NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

MODEL-BASED ASSESSMENT OF ADAPTIVE AUTOMATION'S UNINTENDED CONSEQUENCES

by

Charles P. Rowan

June 2023

Dissertation Supervisor: Lawrence G. Shattuck

Research for this dissertation was performed at the MOVES Institute.

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE				*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.				
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2023	**3. REPORT TYPE AND DATES COVERED** Dissertation		
4. TITLE AND SUBTITLE MODEL-BASED ASSESSMENT OF ADAPTIVE AUTOMATION'S UNINTENDED CONSEQUENCES				**5. FUNDING NUMBERS** RNK76
6. AUTHOR(S) Charles P. Rowan				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000				**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Commanding Officer, U.S. Army Combat Capabilities Development Command, Aviation and Missile Center Agency, Redstone Arsenal, Alabama 35898-5000				**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.				**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Recent technological advances require development of human-centered principles for their inclusion into complex systems. While such programs incorporate revolutionary hardware and software advances, there is a necessary space for including human operator design considerations, such as cognitive workload. As technologies mature, it is essential to understand the impacts that these emerging systems will have on cognitive workload. Adaptive automation is a solution that seeks to manage cognitive workload at optimal levels. Human performance modeling shows potential for modeling the effects of adaptive automation on cognitive workload. However, the introduction of adaptive automation into a system can also present unintended negative consequences to an operator. This dissertation investigated potential negative unintended consequences of adaptive automation through the development of human performance models of a multi-tasking simulation. One hundred twenty participants were enrolled in three human-in-the-loop experimental studies (forty participants each) that collected objective and subjective surrogate measures of cognitive workload to validate the models. Results from this research indicate that there are residual increases in operator workload after transitions in system states between manual and automatic control of a task that need to be included in human performance models and in system design considerations.

14. SUBJECT TERMS			**15. NUMBER OF PAGES**
modeling and simulation, human systems integration, adaptive automation, human performance modeling, cognitive workload			273
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

MODEL-BASED ASSESSMENT OF ADAPTIVE AUTOMATION'S UNINTENDED CONSEQUENCES

Charles P. Rowan
Lieutenant Colonel, United States Army
BS, United States Military Academy, 2005
MS, Human Systems Integration, Naval Postgraduate School, 2014

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN MODELING, VIRTUAL ENVIRONMENTS, AND SIMULATION

from the

NAVAL POSTGRADUATE SCHOOL
June 2023

Approved by: Lawrence G. Shattuck
Department of Research
Dissertation Supervisor

Imre L. Balogh
MOVES Institute
Dissertation Chair

Rudolph P. Darken
Department of
Computer Science

Glenn A. Hodges
MOVES Institute

Panagiotis Matsangas
Crew Endurance Team

Approved by: Gurminder Singh
Chair, Department of Computer Science

Joseph P. Hooper
Vice Provost of Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Recent technological advances require development of human-centered principles for their inclusion into complex systems. While such programs incorporate revolutionary hardware and software advances, there is a necessary space for including human operator design considerations, such as cognitive workload. As technologies mature, it is essential to understand the impacts that these emerging systems will have on cognitive workload. Adaptive automation is a solution that seeks to manage cognitive workload at optimal levels. Human performance modeling shows potential for modeling the effects of adaptive automation on cognitive workload. However, the introduction of adaptive automation into a system can also present unintended negative consequences to an operator. This dissertation investigated potential negative unintended consequences of adaptive automation through the development of human performance models of a multi-tasking simulation. One hundred twenty participants were enrolled in three human-in-the-loop experimental studies (forty participants each) that collected objective and subjective surrogate measures of cognitive workload to validate the models. Results from this research indicate that there are residual increases in operator workload after transitions in system states between manual and automatic control of a task that need to be included in human performance models and in system design considerations.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AA	adaptive automation
ACT-R	Adaptive Control of Thought – Rational
AFC	Army Futures Command
ANS	autonomic nervous system
BWRS	Bedford Workload Rating Scale
COMM	MATB-II communications task
CSWAG	Continuous Subjective Workload Assessment Graph
CTA	Cognitive Task Analysis
DOD	Department of Defense
ECG	electrocardiography
EEG	electroencephalography
EMA	exponential moving average
FAA	Federal Aviation Administration
fMRI	functional magnetic resonance imaging
fNIRS	functional near-infrared spectroscopy
FOM	figure of merit
FVL	Future Vertical Lift
GIMP	GNU Image Manipulation Program
GSR	galvanic skin response
Hb	deoxygenated hemoglobin
HbO	oxygenated hemoglobin
HCI	human-computer interaction
HFE	human factors engineering

HIP	Human Information Processing
HITL	human in the loop
HRV	heart rate variability
HSA-DM	Holistic Situational Awareness and Decision Making
HSI	Human Systems Integration
HSIL	Human Systems Integration Laboratory
HTA	Hierarchical Task Analysis
HTI	Human Technology Integration
IBI	interbeat interval
IMPRINT	Improved Performance Research Integration Tool
IRB	Institutional Review Board
ISA	Instantaneous Self-Assessment
JTA	Job Task Analysis
LOA	Levels of Automation
LSL	Lab Streaming Layer
M&S	modeling and simulation
MACE	Malvern Capacity Estimate
MAT	Multi-Attribute Task
MATB-II	Multi-Attribute Task Battery II
MCH	Modified Cooper-Harper
MEDEVAC	medical evacuation
MLCC	multi-level cognitive cybernetics
MRT	Multiple Resource Theory
MWL	mental workload

NASA	National Aeronautics and Space Administration
NASA-TLX	National Aeronautics and Space Administration Task Load Index
NPS	Naval Postgraduate School
OOTL	out of the loop
OSM	operator state monitoring
PFC	pre-frontal cortex
PNS	parasympathetic nervous system
RESMAN	MATB-II resource management task
RI	resource-interface
RMSSD	Root mean square of successive differences
SA	situation awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SMA	simple moving average
SPAM	Situation Present Assessment Method
SWAT	Subjective Workload Assessment Technique
SYSMON	MATB-II system monitoring task
TA	task analysis
TRACK	MATB-II tracking task
USAARL	United States Army Aeromedical Research Laboratory
V&V	verification and validation
VACP	visual, auditory, cognitive, psychomotor
VV&A	verification, validation, and accreditation
WP	Workload Profile

XDF	Extensible Data Format
XML	Extensible Markup Language

EXECUTIVE SUMMARY

Technological advances that seek to address future operational challenges abound in the Department of Defense (DOD). An example of this technology evolution is the U.S. Army's Future Vertical Lift (FVL) Program. The FVL Program is seeking to modernize the Army's rotary air assets to fly faster and farther than ever before. While this capability will incorporate numerous capabilities, some of which are not yet realized, there is an important space for human design considerations. One human operator design consideration worthy of investigation is cognitive workload. Current aviation systems lead to increased operator workload values above accepted levels in predictive models (Militello et al., 2019). Therefore, the development of FVL lends itself to addressing operator cognitive workload levels in its early system acquisition stages. One solution that has sought to improve cognitive workload management is adaptive automation (AA), or the ability to dynamically change the level of automation in response to varying system demands (Sheridan, 2011). The implementation of AA seeks to increase or decrease operator cognitive workload within an optimal range (Inagaki, 2003; Scerbo, 1996). While AA has the potential to achieve these ends, there also exists the potential of AA introducing unintended negative consequences into a system (de Visser & Parasuraman, 2011; Kaber & Endsley, 2004; Smith & Baumann, 2020).

This dissertation sought to investigate the unintended negative consequences of AA through an adaption of Cassenti, Cox, and Bakdash's (2017) Multi-Level Cognitive Cybernetic framework and Wickens' (2008) Multiple Resource Theory. This research used a novel, model-based approach to assess the impacts of AA on cognitive workload. This assessment was conducted by modeling tasks in NASA's Multi-attribute Task Battery-II (MATB-II) using the Army's Improved Performance Research Integration Tool (IMPRINT). The MATB-II simulation served to replicate aspects of a flight task. The four tasks included in MATB-II are system monitoring, tracking, communications, and resource management.

Cognitive workload, situation awareness (SA), and performance are three constructs that have shown to be inextricably linked in various settings (Endsley, 2021;

Ernst et al., 2020). All three constructs can be impacted by the introduction of AA into a system. The current effort sought to investigate the relationships among these three constructs over the course of the three studies using MATB-II, while also attempting to validate cognitive workload modeling predictions in IMPRINT.

Cognitive workload prediction models of the MATB-II scenarios were modeled by the researcher for every scenario presented in the three studies. The researcher used the default workload demand values provided in IMPRINT to build initial workload models. Additional models were constructed after conducting cognitive walk-throughs with three MATB expert users. The expert users had each operated an adapted version of the MATB for at least 20 hours. The results of this approach yielded models that followed the researcher-derived models in workload demand, suggesting that the validated metrics provided in IMPRINT serve as reliable anchors to adjust baseline workload predictions.

Three human-in-the-loop studies were conducted at the Naval Postgraduate School (NPS). The studies were approved by the NPS Institutional Review Board (IRB). A total of 133 participants volunteered for the three studies. Participants included staff, faculty, and students at NPS. After consideration for emergent behaviors, incomplete data sets, and other data collection issues, 40 participants were included in the analysis of each of the three studies. Participants could only participate in one study in an effort to mitigate learning effects.

The first study examined how objective and subjective surrogate cognitive workload measures related to cognitive demand. Investigation into different experience levels was included by using two training progressions to create a novice and experienced group of MATB-II operators. After completing their respective training progressions, all participants were instrumented with an eye tracking device, a heart rate monitor, and a functional near infrared spectroscopy (fNIRS) headband to obtain physiological measurements during their trial runs. The trial runs consisted of a 10-minute-low and a 10-minute-high workload condition, with presentation counterbalanced across participants. There was a break between the 10-minute trial runs to recalibrate the eye tracker and change MATB-II scenarios. Participants were asked to subjectively rate their cognitive workload as a percentage of their maximum workload every minute during the trials using the Continuous Subjective Workload Assessment Graph (CSWAG) (Shattuck & Miller, 2006). Following the trial runs, participants completed the NASA Task Load Index (NASA-TLX) and the Situation Awareness Rating Technique questionnaire. Results from the first study indicated that there were significant differences between workload conditions when assessed against mean R-R HRV intervals, pupil diameter, and CSWAG results. Performance data also supported differences in experience levels and workload conditions.

The results of Study 1 informed Study 2, where a higher level of automation was introduced in the tracking task of MATB-II. This task can be likened to flying an aircraft in manual or in autopilot. Study 2's participants used the same procedures and experimental setup as those in Study 1, except for only completing one level of workload with two levels of automation. Results from Study 2 indicated increased MATB-II performance scores, increased HRV, and decreased pupil diameter during higher levels of automation. Further, CSWAG results decreased during lower workload and higher levels of automation conditions. These results followed the same pattern as Study 1. Additionally, a significant inverse correlation was found between post-trial NASA-TLX and SART ratings. These results indicated that participants reported lower perceived SA with higher self-assessed workload.

Study 3 incorporated the insights from Studies 1 and 2 in attempt to determine if workload forecasts could be made based on cognitive workload measure changes with dynamically changing levels of automation. Of note, fNIRS was not used in this study because no differences were found across experience groups or experimental conditions in Studies 1 and 2. There were also no significant differences in subjective or objective cognitive workload surrogate measures between novice and experienced participants in the first two studies. Therefore, the decision was made to proceed with the experienced training progression in Study 3 to account for any learning effects and to focus more on the impacts of dynamically changing levels of automation on cognitive workload. After completing their training progressions, participants completed a continuous 20-minute trial with low and high workload levels and low and high levels of automation presented in 5-minute, counterbalanced segments. Results from Study 3 followed the patterns seen in the first two studies with respect to significant differences in performance, mean HRV, mean pupil diameter, and CSWAG percentages between workload and tracking conditions. There was also a significant inverse correlation between NASA-TLX and SART ratings again.

Key findings from Study 3 that addressed this dissertation's objectives were the significant differences in both mean pupil diameters in the 5-, 15-, and 30-second windows before and after the system state transitions that occurred at the 5-, 10-, and 15-minute marks during the trials. This finding suggests that there was an associated workload cost during the transitions to different system states. While adaptive automation was modeled to lower workload immediately, the results indicated that refinement needed to be made to the models to account for this transition period. Additionally, subjective CSWAG ratings were also significantly different in the 30 second window before and after a system state transition. Based on these results, new IMPRINT models were developed that showed increases in overall workload demands during the transition periods of the four trial run conditions.

The results from this dissertation highlight the impacts of AA on cognitive workload. While the results are limited to the sampled population and controlled laboratory setting, the approach of pairing cognitive workload model predictions with human-in-the-loop validation studies provided insights into the impacts of AA on workload. The results of this research indicated that further investigation into other unintended negative consequences of AA could benefit from this approach. These research efforts can help inform human performance modeling and system design considerations early in a system's life cycle. While we develop technologies to address future operational environments, we should address human cognitive workload considerations in parallel with those efforts.

References

Cassenti, D., Gamble, K. R., & Bakdash, J. Z. (2017). Multi-level cognitive cybernetics in human factors. In K. S. Hale & K. M. Stanney (Eds.), *Advances in Neuroergonomics and Cognitive Engineering* (pp. 315–326). Springer International Publishing. https://doi.org/10.1007/978-3-319-41691-5_27

de Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 209–231. https://doi.org/10.1177/1555343411410160

Endsley, M. R. (2021). Situation awareness. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (1st ed., pp. 434–455). Wiley. https://doi.org/10.1002/9781119636113.ch17

Ernst, K., Wilson, M., Durbin, D., Roth, R., Militello, L., & Copeland, R. (2020). Human factors in future vertical lift, the U.S. Army's next generation rotorcraft. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 149–153.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. *Handbook of Cognitive Task Design*, *8*, 147–169.

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. https://doi.org/10.1080/1463922021000054335

Militello, L. G., Roth, E., Scheff, S., Ernst, K., Sushereba, C., Diiulio, J., & Klein, D. (2019). *Toward an optimally crewed future vertical lift vehicle: Crewing strategy and recommendations*. U.S. Army Combat Capabilities Development Command.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In *Automation and human performance: Theory and applications*. CRC Press.

Shattuck, L. G., & Miller, N. L. (2006). Extending naturalistic decision making to complex organizations: A dynamic model of situated cognition. *Organization Studies*, *27*(7), 989–1009. https://doi.org/10.1177/0170840606065706

Sheridan, T. B. (2011). Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, *41*(4), 662–667. https://doi.org/10.1109/TSMCA.2010.2093888

Smith, P., & Baumann, E. (2020). Human-automation teaming: Unintended consequences of automation on user performance. *2020 AIAA/IEEE 39th Digital Avionics Systems Conference* (DASC), 1–9.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

ACKNOWLEDGMENTS

I thank God for His amazing love and grace.

I often joked that the acknowledgements section would be the largest part of this dissertation. I still feel that way and could go on for a long time about all of those that so graciously gave of their time and talents to help me. I am forever humbled and grateful for this wonderful community of people that helped me realize this dream.

I would like to thank my committee for their investment in me and my work. I am biased, but I cannot have imagined a more formidable group of professionals to shepherd me through this process. What an honor it was to be mentored and supported by you these past three years.

Dr. Balogh, thank you for graciously volunteering to serve as my committee chair, and for helping me keep my eyes focused on how this work can assist the modeling and simulation community of tomorrow. These were invaluable lessons for me.

Dr. Darken, you are the consummate professional and your ideas challenged me to think more critically as I approached problems. Thank you for your mentorship and wisdom throughout my time at NPS—you are a combat multiplier.

Dr. Hodges, thank you for everything you did to get me here, literally and metaphorically. I am honored to follow in your footsteps as an FA57 and MOVES Ph.D. graduate. Thank you for offering your personal and professional insights along the way.

Dr. Matsangas, you are a great American! I often think that one of the best decisions I made in my life was asking you to join my committee. Your ability to coach, teach, and mentor is first class, and I am eternally grateful for the assistance you provided to me. Thank you, my friend, and this work is just the beginning.

Thank you to Ms. Yamira Santiago-Espada at NASA for working so diligently to allow me to use MATB-II for my research.

To my friends at the U.S. Army Aeromedical Research Laboratory, thank you will never be enough. To Dr. Paul St. Onge, Dr. Kevin O'Brien, Mr. Jon Vogl, Mr. Chris

Aura, and all the great teammates at USAARL—I would not be here without you. You are incredible professionals, and I cannot thank you enough for the time and energy you poured into this dissertation. I only hope to be fortunate enough to be able to help the next one in line as much as you helped me.

