workload*

| | Low Workload | | | High Workload | | |
|---|---|---|---|---|---|---|
| | HIT | PAS High Expectation | PAS All or None Thinking | HIT | PAS High Expectation | PAS All or None Thinking |
| **NASA-TLX** | | | | | | |
| Global Workload | -.349 | -.248 | .100 | -.223 | -.243 | .164 |
| Mental Demand | -.045 | -.173 | .117 | -.134 | -.149 | .316 |
| Physical Demand | -.515** | -.262 | .176 | -.375* | -.380* | -.020 |
| Temporal Demand | -.285 | -.067 | .285 | -.153 | .071 | .189 |
| Effort | -.235 | -.136 | .273 | -.096 | .317 | .123 |
| Frustration | -.132 | -.229 | -.202 | -.035 | -.136 | -.125 |
| Performance | -.206 | -.075 | -.139 | -.202 | .041 | .328 |
| **MRQ** | | | | | | |
| Auditory Emotional | -.130 | -.008 | -.295 | .165 | .300 | .082 |
| Auditory Linguistic | .099 | -.099 | -.187 | .330 | .175 | .263 |
| Manual Process | -.055 | -.020 | -.048 | .101 | .227 | .026 |
| Short Term Memory | .024 | .122 | -.136 | -.216 | -.177 | -.096 |
| Spatial Attentive | -.141 | -.081 | -.223 | -.159 | -.089 | -.261 |
| Spatial Concentrative | .286 | .367* | .139 | -.008 | .217 | .066 |
| Spatial Categorical | .171 | .042 | .061 | .156 | .009 | .067 |
| Spatial Emergent | .319 | .092 | .091 | .092 | .060 | -.074 |
| Spatial Positional | .153 | .079 | -.141 | .102 | .043 | -.191 |
| Spatial Quantitative | .196 | .220 | .163 | .517** | .381* | .122 |
| Visual Lexical | .110 | .102 | .327 | .058 | -.062 | .144 |
| Visual Phonetic | .426* | .211 | .075 | .317 | .177 | .075 |
| Visual Temporal | .111 | -.014 | -.043 | .191 | .220 | .032 |
| Vocal Process | -.036 | .045 | .277 | .195 | .176 | .255 |
| **ISA** | | | | | | |
| ISA Checking | -.367 | .128 | .429* | -.048 | .279 | .421* |
| ISA Detection | -.400* | -.081 | .374 | -.239 | -.103 | .260 |
| ISA Response | -.131 | .094 | .298 | .091 | .125 | .405* |
| ISA Load Following | -.174 | .066 | .449* | -.135 | .015 | .304 |

**$p < .01$, *$p < .05$

### 7.4.3 Individual Difference Measures and Physiological Measures

Table 21 shows the correlations between dispositional trust measures and physiological responses measured by EEG, ECG, and fNIRS in two workload conditions. Most of the significant correlations came from HIT and EEG beta band in various task types. Specifically, HIT is positively associated with frontal beta in checking tasks (low workload, $r = .440$, $p < .05$; high workload, $r = .515$, $p < .01$), detection tasks (low workload, $r = .408$, $p < .05$; high workload, $r = .382$, $p < .05$), response implementation tasks (low workload, $r = .471$, $p < .05$; high workload, $r = .501$, $p < .01$), and load following tasks (low workload, $r = .495$, $p < .01$; high workload, $r = .538$, $p < .01$). All or None Thinking was found to be predictive of a few EEG and ECG measures, but only in high workload conditions. For example, All or None Thinking was negatively associated with frontal beta and IBI in detection tasks (beta $r = -.372$, $p < .05$; IBI $r = -.431$, $p < .05$), and positively associated with heart rate in response tasks ($r = .443$, $p < .05$).

**Table 21**
*Correlations between Individual Difference Measures and Physiological Measures*

| | Low Workload | | | High Workload | | |
|---|---|---|---|---|---|---|
| | HIT | PAS High Expectation | PAS All or None Thinking | HIT | PAS High Expectation | PAS All or None Thinking |
| **EEG** | | | | | | |
| Alpha F Checking | -.156 | -.120 | -.208 | -.042 | -.069 | -.210 |
| Alpha F Detection | -.183 | -.049 | -.144 | -.019 | -.151 | -.247 |
| Alpha F Response | -.088 | -.072 | -.208 | -.075 | -.021 | -.150 |
| Alpha F Load | -.144 | -.140 | -.191 | -.031 | -.155 | -.297 |
| Beta F Checking | .440* | .123 | -.283 | .515** | .125 | -.248 |
| Beta F Detection | .408* | .194 | -.288 | .382* | .083 | -.372* |
| Beta F Response | .471* | .109 | -.322 | .501** | .122 | -.136 |
| Beta F Load | .495** | .157 | -.234 | .538** | .115 | -.238 |
| Theta F Checking | -.192 | -.056 | -.142 | -.214 | -.130 | -.126 |
| Theta F Detection | -.352 | -.080 | -.014 | -.045 | -.169 | -.264 |
| Theta F Response | -.314 | -.051 | .007 | -.204 | .114 | .011 |
| Theta F Load | -.257 | -.085 | -.084 | -.048 | -.168 | -.265 |
| **ECG** | | | | | | |
| Heart Rate Checking | .250 | .170 | .240 | .144 | .085 | .308 |
| Heart Rate Detection | .272 | .171 | .126 | .220 | -.003 | .349 |
| Heart Rate Response | .099 | -.146 | .102 | .003 | .220 | .443* |
| Heart Rate Load | .337 | .202 | .211 | .237 | .036 | .273 |
| HRV Checking | -.307 | -.185 | -.068 | -.233 | -.072 | -.180 |
| HRV Detection | -.061 | .107 | -.051 | -.096 | -.009 | -.259 |
| HRV Response | .159 | .438 | -.284 | .036 | -.177 | -.260 |
| HRV Load | -.259 | -.235 | -.180 | -.231 | -.199 | -.140 |
| IBI Checking | -.263 | -.272 | -.178 | -.222 | -.206 | -.199 |
| IBI Detection | -.199 | -.228 | -.125 | -.138 | -.095 | -.344 |
| IBI Response | -.064 | .078 | -.182 | .066 | -.256 | -.431* |
| IBI Load | -.229 | -.312 | -.245 | -.272 | -.185 | -.174 |
| **Oximeter** | | | | | | |
| $rSO_2$ Left Checking | -.156 | .004 | -.259 | -.365 | -.143 | -.183 |
| $rSO_2$ Left Detection | -.174 | .002 | -.179 | -.256 | .022 | -.111 |
| $rSO_2$ Left Response | -.259 | -.160 | -.170 | -.336 | -.192 | -.254 |
| $rSO_2$ Left Load | -.195 | -.011 | -.198 | -.330 | -.072 | -.138 |
| $rSO_2$ Right Checking | -.246 | -.019 | -.133 | -.311 | -.069 | -.185 |
| $rSO_2$ Right Detection | -.272 | -.130 | -.125 | -.237 | -.060 | -.098 |
| $rSO_2$ Right Response | -.335 | -.196 | -.084 | -.289 | -.050 | -.228 |
| $rSO_2$ Right Load | -.261 | -.030 | -.085 | -.294 | -.001 | -.110 |

**$p < .01$, *$p < .05$

### 7.4.4 Individual Difference Measures and Task Accuracy

Table 22 shows the correlations between dispositional trust measures and task accuracy measures from primary and secondary tasks in two workload conditions. Overall, most significant correlations were PAS, especially in power change tasks. Secondary task accuracy was associated with HIT (low workload, $r = .489$, $p < .01$; high workload, $r = .378$, $p < .05$), but not with PAS scores. PAS was more predictive of primary task performance. High expectation and All or None Thinking were negatively associated with multiple measures for correct actions in both workload conditions and positively associated with some incorrect action measures. Generally, the correlations showed a similar pattern, in terms of directions, across workload manipulations but in mixed strength.