To Matt Shivers and all the wonderful professionals whom I met around the HSA-DM Program, thank you for your generous support for this research. It was an honor to learn from you all and inspiring to be around so many dedicated people wanting to make a difference for our warfighters.

To Ms. Michelle Morneau at the NPS Graduate Writing Center, thank you for the many hours of reviewing my writing and helping me to communicate my ideas more effectively. You are a dedicated professional, and I know that the work that you and your teammates do for all of us at NPS truly elevates our institution.

Thank you to Bob Sargent for being a masterful teacher and friend. Your knowledge of IMPRINT reduced my cognitive workload considerably. Thank you for the hours of troubleshooting and model development assistance.

To Matt Morse and Scotty Black, thank you for your friendship and being excellent teammates in the Ph.D. office. I learned such an incredible amount from you both and will forever treasure the opportunity to work with you. The Marine Corps is in great hands.

A very special thank you to Mike O'Neil and Cody Nicholson for your friendship and assistance with data collection. It is hard for me to imagine a world where any of this gets done without you. You are incredible professionals, and I am thankful to have gotten to work with you (and to work with Mike for another round!).

Thank you to my family for their support. To my mother, I'm finally a doctor.

To my friends who suffered through years of my stress—thank you. You all never faltered in your belief of me and always picked me up whenever I got down. You are my family, and I am more than blessed to have had you all along for the ride with me.

Whether it was the daily texts in the group, Friday night film sessions, or (of course) Army Football…you always were there. Forward my brothers and sisters—thank you.

Thank you to COL (R) Larry Shattuck, Ph.D., for everything. From my senior capstone project at West Point, to my master's thesis at NPS, and now my dissertation, you hold the distinction (or risk) of being responsible for all my adult education. You have always been in my corner and have always known exactly how to get me to realize my potential. Thank you, sir, for everything the last 20 years. I will forever count myself as blessed to be a part of the "Shattuck Coaching Tree" and will strive to be there for the next generation as you were for me.

To my amazing wife, Caitlin: we did it, babe! Thank you for always reminding me that I was the only one who didn't know that I was Ric Flair. You never stopped believing in me and have supported me through the all the highs and lows of this journey. You were my inspiration to pursue this degree, and you sacrificed incredibly to allow me to do so. I cannot express what that will forever mean to me, and I am so honored to be your husband. I love you more than anything in this world.

To our beloved, Sancho, who went to heaven (like all dogs do) about a month before I defended: we'll always love you, buddy.

Sancho Panza Rowan

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

The U.S. Army's Future Vertical Lift (FVL) program is set to replace legacy rotary lift vehicles to provide enhanced capabilities to meet operational requirements. The FVL program represents a technological evolution for the Army as new materiel solutions are developed with the goal of meeting future challenges. While many FVL developmental efforts center on hardware and software advancements, there is a necessary space for addressing human operator design considerations including cognitive workload considerations. To do so, the FVL program created the Holistic Situational Awareness and Decision Making (HSA-DM) subsidiary program to identify the drivers of cognitive workload and to develop cognitive workload management capabilities (Department of Defense, 2020a). The HSA-DM program is part of a vast research effort seeking to investigate the integration of adaptive automation to address cognitive overload.

As the Army contemplates novel solutions to address emerging challenges within the FVL program, technologies continue to develop in both capability and complexity. This evolution creates the potential for increased cognitive demand on aircrews as they are presented with additional information and tasks that increase the probability of overload (Evans & Fendley, 2017). Designers can introduce complexity by adding information, changing the level of system functionality, and introducing decision options that require more sensory inputs and cognitive resources from a human operator (Rowe et al., 1998). Cognitive overload, loss of situation awareness (SA), readiness degradation, and poorly performed tasks are among the detrimental factors that may result from poorly designed FVL systems (Department of Defense, 2020a; Durbin & Hicks, 2009).

B. PROBLEM STATEMENT

The evolution of the Army's rotary lift capability presents an opportunity to not only integrate new technology but also address acknowledged issues in introducing new technology. As technology matures and provides increased benefits in one area, it can ironically cause unexpected issues in another that counteract the benefits (Bainbridge, 1982a; Hollnagel & Woods, 2005). Similarly, the Law of Stretched Systems posits that systems are stretched to operate at maximum capacity (Woods, 2006). As soon as some incremental improvement is realized, the system will be exploited to meet the new capacity threshold. The Law of Stretched Systems also addresses the increased demand on human operators when technological advances are realized (Woods, 2002, 2006; Hoffman & Woods, 2011). Increased levels of automation (LOA) can also cause poor situation awareness and reduce the probability of successful manual control resumption when needed (Endsley, 2017a). Furthermore, as technology continues to mature, the intended operational environment is not clearly defined or unexpectedly changes. This is known as the envisioned world problem where the work domain in which the technology will be used is not yet realized (Woods & Dekker, 2000). These considerations are directly applicable to FVL's system design.

As technology matures, more systems are being developed that remove the human from many aspects of system operation. The very nature of work has transitioned from a primarily physical activity to an automation-assisted activity in many work domains including aviation, medicine, driving, and nuclear power plant operations. However, human workload may increase and performance may degrade due to added system complexity that was intended to alleviate demands on humans (Evans & Fendley, 2017; Mehler, Reimer, Coughlin, & Dusek, 2009). Many systems deliver a static automation solution that require human operators to determine when to engage or disengage the automated assistance. The act of deciding when to implement static automation to assist an operator has been shown to create issues and even increase reported workload (Wiener, 1988; Wiener & Curry, 1980).

The ability to anticipate and manage changing levels of cognitive workload with adaptive technology solutions may be a way to increase total system performance and provide additional resources during a range of operational conditions. Adaptive automation (AA) is the dynamic change in system control between a human and machine operator as the demands of the system exceed the resources available to address the task demand level (Kaber, Riley, Tan, & Endsley, 2001). Conversely, AA can be used to draw human operators out from a low state of cognitive demand where performance may degrade due to boredom into a level of appropriate arousal. Adaptive automation systems can potentially perform difficult tasks with greater ease and efficiency, leaving the human operator to standby as a monitor and decision maker. However, assigning these roles to a human operator can be problematic given a history of accidents occurring from humans not attending to tasks or missing critical information that they would have been directly responsible for in manually controlled settings (Wiener & Curry, 1980). Adaptive automation is a growing field of investigation that seeks to address issues associated with human error and increased workload. Adaptive automation has shown promising capability to lower operator cognitive workload (Brand & Schulte, 2017; Inagaki, 2003; Scerbo, 2008). However, adaptive automation can also introduce adverse effects on an operator's workload (de Visser & Parasuraman, 2011; Kaber & Endsley, 2004; P. Smith & Baumann, 2020).

If FVL is to be successful, principled guidelines that address the negative unintended consequences of adaptive automation are needed to properly inform design decisions. This dissertation is a first step towards developing those guidelines. This research leverages a model-based approach to assess the unintended negative consequences of adaptive automation on cognitive workload. This dissertation will investigate those issues, measure them, and asses their impacts on cognitive workload.

C. RESEARCH APPROACH

This research sought to develop and assess a model of cognitive workload through examination of the positive and negative impacts of adaptive automation's use. This effort was broken into phases that support the validation of an overarching model. The tasks that comprise NASA's Multi-Attribute Trask Battery-II (MATB-II) Version 3.5 served as inputs for a task analysis that were built into models using the U.S. Army's Improved Performance Research Integration Tool (IMPRINT). These models provided the basis to analyze performance and workload during interactions with fully manual operator control and dynamically changing levels of automation during three experimental studies. The models were then compared with objective and subjective data collected during simulation experiments with human participants to determine the model's validity and its predictive capability of operator workload. This approach of leveraging workload data from humans in the loop as assessed by psychophysiological measures has been suggested as a line of effort to improve workload modeling tools like IMPRINT (Rusnock & Geiger, 2017). Resulting data informed the modification of the original model for future recommendations and incorporation in design processes. Additionally, situation awareness data were collected using a self-report measure to investigate the correlation between cognitive workload, performance, and SA.

D. PROJECT SCOPE AND GOALS

The approach for this research effort was broken into sequential phases that built on each other to refine and validate a model of cognitive workload when using adaptive automation. This approach was used to assess workload during interaction with adaptive automation in a simulated task.

An adapted model of multi-level cognitive cybernetics (MLCC) serves as a baseline for this dissertation's investigation into cognitive workload when using adaptive automation as seen at the top of Figure 1 (Cassenti & Veksler, 2018). The MLCC model depicts the co-adaptation between a human and a computer in a closed-loop system. The model of the dissertation research is located below the adapted MLCC graphic. Areas of particular emphasis of the dissertation and their mapping to address aspects of the MLCC are shown in blue arrows. The model begins with the assertion that psychophysiological measures correspond to cognitive demand. The results of Study 1 addressed this measurement process and its corresponding values at different levels of cognitive workload. Then, as levels of automation change, so too do psychophysiological workload measures. Study 2 incorporated the workload measures from Study 1 to investigate the relationship between workload and LOAs. Additionally, this work sought to determine what objective measures might be predictive of performance when using adaptive automation. Study 3 facilitated the examination of this final portion of the model by investigating adaptive automation's impacts as it assists but also impedes task completion. There is a continuous feedback verification and validation (V&V) loop that highlights the sequential and iterative refinement of each study. Measures used in all three humans in the loop (HITL) studies are listed as either objective or subjective. Throughout the model, external and internal factors inject variability on the impact that AA has on workload measurement and task performance. These considerations are almost endless. Therefore, this dissertation's scope was focused on individual operators interacting with a simulated system in controlled settings where various internal and external confounds were mitigated.

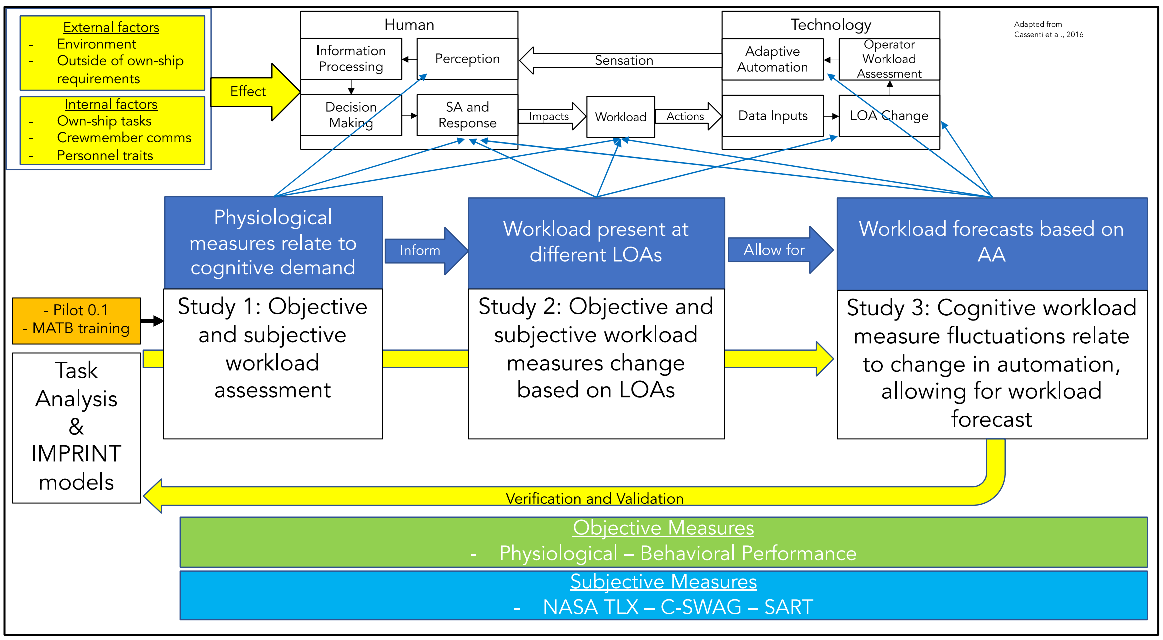


Figure 1. Adaptation of a model of cognitive workload interacting with AA, with current research efforts overlayed.

Generally, this work was broken down into a modeling effort and an experimental effort. The modeling efforts using IMPRINT and subsequent HITL studies supported the validation of the high-level model of cognitive workload when using adaptive automation. As workload modeling predictions were made based on the task analysis and construction of IMPRINT models, studies to investigate those predictions were conducted as part of verification and validation activities.

Modeling Effort

A task analysis was conducted with subject matter experts to determine the key tasks necessary to accomplish the MATB-II scenarios in the experimental studies. The results of this task analysis served as the basis for inputs modeled using IMPRINT. These models were verified in similar fashion to the task analysis. Three HITL studies were conducted using simulation to validate the models. The scenarios modeled in IMPRINT and in the simulation HITL studies used MATB-II as the referent.

Experimentation

The concepts and approaches for the three HITL experimental studies follow. Forty to fifty volunteers served as participants for each of the three studies. The number of participants recruited for the study ranged from 40–50 to support experimental power analysis while accounting for potential study dropouts and unforeseen data collection issues.

Study 1. The first study sought to induce stress to gauge participants' cognitive demand based on psychophysiological measures in two different workload conditions. The approach to achieve this leveraged MATB-II to replicate aspects of the task modeled in IMPRINT. Participants completed the MATB-II simulation during a 10-minute low workload condition and a 10-minute high workload condition to gauge objective (psychophysiological and performance) and self-reported subjective workload measures. Situation awareness data were also subjectively collected from the participants using the Situation Awareness Report Technique (SART) at the end of the trial runs. The results from this study served to validate the first portion of the dissertation's overarching model that gives psychophysiological baseline measurements for cognitive workload at two different workload levels.

Study 2. From the baseline study, a follow-on study introduced different levels of automation during a MATB-II simulation. The study sought to determine the levels of workload present at the different levels of automation. Study 2 used the same multitasking simulation in Study 1 but modified the behavior of the system to include different invocations of automation to support the operator in completing the task. The levels of automation (LOAs) for this study followed previous work from Evans and Fendley (2017) using Sheridan and Verplank's (1978) LOAs 2 (highly manual) and 9 (highly automated). Study 2 was used to address the next portion of the model that investigated how workload measures change based on LOAs.

Study 3. The results from the first two experimental efforts led to a study that assessed the effects of adaptive automation when triggered at specific times using a critical event strategy approach (Aricò et al., 2016). The study design incorporated MATB-II tasks again to determine the effects of what happens when those tasks are delegated to adaptive automation. Key to this study was determining if the workload predictions made based on the psychophysiological changes collected during the first two studies allowed for accurate forecasting of operator workload. The tasks associated with MATB-II were modeled using IMPRINT. The resulting cognitive workload prediction values served as the basis for comparison with the HITL studies. The validation process included workload prediction values by MATB-II SMEs. This final study focused on the ability to make cognitive workload predictions given multiple measures to find a way to provide real-time measurement.

Numerous objective, subjective, and performance measures were collected throughout the three studies. Situation awareness data were collected using the Situation Awareness Rating Technique (SART) (Taylor, 2011). Subjective workload assessment data was collected using the NASA-TLX (Hart & Staveland, 1988) and CSWAG (Shattuck & Miller, 2006). Objective measures were collected from eye tracking, heart rate variability, and fNIRS data. These data were contrasted against the IMPRINT model to determine the goodness of fit of the newly proposed workload model and to provide validation of the model. This approach provides a framework to assist with the validation of IMPRINT models and with identifying some limitations of the program. Further, this methodology provides recommendations to address the limitations of current human performance modeling techniques.

E. CONTRIBUTIONS

While the effects of adaptive automation are referred to often in the literature, there is not a clear effort to model those effects and measure them to provide validation of the models. This dissertation attempted to use a novel model-based approach to assess workload measurements when using adaptive automation. Specifically, a key item of interest in this dissertation was the modeling and measuring of the negative unintended consequences that emerge when using AA. Another aim of this research was to contribute to the cognitive workload body of knowledge with measurement and assessment techniques particularly developed to study adaptive automation. This research also sought to provide system design and configuration recommendations to assist with cognitive workload management for consideration in future adaptive automation systems including the Army's FVL platforms.

F. DISSERTATION ORGANIZATION

This dissertation is organized into eight chapters. Chapter I serves as the introduction and begins with an operational context that serves as the basis for the dissertation work. A primary focus of this work is on supporting the next generation of U.S. Army aviation capabilities. The chapter then introduces the dissertation's problem space involving adaptive automation and cognitive workload. Additionally, the introduction describes the research approach and supporting efforts associated with the dissertation. The introductory chapter introduces key concepts that will be used throughout the document. Finally, the chapter lists the forecasted novel contributions of this research effort.

Chapter II reviews the literature concerning important elements of the dissertation's scope. The chapter begins with an overview of the main topics addressed its subsections along with a discussion of an adapted MLCC framework that helps provide context to the current dissertation effort. Discussion includes key concepts such as cognitive workload, psychophysiological and subjective workload measures, automation in support of operations, adaptive automation, situation awareness, modeling & simulation (M&S) tools, and important M&S activities. The chapter concludes with the research questions and hypotheses that followed from the investigation of the relevant literature.

Chapters III, IV, and V describe each of the three studies conducted in support of this dissertation. Each chapter provides an overview of the study conducted as well as the methods that were used. These three chapters build on one another to highlight the progression of the studies in answering the research questions. Initial results are provided for each study that later serve as the basis for more comprehensive analysis and discussion.

Chapter VI provides a discussion the results of the three studies in context with each other. The chapter also provides analyses of the measures of workload and performance individually and together. This analysis is used to validate the cognitive workload model developed by this dissertation. This chapter also analyzes the various relationships between workload, performance, and situation awareness.

Chapter VII discusses future research recommendations based on the results of the dissertation and trends that emerged during data analysis. Further, applications of the research approach to other domains are recommended for investigation for systems using AA. The dissertation concludes with key takeaways in summation of the total effort. Final discussion regarding the research's end states provides the basis for analysis of the impacts of AA in systems designed to manage cognitive workload.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. OVERVIEW

The purpose of this literature review is to provide an overview of key concepts that will be used in this research. The following sections will provide an overview of these concepts and how they will be used in the overall dissertation effort. The review begins with a discussion of multi-level cognitive cybernetics, cognition, cognitive workload, and situation awareness. These concepts are discussed independently and in relation to each other. A discussion on automation follows with examination into levels of automation and adaptive automation. Specifically, a discussion on automation's impacts on workload is presented with particular emphasis placed on the unintended negative consequences that these capabilities can introduce into a system. The chapter continues with an overview of human performance modeling and a discussion of a multi-tasking simulation tool used for investigation in support of the research effort. The literature review concludes with a summary of the main concepts that emerged and how they lead to identification of the gaps that will be addressed by the dissertation.

B. MULTI-LEVEL COGNITIVE CYBERNETICS

Systems are often developed with the assumption that humans are adaptive in different contexts (Benyon, 1993). Research efforts have also attempted to develop technology to behave in a similar adaptive manner (Benyon, 1993; Hou et al., 2014; Karwowski et al., 2006). One approach to do so is the multi-level cognitive cybernetics (MLCC) approach seen in Figure 2 (Cassenti & Veksler, 2018). Cybernetics is the study of how humans and machines interact within an environment (Wiener, 1948). Communication between entities is a prevailing theme of cybernetics, as the interactions between a human and machine provide inputs for behavioral modifications in close-looped systems (Cassenti et al., 2017). Wiener's cybernetics conceptualization included a description of how humans adjust their behavior based on environmental feedback. The MLCC approach builds on Wiener's work by suggesting a continuous co-adaptation between humans and technology. As humans proceed through information processing stages and behavioral actions, adaptive automation is calibrated using those actions to determine the best ways to assist an operator. It is also important for adaptive automation to be triggered at appropriate times to not introduce more complexity for the operator. The MLCC approach differs from traditional AA triggering approaches by positing that cognitive variables can serve to trigger automation rather than time or performance metrics (Cassenti et al., 2017). The outputs of the adaptive automation serve as further sensory inputs for the human, resulting in another cycle of the MLCC approach.

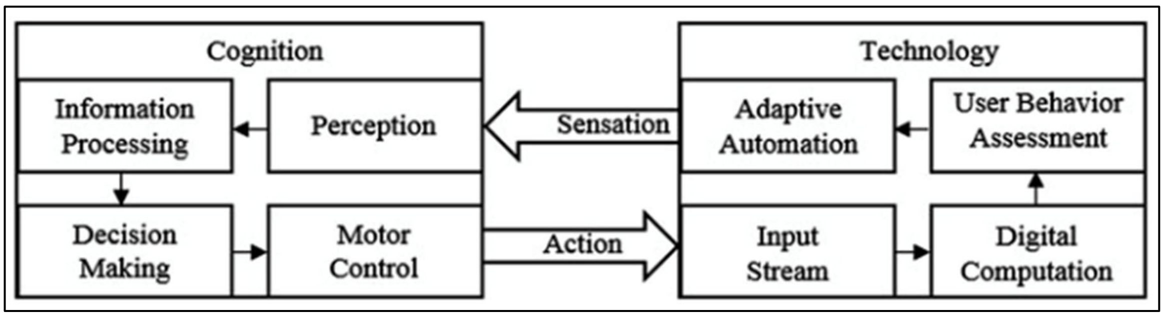


Figure 2. Stages of cognition and adaptive technology in the Multi-Level Cognitive Cybernetics. Source: Cassenti et al. (2017).

An adapted model of Cassenti et al.'s (2017) MLCC approach is used to frame this dissertation and is depicted in Figure 3. The adapted MLCC framework was chosen as a foundational concept in this effort because of its intended outcomes of integrating subjective, objective, and cognitive workload modeling to inform AA design. This model is bounded by the human operator and the technology used within that closed-loop system. System-specific considerations are modeled through the interface interactions that the human has with them. This assumption assists with scoping the model to focus on the human operator. While external considerations and variables interact at the boundaries of the system, they are not of direct interest when examining the relationship between a human operator's cognitive workload and AA.

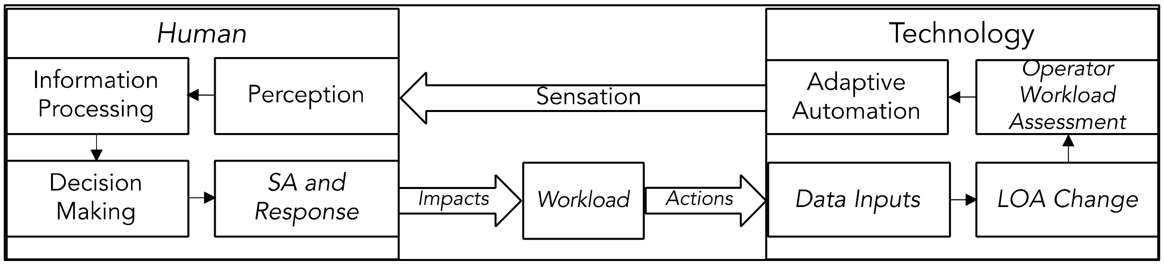


Figure 3. Adapted version of Cassenti et al.'s (2017) MLCC framework with modifications italicized.

Starting with the left-hand side of the diagram, the human is considered more broadly than with specific focus on cognition in the original framework. As the human perceives sensory inputs, they proceed through stages of information processing that lead to decision making. The outputs of their decisions affect situation awareness and responses. The resulting impacts of these responses and level of SA impact cognitive workload. The degree to which cognitive workload is influenced leads to actions by an operator. These actions transition to the technology portion of the cybernetic loop and serve as data inputs that can drive changes in levels of automation. Technology should then conduct an operator workload assessment based on established criteria to determine what the changes in adaptive automation should be. These modifications in adaptive automation provide feedback for the human operator through sensory inputs that allows the system to continue back to the perception of the human operator through another iteration of the cycle.

This dissertation will investigate specific aspects of the adapted MLCC framework. The subsequent sections of this literature review will discuss the human considerations from perception, cognition, cognitive workload, situation awareness, and performance. The literature review will also address technological capabilities and their impacts on the human operator using adaptive automation. The focus on these concepts allows for mapping of experimental objectives to the adapted MLCC framework to model and validate cognitive workload predictions through the dissertation's three experimental studies. Given that the MLCC framework is focused on human's cognition when interacting with technology, it is important to foundationally discuss cognitive workload in the proceeding sections.

C. COGNITIVE WORKLOAD

1. Cognition Introduction

Cognition can be described as what goes on inside the human mind when going about life (Preece et al., 2002). Cognition also can be generally categorized as the ways people think, perceive, remember, and process information. It is what happens after the world is sensed and perceived, and responses are elicited in accordance with Wickens' Model of Human Information Processing (HIP) (Lee et al., 2017). This framework provides a way to investigate the movement of sensory inputs into long-term memory and actions of a human. This flow from inputs to outputs can have significant implications for efficient and effective cognitive design that leads to better performance. Research into externalized cognition posits that cognition can happen outside of the human mind through such tasks as off-loading ideas onto notes and using paper and pencil to solve mathematical problems (Fiore & Wiltshire, 2016).

Humans possess a finite ability to sense, perceive, and process information. Models of human information processing, such as Wickens' HIP seen in Figure 4, theorize that humans' information processing capabilities come with organic filtering mechanisms to attend to the most saliently perceived sensations. Without this capability, humans would be in a constant overload state. Further, differentiating between stimuli that enter a human's information processing system yields necessary discussion to explain why multiple tasks can be performed at the same time.

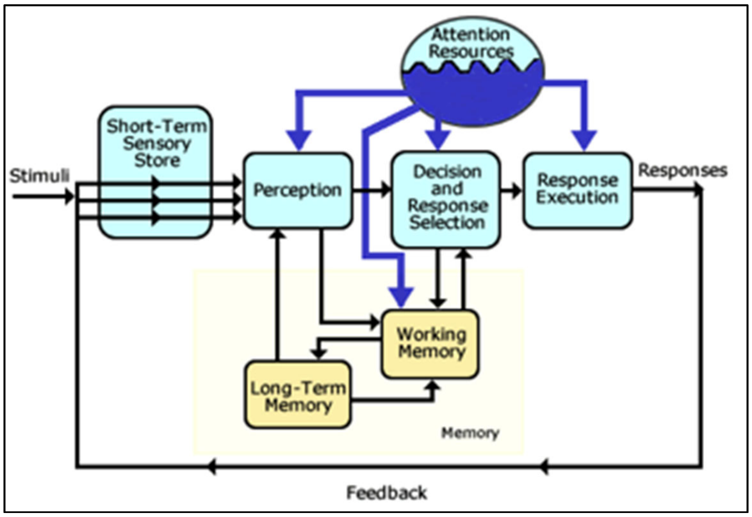


Figure 4. Wickens' Model of Human Information Processing. Source: Lee et al. (2017).

Wickens' (1981, 2002, 2008a) proposed Multiple Resource Theory (MRT) seen in Figure 5 to address how attentional resources can be divided across sensory modalities, enabling multi-tasking. Wickens' MRT further suggests that presenting information through various modalities in multiple displays will facilitate more effective information processing (Wickens et al., 1984). These modalities can include verbal, spatial, visual, and auditory formations. This proposition has implications for various aspects of a system, including design, measuring performance, and predicting cognitive workload. Multiple Resource Theory explains that tasks can be aligned such that multi-tasking can occur to allow for better performance if processing channels are separate and distinct (e.g., a visual task and a separate auditory task). Further, the tasks should not converge into a single mental model. Otherwise, using the same resources for the task may be required and prevent from separate and effective processing. This can work conversely, however, if a task uses two channels to process the information (Wickens et al., 1984). If a task uses distinct value decisions about a display, then optimal performance will be realized when spatial and verbal analogs are used independently. (Wickens et al., 1984).

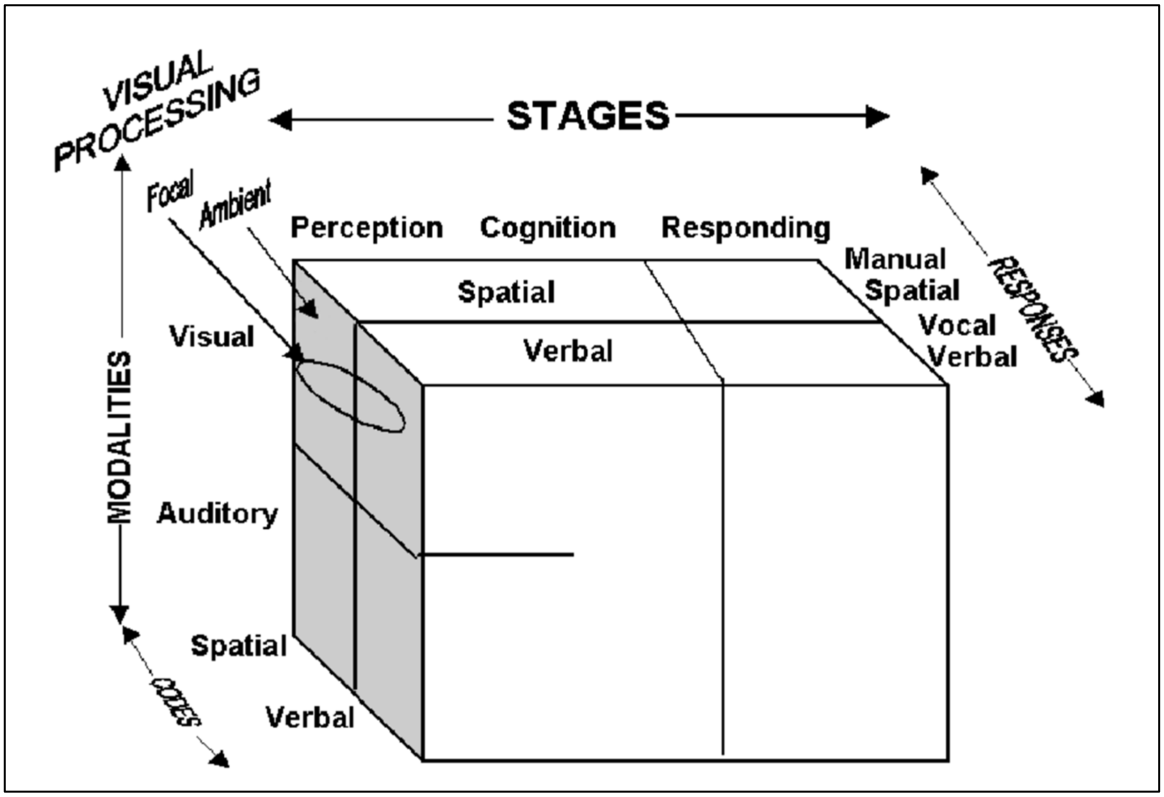


Figure 5. Wickens' Multiple Resource Theory model. Source: Wickens (2008a).

Much like other environments that humans encounter, the cognitive environment is an important domain for designers of any system to consider. Designers should address how quickly the cognitive environment changes, the level of familiarity with the environment, and the extent to which features in the environment provide context and cues for interaction within it (Lee et al., 2017). These considerations can help provide a foundational framework to investigate demands and resources available to address them.

2. Cognitive Workload

Unlike psychophysiological attributes such as blood pressure or body temperature, cognitive workload is not directly measurable. However, it can be characterized by the resource demand that tasks require and the available cognitive resources available to address them. (Embrey et al., 2006; Matthews et al., 2014). In this sense, cognitive workload measurement attempts to use surrogate measures through a measure or combination of measures to infer levels of cognitive workload. This approach to cognitive workload measurement mirrors similar attempts to use correlates for things that we cannot measure directly, For example, heart rate can provide insight into cardiovascular disease (Perret-Guillaume et al., 2009). The approach to determine correlations between surrogate measures and phenomenon is not new. Further, an essential item to consider when examining surrogate measures is finding those metrics that best correlate to things that are not directly measurable, including cognitive workload.

Cognitive workload can be considered as being either demanded in single or multiple levels (Moray, 2013). There are generally two regions of task demand level. The first region deals with where the task demand is less than resources available. This is an ideal state that leaves a buffer for a user to operate, should demand resources be needed. The second region is where task demand exceeds resources available. This is not ideal as performance will break down when this state persists (Wickens, 2008). It is important to note that workload is a uniquely personal experience that reveals the relationship between cognitive resources available and the cognitive work demands presented (Vogl et al., 2020).