**Table 22**

*Correlations between Individual Difference Measures and Performance Measures*

| | Low Workload | | | High Workload | | |
|---|---|---|---|---|---|---|
| | HIT | PAS High Expectation | PAS All or None Thinking | HIT | PAS High Expectation | PAS All or None Thinking |
| **Checking Task** | | | | | | |
| Correct Actions | .085 | -.060 | -.101 | .272 | -.187 | -.107 |
| Correct Hits | .090 | -.103 | -.038 | .241 | -.226 | -.120 |
| Correct Rejects | -.008 | .12 | -.176 | .320 | .031 | -.021 |
| Misses | -.242 | .238 | .055 | -.406* | .044 | .266 |
| False Alarms | .008 | -.112 | .176 | -.328 | -.247 | .002 |
| **Detection Task** | | | | | | |
| Correct Hits | .251 | .041 | -.125 | .172 | -.221 | -.235 |
| Correct Hits Auto | .023 | .175 | .235 | .207 | .086 | .228 |
| Correct Hits Manual | .188 | -.091 | -.268 | -.050 | -.243 | -.377* |
| Misses | -.289 | -.052 | .166 | -.299 | .097 | .330 |
| False Alarms | -.233 | -.045 | .035 | -.440 | .042 | .200 |
| Additional Actions | -.087 | .154 | .201 | -.038 | .201 | -.188 |
| Invalid Actions | .005 | .148 | .011 | -.452* | .207 | .121 |
| **Response Task** | | | | | | |
| Correct Hits | .250 | .042 | -.027 | .131 | -.340 | -.286 |
| Correct Hits Auto | .161 | .181 | .228 | .159 | .023 | -.277 |
| Correct Hits Manual | .100 | -.197 | -.350 | -.040 | -.395* | .006 |
| Misses | -.250 | -.042 | .027 | -.103 | .195 | .454* |
| Additional Actions | -.205 | .021 | -.074 | -.076 | -.046 | -.313 |
| Invalid Actions | .186 | -.274 | -.386* | -.057 | -.270 | -.391* |
| **Power Change Task** | | | | | | |
| Correct Actions | .000 | -.427* | -.225 | .081 | -.282 | -.078 |
| Correct Hits | -.075 | -.336 | -.269 | .098 | -.287 | -.087 |
| Correct Rejects | .112 | -.469* | -.112 | -.042 | -.216 | -.008 |
| Misses | .075 | .336 | .269 | -.092 | .153 | .358 |
| False Alarms | -.112 | .469* | .112 | .259 | -.050 | -.043 |
| Additional Actions | -.124 | -.338 | -.196 | -.114 | -.329 | -.170 |
| Invalid Actions | .139 | .330 | -.228 | -.030 | .003 | -.287 |
| Correct OperAuto | -.135 | -.405* | -.337 | -.118 | -.439* | -.322 |
| Correct Go | -.178 | -.353 | -.354 | -.134 | -.403* | -.439* |
| Correct Reset | -.040 | -.483* | -.122 | -.144 | -.549** | -.040 |
| Miss OperAuto | .135 | .405* | .337 | .185 | .077 | .067 |
| **Secondary Task** | | | | | | |
| Correct Hits | .489** | -.006 | -.212 | .378* | -.117 | -.202 |

**$p < .01$, *$p < .05$

### 7.4.5 Individual Difference Measures and Reaction Times

Table 23 shows the correlations between dispositional trust measures and reaction times in primary and secondary tasks in two workload conditions. Generally, the correlations between dispositional trust on automation and reaction time were weak. High expectation was found to be negatively associated with the reaction time when participants took an incorrect action in checking task type (misses) in high workload conditions, $r = -.803$, $p < .05$. Additionally, HIT showed a moderate correlation with the reaction time when participants took an incorrect action in response implementation task type (misses) in low workload conditions, $r = -.482$, $p < .05$.

**Table 23**

*Correlations between Individual Difference Measures and Reaction Times*

| | Low Workload | | | High Workload | | |
|---|---|---|---|---|---|---|
| | HIT | PAS High Expectation | PAS All or None Thinking | HIT | PAS High Expectation | PAS All or None Thinking |
| **Checking Task** | | | | | | |
| Hit RT | -.003 | -.047 | .077 | -.041 | .211 | .073 |
| Reject RT | .015 | -.128 | -.181 | -.023 | -.139 | -.057 |
| Miss RT | -.046 | -.300 | -.410 | -.155 | -.803* | -.472 |
| **Detection Task** | | | | | | |
| Hit RT | -.058 | .033 | -.113 | .052 | .085 | -.068 |
| **Response Task** | | | | | | |
| Hit RT | .056 | -.184 | .007 | -.085 | -.177 | -.336 |
| Miss RT | -.482* | -.407 | -.020 | .153 | -.102 | -.227 |
| **Load Following Task** | | | | | | |
| Hit RT | -.364 | -.191 | -.121 | -.279 | .069 | -.164 |
| Reject RT | -.309 | -.069 | -.251 | -.265 | -.201 | -.168 |
| Correct Go RT | .058 | -.055 | .147 | .065 | -.042 | .171 |
| Correct Reset RT | -.057 | -.165 | .028 | -.022 | -.186 | -.039 |
| **Secondary Task** | | | | | | |
| Hit RT | -.210 | .172 | .245 | .142 | -.055 | .199 |

$^{**}p < .01$, $^{*}p < .05$

### 7.4.6 Primary and Secondary Task Accuracy

Correlations between performance measures in primary and secondary tasks were calculated to test the potential tradeoff between the tasks (Table 24). The results revealed no tradeoff effect between primary and secondary tasks. The accuracy in secondary tasks was significantly correlated with correct actions in checking ($r = .501$, $p < .01$), detection ($r = .397$, $p < .05$), and response implementation ($r = .419$, $p < .05$) tasks.

**Table 24**
*Correlations between Primary and Secondary Task Accuracy*

| | Low Workload Secondary Accuracy | High Workload Secondary Accuracy |
|---|---|---|
| Checking | .307 | .501** |
| Detection | .160 | .397* |
| Response | .248 | .419* |
| Power Change | .268 | .323 |

**$p < .01$, *$p < .05$

### 7.4.7 Primary and Secondary Reaction Time

Table 25 shows the correlations between reaction times in primary and secondary tasks in two workload conditions. The reaction time in secondary tasks was negatively associated with reaction time in checking tasks in high workload conditions, suggesting a potential tradeoff effect. But the correlation was weak and not statistically significant. In low workload conditions, the reaction time in secondary tasks was positively associated with reaction time in detection tasks, $r = .496$, $p < .05$.

**Table 25**
*Correlations between Primary and Secondary Reaction Time*

| | Low Workload Secondary RT | High Workload Secondary RT |
|---|---|---|
| Checking | .227 | -.366 |
| Detection | .496* | .445 |
| Response | .243 | -.074 |
| Power Change | .133 | .156 |

**$p < .01$, *$p < .05$

### 7.4.8 CTPA and Task Performance

Correlations between CTPA and performance measures were run to verify the association between trust in automation and task performance (Table 26). Generally, the correlations were weak. The only significant result came from the correlation between CTPA and reaction time in load following task in low workload conditions, $r = -.448$, $p < .05$. The relationship between trust in automation and automation dependence showed an opposite trend in low and high workload conditions (except in response implementation tasks), although only in weak correlations. This may suggest that the impact of subjective trust on automation dependence is moderated by workload factors.

**Table 26**
*Correlations between CTPA and Task Performance*

| | Low Workload CTPA | High Workload CTPA |
|---|---|---|
| **Accuracy** | | |
| Checking | -.089 | .204 |
| Detection | .225 | .259 |
| Response | -.080 | .067 |
| Load Following | -.013 | .008 |
| Secondary | .236 | .236 |
| **Reaction Time** | | |
| Checking | -.315 | -.056 |
| Detection | .044 | .252 |
| Response | -.361 (p = .05) | -.008 |
| Load Following | -.448* | -.142 |
| Secondary | -.023 | .135 |
| **Automation Dependence** | | |
| Checking | -.089 | .204 |
| Detection | -.105 | .261 |
| Response | -.150 | -.031 |
| Load Following | -.013 | .008 |

**$p < .01$, *$p < .05$

# SECTION 8: EXPERIMENT 7 DISCUSSION

Experiment 7 investigated impacts of task type and a workload manipulation on multiple outcome measures, in the context of a simulated EOP supported by automation. Outcomes included workload, performance, and trust. Several significant effects of task type on workload measures were found, with the physiological indices being more sensitive to task type effects than the subjective ratings. Performance accuracy was lower for detection than for other tasks, and psychophysiological measures suggested that the task produced both higher workload and reduced cortical arousal relative to the others. This study was the first in the HPTF research series to include a task that would represent the sort of activities performed during a load following procedure. This task did not appear to be demanding for participants. The workload manipulation used for the secondary task produced only modest impacts on workload measures and had no effect on trust in the system. We found some significant associations between dispositional trust in automation and outcomes, but these varied across trust scales. In the remainder of the discussion, we evaluate the findings on workload, performance, and trust in more detail, and advance some general conclusions.

Hypothesis 1a proposed that the power change task meant to mimic the sort of work performed by the Balance of Plant Operator during load-following would impose higher workload than other task types. This hypothesis was not supported. The power change task generally showed workload patterns similar to checking and response implementation tasks across both subjective and physiological measures. ECG data showed some evidence of elevated workload, but this effect was modest and not reflected in other measures. High performance accuracy on the power change tasks (>95%) further suggests these tasks were not particularly demanding.

On the other hand, higher accuracy does not necessarily imply lower workload. According to Attentional Control Theory (Eysenck et al., 2007), individuals may maintain high levels of performance under increased cognitive demand by reallocating attentional resources. This can come at the cost of efficiency or resilience to distraction. Thus, the high accuracy observed may reflect compensatory control strategies rather than inherently low task demand. Hypothesis 1b proposed that a higher workload manipulation would exacerbate task-type differences, particularly for detection tasks. This hypothesis was also not supported.

Hypothesis 2a predicted that detection tasks would have the lowest performance accuracy, which was supported. Detection tasks consistently showed lower accuracy (83.19%) compared to other primary tasks. While this may reflect the sustained attention demands of monitoring for threshold changes, we avoid characterizing this as a vigilance decrement given the relatively short task duration. Hypothesis 2b predicted that power change tasks were also expected to show reduced accuracy, which was not supported by the data. Power change tasks had the highest accuracy (97.33%) compared to the other primary tasks. The power change task in this study required using the same I&C, and it is possible that participants learned how to effectively perform this task across repeated steps of the EOP. That is, the high accuracy for the power changing may reflect the procedural consistency of the task.

The third hypothesis that automation failures would lead to reduced trust, especially at higher workload levels, was not supported. CTPA trust scores did not significantly differ between workload conditions despite participants encountering automation failures in both conditions. This suggests that occasional automation failures did not substantially impact operator trust in the short term.

## 8.1 Effects of Task Type and Workload Manipulation on Subjective Workload

The study included a power level change task, meant to simulate load-following, as well as checking, detection, and response implementation tasks. We hypothesized that the power change task would impose higher levels of workload than the other task types. The ISA was used to compare subjective workloads of the tasks, but no significant differences were found. However, task type effects were observed in multiple physiological workload measures. In the EEG data, beta power was sensitive to task type, with detection showing reduced beta activity relative to the other three tasks across frontal, parietal, and occipital regions. While beta elevation is generally associated with increased cognitive workload, reduced beta during detection may not necessarily indicate lower workload. Instead, it could reflect early signs of task disengagement or a shift in cognitive strategy, particularly given the sustained attention demands of detection tasks. This interpretation aligns with prior HPTF studies (Hughes et al., 2023a; Lin et al., 2024), which also reported reduced beta during detection. Lin et al. additionally found reduced alpha power during detection, a trend we observed in the right hemisphere, though it was non-significant in the current study. We also found lower theta power during detection compared to the other tasks, especially in the right hemisphere. These findings suggest that psychophysiological measures may be more sensitive to task type effects than subjective ratings, particularly in novice samples.

The analysis of fNIRS data also suggested that brain activity during detection differs from the other three tasks. The detection task showed the highest level of brain oxygenation ($rSO_2$), in both hemispheres. These data concur with those obtained from the first three HTPF experiments (Hughes et al., 2023c), although Lin et al. (2024) did not find a significant task type effect. By contrast, ECG data showed that mean heart rate was lower for the detection task, and higher for response implementation and load following. Task type effects on mean heart rate have been somewhat inconsistent across previous HPTF studies (Hughes et al., 2023c; Lin et al., 2024).

The data show a nuanced picture of task type effects on workload measures consistent with previous findings, illustrating the value of a multivariate assessment approach. As found in previous studies (Hughes et al., 2023c; Lin et al., 2024), the detection task appeared to impose the highest demands when workload was assessed by fNIRS, but EEG beta, theta and ECG appeared to show the opposite trend. The conflicting results can be resolved by viewing the detection task as a vigilance task requiring sustained attention over time (Reinerman-Jones et al., 2016). On the one hand, performance on vigilance tasks produces loss of cortical arousal, evident in autonomic arousal measures such as heart rate. Decreased beta (along with increased slow-wave EEG activity) is attributable to the concurrent loss of alertness and reduced cognitive activity during vigilance (Borghini et al., 2014). By contrast, fNIRS appears to be more sensitive to workload than to arousal change; temporal changes in rSO2 during tasks requiring sustained attention are rather task-specific (Warm et al., 2012). Overall, the data concur with the findings from Hughes et al.'s (2023c) review of HPTF studies that fNIRS and EEG beta are the most sensitive psychophysiological measures for reliable discrimination of different task types.

We independently manipulated workload by varying the precision of automation support for navigation. This turned out to be a relatively weak manipulation. Subjective data showed an increase in NASA-TLX mental demand in the higher workload condition, but no global workload difference between conditions. Means for the MRQ showed that, as expected, tasking imposed moderate levels of demand on manual, spatial, visual and short-term memory scales. The workload manipulation produced higher levels of demand on some of these scales, but differences were non-significant. There were some trends that might prove significant with a larger sample. While non-significant, MRQ indicated short term memory and visual temporal demands were both higher for the higher workload version of the task, with small to medium effect sizes. By contrast, effect sizes for the spatial demand ratings were in most cases trivially small. Thus, although we expected the workload manipulation to increase spatial demands, it actually appeared to produce a modest increase in more general aspects of workload.

The workload manipulation also did not produce significant effects on performance accuracy and response speed, although the effect on RT showed a trend in the expected direction. Response to the secondary monitoring task, in particular, appeared to be sensitive to workload. Mean response time was longer in the higher workload condition, suggesting that this metric may be of value as a behavioral index. There were no significant interactions between the task type and workload manipulations (other than their effect on automation dependence), so that the task type effects were quite robust across conditions.

The workload findings discussed should be considered in the context of the overall workload of the tasks, which was very low. Mean NASA-TLX workload was around 17 (on a 100-point scale), similar to that found by Lin et al. (2024) but considerably less than values with a median of around 30 for the conventional plant simulation reported by Hughes et al. (2023a, 2023c). ISA values were also low for all tasks. Very low workloads may signal monotony, loss of cortical arousal and potential underload stress (Hancock & Warm, 1989; Saxby et al., 2013). These psychological states can lead to loss of vigilance during operation of automated systems (Greenlee, DeLucia, & Newton, 2024), which in turn can lead to impairments of situation awareness (Endsley, 1996). There are concerns that CBPs may impair situation awareness in

NPP operators (Hall, 2023; Qing et al., 2021; Seong et al., 2013). Further research is needed to determine relationships between situation awareness, workload, and arousal.

Means for the psychophysiological data are consistent with an underload interpretation of the task environment. The EEG data showed increases in power in the slower wave bands, theta and alpha, especially in the right hemisphere, which are typically linked to declines in cortical arousal and loss of vigilance (Kamzanova, Kustubayeva & Matthews, 2014). Unusually, frontal oxygenation measured by fNIRS showed consistent declines from baseline, suggesting declining frontal metabolic activity consistent with very low workload. On the one hand, it is encouraging that the metrics diagnostic for task type effects are robust across different versions of the simulation. On the other hand, the combination of low workload and low arousal suggested that using automation-supported CBPs may produce the suboptimal neurocognitive states typically attributed to automation such as complacency and mental disengagement (Matthews, Warm & Smith, 2017).

**Regulatory significance:** These findings inform regulatory reviews in terms of workload assessment in advanced control room designs. The very low subjective workload observed suggests that underload, rather than overload, may be a primary concern in highly automated environments. Regulatory reviews should consider how human factors validation testing addresses the potential for boredom, vigilance decrement, and diminished situation awareness under low workload conditions, particularly for advanced reactor designs that may rely heavily on automation.

## 8.2 Effects of Task Type and Workload Manipulation on Performance and Automation Dependence

We hypothesized that there would be both task type and workload effects on performance quality, especially as expressed in greater accuracy. The anticipated effect of task type was confirmed. As in previous studies, performance accuracy was notably lower for detection than for the other tasks, which may reflect the demands of sustaining vigilance even over short time intervals. Reinerman et al. (2016) showed temporal decrement in detection rate in a fine-grained minute-by-minute analysis of the task, even though duration was only five minutes. This was accompanied by changes in EEG measures (i.e., increasing alpha power) which have been linked to vigilance decrement. However, in the absence of converging behavioral evidence, these neural changes may also reflect early signs of task disengagement rather than a definitive loss of vigilance. Contrary to expectation, accuracy on the power change task was very high and close to ceiling. Performance on this task may reflect the consistency of response needed to execute it. Power changing always required use of the same I&C so demands of navigation were low and participants may have been able to learn how to perform the task effectively across the repeated steps of the EOP.

Accuracy on the secondary gauge-monitoring task was substantially lower than primary task performance, which would be an operational concern if reproduced in a real control room. This task too requires vigilance, and participants appear to have neglected it. Workload mitigation through automation is considered necessary for novel plant designs that require multi-tasking such as SMRs (Blackett et al., 2022). In the present case, workload was low, but participants were still poor at dividing their attention across primary and secondary tasks. Careful consideration to automation design to support multi-tasking may be necessary (e.g., Bahnsen et al., 2024).