Cognitive workload has been defined in many ways. Often, these definitions of cognitive workload rely on a singular consideration such as subjective or objective metrics. These characterizations are rarely considered together (Longo et al., 2022). Van Acker, Parmentier, Vlerick, and Saldien (2018, p. 9) provide this conceptual definition: "Mental workload is a subjectively experienced psychophysiological processing state, revealing the interplay between one's limited and multidimensional cognitive resources and the cognitive work demands being exposed to." Within the aviation domain, Ellis and Roscoe defined workload as "the integrated physical and mental effort generated by the perceived demands of a specified piloting task" (1982, p. 11). These definitions highlight the complexity of cognitive workload across multiple considerations.

There have been recent attempts to provide a comprehensive definition of cognitive workload. Longo, Wickens, Hancock, and Hancock (2022, p. 18) propose that cognitive workload: "represents the degree of activation of a finite pool of resources, limited in capacity, while cognitively processing a primary task over time, mediated by external stochastic environmental and situational factors, as well as affected by definite internal characteristics of a human operator, for coping with static task demands, by devoted effort and attention." This definition was developed based on an extensive analysis of cognitive workload literature. Longo et al.'s definition addresses basic concepts such as resources demanded versus resources available while attempting to account for internal and external factors that a person may experience in completing a task. Longo et al.'s definition considers multiple aspects that influence cognitive workload. Their description of cognitive workload addresses cognitive resources and their limited capacity, and how those resources can be activated for completing a primary task. Additionally, Longo et al. account for individual differences in personnel traits and strategies. Their definition also acknowledges that environmental factors can influence task completion. Finally, their definition highlights the role that an individual's effort and attention have in accomplishing a task. This definition serves as a guiding principle for this dissertation effort because of its multidimensional considerations in describing the complex nature of cognitive workload.

Cognitive workload became a major design consideration in the aviation community as automation and flight support tools were used to replace flight engineers (Parasuraman et al., 2008). Workload was a key factor in the Federal Aviation Administration's (FAA) decision to certify reduced manned aircraft. Measures analogous to the Cooper-Harper rating scale for quantifiable handling were used in these studies (Parasuraman et al., 2008). However, just because two flight crew members are experiencing high levels of cognitive workload does not mean both will perform at the same level (Guastello et al., 2015). For instance, an expert pilot may be demonstrating the same behavioral and performance outputs as a novice, but the expert pilot may have significantly more workload resources available for task allocation than the novice. This difference is an important consideration for both training and system design. Differences in aviators' performance levels suggest that the cockpit should be designed to account for the differences between expert performers and novices. Novices rely on context-free rules and perceive situations superficially, whereas experts rely on intuition gained over time to react to situations more efficiently due to gaining more focused experience (Dreyfus, 2004; Hutton & Klein, 1999). Technology should provide commensurate affordances and constraints to support an array of human operators who have different capabilities and limitations (Shattuck, 2017).

Cognitive workload has generally been measured in three ways: subjectively through self-assessment, employing task performance metrics, and through psychophysiological signals (Cain, 2007; Eggemeier et al., 1991). Not one of these methods has been shown to measure cognitive workload directly. Each approach is sensitive to different aspects of workload, and not all these approaches measure the same thing (Hancock & Matthews, 2019). Additional components to workload include task difficulty, task context, and the task performer's proficiency level (Young et al., 2015). Therefore, it is important to approach cognitive workload measurement from multiple dimensions that include each of the aforementioned areas. This dissertation will follow this multi-dimension approach using certain objective, subjective, and performance measures to find relationships between them and cognitive workload. While there are many measures that have been identified as sensitive to changes in workload, it is beyond the scope of this effort to investigate them all.

3. Subjective Cognitive Workload Measurement

Cognitive workload demanded by a task can impact performance (Brand & Schulte, 2017; de Greef & Arciszewski, 2007; Hart, 2006; Inagaki, 2003; Kaber et al., 2001; Kaber & Endsley, 2004; Kanaan & Moacdieh, 2021; P. Smith & Baumann, 2020; Vagia et al., 2016). Given this relationship, measuring workload is an essential step to bring about improved performance. Because cognitive workload is experienced subjectively by individuals, it can be described and assessed through introspection subjectively with reliability (Cain, 2007). Generally, a person will be able to identify when they are experiencing higher levels of cognitive workload or stress and be able to articulate that through self-reporting. However, the use of subjective self-report measures by an operator may be unreliable. Some external factors such as unannounced strategy, effort changes, anxiety, or emotional intelligence may influence subjective workload ratings (Cain, 2007; Guastello et al., 2015). Therefore, the use of multiple subjective scales in conjunction with objective data provide more reliability in determining levels of cognitive workload (Cain, 2007; Lohani et al., 2019; O'Donnell & Eggemeier, 1986; Vogl et al., 2020).

There are several other subjective methods to assess cognitive workload, including crew status survey (Ames & George, 1993), Malvern Capacity Estimate (MACE) (Goillau & Kelly, 2017), Modified Cooper-Harper (MCH) (Cooper & Harper, 1969), Bedford Workload Rating Scale (BWRS) (Roscoe, 1984), Subjective Workload Assessment Technique (SWAT) (Potter & Bressler, 1989), and Workload Profile (WP) (Tsang & Velazquez, 1996). These methods represent approaches to capturing workload either during or after a task. Other methods that have been used for cognitive workload measurement in adaptive interface systems include using language complexity analysis techniques. Some of these linguistic measures include Lexical Density, Complex Word Ratio, and the Gunning Fog Index (Khawaja, Chen, & Marcus 2010; Khawaja, Chen, Owen, & Hickey, 2009) (Khawaja et al., 2009, 2010).

Another way to subjectively measure cognitive workload is through the NASA Task Load Index (NASA-TLX). The NASA-TLX is a multidimensional scale examining the six dimensions of workload seen in Figure 6: mental workload, physical workload, temporal workload, performance, effort, and frustration (Hart, 2006). Individuals complete the NASA-TLX by rating each of these dimensions from low to high using an analog scale. Each scale consists of 21 marks that individuals use to rate their experienced level of demand. The NASA-TLX is a widely used workload assessment tool in the field of human factors and is typically administered after the completion of a task.

Title	Endpoints	Descriptions
MENTAL DEMAND	Low /High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low /High	How much physical activity was required (e.g.. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous. restful or laborious?
TEMPORAL DEMAND	Low/ High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	good/poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVELi	Low /High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Figure 6. NASA-TLX rating scale definitions. Source: Hart and Staveland (1988).

Another tool available for cognitive workload measurement is the Continuous Subjective Workload Assessment Graph (CSWAG) (Miller & Shattuck, 2004). Research using the CSWAG seeks to have participants report their workload as a percentage of their maximum cognitive workload. Participants are asked to rate their cognitive workload percentage at given time intervals. The intervals used in a study must balance the tradeoff between asking a participant their workload too often versus not enough. Asking a participant every five seconds during a 10-minute task may be too distracting and may introduce an increased source of cognitive workload. Conversely, asking a participant every five minutes during a 10-minute task may not capture an accurate representation of their experienced cognitive workload during that period. Previous studies have asked participants for their CSWAG percentage every minute during a task (Brown et al., 2021). Similar approaches have been used with the SWAT at 30 second intervals and with the Instantaneous Self-Assessment (ISA) of workload at two minute intervals (Brennan & Jordan, 1992; Zak et al., 2020). The CSWAG cognitive workload percentages can then be generally classified in three bins: lowest workload (0%-33%), just about right (34%-66%), and highest workload (67%-100%) (Miller & Shattuck, 2004). A researcher instructs a participant that they will be asked for their workload as a percentage of their maximum workload. The researcher informs the participant that they will be asked for their assessment when they hear "workload." The use of brevity in asking for the participants' workload and training them on the CSWAG allows for operator workload assessment during task completion. This approach also helps mitigate disruptions to the operator's primary task to a negligible level. The CSWAG approach is complementary to post-trial assessments.

The NASA-TLX and CSWAG were selected for the current research effort due to the ease in the ability to administer subjective workload assessments during and after the trials while minimizing disruptions to the primary task. Additionally, the use of more continuous inquiries into an assessed state at set intervals have shown sensitivity to changes in perceived cognitive workload (Brennan & Jordan, 1992; Brown et al., 2021; Zak et al., 2020). While both metrics have been used previously, they also have issues as with all subjective workload measures. The CSWAG provides real-time workload measurement but may interfere with task completion due to the interruption of being asked to assess one's workload at that moment. The NASA-TLX relies on a participant to reflect on their subjective experience, which may not be representative of the whole task since memories may be forgotten. Further, a recency effect may influence a participant's NASA-TLX ratings. Participants may remember their most recent experience and base their responses based on that time interval instead of the whole period (Guastello et al., 2015). Additionally, participants may try to average their total experience if there are multiple conditions. This averaging approach may lead to biased values in their NASA-TLX ratings. However, when used together and in complementary fashion with the other measures of workload, these measures can aid in the correlational analysis of cognitive workload.

4. Objective Cognitive Workload Measurement

Historically, unobtrusive measures of cognitive workload relied on somewhat unreliable subjective techniques, such as task omission rates as well as primary and secondary task performance (O'Donnell & Eggemeier, 1986). While cognitive workload can be measured via objective performance and subjective assessments, these metrics are not sensitive enough to determine differences between normal and overloaded workload conditions (O'Donnell & Eggemeier, 1986). Performance metrics also do not provide for identifying which resources are being used in accordance with MRT. However, psychophysiological measurement provides a robust and continuous means to gauge workload without any behavioral outputs needed (Hughes et al., 2019). Aricò et al. (2016) propose that there is value in using psychophysiological measurement in conjunction with subjective measures when investigating the relationship between humans and adaptive automation. Multiple ways exist to measure workload objectively that provide more fidelity on cognitive workload. Additionally, a host of psychophysiological measures are available which seek to objectively assess workload in real-time. Oculometrics (including pupillometry and blink data), electrocardiography (ECG) and cardiovascular measures, respiration, catecholamines and hormonal responses, electrodermal activity (EDA), electroencephalography (EEG), and functional near-infrared spectroscopy (fNIRS) are a few examples of these methods (Hancock et al., 2013; Neubauer et al., 2020; Vogl et al., 2020).

a. Eye Tracking

The human eye can provide insight into brain activity (Janisse, 1977; Lohani et al., 2019). Eye tracking has been used in numerous studies to gain insight into the human state and cognitive activity (Matthews et al., 2014; Palma Fraga et al., 2020; Pfleging et al., 2016). Eye tracking has evolved from an unreliable and intrusive technique to its current state that allows for relatively simple and unobtrusive oculometric data collection (Vogl et al., 2020). Further, modern eye tracking solutions even present the possibility for inclusion in applied operational settings instead of solely in a controlled laboratory environment (Aura et al., 2021).

Eye tracking provides insight into where humans are focusing attention by analyzing eye movements and pupil dilation, which is coupled with autonomic nervous system (ANS) activity (Neubauer et al., 2020). The correlation of pupil size to the ANS provides an avenue for validation and study of the relationship between other psychophysiological measures. The pupil's constrictions and dilations allow for the management of light entering the retina. The more light that is presented to the eye yields smaller pupil diameter, while less light yields larger pupil diameter (Aura et al., 2021). Ellis (1981) refers to this phenomenon as pupillary light reflex. Thus, pupil size is constantly changing due to constriction and dilation brought about due to environmental changes, giving humans the ability to see in rapidly changing conditions. In general, the size of human pupils vary between 2 and 8 mm (Watson & Yellott, 2012).

When luminance conditions are controlled or held constant in a laboratory setting, pupil size has been shown to provide significant insights into psychosensory effects, including cognitive workload (Neubauer et al., 2020; Kahneman & Beatty, 1966). As a human operator's experienced cognitive workload increases, the pupil's diameter will also grow (Aura et al., 2021; Beatty & Lucero-Wagoner, 2000; Pfleging et al., 2016). The dilation associated with increasing cognitive workload is also referred to as task-evoked pupillary response (Pfleging et al., 2016). Task-evoked pupil diameters should be evaluated against a baseline period to determine the differences in pupil size following initiation of the task (Krejtz et al., 2018). Individual differences, such as age, can account for differences in pupillary response as well as cognitive activity and environmental conditions. The use of a within-subjects experimental design can mitigate and account for the impact of individual differences since participants complete all conditions (Greenwald, 1976).

Previous work has investigated ways to measure cognitive activity as a function of pupil size (Duchowski et al., 2018; Marshall, 2002). Studies have also examined pupil size in simulated and live operating room settings to determine effects of workload on pupil size for novice and expert surgeons. Findings suggested that experts' pupil sizes were significantly smaller than those of novice surgeons in both simulated and live operating rooms (Richstone et al., 2010). These findings indicate that pupil measurement can be used to assess cognitive activity and workload. Further, the findings suggest that there may be differences occurring in neural pathways that explain different pupil sizes between right and left eyes.

Two additional eye metrics that are related to cognitive workload and fatigue are blink rate and duration (Benedetto et al., 2011). As workload increases, blink rate and duration decrease and are indicative of focused attention on a task (Ahlstrom & Friedman-Berg, 2006). Eye tracking has been demonstrated to be an effective method for understanding the effects of interruptions on how humans apply their cognitive resources in varying conditions, as measured by workload (Kanaan & Moacdieh, 2021). Higher workload manifests in slower scan patterns on displays than in conditions with lower workload (Kanaan & Moacdieh, 2021). This phenomenon of focused attention during periods of higher workload is the result of humans leveraging the limited resources available on accomplishing the most salient task (Dehais et al., 2011; Tao et al., 2019; Wickens et al., 2015).

Previous research has found that visual physiology assessment is a viable candidate to assess cognitive workload within fixed levels of automation (Evans & Fendley, 2017). Evans and Fendley (2017) conducted an experiment using an open-source real-time strategy game that involved participants interacting with the game using Sheridan and Verplank's LOAs 2, 4, and 9 listed in Table 1. Their results showed significant differences in NASA-TLX subjective workload ratings, time to complete a run of the game, and differences in visual fixation rates among each of the levels. These findings suggest that visual physiology may be valuable in evaluating cognitive workload within static automation levels. However, investigation into cognitive workload when using dynamic autonomy levels is limited and is perhaps a valuable line of research to pursue given the findings from static levels of automation and the potential to increase performance (Evans & Fendley, 2017; O'Neill et al., 2020). A multimodal approach that couples neuroimaging with eye movements and observable behavior may be beneficial to the present effort for analyzing cognitive workload increases brought on by task difficulty and complexity (Palma Fraga et al., 2020).

Table 1. Sheridan and Verplank's levels of automation. Sources: Beer et al. (2014); Parasuraman et al. (2000); Sheridan and Verplank (1978).

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

While pupil diameter and blink rates have shown to be related to changes in experienced cognitive workload, challenges exist in their measurement in applied settings. Many studies have used controlled, static means to evaluate oculometrics in a laboratory setting with great success. However, environmental conditions and technology capabilities make oculometrics in applied settings more challenging. Solutions are improving as iterations of field-worthy systems are being developed, which will allow for more real-time assessment of oculometrics in applied settings. Additionally, using oculometric data along with multiple cognitive workload metrics will help provide a more complete analysis of the workload levels over time. Pupil diameter has shown to reliably discern differences between workload conditions in different luminance environments (Aura et al., 2021). Because of its sensitivities to varying workload and environmental conditions, pupil diameter will be leveraged as one potential workload correlate in the present effort.

b. Functional Near-Infrared Spectroscopy (fNIRS)

Parasuraman and Rizzo (2008, p. 239) define neuroergonomics as "the study of the brain and behavior at work." The aim of neuroergonomics is to help understand how humans interact with the environment through leveraging neuroscience. One method to assist with this understanding is functional near-infrared spectroscopy. Functional near-infrared spectroscopy is a wearable brain-based technology that can measure brain activity in a non-invasive and transportable manner (Li et al., 2022; Reddy et al., 2022).

Changes in blood oxygenation levels are associated with cognitive workload. For instance, the higher the change in oxygenation, the higher the associated cognitive workload (Ayaz et al., 2012; Causse et al., 2017; Herff et al., 2014). The use of fNIRS technology provides a method for measuring oxygenation levels in the brain (von Lühmann et al., 2015). To do so, fNIRS systems use a signal and detection methodology for examining infrared (IR) light at specific regions of the brain. The fNIRS system will send an IR light that is then absorbed by different components of the body. Detectors in the fNIRS system will then measure the resulting IR light (von Lühmann et al., 2015). As such, fNIRS systems examine the manner in which light passing through cortical tissues relate to changes in oxygenated (HbO) and deoxygenated hemoglobin (Hb) concentration levels (Reddy et al., 2022; Scerbo, 2008). Much like functional magnetic resonance imaging (fMRI) technology, fNIRS provides a view of how oxygen supply in the blood changes for different regions of the brain. Unlike fMRI, however, fNIRS systems can be used in an operational setting due to lower resource and space demands. Research efforts using fNIRS to examine cognitive workload have grown in recent years due to fNIRS systems becoming more reliable, affordable, and portable (Herff et al., 2014).

Research using fNIRS has demonstrated the potential to continuously and objectively measure cognitive workload by leveraging blood oxygenation levels as a surrogate to cognitive workload (Harrison et al., 2014). Cognitive workload assessment using fNIRS has been successful when measuring hemodynamic activity in the prefrontal cortex (Herff et al., 2014). Because of its ability to continuously measure activity without reliance on any overt human behavior, a brain-based system like fNIRS can potentially benefit AA systems (Scerbo, 2008). Operator state monitoring approaches using fNIRS can stand to benefit by providing another low-intrusive sensor to provide real-time cognitive workload metrics. While fNIRS has been shown to provide insights into cognitive workload, challenges exist with its use in applied settings due to ambient light and operator movement artifacts, creating noise in collected data (Girouard et al., 2010). These confounding variables can be eliminated in large part with mitigating techniques during data collection such as light blocking caps or with post hoc data process filtering (von Lühmann et al., 2015).

c. *Heart Rate Variability*

Cardiac activity has shown to be sensitive to variations of cognitive workload, particularly task workload demands and duration, as well as event presentation rate (Hughes et al., 2019). Heart Rate Variability (HRV) provides a more sensitive measure of cognitive workload than heart rate alone and has been used in numerous studies as a cognitive workload metric (Backs, 1995; Hughes et al., 2019; Matthews et al., 2014). Heart rate variability examines the changes in inter-beat intervals (IBI), commonly referred to as the R-R interval shown in Figure 7, over some period. The variation in timing between heartbeats provided by HRV data is caused by sympathetic and parasympathetic nervous system activity. Heart rate variability is also a cognitive workload metric that can be derived from ECG data. It has been used to measure cognitive workload in multiple domains, including aviation (Backs, 1995; Vogl et al., 2020). For these reasons, HRV is of primary interest in this study to aid in determining cognitive workload changes.

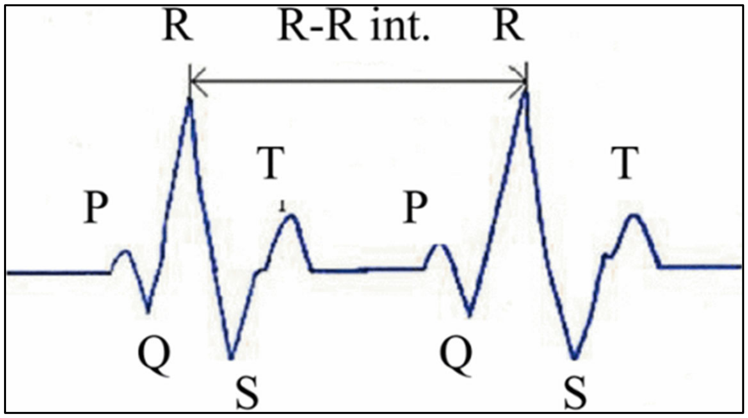


Figure 7. A heartbeat outline showing an R-R interval. Source: (Murai et al., 2015).

Inter-beat intervals measure the time between contractions of the heart muscle. This is an important measurement when calculating an HRV metric. There are numerous peak detection algorithms available that assist with detecting R-wave spikes produced by ECG signals. The R-wave spike detection allows for determining R-R IBIs for a given time. One approach to analyzing the R-R IBIs leverages time domain analysis. Another common time domain analysis is the Root Mean Square of Successive Differences (RMSSD). The RMSSD is calculated by taking the square root of the mean of the squared differences between successive R-R intervals for a given period (Tao et al., 2019). As cognitive workload increases, the IBI and RMSSD decrease highlighting the inverse relationship between cognitive workload and HRV (Fairclough et al., 2005; Tao et al., 2019). Mean R-R intervals have shown to be sensitive to the inverse relationship that accompanies changes in cognitive workload (Delliaux et al., 2019).

The ability to leverage ECG signals to determine HRV while minimizing intrusiveness provides promising real-time assessment capability in future operational environments. As many commonly worn wearable technologies provide these data, low costs and operator familiarity with the systems may also break down barriers to acceptance in more applied settings. Additionally, the relatively low overhead to collect HRV may help achieve real-time operator state monitoring needed to properly employ AA systems.

5. Cognitive Workload and Performance

A common depiction of the relationship between workload and performance can be seen in the inverted "U" curve presented in the Yerkes-Dodson Law (Yerkes & Dodson, 1908). An adaptation of this law is seen in Figure 8, which depicts the relationship of workload and resources available with task performance. If workload is too low, increased workload can increase performance. However, too much workload can then lead to a decrease in performance. When a person is under-aroused and bored, performance can decrease, although a person may have the resources available to attend to a situation.

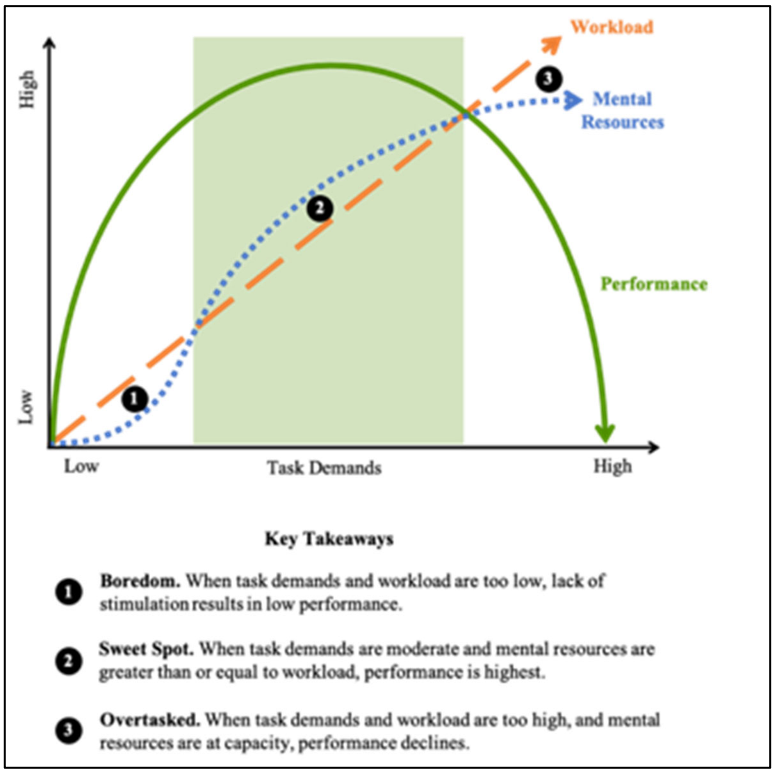


Figure 8. Relationship of performance, mental resources, and workload. Source: Ernst et al. (2020).

Operators may drift or lose interest in the task, and thus suffer a decrement in performance (Parasuraman, 1987). Time increases workload, as temporal demands trigger an external pressure to complete the task. Errors in task performance also become more prevalent as workload approaches the 80th percentile capacity of the individual. This concept is known as the workload redline (Young et al., 2015). Designers should consider the tasks required to complete an operation and the demands that those tasks place on the humans in the system. In other words, it is important to account for resources demanded versus the resources available to complete a task when assessing cognitive workload (McKendrick et al., 2019; Welford, 1978; Young et al., 2015).

D. SITUATION AWARENESS

Situation awareness (SA) can be conceptualized as understanding the state of an environment, or more simply as knowing what is happening (Endsley, 2021; Wickens, 2008b). Situation awareness can be formally defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 2021, p. 435). There are also internal and external factors that interact with SA to help provide a more developed description of it as seen in Figure 9. As the state of the environment changes, achieving Level 1 SA involves a human perceiving relevant elements through one or multiple sensory mechanisms. Level 2 SA describes comprehension of those elements perceived in Level 1 as well as an analysis of those elements in relation to the operator's goals. This concept can be likened to comprehending words rather than just reading them. Level 3 SA builds on the previous two levels and occurs when an operator can make predictions on future states of the environment.

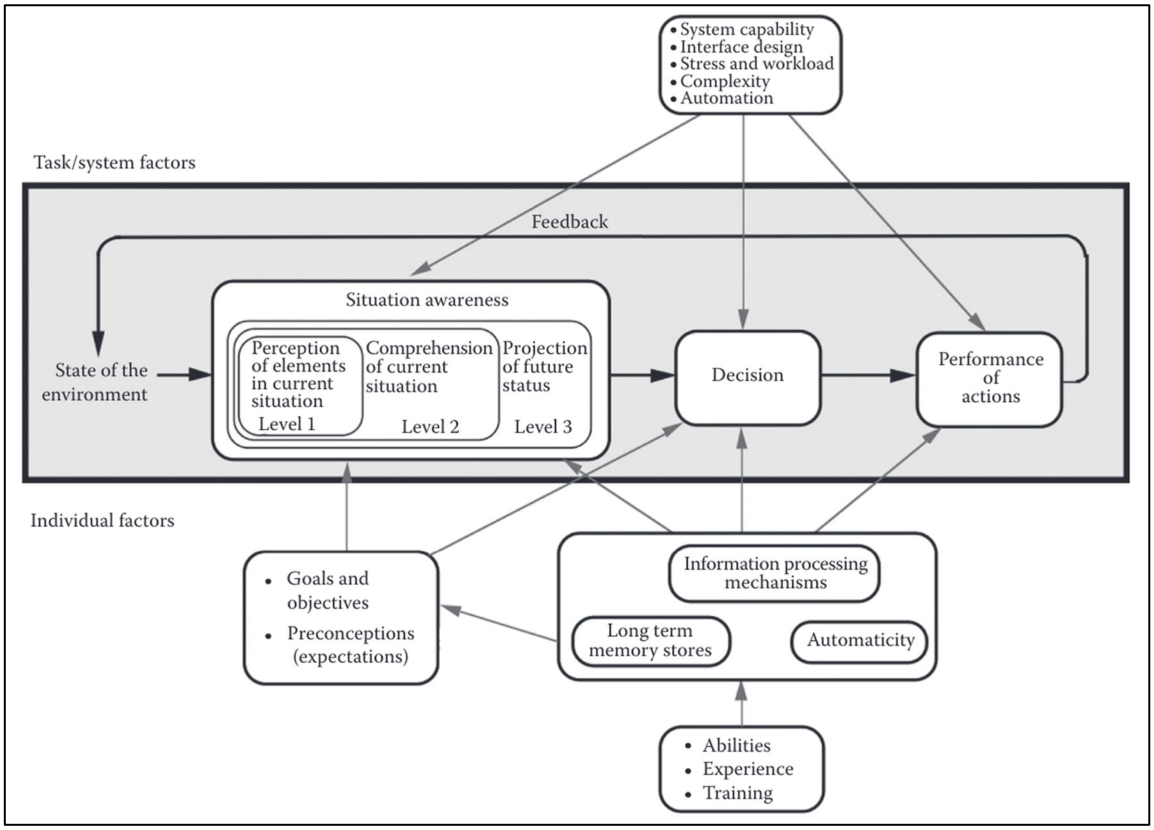


Figure 9. Model of SA in dynamic decision-making. Source: Endsley and Jones (2011).

An operator's decision and subsequent actions are made based on perception, comprehension, and prediction that feed back into the SA cycle. While SA is susceptible to the limitations of human memory processes, there are individual characteristics that can help humans achieve higher levels of SA. For instance, operators can leverage goal-directed processing, automaticity in actions, and long-term memory stores to help match patterns and schemas to currently experienced phenomena. These enhancing mechanisms are developed through the abilities acquired from training and experience.

External factors in Figure 9 include a system's capabilities, interface design, workload, and automation. These factors can afford or constrain the progression from situation awareness to action based on the number of resources the external factors demand. For example, a system that is highly automated and that has an intuitive user-interface design may allow the sub-conscious progression through the Levels of SA without interference. As a result, operators can make informed decisions and achieve higher levels of performance. Conversely, high complexity and increased workload can take an operator away from projecting the future consequences of their actions and lead to an incomplete understanding of a situation. A final key factor of SA is the inclusion of a feedback loop that helps inform the changed state of the environment as an operator progresses through a situation.

Situation Awareness has been assessed in different ways in aviation environments (Nguyen et al., 2019). Some of the different measurement techniques for assessing SA in aviation have included the Situation Awareness Global Assessment Tool (SAGAT) (Endsley, 2017b), Situation Present Assessment Method (SPAM) (Durso et al., 1999), and Situation Awareness Rating Technique (SART) (Taylor, 2011). Each measurement technique possesses advantages and disadvantages.

SAGAT is a widely used technique that will seek to assess an operator's SA at a specified time during a task (e.g., pausing a flight simulation mid-flight). SAGAT can then provide real-time assessment of SA and mitigate issues such as forgetfulness that manifest in post-trial assessments. However, this approach can be problematic in collecting data on the primary variables of interest by disrupting the task (Sarter & Woods, 1995). SAGAT also allows for potential retrieval from LTM stores to provide a picture of SA retroactively rather than in real-time. This phenomenon is problematic given that SAGAT questions are dependent on memory information to assess SA.

SPAM addresses the reliance on memory that SAGAT presents by assessing the latency in which operators report their SA. For instance, SPAM is not concerned with exact data being given, but rather on an operator's ability to find it in a timely manner to report it. This approach is thus an easy and fast approach to SA measurement that is not reliant on freezing a task. However, SPAM can still disrupt task performance as it is conducted during a task. There is also more overhead in determining the questions to be asked of the participants to address the SA requirements in question (Durso et al., 1999).

SART relies on subjective assessment and is thus susceptible to operators forgetting or reporting on only perceived SA during what they deemed as most important or when they performed best. However, SART provides a fast and efficient way to evaluate SA in a less intrusive way. It has been used in a variety of domains and provides quantitative data for analysis (Taylor, 2011).

One key feature of situation awareness is that it can be diagnostic of operator states. For example, operator accuracy can be continuously diagnosed and assessed against ground truth. Therefore, it can help guide human factors solutions in situations where operator SA is lacking (Parasuraman et al., 2008). Situation awareness represents a continuous diagnostic state of a dynamically changing environment. Because of this, there is a ground truth that can (and should) be used to assess levels of SA. Situation awareness is not performance, but is rather a psychological construct (Parasuraman et al., 2008). Further, SA is not just general knowledge retrieved from long-term memory (LTM). Situation awareness applies to events that are ongoing and unfolding in (near-) real time, as opposed to events that occurred in the past. While long-term memories help inform people's SA of the environment, LTM and SA are distinct concepts.

Systems should be designed in a manner that supports human SA during operations. Humans in the loop have control and direct interaction with a system (Cummings & Thornburg, 2011). Tasks with HITL allow for greater situation awareness, given that the human is constantly engaged. Humans on the loop brings the human from direct interaction to more of a supervisory role (Stilgoe, 2018). Humans out of the loop (OOTL) takes the human away from a task and leaves the completion of the task to automation. This separation can result in a loss of SA (if the systems is not transparent) and can lead to a degradation of manual skill over time (Endsley & Kiris, 1995). Additionally, systems should be designed to an appropriate level of automation that reduce negative effects on operator SA. Implementing SA support tools while giving a human a high level of control allows for the human to remain appropriately engaged with the task being accomplished. It may be worth including periodic intervention by humans to help them maintain a proper SA picture (Endsley & Kiris, 1995).

Cognitive workload and SA have similar characteristics, and both can impact an operator's performance. They are constructs that are different from behavior and performance (Parasuraman et al., 2008). Workload and SA are also impacted by adaptive automation through various relationships. The relationship between SA and cognitive workload is dynamic and at times they may even be inversely linked in an operational system. For instance, adaptive automation may help increase SA by decreasing workload. However, if workload is reduced by taking the humans out of the loop, SA may decrease even though workload is lowered (Endsley, 2021; Kaber & Endsley, 2004). A depiction of the relationship between cognitive workload and SA as they relate to levels of automation is presented in Figure 10. The x-axis represents increasing levels of automation. The y-axis represents positive system attributes that would be present for cognitive workload and SA. If cognitive workload and SA are linear as seen in Figure 10a and the system performance of the human and the automation are equal, then an optimal level of SA can be determined based on the weight that designers give to each construct. Figure 10b shows the relationship of one or both functions if they are exponentially increasing. The optimal level of SA will then exist at the high or low LOAs. Similarly, if the relationship between SA and workload follows a logarithmic function, then the optimal LOA towards the center of the LOA scale. This optimal LOA will be achieved in logarithmic functions where increases or decreases have reached an asymptotic level as shown in Figure 10c. Figure 10 shows workload reductions increasing as LOAs increase, with SA degrading as LOAs increase. These relationships highlight key design considerations as the interaction of these concepts can have significant impact on system performance. These three concepts are important aspects of this dissertation effort and are each evaluated due to their inextricable relationships.