A concern with introducing a secondary task is that it may introduce tradeoffs between primary and secondary task performance depending on how the participant chooses to allocate their attention (Wickens, 1991). Such tradeoffs can affect the validity of secondary RT as a workload measure. If tradeoffs are present, there should be negative correlations between primary and secondary task performance measures. The correlational analysis did not find any significant trade-offs. Primary and secondary task accuracy were consistently positively correlated, especially at the higher level of workload. These findings suggest individual differences in overall performance competence, such as variation in overall availability of attentional resources (Matthews et al., 2000).

Response time data are of lesser importance, and not directly comparable across task types. The slowness of response on some tasks such as response implementation may reflect the time taken to locate I&C by novice participants, even with assistance from the automation. However, there was no imperative for rapid response.

We also found considerable variation across tasks in automation dependence, defined as the percentage of trials on which the person approved the recommendation from the automation. There were very high levels of dependence for checking and load following, potentially contributing to the high overall levels of accuracy on these tasks. Dependence was lower for response implementation, although accuracy remained high, implying that participants were quite good at spotting automation errors on this task. Paradoxically, although the detection task was the most difficult, participants were least likely to depend on automation. Dependence was higher for the task in the higher workload condition. Previous studies have suggested that the fatigue associated with sustained attention may impair participants' ability to utilize automation effectively (Lin et al., 2020; Matthews et al., 2017).

**Regulatory significance:** The poor performance on secondary tasks despite low overall workload raises areas of significant interest for regulatory reviews. For advanced reactor designs where operators may need to monitor multiple units or systems simultaneously, these findings suggest that current assumptions about spare attentional capacity may be overly optimistic. Human factors engineering reviewers may need to consider whether secondary monitoring tasks receive appropriate attention in design and validation testing, particularly for highly automated systems.

### 8.3 Trust Calibration

Trust calibration and optimization is a major issue for design and evaluation of automated systems; both over- and under-trust are potentially harmful in operational settings (Lee & See, 2004; Parasuraman & Manzey, 2010). In this study we investigated trust primarily through assessment of individual differences in trait trust – general propensity to trust automation – and in state trust – immediate trust in the system operated. Mean levels of state trust on the CPTA state measure (Jian et al., 2000) were relatively low for a high-reliability system. Lin et al. (2024) also found limited trust in the automation in their study. We found that trait and state trust measures were only modestly associated, although the HIT trait measure (Lyons & Guznov, 2019) was significantly positively correlated with the CPTA (Jian et al., 2000) under high workload. Lin et al. (2024) similarly found that the HIT was a better predictor of state trust than the PAS (Merritt et al., 2015). However, most of the variation in state trust appears to reflect the participant's immediate experience with the automation of the CBP rather than dispositional attitudes. State trust was not strongly associated with objective automation dependence, supporting Patton and Wickens' (2024) contention that the two kinds of measure can dissociate.

We hypothesized that automation failure would lead to reduced trust, especially at higher workload levels, to the extent that performance challenges lead to negative misattributions of the utility of the automation. Similarly, high trust might lower workload, if the participant believes that the automation will handle performance without their intervention. The workload manipulation did not affect CPTA state trust scores, although we have noted the modest impact of the manipulation. Similarly, significant correlations between the trust scales and subjective workload were relatively few, and there was no general trend towards a negative workload – trust association. The PAS All-or-none Thinking scale was quite consistently correlated with higher subjective workload as measured by the ISA. Sensitivity to automation failure – assuming the automation was perceived to be imperfect – may have increased attention to the task. Correlations between the trust scales and the physiological workload measures were also rather weak, in general. There was a consistent set of significant positive correlations between the HIT scale (Lyons & Guznov, 2019) and EEG beta power. Possibly, high HIT scores signal an interest in automated systems that is reflected in greater cognitive activity during performance. Given the low levels of significance overall, these relationships require replication.

Correlations between trust scales and task performance were also modest. There was a fairly consistent set of negative associations between the PAS High Expectations scale and power change accuracy, especially under high workload. Possibly, in this case, excessive positive beliefs about the automation may have led some participants to neglect the task. By contrast, there was a significant positive correlation between HIT scores and accuracy on the secondary gauge monitoring task. Confidence in primary task automation may have encouraged high HIT scorers to allocate more attention to the secondary task. There was also a trend towards associations between the HIT and better performance on the detection task under high workload. Generally, greater intention to rely on automation, indexed by the HIT, appeared to be beneficial. Similarly, high scorers on the scale tend to show higher state trust, measured by the CPTA, and lower workload on the detection task, measured by the ISA.

Overall, the results did not provide evidence that individual differences in trust have major consequences for workload and performance. The correlations of the HIT and PAS scales discussed in this section should be investigated further. Matthews et al. (2024) distinguished "active trust" related to beliefs that active, effortful engagement with the system will be effective in accomplishing task goals from "passive trust" associated with broad optimism about automated task outcomes. Elevated trust may be beneficial to the extent that it reflects active trust, which may maintain task engagement and commitment, whereas passive trust breeds complacency.

With fairly reliable automation, as implemented in the present study, over-trust and over-reliance on automation is likely to be the major operational risk. Operators may neglect to monitor and check automated functions because they are over-confident in the system (Parasuraman & Manzey, 2010; Parasuraman & Riley, 1997). The current data show no general tendency towards either performance impairment or low workload (which might indicate complacency) in more trusting individuals except in the case of high expectations and load following impairment. The general lack of a downside to high trust may reflect some initial distrust of an initially unfamiliar system, evidenced by the CPTA trust means falling some way short of the top of the scale. Over-trust may become more of an issue as operators gain greater confidence and familiarity with the system over time.

**Regulatory significance:** The findings on trust calibration have implications for training and qualification programs. The observed variation in trust and the generally modest trust levels despite high system reliability suggest that operator training may benefit from addressing appropriate trust calibration for automated systems. Reviewers may consider evaluating if training programs include components that are intended to 1) help operators develop appropriate levels of trust in automation and, 2) help operators understand the reliability characteristics of different system functions.

## 8.4 Study Limitations

Three study limitations should be noted. First, the sample size may have limited statistical power to detect significant effects. Power is more of an issue for the correlational analyses than for the *t*-tests and ANOVAs. With an N of 31, the power for a two-tailed repeated measures *t*-test to detect a medium effect size ($d = .5$) is .77, which is close to the conventional standard of .80. As stated above though, the need for Bonferroni correction to mitigate Type 1 errors lowers the power of the analyses. In the tests for effects of task type, using repeated measures ANOVA, power to detect a medium sized mean difference ($f = .25$) was estimated as .8 - .9 (depending on inter-correlations of measures). However, for the correlations, power for a two-tailed medium effect size $r$ of .3 was only .38, implying a significant risk of Type II error.

Second, the study used a novice sample, raising questions about generalization to operators. Previous HPTF work has justified the use of novices to investigate workload issues that may also be experienced by operators (Hughes et al., 2023a, Ulrich et al., 2022, Ulrich et al., 2023, & Park et al., 2020). It is assumed that when participants perform proceduralized tasks that do not require deep understanding of reactor functioning and control, cognitive demands experienced by novices may resemble those for trained operators. This argument is supported by data from the HPTF studies showing qualitatively similar workload responses in the two groups, with some caveats (Hughes et al., 2023b). However, further work is needed to test whether findings on trust in automation obtained in novice participants generalize to experienced operators. An argument in favor of investigating novices is that when plants utilize new designs, operators will generally lack experience. If the pace of innovation increases, anticipating human factors issues for operators with limited experience will become increasingly important.

It is accepted that expertise may influence trust in automation though results vary across studies (Hoff & Bashir, 2015). System knowledge is expected to enhance calibration of trust to the actual capability of the system (Lee & See, 2004; Schaefer, Chen, Szalma & Hancock, 2016). Broadly, with a well-designed system, experienced operators tend to trust automation more and to place more reliance on it (Hoff & Bashir, 2015). Competence in using the system and experience also predict trust in AI systems (Kaplan, Kessler, Brill & Hancock, 2023), which may become an increasingly important issue in future plant HSI (Lu et al., 2020).

Third, allied to the previous point, trust is a dynamic process that varies depending on the operator’s history of experience with the system (de Visser, Pak & Shaw, 2018). A single task session does not capture the processes that may build or erode trust over time. As discussed above, CPTA scores suggested that participants were only moderately trusting of the system. Greater experience with the system may enhance trust, but potentially also increase the likelihood of complacency. If operators experience increasing boredom and fatigue with repeated use of the automated system, they may use automation less effectively (Matthews et al., 2017).

## 8.5 Conclusion

The findings support three general conclusions about the operational impacts of automation and unexpected automation failures during CBP-based execution of an EOP.

- Automation provides (mostly) effective support for human performance. Performance data showed that novice participants were generally able to execute the EOP competently despite their unfamiliarity with the task. Performance remained high despite occasional, unexpected automation failures. Similarly, workload remained low. However, despite automation, detection task performance remained moderate. Automation cannot necessarily counter human failings in maintaining vigilance (unless the task can be fully automated). Performance on the secondary gauge-monitoring task was poor, pointing towards a need to investigate possible concerns with multi-tasking for advanced reactor designs. Since advanced reactors will be highly automated, an operator's limited ability to monitor multiple types of monitoring tasks could be significant. In this instance, there was no support from automation.

- Underload may be an issue for operators of automated systems. Both subjective and objective workload measures indicated that workload was exceptionally low, despite the novelty of the task to participants and the moderate LOA. Automation appears to produce a low-arousal state in which the participant is potentially vulnerable to boredom, fatigue and task disengagement. In this study, underload did not appear to produce significant performance impairments. However, the findings signal a vulnerability that may be expressed in an increase in error likelihood over longer work shifts.

- Trust and automation-dependence are possible operational concerns. Consistent with automation being generally beneficial, findings did not indicate any general tendency towards complacency or over-reliance on automation. However, data do raise some questions about the role of trust in system operation. There was large variation across tasks in automation-dependence, for reasons which are unclear; under-reliance on automation may have contributed to performance deficits in detection. The different trust measures showed different sets of correlations with the outcome variables, indicating that improved measurement of trust is needed. In the present study, intentions to rely on automation appeared to be generally beneficial. Stronger psychometric models for trust would contribute to distinguish components of trust that aid performance from those that may contribute to excessive automation-dependence.

# SECTION 9: GENERAL DISCUSSION

## 9.1 Summary of Key Findings

This research program investigated the effects of automation and computerized procedures on operator workload, performance, and trust across two complementary studies. Experiment 6 compared two intermediate levels of automation (management-by-consent and management-by-exception) in a simulated emergency operating procedure, while Experiment 7 examined the impact of workload manipulations and different task types, including power change tasks relevant to advanced reactor designs.

Key findings across both studies include:

- Detection tasks showed unique workload patterns, with higher brain oxygenation but lower cortical arousal, suggesting vigilance demands rather than traditional workload challenges.

- Higher levels of automation (management-by-exception) produced faster response times but unexpectedly higher temporal demand and physiological workload indicators.
- Performance on secondary monitoring tasks was poor (~60% accuracy) despite low primary task workload, raising concerns for multi-unit monitoring scenarios.
- Reliance on automation varied significantly by task type, with detection tasks showing the lowest dependence (~47%) compared to checking tasks (~94%).
- State trust in automation was moderate despite high system reliability, indicating potential undertrust rather than complacency.

## 9.2 Implications

### 9.2.1 Contradictory Findings Across Measurement Approaches

A notable pattern across both studies was the divergence between different workload measures, particularly for detection tasks. While fNIRS data consistently showed higher brain oxygenation during detection tasks, EEG data showed reduced beta and theta activity, suggesting lower cortical arousal. These seemingly contradictory findings can be reconciled through an understanding of the vigilance decrement phenomenon.

Detection tasks in nuclear operations require sustained attention over time with relatively infrequent signal occurrences. This creates a vigilance situation where:

1. The task imposes high cognitive demand (reflected in fNIRS data showing increased metabolic activity),

2. It simultaneously induces reduced cortical arousal over time (reflected in EEG patterns).

3. This combination creates a paradoxical state where operators feel subjectively underloaded yet are objectively at higher risk for performance failures.

These findings have important methodological implications for workload assessment in nuclear environments. Relying on any single workload measure (whether subjective, performance-based, or physiological) may lead to incomplete or misleading conclusions about task demands. Multivariate assessment approaches that capture both metabolic demand and arousal state are necessary to fully characterize operator states during automated operations.

### 9.2.2 Automation Challenges Across Studies

Both studies revealed nuanced challenges with automation that contradict simplistic assumptions about automation benefits. Higher levels of automation did not universally reduce workload as often assumed, and automation dependence varied significantly by task type. This suggests that operator-automation interaction is complex and context-dependent, requiring careful consideration in system design.

Across both studies, detection tasks emerged as particularly vulnerable to automation effects. Despite automation support, detection accuracy remained lower than other task types, and operators were least likely to rely on automation for these tasks. This pattern suggests that automation may not effectively mitigate the fundamental vigilance challenges inherent in monitoring tasks.

#### 9.2.3 Trust and Workload Relationship

The relationship between trust and workload was not straightforward across either study. While there was some evidence that individual differences in trust disposition influenced workload experiences, there was no consistent pattern of higher trust leading to lower workload or vice versa. This challenges assumptions that increased trust in automation necessarily reduces operator workload and suggests that trust calibration is a complex process influenced by multiple factors beyond system reliability.

### 9.3 Regulatory Significance of Findings

The findings from these studies contribute to the general understanding and development of a technical basis that may support reviews related to automation and human-system interfaces in nuclear power plants.

#### 9.3.1 Implications for Human Factors Engineering Reviews

The findings from Experiment 6 and Experiment 7 relate to several elements of human factors engineering reviews:

- Results demonstrate that detection tasks may pose unique vigilance challenges even with automation support. Task analyses may benefit from evaluating vigilance requirements for detection tasks and ensuring appropriate mitigation strategies.

- The contradictory effects of different automation levels on workload suggest that reviews may benefit from looking for evidence that proposed automation levels have been systematically evaluated for both intended workload benefits and unintended consequences on operator attention and task management.

- The multivariate approach to workload assessment revealed that different measures can provide contradictory conclusions. This suggests that using multiple, complementary workload assessment techniques during validation testing rather than relying solely on subjective measures like NASA-TLX can support a more robust understanding of the metric.

#### 9.3.2 Implications for Human Reliability Analysis

The findings have direct relevance to Human Reliability Analysis methodologies:

- Results suggest that automation dependence varies substantially by task type, with detection tasks showing the lowest automation dependence. This suggests that HRA methods may benefit from incorporating automation dependence as a performance shaping factor that varies by task type.

- The observed underload conditions and poor secondary task performance suggest that error probabilities may increase in highly automated environments. This increase appears due to vigilance decrements and attention allocation issues rather than overload. HRA methods may consider how to better account for these error mechanisms.

### 9.3.3 Implications for Advanced Reactor Oversight

For advanced reactor designs, including SMRs and microreactors:

- Findings regarding poor secondary task performance suggest that assumptions about operators' ability to monitor multiple units simultaneously may need to be evaluated carefully during licensing reviews. One possibility would be to emphasize simulator evaluations that test operator performance under realistic multi-unit conditions.

- Although the power change tasks showed high performance in the current studies, this was in a setting where the task was consistent and predictable. It is important to consider how the tasks required to manage load-following in variable environments are specifically validated under more diverse and unpredictable conditions that more closely approximate real-world implementation.

These implications contribute to the general body of knowledge in human factors engineering and provide a more comprehensive understanding of how automation affects performance in operational settings.

## 9.4 Limitations

Limitations should be considered in interpreting these results. In terms of the sample, both studies utilized novice participants rather than experienced operators. While previous HPTF research has shown qualitatively similar workload responses between novices and experienced operators, some findings may not fully generalize to expert populations, particularly those related to trust and automation dependence. For study design, the experimental sessions were relatively brief, which may not capture fatigue effects or long-term vigilance decrements that could occur during extended operations. Operational settings involve much longer durations that may exacerbate the vigilance challenges observed. In addition to the duration, while the simulator provided a realistic approximation of nuclear power plant operations, it necessarily simplified some aspects of the operational environment. The absence of team dynamics, in particular, represents a significant departure from actual control room operations. Beyond this, trust is a dynamic process that develops over time with system experience. The single-session design of these studies captures only initial trust formation rather than the evolution of trust over extended system use.

## 9.5 Future Directions

Future research should address key areas to build on the findings from these studies. Regarding trust development, studies examining how trust in automated systems develops over extended periods would provide valuable insights into long-term automation dependence patterns. Replicating key findings from these studies with experienced operator samples would strengthen the generalizability of the conclusions, particularly regarding trust calibration and automation dependence. In terms of operator awareness, development and testing of interventions to mitigate the vigilance decrements observed in automated environments should be a priority for supporting safe operations. Follow-up research is planned to examine the effects of awareness in setting where operators are monitored for longer intervals.

As artificial intelligence and advanced automation are increasingly incorporated into nuclear operations, research should examine operator interaction with these technologies, including transparency, explainability, and appropriate trust calibration. Advanced systems are anticipated to behave more like an independent system or collaborator than a tool. Given this shift, future research should examine how automation affects team dynamics in multi-operator control rooms, including shared situation awareness, communication patterns, and collective trust calibration.

## 9.6 Conclusions

This research program provides empirical evidence that challenges several common assumptions about automation in safety-critical environments:

1. Automation does not universally reduce operator workload; higher levels of automation may introduce new monitoring demands that offset workload benefits.

2. Detection tasks remain vulnerable to vigilance decrements even with automation support, suggesting fundamental cognitive limitations that automation cannot fully mitigate.