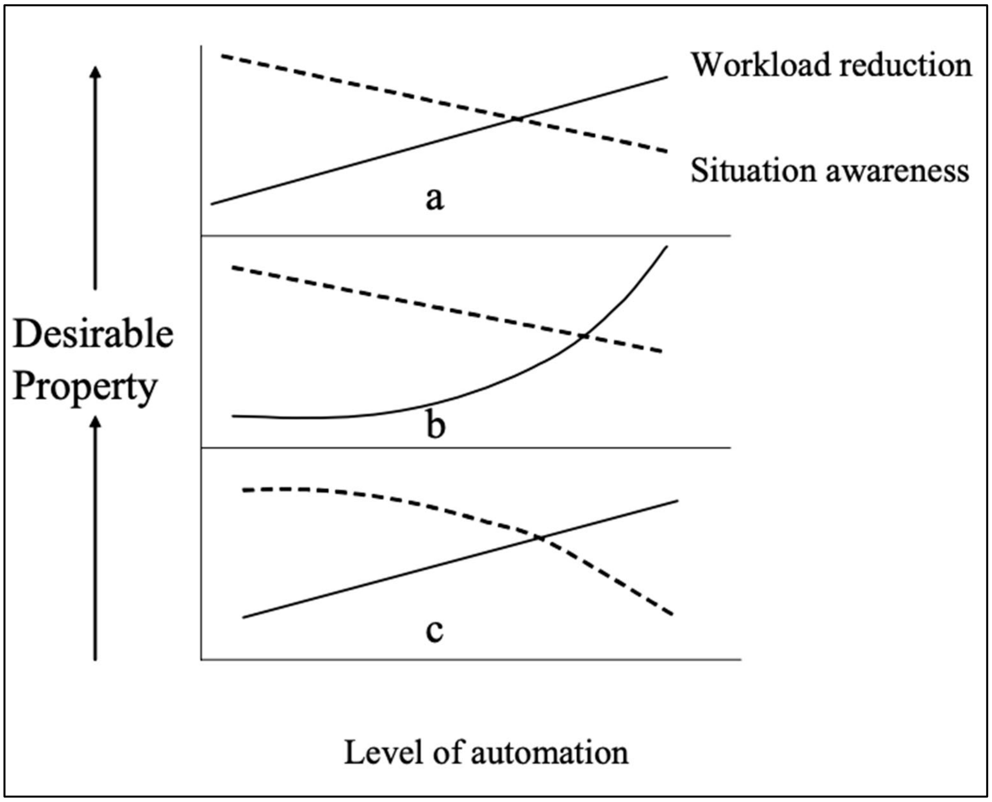


Figure 10. Hypothetical functions of workload and SA as LOAs are increased. Source: Wickens (2008b).

While Endsley's definition of SA has been discussed as a guiding concept in this research effort (and for a myriad of other lines of research for over two decades), criticisms of SA have abounded. Because of the contested nature of SA, there is no single universally accepted definition of it (Stanton et al., 2017). Criticisms of SA have ranged from disagreements on the ability and methods to measure SA (as described in the advantages and disadvantages of the SAGAT, SPAM, and SART methods) as well as the validity of SA as a construct overall. Opponents to Endsley's SA construct suggest that SA is redundant in the face of existing concepts like attention (Dekker & Hollnagel, 2004; Wickens, 2008b). This kind of simplification of SA lacks the contextual application that Endsley's SA model can provide. Situation awareness can assist in identifying where in a complex environment a person's knowledge of specific salient factors should be focused. Further, being able to operate in Levels 2 and 3 of SA allow for more exploration of the environment to find Level 1 items. This process can lead to continual SA development which distinguishes Endsley's model from attention conceptually. This process seen in Figure 11 highlights SA as an iterative process in focusing attention (Endsley, 2015).

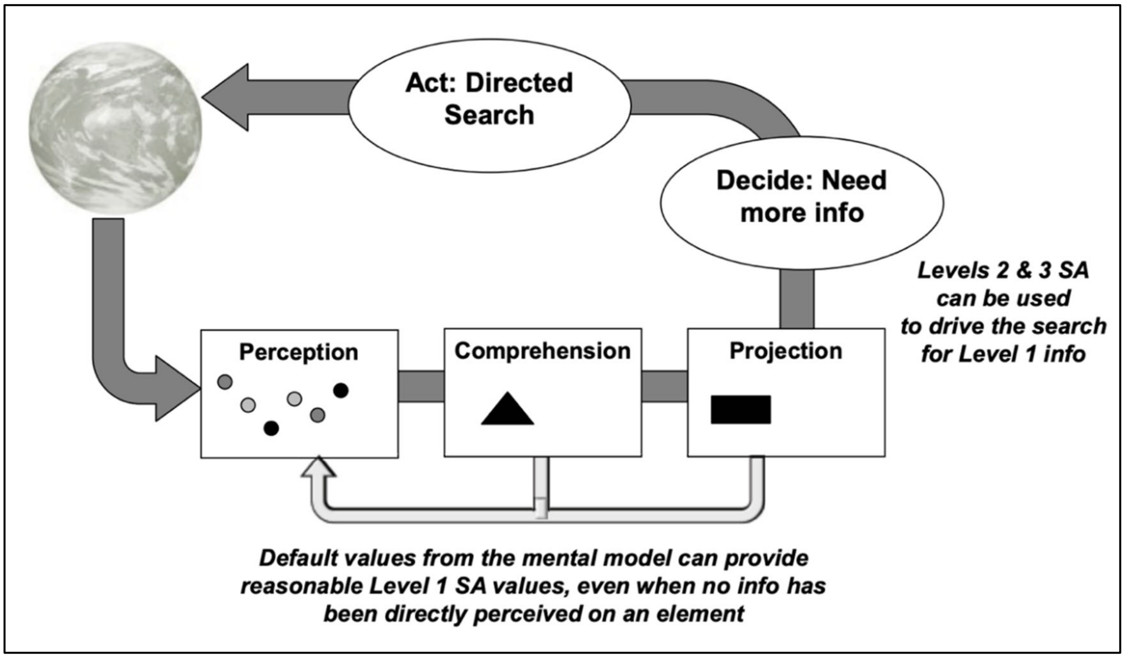


Figure 11. Leveraging higher levels of SA to find new information. Source: Endsley (2015).

Smith and Hancock have defined SA as "adaptive, externally directed consciousness" (1995, p. 138). Smith and Hancock's approach to SA attempts to extend an existing concept of consciousness rather than to categorize it as either knowledge or a process. In doing so, they propose that SA is at the interface between a human and their interaction with the environment, and not just an investigation into a human's performance. As such, they liken SA to other constructs such as cognitive workload as aspects of consciousness. One of the weaknesses that Smith and Hancock identify with Endsley's definition of SA is that the dynamic nature of interactions between the human and the environment are not explicitly identified. However, Endsley's model in Figure 11 provides an important feedback loop that accounts for this dynamically changing interaction with the environment, allowing for on-going SA updates in many dynamic environments (Endsley, 2015).

Others agree with using SA as a description of observed phenomenon. However, they caution against the use of SA as being a causal agent that can lead to circular reasoning when determining root causes of incidents (Flach, 1995). For instance, accident investigations may cite the loss of SA as the leading cause of an operator's fatal error. Critics of SA might ask how one can determine if SA was lost and why did the operator respond inappropriately. In both instances, the answer can be that the operator lost SA, causing an unclear determination into the actual cause of the incident. Further, Flach argues that using SA to describe the phenomenon to further understand its role in causality can be beneficial in analyzing performance in human-machine systems. Using Endsley's conceptualization mitigates Flach's causality concern by identifying the factors that contributed to an operator losing SA rather than trying to mark a specific time when the SA loss may have occurred.

Endsley's model of SA provides a way to assess the temporal aspects of a situation through the dynamic cycle of interactions depicted therein. This approach means that SA is not a static process, but rather a dynamically changing one. Because of this consideration, Endsley's model serves as a prime candidate for investigation into dynamically changing environments, such as those with dynamic levels of automation (Endsley, 2015). As such, Endsley's definition and model of SA are used in the current effort to analyze the relationships between workload, performance, and SA.

E. AUTOMATION IN SUPPORT OF SYSTEM PERFORMANCE

Automation refers to the process of substituting task control by a human with a system (Parsons, 1985). Sheridan and Verplank's (1978) seminal work on levels of automation (LOA) of decision and action selection depicts a range of automation from low to high, with lower levels associated with more human control and higher levels corresponding to more autonomous control by automation. Parasuraman, Sheridan, & Wickens (2000) expanded on this framework by proposing that levels of automation could be associated with equivalent system functions. The four system function classes are information acquisition, information analysis, decision and action selection, and action implementation (Parasuraman et al., 2000). Each of the functions could be automated to varying degrees. These functions are key design considerations for systems that have tasks that could benefit from automating them to improve system performance. The model in Figure 12 depicts the four system function classes with human performance, automation reliability, and action costs included for design considerations. The model highlights that at different points of information processing, humans can benefit from automation assistance. Further, the model helps to show that different aspects of information processing can be allocated between humans and automation (Save, Feuerberg, & Avia., 2012).

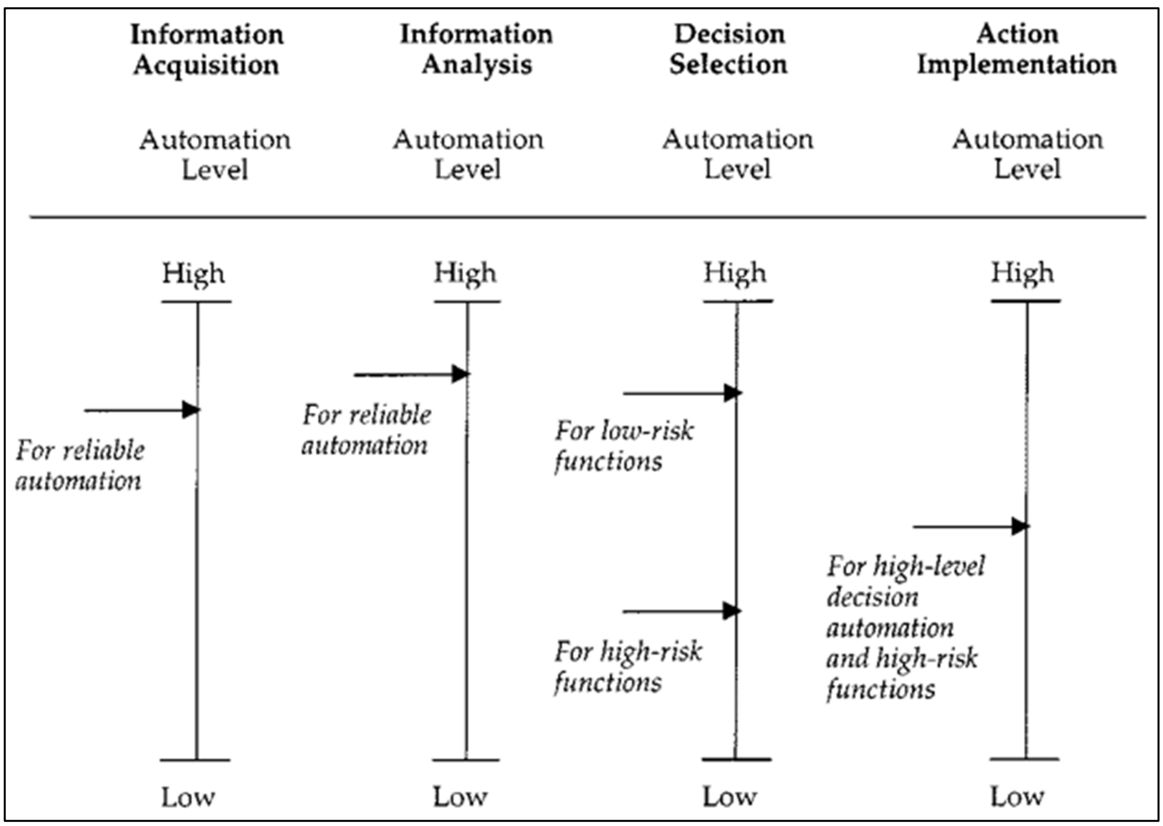


Figure 12. Types and levels of automation for ATC systems. Source: Parasuraman et al. (2000).

An adaptation of Sheridan and Verplank's (1978) levels of automation and Parasuraman, Sheridan, and Wickens' (2000) levels of autonomy framework can be applied to autonomy levels as well, as seen below in Figure 13 (O'Neill et al., 2020). As computers gain higher levels of agency (post-level 4), they begin to be considered more as teammates. The difference between automation and autonomy is the extent of the degree they exist on the LOA scale. For instance, at level 10, a computer agent is acting autonomously in that it is deciding and acting without any consideration from a human agent. Conversely, at level 2, the computer agent aids a human agent, but the human is ultimately in control of the system. Cognitive workload is one of the many factors impacted by different levels of automation, and its management is a key consideration to ensure operator safety and mission completion. The extent to which workload varies within each static level of automation can be significant (Evans & Fendley, 2017). This impact of automation on workload suggests that cognitive workload model predictions should change based on dynamic shifts between LOAs.

Automation Level	Agent Autonomy Level	Automation or Autonomous Agent Role and Capability
High	High agent autonomy	10. The computer decides everything and acts autonomously, ignoring the human. 9. The computer informs the human only if it, the computer, decides to. 8. The computer informs the human only if asked, or 7. The computer executes automatically, then necessarily informs the human, and
	Partial agent autonomy	6. The computer allows the human a restricted time to veto before automatic execution, or 5. The computer executes that suggestion if the human approves, or
	No autonomy / Manual control	4. The computer suggests one alternative, or 3. The computer narrows the selection down to a few, or 2. The computer offers a complete set of decision/action alternatives, or 1. The computer offers no assistance; the human must take all decisions and actions
Low		

Note. Adapted from Parasuraman et al. (2000) with permission from the Copyright Clearance Center and Rights Link/IEEE.

Figure 13. Adapted levels of automation to autonomy levels. Source: O'Neill et al. (2020).

F. ADAPTIVE AUTOMATION

Adaptive automation (AA) has been identified as "one of the most important ideas in the history of human factors/ergonomics" (Hancock et al., 2013, p. 11). It abounds in many fields, including aviation medicine, transportation, and logistics. Within the aviation domain, advances in automation afford more efficient and effective flight operations by augmenting a pilot's cognitive abilities (Calhoun, 2021; Scerbo, 1996; Wiener, 1989).

Adaptive automation is the ability to modify the automation level in real time in response to demands on the system (Sheridan, 2011). The concept emerged in the 1970s in the field of artificial intelligence to determine the blend of task allocation between humans and machines and later would extend to finding ways to utilize adaptive systems in the aviation field (Rouse, 1988). Adaptive automation is differentiated from adaptable automation in that adaptable automation relies on an operator selecting and initiating the level of automation. With adaptive automation, humans and machines share the ability to determine changes in the automation's state (Scerbo, 1996).

One of the primary objectives of adaptive automation is to manage operator workload at an optimal level (Parasuraman et al., 1992; Hilburn et al., 1993; Endsley, 2017a). An assumption accompanying this perspective is that workload levels can be defined and specified through surrogate measures to serve as a starting point for adjusting automation. Adaptative automation has shown to lessen human error and performance variability (Scerbo, 1996). It also can enhance control of increasingly complex systems while mitigating operator performance variability, leading to reduced error rates (Scerbo, 1996; Woods, 1996). Additional intended outcomes of adaptive automation include keeping humans within a "band of proper workload" (de Greef & Arciszewski, 2009), assisting operator cognitive processes (Inagaki, 2003; Hancock et al., 2013; Kaber et al., 2006), and enhancing system performance (Brand & Schulte, 2017). These findings of reductions in cognitive workload and increases in performance suggest that the use of adaptive automation can serve as a viable intervention in high workload tasks (Endsley, 2017a; Ernst et al., 2020). Further, changes in operator performance should be evident at different LOAs through their dynamic changes when using AA systems.

Adaptive automation also seeks to reduce human OOTL processing issues such as complacency, which can manifest when an operator is over-reliant on the automation to conduct a task (Kaber et al., 2006). This characteristic helps facilitate two other intended outcomes: enhancing performance and safety (Kaber et al., 2006). Further, AA solutions need to balance changes in concert with factors happening in the periphery (Vagia et al., 2016) and do so at the right time to provide the appropriate assistance at the point of need (Hancock et al., 2013).

While adaptive automation solutions provide benefits, designers can inadvertently create detrimental issues when developing AA systems. For instance, adaptive automation should only provide assistance when required so as to prevent unnecessary or ill-timed assistance that can lead to degraded performance (Scerbo, 1996). Adaptive automation creates increased system complexity with new technical challenges that are introduced with novel systems (Christoffersen & Woods, 2002; Endsley, 2017a; Scerbo, 1996; Woods, 2016). Adaptive automation systems may actually increase workload demands on humans when the human operators have to manage function allocation while also performing routine tasks (Kaber et al., 2001; Bainbridge, 1982a; Endsley, 2017a; Hollnagel & Woods, 2005). Adaptive automation can introduce additional monitoring and processing demands on operators due to increasing information and visual displays that depict the actions and statuses of AA (Kaber et al., 2001; Kaber & Endsley, 2004). These are just some of the issues that may contribute to the manifestation of unintended consequences, negatively impact system performance, and negate the intended benefits AA should bring to a system.

Unintended consequences are essential considerations in an evolving technological world (Tonn & Stiefel, 2019). As AA systems grow in complexity, there is a potential for them to introduce unintended consequences. Some of these consequences may be positive. For instance, humans may be able to focus more attention and complete tasks that are directly supporting task accomplishment. Operators may also find new ways to use existing systems for additional purposes to facilitate more efficient operations. Additionally, there exists the potential for digital technologies such as AA to become more sustainable in their development to support the human, rather than having to develop multiple methods to do so (Bohnsack et al., 2021).

However, AA can also present risk and negative unintended consequences on automation systems (P. Smith & Baumann, 2020). These negative impacts include knowledge and skill degradation, overreliance on automation to identify operational issues, fewer attentional resources to address system errors, non-compliance with system state indicators, and adverse impacts on team SA (P. Smith & Baumann, 2020). Unintended consequences introduced by humans operating complex systems are thus susceptible to multiple avenues of failure if the automation itself fails. These systems are further subject to operational issues if the human interaction with it is suboptimal. One consideration with the capabilities adaptive automation systems offer is the inadvertent increase in workload that is created when the operator must interact with a computer or manage function allocations between them and the adaptive systems. This increase in workload proposition follows previous findings that place more demands on human operators as they are responsible for managing the increasing amount of technology and the corresponding amount of increased information in their workspace (Bainbridge, 1982a; Endsley, 2017a; Hollnagel & Woods, 2005).

Additionally, determining when to initiate adaptive automation solutions without negatively impacting an operator's performance is a critical consideration that remains paramount in attempts to employ this technology (Aricò et al., 2016; Rusnock & Geiger, 2017). Adaptive automation can lead to decreased trust due to surprises that result from its actions (de Greef & Arciszewski, 2009; Inagaki, 2003; Parasuraman & Riley, 1997; Woods et al., 2021). Adaptive automation can provide assistance at the wrong time or not at all, which can lead to unintended increases in operator workload (Hancock et al., 2013). Further, operators may experience decreased situation awareness and have to address adaptive shortfalls that require diverting their attention from their primary tasks (Inagaki, 2003; Kaber et al., 2001; Woods, 2016).

As adaptive automation solutions are embedded in more complex systems, new challenges and requirements are introduced that need to be addressed to mitigate potential mismatches in simulated system demonstration and real-world performance. Designers must confront new challenges to mitigate the potential for detrimental surprises. This consideration follows the concept of Doyle's Catch, whereby new systems create a potential mismatch between a simulated capability demonstration and the real world (Woods, 2016).

Adaptative automation is generally described as being triggered using three different approaches: the critical-event strategy where analysis conducted ahead of an event seeks to identify high workload event times (Inagaki, 2003); the performance-measurement strategy whereby an operator's performance during tasks is used to estimate current and future state and workload (Aricò et al., 2016; Inagaki, 2003); and the neurophysiological measurement strategy that uses various signals from an operator including EEG, ECG, and GSR to make inferences on mental workload in near-real time (Scerbo et al., 2001).

Revoking adaptive automation is a key consideration for when and how to return control of a task back to a human. Rusnock and Geiger (2017) posited that the revocation of adaptive automation should be done after cognitive workload returns to manageable levels. Strategies to revoke adaptive automation include the inverse of the invoking strategies and user-initiated deactivation. Because AA is reliant on real-time operator state data for its invoking and revoking, further study into psychophysiological workload measurement as related to workload predictions is warranted (Rusnock & Geiger, 2017). Cognitive workload could then be measured and managed more effectively in keeping with the goals of AA. Therefore, it is essential to have reliably accurate measures of an operator's cognitive workload to facilitate proper use of AA systems (Ayaz et al., 2012; Parasuraman, 2003).

Psychophysiological measurements, such as brain activity, have been shown to provide accurate representations of an operator's workload in air traffic control tasks (Aricò et al., 2016). Subjective measures such as the NASA-TLX have shown to correlate with these data to provide validation of cognitive workload models (Rusnock & Geiger, 2017). Taken together with performance metrics, these measures can facilitate better understanding of an operator's cognitive workload. Further, assessing cognitive workload model predictions with these subjective and objective metrics might yield relationships that allow for workload forecasting.

There is an extensive body of research that has identified and investigated the potential impacts on operators when using adaptive automation. Because of these relevant considerations, adaptative automation serves as an integral element for this research. However, studies that seek to model and measure the extent to which AA negatively and unintentionally impacts operator workload are lacking. Research that uses psychophysiological measures, subjective ratings, and task performance metrics is needed to provide more robust analysis on the impacts of using AA. This approach also supports using real-time psychophysiological and performance metrics to predict an operator's state, which is critical for the use of AA. This dissertation addresses this gap in the literature by modeling and revealing the identified potential negative unintended and unanticipated consequences that emerge when using AA systems.

G. TASK ANALYSIS

Task analysis (TA) can help describe how an operator accomplishes an end state. One definition of TA is: "the study of what an operator (or team of operators) is required to do, in terms of actions and/or cognitive processes, to achieve a system goal" (Kirwan & Ainsworth, 1992, p. 1). There are different types of task analyses such as Hierarchical Task Analysis (HTA), Job Task Analysis (JTA), and Cognitive Task Analysis (CTA).

Task analyses allow us to allocate functions in a system to a human, to a machine, or to a combination of both. They also allow for workload assessment to determine the load a task may place on someone (Kirwan & Ainsworth, 1992). Task analyses also impact job specifications, manning numbers, personnel selection considerations, and task procedures by identifying the number and type of people a system might require for effective operation. These factors are important when conducting tradeoff analysis among the domains of human systems integration (HSI) and in the system design process.

Task analysis can also inform user interface design (UID) and be leveraged to reduce workload and error probability through proper employment of its results (Kirwan & Ainsworth, 1992; Rosala, 2020). Knowing what tasks are most important in a system can drive the design of a display to place those critical items in high visibility areas. Further, the content presented on those displays can be oriented and presented in such a way that facilitates task completion. This is a key tenet of human factors engineering (HFE) and is another example of how TA can be leveraged in system design to present salient information in efficient ways.

Different task analysis methods can be applied at different stages and mixed at different times to answer questions when designing training. Task analysis is not a specific method, but rather a concept or goal to understand what a user is required to do to complete a task (Adams et al., 2012). Understanding the task is not as simple as this view, however, as a task can be multi-faceted (cognitive, physical, goals, etc.). Therefore, it is important to look at task analysis methods as approaches that address dynamic task aspects and that account for the changing nature of work from physical to cognitive. Additionally, task analysis methods should be further viewed as approaches that emphasize safety-focused operations instead of efficiency (Adams et al., 2012). Designing for adaptive aiding through automation highlights the importance of task analyses because it requires understanding the context of activities and the functions necessary to complete a task (Adams et al., 2012; Inagaki, 2003; Hancock et al., 2013).

H. MODELING AND SIMULATION

Modeling and simulation (M&S) is an interdisciplinary field that is widely used across an array of domains. The Department of Defense (DOD) defines a model as a "physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process" (Department of Defense, 2018, p. 9). Some examples include a mathematical equation, a simple toy model, or representations of human behavior. A model can also be described as an abstract representation or idealization of some phenomenon (Weisberg, 2013). Simulations are described as the dynamic application of a model, or a model run over time (Department of Defense, 2018). Modeling and simulation's applications can be seen in training, healthcare, test and evaluation, human factors, and many other domains. For the DOD, human performance modeling represents an acknowledgment of the intersection of technological capabilities to support a system's life cycle and the importance of warfighter consideration early in that process.

1. Cognitive Architectures and the Improved Performance Research Integration Tool (IMPRINT)

Two ways of modeling human performance and workload are cognitive architectures and task network modeling. A cognitive architecture can be described as "a scientific hypothesis about those aspects of human cognition that are relatively constant over time and relatively independent of task" (Ritter & Young, 2001, p. 3). They are broad theories of human cognition based on various human data and implemented as a simulation. Cognitive architectures can help with physical model development and also seek to simulate human intelligence in ways that emulate humans (Byrne, 2009). As we learn about the human mind through cognitive architectures, we can create models of those behaviors and then continue in facilitating human-centered design of them to bring about more optimal systems. As the system learns from the human, those inputs can be translated into changing software requirements that are geared towards meeting dynamic mental models as the user gains experience, training, and interaction with the system. Examples of cognitive architectures include ACT-R (Carnegie Mellon University, 2013) and Soar (University of Michigan, 2021). Cognitive architectures facilitate a bottom-up approach of modeling human performance by combining multiple cognitive individual knowledge elements into a complex process (Lebiere et al., 2005).

Task network modeling techniques represent a top-down approach to human performance and workload modeling (Lebiere et al., 2005). Task network modeling, or discrete event simulations, can be used to conduct human performance modeling. To do so, task analyses are conducted to decompose an operator's functions into tasks and then develop the task network in context. Once the task analysis is complete, the process of developing a network based on the analysis is relatively straightforward and can be executed with low overhead. An example of this is the U.S. Army's Improved Performance Research Integration Tool (IMPRINT).

As human considerations became a more prevalent driving force in system design in the 1980s, the U.S. Army established the Manpower and Personnel Integration (MANPRINT) program. To support the program, the Army sought tools, techniques, approaches, and methods to provide more empirically robust foundations for the program. A major and long-lasting solution born from addressing that capability gap was the development of a human performance modeling tool called Hardware vs. Manpower III (HARDMAN III), which was later recast as IMPRINT (Allender et al., 2005).

IMPRINT is a dynamic, stochastic, discrete event modeling and simulation tool that assists in assessing human interaction with a system and the resulting workload experienced in completing a mission (Department of the Army, 2019; Mitchell, 2000, 2009; Wojciechowski, 2004). IMPRINT was the first human performance modeling tool accredited by the U.S. Army (Allender et al., 2005). In system design, IMPRINT can assist in identifying operator-driven constraints, setting realistic requirements, and allowing for analysis of manpower and personnel requirements to operate a system. In research efforts, IMPRINT can provide a means to analyze task analyses, workload modeling, and performance-shaping functions (Samms, 2010).

IMPRINT provides predictions of mental workload through both the Visual, Auditory, Cognitive, and Psychomotor (VACP) Theory (McCracken & Aldrich, 1984) and the Multiple Resource Theory (MRT) (Wickens, 2002). To do so, a task analysis must be conducted first to decompose an operator's functions into tasks and then a network of the task sequence in context is developed. Once the initial task analysis is completed, a task network model can be executed with relatively low overhead. IMPRINT can then run the inputs from the modeler to assess different factors in a variety of conditions as see in Figure 14.

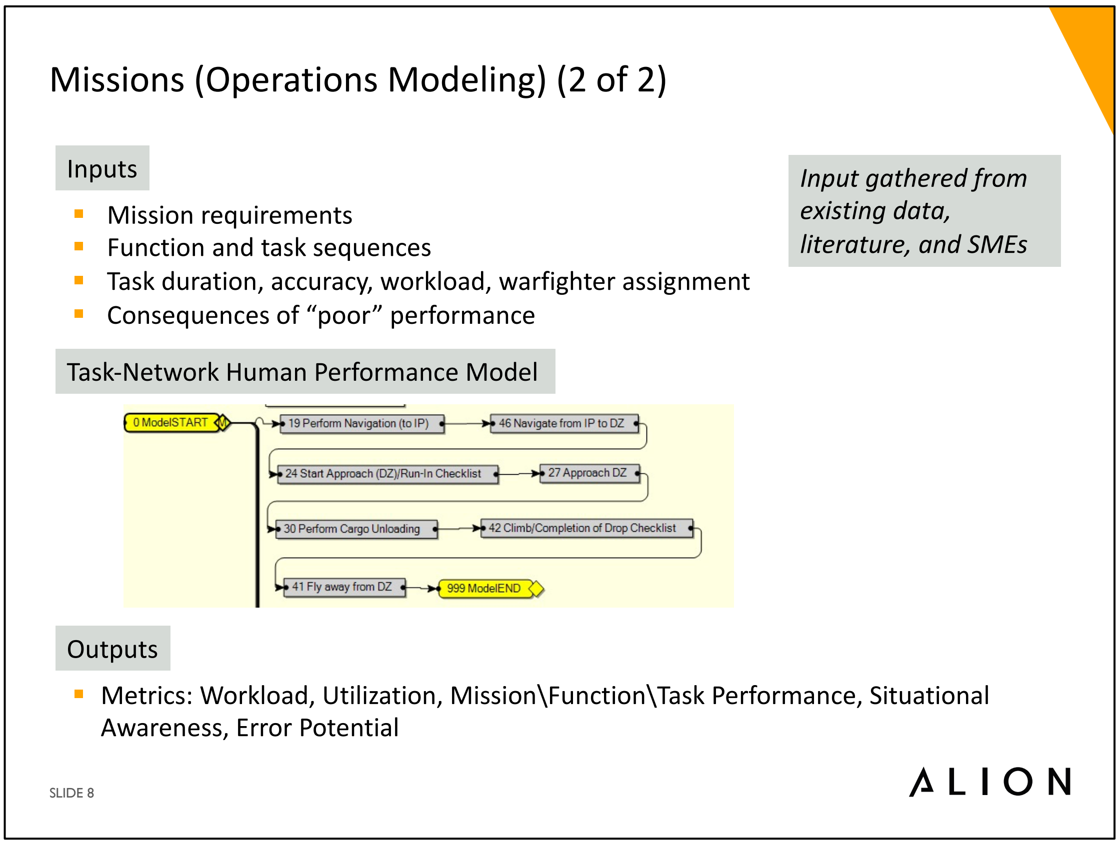


Figure 14. Example of IMPRINT model flow. Source: Alion Science (2021).

Modelers using IMPRINT can link tasks with the mental resources required to accomplish them. They can then assign quantitative demand values based on the VACP scale to each mental resource for the task, with descriptions for each demand level provided as seen in Figure 15. Hardware and software sub-models can be incorporated into the model to show how the human, machine, and environment are represented during a closed-loop cycle of an operation (Dahn & Laughery, 1997).