3. Secondary task performance may be compromised in automated environments despite low primary task workload, raising concerns for multi-unit monitoring scenarios.

4. Trust calibration remains a challenge even with highly reliable systems, with evidence of undertrust rather than the complacency often assumed with automation.

These findings have significant implications for regulatory reviews, human reliability analysis, and advanced reactor oversight. They suggest the need for multivariate assessment approaches, task-specific consideration of automation effects, and careful scrutiny of assumptions about operator capacity in multi-tasking environments. The multivariate approach is particularly important for reviewers because it can uncover issues a survey alone may underestimate. As nuclear facilities continue to modernize and incorporate advanced automation, addressing these human factors challenges will be essential for maintaining safe operations.

# SECTION 10: REFERENCES


Ajenaghughrure, I. B., Sousa, S. D. C., & Lamas, D. (2020). Measuring trust with psychophysiological signals: a systematic mapping study of approaches used *Multimodal Technologies and Interaction*, *4*(3), 63.

Anokhin, A., Lourie, V., Dzhumaev, S., Golovanev, V., & Kompanietz, N. (2010). Upgrade of the Kursk NPP main control room (case study). *International Control Room Design Conference Proceedings 2010,* 207-214.

Ayaz, H., Shewokis, P. A., Curtin, A., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2011). Using MazeSuite and functional near infrared spectroscopy to study learning in spatial navigation. *Journal of Visualized Experiments*, *56*, 3443.

Bahnsen, K. L., Tiemann, L., Plabst, L., & Grundgeiger, T. (2024, May). Augmented reality cues facilitate task resumption after interruptions in computer-based and physical tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1-16).

Balfe, N., Sharples, S., & Wilson, J. R. (2015). Impact of automation: Measurement of performance, workload and behaviour in a complex control environment. *Applied Ergonomics*, *47*, 52-64.

Barg-Walkow, L. H., & Rogers, W. A. (2016). The effect of incorrect reliability information on expectations, perceptions, and use of automation. *Human Factors*, *58*(2), 242-260.

Blackett, C., Eitrheim, M. H. R., McDonald, R., & Bloch, M. (2022, June). Human performance in operation of small modular reactors. In *Proceedings of the 16th International Conference on Probabilistic Safety Assessment and Management (PSAM 16), Honolulu, Hawaii* (Vol. 26).

Boles, D. B. (1991). Factor analysis and the cerebral hemispheres: Pilot study and parietal functions. *Neuropsychologia*, *29*(1), 59-91.

Boles, D. B. (1992). Factor analysis and the cerebral hemispheres: Temporal, occipital and frontal functions. *Neuropsychologia*, *30*(11), 963-988.

Boles, D. B. (1996). Factor analysis and the cerebral hemispheres: “Unlocalized” functions. *Neuropsychologia*, *34*(7), 723-736.

Boles, D. B. (2002). Lateralized spatial processes and their lexical implications. *Neuropsychologia*, *40*(12), 2125-2135.

Boles, D. B., & Adair, L. P. (2001). The multiple resources questionnaire (MRQ). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(25), 1790-1794.

Boles, D. B., Bursk, J. H., Phillips, J. B., & Perdelwitz, J. R. (2007). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). *Human Factors*, *49*(1), 32-45.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58-75.

Boring, R. L. (2023). Human Factors for Advanced Reactors. In T. Ahram (Ed.), *Human Factors in Software and Systems Engineering: Vol. 94.* AHFE International, USA.

Boring, R. L., & Gertman, D. I. (2012). *Considerations for the treatment of computerized procedures in Human Reliability Analysis* (No. INL/CON-12-26092). Idaho National Lab (INL), Idaho Falls, ID.

Boring, R. L., Ulrich, T. A., & Lew, R. (2023, September). Levels of digitization, digitalization, and automation for advanced reactors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 67, No. 1, pp. 207-213). Sage CA: Los Angeles, CA: SAGE Publications.

Bye, A. (2023). Future needs of human reliability analysis: The interaction between new technology, crew roles and performance. *Safety science*, *158*, 105962.

Chance, B., Zhuang, Z., UnAh, C., Alter, C., & Lipton, L. (1993). Cognition-activated low-frequency modulation of light absorption in human brain. *Proceedings of the National Academy of Sciences*, *90*(8), 3770-3774.

Clayton, D., & Wood, R. (2010, November). The role of instrumentation and control technology in enabling deployment of small modular reactors. In *Proceeding of the Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies* (Vol. 615).

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From ‘automation’ to ‘autonomy’: The importance of trust repair in human–machine interaction. *Ergonomics*, *61*(10), 1409-1427.

De Waard, D., & Brookhuis, K. A. (1996). *The measurement of drivers' mental workload.* The Traffic Research Centre VSC, University of Groningen.

de Winter, J. C. (2014). Controversy in human factors constructs and the explosive use of the NASA-TLX: a measurement perspective. *Cognition, Technology & Work*, *16*(3), 289-297.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Multiple-task performance. *Workload Assessment in Multi-Task Environments*, 207-216.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (2020). Workload assessment in multi-task environments. In D.L. Damos (Ed.), *Multiple task performance* (pp. 207-216). Boca Raton, FL: CRC Press.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163–181). Hillsdale, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors, 59*(1), 5-27.

Gao, Q., Wang, Y., Song, F., Li, Z., & Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, *56*(7), 1070-1085.

Gao, Q., Yu, W., Jiang, X., Song, F., Pan, J., & Li, Z. (2015). An integrated computer-based procedure for teamwork in digital nuclear power plants. *Ergonomics*, *58*(8), 1303-1313.

Gardony, A. L., Okano, K., Hughes, G. I., Kim, A. J., Renshaw, K. T., & Sipolins, A. (2022). Aided target recognition visual design impacts on cognition in simulated augmented reality. *Frontiers in Virtual Reality*, *3*, 982010.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 113-131.

Greenlee, E. T., DeLucia, P. R., & Newton, D. C. (2024). Driver vigilance decrement is more severe during automated driving than manual driving. *Human factors*, *66*(2), 574-588.

Hall, A., Boring, R. L., Ulrich, T. A., Lew, R., Velazquez, M., Xing, J., ... & Makrakis, G. M. (2023, November). A comparison of three types of computer-based procedures: an experiment using the Rancor Microworld Simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 67,* 2552–2557.

Hancock, P. A. (2023). Are humans still necessary?. *Ergonomics*, *66*(11), 1711-1718.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human factors*, *31*(5), 519-537.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). Amsterdam, the Netherlands: North-Holland.

Henelius, A., Hirvonen, K., Holm, A., Korpela, J., & Muller, K. (2009). Mental workload classification using heart rate metrics. In *Engineering in Medicine and Biology Society, 2009: Proceedings of the Annual International Conference of the IEEE* (pp. 1836–1839). New York, NY: IEEE.

Hildebrandt, M., Hughes Green, N. M., Blackett, C., Boring, R., Egli, S., Green, B., & Zhang, S. (2023, October). Human factors and new nuclear: Advancements and future possibilities. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1) 1002–1006.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407-434.

Huang, F. H., Hwang, S. L., Yenn, T. C., Yu, Y. C., Hsu, C. C., & Huang, H. W. (2006). Evaluation and comparison of alarm reset modes in advanced control room of nuclear power plants. *Safety Science*, *44*(10), 935-946.

Hughes, N., D'Agostino, A., Dickerson, K., Matthews, M., Reinerman-Jones, L., Barber, D., Mercado, J., Harris, J., & Lin, J. (2023a). *Human Performance Test Facility (HPTF).*

*Volume 1 - Systematic human performance data collection using nuclear power plant simulator: A methodology.* Research Information Letter RIL 2022-11. Washington, DC: United States Nuclear Regulatory Commission.

Hughes, N., Lin, J., Matthews, G., Barber, D., & Dickerson, K. (2023c). *Human Performance Test Facility (HPTF). Volume 3 – Supplemental exploratory analyses of sensitivity of workload measures.* Research Information Letter RIL 2022-11. Washington, DC: United States Nuclear Regulatory Commission.

Hughes, N., Reinerman-Jones, L., Lin, J., Matthews, G., Barber, D., & Dickerson, K. (2023b). *Human Performance Test Facility (HPTF). Volume 2 - Comparing operator workload and performance between digitized and analog simulated environments*. Research Information Letter RIL 2022-11. Washington, DC: US Nuclear Regulatory Commission.

Hwang, S. L., Yau, Y. J., Lin, Y. T., Chen, J. H., Huang, T. H., Yenn, T. C., & Hsu, C. C. (2008). Predicting work performance in nuclear power plants. *Safety Science*, *46*(7), 1115-1124.

International Atomic Energy Agency (2009). *Implementing digital instrumentation and control systems in the modernization of nuclear power plants.* IAEA Nuclear Energy Series No. NP-T-1.4. Vienna, Austria: IAEA.

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. *International Journal of Human-Computer Studies*, *131*, 99–107.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International journal of Cognitive Ergonomics*, *4*(1), 53-71.

Joe, J. C., & Kovesdi, C. R. (2018). *Developing a strategy for full nuclear plant modernization* (No. INL/EXT-18-51366-Rev000). Idaho National Lab. (INL), Idaho Falls, ID (United States).

Joe, J. C., Boring, R. L., & Persensky, J. J. (2012). Commercial utility perspectives on nuclear power plant control room modernization. *8th International Topical Meeting on Nuclear Power Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT)*, 2039-2046.

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology*, *34*(2), 237–257.

Jou, Y. T., Yenn, T. C., Lin, C. J., Yang, C. W., & Chiang, C. C. (2009). Evaluation of operators' mental workload of human–system interface automation in the advanced nuclear power plants. *Nuclear Engineering and Design*, *239*(11), 2537–2542.

Jou, Y. T., Yenn, T. C., Lin, C. J., Yang, C. W., & Lin, S. F. (2009, March). Evaluation of mental workload in automation design for a main control room task. In *2009 International Conference on Networking, Sensing and Control* (pp. 313-317). IEEE.

Kaber, D. B. (2018). Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 7-24.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113-153.

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *10*(4), 409-430.

Kaber, D. B., Riley, J. M., Tan, K. W., & Endsley, M. R. (2001). On the design of adaptive automation for complex systems. *International Journal of Cognitive Ergonomics*, *5*(1), 37-57.

Kamzanova, A., Kustubayeva, A.M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors, 56*, 1136-1149.

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, *65*(2), 337-359.

Kim, Y., & Park, J. (2018, October). Envisioning human-automation interactions for responding emergency situations of NPPs: A viewpoint from human computer interaction. *Proceedings of Transactions of the Korean Nuclear Society Autumn Meeting*.

Kim, Y., Jung, W., & Kim, S. (2014). Empirical investigation of workloads of operators in advanced control rooms. *Journal of Nuclear Science and Technology*, *51*(6), 744-751.

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological bulletin*, *144*(6), 557.

Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y. C., & Shaw, T. H. (2021). Measurement of trust in automation: A narrative review and reference guide. *Frontiers in psychology*, *12*, 604977.

Laarni, J. (2021). Multitasking and interruption handling in control room operator work. In A. Teperi & N. Gotcheva (Eds.), *Human Factors in the nuclear industry* (pp. 127-149). Woodhead Publishing.

Le Blanc, K. L., & Oxstrand, J. H. (2013, September). Computer–Based Procedures for Nuclear Power Plant Field Workers: Preliminary Results from Two Evaluation Studies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1722-1726). Sage CA: Los Angeles, CA: SAGE Publications.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80.

Lew, R., Boring, R. L., & Ulrich, T. A. (2018). A computerized procedure system framework for US utilities. In S. Haugen, A. Barros, C. van Gullijk, T. Kongsvik, & J. E. Vinnem (Eds) *Safety and reliability–safe societies in a changing World* (pp. 427-432). CRC Press.

Lew, R., Boring, R. L., & Ulrich, T. A. (2018, September). Transitioning nuclear power plant main control room from paper based procedures to computer based procedures. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 1605-1609). Sage CA: Los Angeles, CA: SAGE Publications.

Lin, C. J., Yenn, T. C., & Yang, C. W. (2010). Optimizing human–system interface automation design based on a skill-rule-knowledge framework. *Nuclear Engineering and Design*, *240*(7), 1897-1905.

Lin, C. J., Yenn, T. C., Jou, Y. T., Hsieh, T. L., & Yang, C. W. (2013). Analyzing the staffing and workload in the main control room of the advanced nuclear power plant from the human information processing perspective. *Safety Science*, *57*, 161–168.

Lin, J., Matthews, G., Hughes, N., & Dickerson, K. (2022). Novices as models of expert operators: Evidence from the NRC Human Performance Test Facility. In J. Wright & D. Barber (Eds.), *Human Factors and Simulation. The 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022). AHFE Open Access, vol 30* (pp. 1-10). New York: AHFE International.

Lin, J., Matthews, G., Wohleber, R. W., Funke, G. J., Calhoun, G. L., Ruff. H. A., Szalma, J. & Chiu, P. (2020). Overload and automation-dependence in a multi-UAS simulation: Task demand and individual difference factors. *Journal of Experimental Psychology: Applied*, *26*(2), 218-235.

Lipner, M. H., & Kerch, S. P. (1994, October). Operational benefits of an advanced computerized procedures system. In *Proceedings of 1994 IEEE Nuclear Science Symposium-NSS'94* (Vol. 3, pp. 1068-1072). IEEE.

Liu, P., & Li, Z. (2016). Comparison between conventional and digital nuclear power plant main control rooms: A task complexity perspective, part II: Detailed results and analysis. *International Journal of Industrial Ergonomics*, *51*, 10-20.

Lu, C., Lyu, J., Zhang, L., Gong, A., Fan, Y., Yan, J., & Li, X. (2020). Nuclear power plants with artificial intelligence in industry 4.0 era: Top-level design and current applications—A systemic review. *IEEE Access*, *8*, 194315-194332.

Lyons, J. B., & Guznov, S. Y. (2019). Individual differences in human–machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, *20*(4), 440-458.

Matthews, G., & Reinerman-Jones, L. (2017). *Workload assessment: How to diagnose workload issues and enhance performance.* Santa Monica, CA: *Human Factors and Ergonomics Society*.

Matthews, G., Ateniese, G., Barbará, D., Thayer, S. C., & Leskovec, N. C. (2024, in press). Usage of an AI-Based password tool: Impacts of security fatigue, age, and individual differences. *Proceedings of the Meeting of the Human Factors and Ergonomics Society* (Vol. 68).

Matthews, G., Davies, D.R., Westerman, S.J., & Stammers, R.B. (2000). *Human performance*: *Cognition, stress and individual differences*. London: Psychology Press.

Matthews, G., Hancock, P. A., Lin, J., Panganiban, A. R., Reinerman-Jones, L. E., Szalma, J. L., & Wohleber, R. W. (2021). Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Personality and Individual Differences*, *169*, 109969.

Matthews, G., Hancock, P. A., Lin, J., Panganiban, A. R., Reinerman-Jones, L. E., Szalma, J. L., & Wohleber, R. W. (2021). Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Personality and individual differences*, *169*, 109969.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human Factors, 57*(1), 125-143.

Matthews, G., Warm, J. S., & Smith, A. P. (2017). Task engagement and attentional resources: Multivariate models for individual differences and stress factors in vigilance. *Human Factors, 59*, 44-61.

Matthews, R. B. (2005). Nuclear safety: Expect the unexpected. *Professional Safety*, *50*(12), 20.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734.

Medema, H., Savchenko, K., Boring, R., & Ulrich, T. (2019). Defining mutual awareness: Results of reactor operator surveys on the emergence of digital technology in main control rooms. In R. Boring (Ed.), *Advances in human error, reliability, resilience, and performance* (pp. 58-67). Springer International Publishing.

Merritt, S. M., Unnerstall, J. L., Lee, D., & Huber, K. (2015). Measuring individual differences in the perfect automation schema. *Human Factors*, *57*(5), 740-753.

Mumaw, R. J. (2024). Not automation failures, but automation interface failures. *Journal of Cognitive Engineering and Decision Making*. Advance online publication.

Nystad, E., Kaarstad, M., & McDonald, R. (2019). Crew decision-making in situations with degraded information. *Proceedings of the 29th European Safety and Reliability Conference.*

O'Hara, J. M., & Fleger, S. A. (2020). *Human-System Interface Design Review Guidelines* (NUREG-0700, Rev.3). Washington, DC: United States Nuclear Regulatory Commission.

O'Hara, J. M., Higgins, J. C., Fleger, S. A., & Pieringer, P. A. (2012). *Human Factors Engineering Program Review Model* (NUREG-0711, Rev.3). Washington, DC: United States Nuclear Regulatory Commission.

O'Hara, J. M., Higgins, J. C., & Brown, W. S. (2009). Identification and evaluation of human factors issues associated with emerging nuclear plant technology. *Nuclear Engineering and Technology*, *41*(3), 225-236.

O'Hara, J. M., Higgins, J. C., Stubler, W. F., & Kramer, J. (2000). *Computer-based procedure systems: Technical basis and human factors review guidance*. Division of Systems Analysis and Regulatory Effectiveness, Office of Nuclear Regulatory Research, US Nuclear Regulatory Commission.

Oxstrand, J., Le Blanc, K. L., & Bly, A. (2015). *The next step in deployment of Computer Based Procedures for field workers: Insights and results from field evaluations at nuclear power plants*. Technical report INL/CON-14-32990. Idaho National Lab (INL), Idaho Falls, ID.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Park, J., Ulrich, T. A., Boring PhD, R. L., Kim, J., Lee, S., & Park, B. (2020). *An empirical study on the use of the rancor microworld simulator to support full-scope data collection* (No. INL/CON-20-57751-Rev000). Idaho National Laboratory (INL), Idaho Falls, ID (United States).

Park, J., & Im, Y. (2021). Visual enhancements for the driver's information search on automotive head-up display. *International Journal of Human–Computer Interaction*, *37*(18), 1737-1748.

Patton, C. E., & Wickens, C. D. (2024). The relationship of trust and dependence. *Ergonomics*, 1-17.

Porthin, M., Liinasuo, M., & Kling, T. (2020). Effects of digitalization of nuclear power plant control rooms on human reliability analysis – A review. *Reliability Engineering & System Safety*, *194*, 106415.

Qing, T., Liu, Z., Tang, Y., Hu, H., Zhang, L., & Chen, S. (2021). Effects of automation for emergency operating procedures on human performance in a nuclear power plant. *Health Physics*, *121*(3), 261.

Reinerman-Jones, L. E., Guznov, S., Mercado, J., & D'Agostino, A. (2013). Developing methodology for experimentation using a nuclear power plant simulator. In D. D. Schmorrow, & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition* (pp. 181-188). Heidelberg, Germany: Springer.

Reinerman-Jones, L. E., Harris, J., Hughes. N., & D'Agostino, A. (2017, June). *Workload response to soft controls presented on two interfaces*. Paper presented at American Nuclear Society 10th International Conference on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies, San Francisco, CA.

Reinerman-Jones, L. E., Lin, J., Matthews, G., Barber, D., & Hughes, N. (2019). *Human performance test facility experiment 4: Former operator workload and performance comparison between two simulated environments.* Rockville, MD: United States Nuclear Regulatory Commission.

Reinerman-Jones, L., Matthews, G., & Mercado, J. E. (2016). Detection tasks in nuclear power plant operation: Vigilance decrement and physiological workload monitoring. *Safety Science, 88*, 978-107.

Reinerman-Jones, L., Matthews, G., & Mercado, J. E. (2016). Detection tasks in nuclear power

Sauer, J., Chavaillaz, A., & Wastell, D. (2016). Experience of automation failures in training: effects on trust, automation bias, complacency and performance. *Ergonomics*, *59*(6), 767-780.

Saxby, D.J., Matthews, G., Warm, J.S., Hitchcock, E.M., & Neubauer, C. (2013) Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied, 19,* 287-300.

Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, *58*(3), 377-400.

Seong, P. H., Kang, H. G., Na, M. G., Kim, J. H., Heo, G., & Jung, Y. (2013). Advanced MMIS toward substantial reduction in human errors in NPPs. *Nuclear Engineering and Technology*, *45*(2), 125-140.

Sethu, M., Kotla, B., Russell, D., Madadi, M., Titu, N. A., Coble, J. B., ... & Khojandi, A. (2023). Application of Artificial Intelligence in detection and mitigation of human factor errors in nuclear power plants: A review. *Nuclear Technology*, *209*(3), 276-294.

Sheridan, T. B. (1988). Task allocation and supervisory control. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 159-173). Amsterdam: North-Holland.

Shivakumar, A., Bositty, A., Peters, N. S., & Pei, Y. (2020). Real-time interruption management system for efficient distributed collaboration in multi-tasking environments. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1), 1-23.

Singh, A. L., Tiwari, T., & Singh, I. L. (2009). Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *Journal of the Indian Academy of Applied Psychology*, *35*(2009), 9-22.

Skjerve, A. B. M., & Skraaning Jr, G. (2004). The quality of human-automation cooperation in human-system interface for nuclear power plants. *International Journal of Human-Computer Studies*, *61*(5), 649-677.

Skraaning Jr, G., Eitrheim, M. H., & Lau, N. (2010). Coping with automation in future plants. In *The 7th American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC&HMIT), Las Vegas, NV*.

Spielman, Z., & Le Blanc, K. (2021). Boeing 737 MAX: Expectation of human capability in highly automated systems. In *Advances in Human Factors in Robots, Drones and Unmanned Systems: Proceedings of the AHFE 2020 Virtual Conference on Human Factors in Robots, Drones and Unmanned Systems, July 16-20, 2020, USA* (pp. 64-70). Springer International Publishing.

Steelman, K. S., McCarley, J. S., & Wickens, C. D. (2017). Theory-based models of attention in visual workspaces. *International Journal of Human–Computer Interaction*, *33*(1), 35-43.

Swain, A. D., & Guttmann, H. E. (1983). *Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report* (No. NUREG/CR--1278). Albuquerque, NM: Sandia National Labs.

Tasset, D., Charron, S., Miberg, A. B., & Hollnagel, E. (1999). *The impact of automation on operator performance. An explorative study* (No. HRP--352/V. 1).

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, *39*(5), 740-748.

Taylor, G., Reinerman-Jones, L. E., Cosenzo, K., & Nicholson, D. (2010). Comparison of multiple physiological sensors to classify operator state in adaptive automation systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(3), 195-199.

Thomson, R., Cassenti, D. N., & Hawkins, T. (2024). Too much of a good thing: How varying levels of automation impact user performance in a simulated intrusion detection task. *Computers in Human Behavior Reports*, *16*, 100511.Tipping, P. (2005, February). Aspects of unexpected events in nuclear power plants. In *Proceedings of a Technical Meeting: Material Degradation and Related Managerial Issues at Nuclear Power Plants* (pp. 125-129), International Atomic Energy Agency.

Ulrich, T. A., Boring, R. L., & Lew, R. (2015, August). Control board digital interface input devices-touchscreen, trackpad, or mouse?. In *2015 resilience week (RWS)* (pp. 1-6). IEEE.

Ulrich, T., Boring, R., & Lew, R. (2022). Studying Control Room Operations on a Shoestring Budget–Reflections on the Rancor Microworld. *Human Factors and Simulation*, *30*(30).

Ulrich, T. A., Boring, R. L., Lew, R., & Whiting, T. A. (2023, September). Rancor Computer-Based Procedures–A Framework For Task Level Human Performance Data Collection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 67, No. 1, pp. 1023-1028). Sage CA: Los Angeles, CA: SAGE Publications.

U. S. Nuclear Regulatory Commission. (1998). *NRC Regulations Title 10, Code of Federal Regulations*. Retrieved from http://www.nrc.gov/reading-rm/doc-collections/cfr/part050/full-text.html

Vicente, K. J., & Rasmussen, J. (1988, October). On applying the skills, rules, knowledge framework to interface design. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 32, No. 5, pp. 254-258). Sage CA: Los Angeles, CA: SAGE Publications.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433–441.

Warm, J.S., Tripp, L.D., Matthews, G., & Helton, W.S. (2012). Cerebral hemodynamic indices of operator fatigue in vigilance. In G. Matthews, P.A. Desmond, C. Neubauer & P.A. Hancock (Eds.), *Handbook of operator fatigue* (pp. 197-207). Aldershot, UK: Ashgate Press.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449-455.

Wickens, C.D. (1991). Processing resources and attention. In D. Damos (Ed.), *Multiple task performance* (pp. 1-34). London: Taylor & Francis.

Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology*, *34*(2), 163–178.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, *12*(1), 3-18.

Wu, X., & Li, Z. (2013). Secondary task method for workload measurement in alarm monitoring and identification tasks. In *Cross-Cultural Design. Methods, Practice, and Case Studies: 5th International Conference, CCD 2013* (pp. 346-354). Berlin: Springer Berlin Heidelberg.

Human Performance Test Facility (HPTF) Volume 6 DATE September 30, 2025

DISTRIBUTION:
JGrasso, RES/DRA/HFRB
HWatkins, RES/DRA/HFRB
NHughes, RES/DRA/HFRB
SPeters, RES/DRA/HFRB
BSmith, RES/DRA
JTappert, RES/DRA
JAnderson, NRR/DRO/IOLB
BGreen, NRR/DRO/IOLB/HFT
LNist, NRR/DRO/IOLB
MScheetz, NRR/DRO/IOLB
TCox, NRR/DRO/IOLB/HFT
DKi, NRR/DRO/IOLB/HFT
KMartin, NRR/DRO/IOLB/HFT
JSteckel, RES/DE/RGGIB
EO'Donnell, RES/DE/RGGIB
CPeres, NRR/DRO/IOLB/HFT

**ADAMS Accession No.: ML25268A030**

| | | | | |
|---|---|---|---|---|
| OFFICE | RES/DRA/HFRB | RES/DRA/HFRB | RES/DRA/HFRB | RES/DRA |
| NAME | JGrasso *JG* | HWatkins *HW* | SPeters *SP* | BSmith *BS* |
| DATE | Sep 25, 2025 | Sep 29, 2025 | Sep 30, 2025 | Sep 30, 2025 |

***OFFICIAL RECORD COPY***