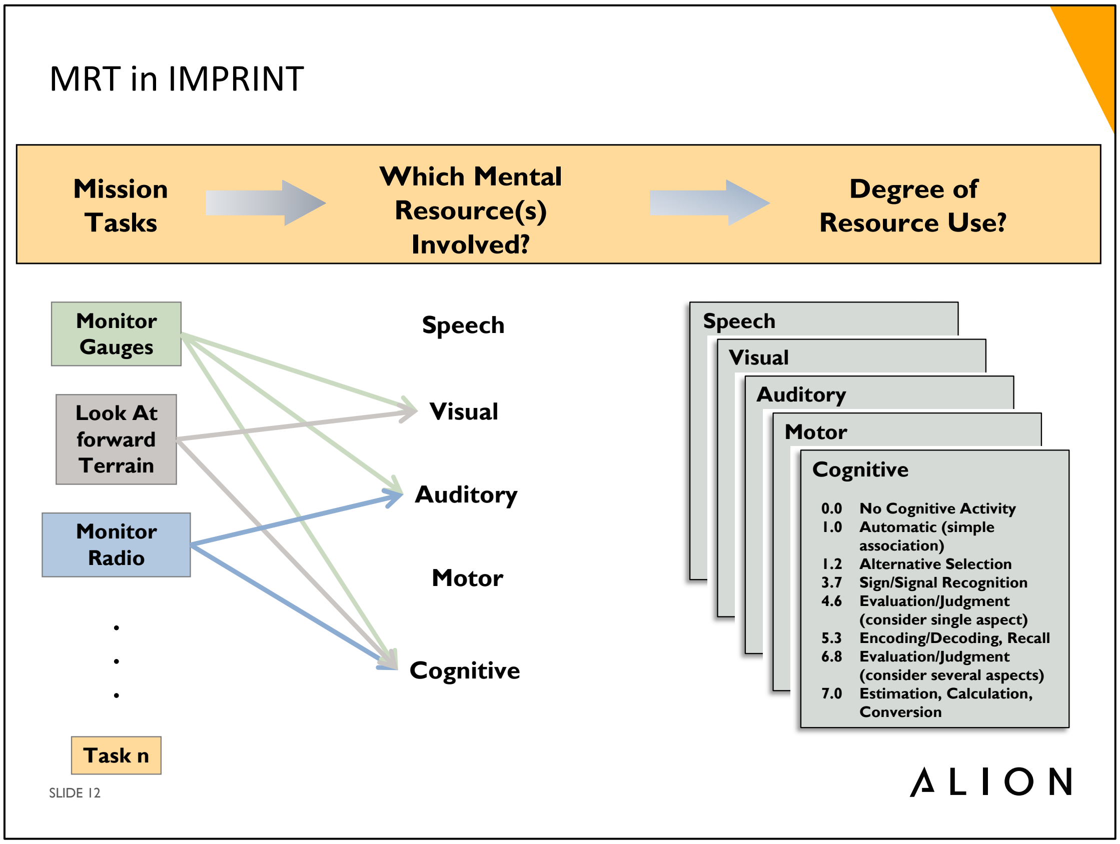


Figure 15. Overview of Multiple Resource Theory application in IMPRINT. Source: Alion Science (2021).

Engineering models of human performance provide values for aspects of performance (e.g., time on task) in an intuitive manner. They should satisfy three criteria: view the human as an information processor, leverage approximate calculations based on a task analysis, and allow for performance predictions of systems while they are still in the design phase though much uncertainty still exists (Card et al., 1983). They should afford designers approximate quantitative predictions of performance for design alternatives (Proctor & Van Zandt, 2018).

IMPRINT meets the aforementioned criteria, and researchers using it have produced models supporting FVL with recommendations regarding crew manning levels and task allocation between crewmembers (Militello et al., 2019). IMPRINT uses previous findings of cognitive workload to build algorithms for use in the simulation and analysis of tasks (Lebiere et al., 2005). Cognitive architectures do not typically have the capability to predict cognitive workload (Jo et al., 2012). Unlike cognitive architectures, IMPRINT does not have any model of cognitive processes and instead relies on the modeler to build those (Lebiere et al., 2005). Further, IMPRINT facilitates looking at the operator and system interactions in an integrated manner, whereas cognitive architectures focus more on the human cognitive processes needed to accomplish a task. Based on of these considerations, IMPRINT will be used in the current effort to build on the line of previous research efforts examining workload and the human-system interactions associated with it from a top-down perspective.

2. NASA Multi-attribute Task Battery II

The NASA Multi-Attribute Task Battery II (MATB-II) is a computer-based solution that allows for evaluation of human workload and performance (Santiago-Espada et al., 2011). MATB-II is derived from the original Multi-Attribute Task (MAT) Battery developed in the early 1990s that was used to investigate workload and human-automation interaction (Comstock & Arnegard, 1992). MATB-II is capable of being used in training and testing modes and has flexibility in its configuration and execution through manipulation of its source files written in the Extensible Markup Language (XML). Numerous studies have used MATB-II to investigate an operator's performance when executing multiple, complex tasks in both aviation and non-aviation domains (Gutzwiller et al., 2014; Kong et al., 2022; Liu et al., 2016; Santiago-Espada et al., 2011; Wusk et al., 2019; Zhang et al., 2020).

MATB-II includes four tasks presented through a user interface as seen in Figure 16: a system monitoring task (SYSMON), a tracking task (TRACK), a communications task (COMM), and a resource management task (RESMAN) (Santiago-Espada et al., 2011). MATB-II presents a scheduling indicator to show users upcoming communications and tracking task requirements within the next eight minutes of the trial. Flow rates to support decision-making on the RESMAN task also appear immediately to the right of the main RESMAN task. The MATB-II interface also displays a Figure of Merit (FOM) beneath the scheduling indicator to inform the operator of their performance at that time. The figure of merit can be calculated using the current score, a simple moving average (SMA), or an exponential moving average (EMA). The exponential moving average calculates the average performance across all four tasks with more weight placed on the current score to address lagging indicators of performance that an SMA would provide. Therefore, the use of the EMA was used in the current studies to allow for examination of a user's performance as they gain experience operating the tasks.

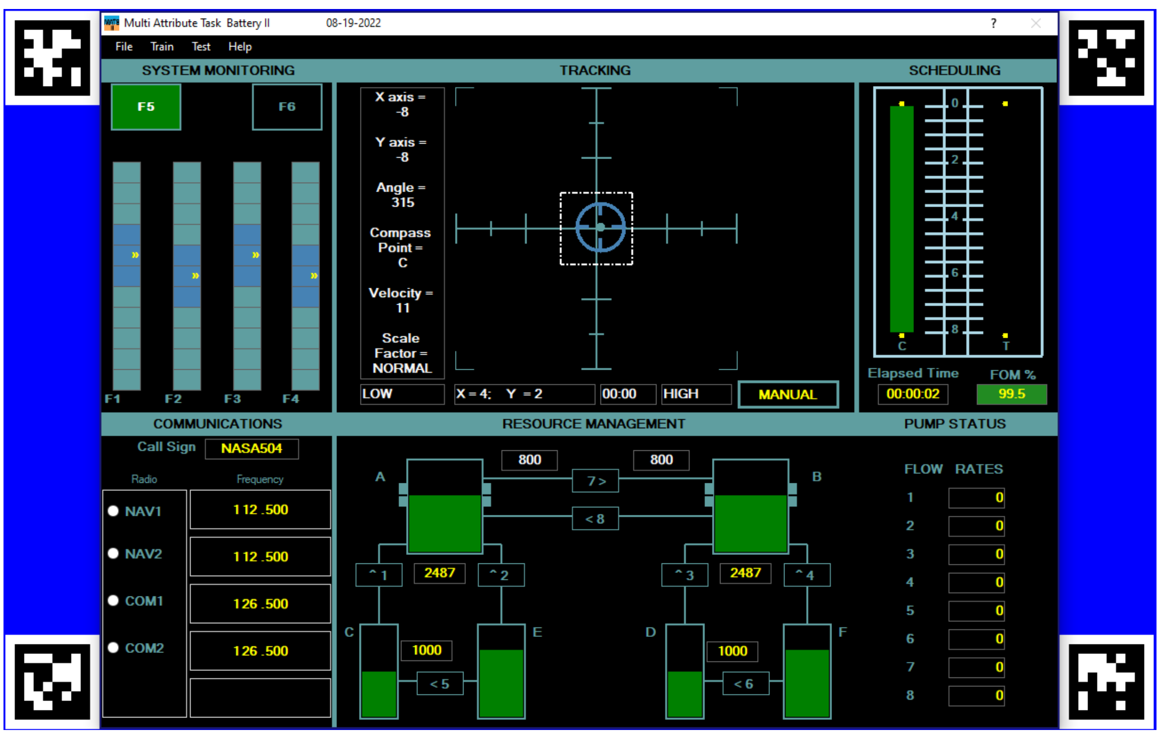


Figure 16. MATB-II user interface with April tags. Source: Santiago-Espada et al. (2011) and Olson (2010).

a. Verification, validation, and accreditation

Verification, validation, and accreditation (VV&A) are integral components of the M&S domain. The DOD verification, validation, and accreditation (VV&A) process is an important component of the modeling and simulation enterprise. The VV&A process allows M&S professionals to analyze and provide recommendations for accepted uses of a model. VV&A helps define the boundaries of a model's application. VV&A also helps close the credibility gap between the developer and the end user. VV&A is an important topic in M&S because even the best models we have are still large abstractions of the referent. Therefore, intended use of the model is important so one can offer potential areas for reuse.

Verification is defined as ensuring "that the computer programming and implementation of the conceptual model are correct" (Sargent, 2013, p. 14). The DOD defines verification as the "process of determining that a model or simulation implementation and its associated data accurately represent the developer's conceptual description and specifications" (Department of Defense, 2018, p. 10). Another definition of verification is that it is "the process of determining that a model implementation and its associated data accurately represent the developer's conceptual description and specifications" (Strickland, 2011, p. 60). Verification methods help show that the M&S solution correctly performs the intended functions. In other words, verification helps to identify if the model was built in the correct manner.

Validation is defined by the DOD as "the process of determining the degree to which a model or simulation and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model" (2018, p. 10). Sargent defines validation as "substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (2013, p. 12). It can also be expressed as "the process of determining the degree to which a model and its associated data provide an accurate representation of the real world from the perspective of the intended uses of the model" (Strickland, 2011, p. 60). Validation methods show how well the M&S represents the real world or the referent. Validation is also known as building the right model, or that the conceptual models align with the referent. Validation ensures that the computational model follows the conceptual model adequately.

Accreditation is defined as an official determination that a model or simulation is acceptable for use for a specific purpose. The appropriate decision maker confirms that the model is suitable for their needs (Department of Defense, 2020b). Accreditation can also be defined as "the official certification that a model, simulation, or federation of models and simulations and its associated data are acceptable for use for a specific purpose" (Department of Defense, 2018, p. 9).

The VV&A activities that correspond with the development of M&S solutions are critical to ensure that M&S professionals can analyze and provide recommendations or acceptable uses for a model. The VV&A process can help define the boundaries of what the model can and cannot do. This scoping of a model's reach can help manage the implementation to ensure that the model is being used in an appropriate context. Models are abstractions of some referent, and often they are large abstractions. Therefore, it is important to clearly define the intended use of the model to ensure that its use in a scenario is appropriate and to provide recommendations for reuse in other domains.

The VV&A process helps reduce a risk incrementally throughout a model's development. Verification helps reduce risk by attempting to ensure that the model has fewer undetected errors in it. Validation reduces the risk that the model or simulation does not match the real-world referent well enough to lose credibility with the stakeholders. Accreditation reduces the risk that inappropriate or unsuitable models and simulations are selected and implemented. Therefore, including the VV&A process in building the current research effort is critical to ensure that the models built to examine an operator's cognitive workload are correctly determining workload values. Further, VV&A helps to ensure that the models are representing the real-world referent accurately and adequately. While the VV&A process is a continuous and iterative process, the current research effort will focus on V&V specifically given that accreditation of a system is not within the scope of the dissertation. Additionally, IMPRINT is an accredited U.S. Army M&S tool which allows for the current research to leverage its capabilities in cognitive workload modeling. The focus on V&V will allow for building more robust models of human operator cognitive workload during interactions with varying levels of automation.

I. SUMMARY

Cognition, cognitive workload, and situation awareness are key design considerations that impact performance independently and through resulting interactions with each other. While cognitive workload and SA are independent concepts, they can be inter-related (Vidulich, 2000). Cognitive workload can be associated with SA in an inverse manner, such as seen with an inverse-U curve. When cognitive workload increases, SA can decrease. Situation awareness can have the similar effects on cognitive processing. Therefore, system designers need to consider methods and approaches to managing this relationship to keep operators within optimal levels of performance. Automation has long been considered a way to accomplish the goal of keeping operators within an optimal level of performance through managing cognitive workload and promoting situation awareness. Adaptive automation attempts to go further by dynamically allocating functions to operators or a system to alleviate demands on human operators. However, this assistance does not always meet the intended outcomes of lowering cognitive workload. In fact, adaptive automation can create additional demands on an operator. These additional demands act in opposition to the intended outcomes of adaptive automation and introduce additional attentional demands, complex system interactions, and more cognitively involved tasks. The resulting unintended consequences of adaptive automation degrade human performance and are worthy of investigation to identify, measure, and mitigate their causes.

Subjective and objective measures of workload, coupled with human performance modeling predictions from tools like IMPRINT, is an area of exploration that may help in alleviating adaptive automation's unintended consequences. Further, incorporation of performance metrics with subjective and objective workload measures has the potential to assist in the development of a cognitive workload framework rather than using any of those metrics individually or in pairs (Longo et al., 2022). Simulation tools abound that can model tasks and assist with analyzing such emergent effects.

J. RESEARCH QUESTIONS AND HYPOTHESES

The following research questions and hypotheses provide guides to the current effort. These questions and hypotheses were derived from gaps and opportunities identified during the literature review.

Research Question 1: Can cognitive workload modeling inform design decisions in AA systems?

Hypothesis 1: Effective cognitive workload modeling will reflect changes that occur as LOAs vary within AA systems.

Research Question 1a: How do operator cognitive workload model predictions, psychophysiological measures, and subjective workload measures correlate in adaptive automation systems?

Hypothesis 1a: There is a relationship between workload model predictions, subjective workload measures, and objective workload measures as LOAs change with AA.

Research Question 1b: What is the relationship between cognitive workload and situation awareness as AA's unintended consequences emerge during a task?

Hypothesis 1b: There is a negative correlation between cognitive workload and SA as LOAs change with AA.

Research Question 2: Do cognitive workload predictions forecast future performance in AA systems?

Hypothesis 2: Cognitive workload measures can be used to predict future performance in AA systems.

Research Question 2a: Will unintended (or unanticipated) design consequences of AA systems emerge in the form of changes in performance?

Hypothesis 2a: Unintended (or unanticipated) design decisions of AA lead to performance changes.

K. CHAPTER REVIEW

Based on the review of the relevant literature, research questions and hypotheses were derived to address a gap in modeling and measuring the impacts of the unintended negative consequences of adaptive automation. While efforts have identified various categories and impacts of these consequences, investigation into measurement of their impact is lacking. Predictive human performance modeling techniques such as IMPRINT do not provide specific measures to allow for specific workload predictions with AA. This research will seek to provide parameters for those predictions based on analysis of subjective, objective, and performance measures collected throughout the research effort. Additionally, the research approach provides for the potential to assess operator workload in real-time. Real-time workload measurement is a critical consideration for AA systems since their operation is predicated on operator state data being communicated constantly to allow for dynamic adaptations.

The next chapter will introduce the first of three experimental studies that were undertaken to address the research questions and associated hypotheses. These research questions and hypotheses were based on the review of relevant literature. The first study investigated cognitive workload at two different workload conditions and served as a foundational experiment for the two subsequent studies. Results and discussions will be included for each study to describe how their designs built on each other.

THIS PAGE INTENTIONALLY LEFT BLANK

III. STUDY 1

A. OVERVIEW

The first study sought to induce stress to gauge participants' cognitive demand based on psychophysiological measures. The study was a 2 x 2 mixed design that compared levels of experience (novice vs. experienced) at different workload levels (low and high). The approach to achieve this study's aim used NASA's MATB-II as a referent for flight operations. The tasks associated with the MATB-II scenarios were also modeled in IMPRINT. Levels of workload were manipulated based on the number of tasks presented to operators. The degrees of difficulty for each MATB-II scenario was based on previous research that determined the number of tasks that would constitute different levels of workload (McCurry et al., 2022). Participants conducted multitasking operations in MATB-II at two difficulty levels (low and high) to gauge objective (psychophysiological and performance) and subjective (self-reported through NASA-TLX and CSWAG) workload measures. Additionally, participants rated their situation awareness using the SART questionnaire. The results from this study served to validate the first portion of the research's model that provides psychophysiological baseline measurements for cognitive workload and inform the follow-on studies that investigated workload with automation incorporated. The mapping of this study to the adapted MLCC framework is depicted in Figure 17.

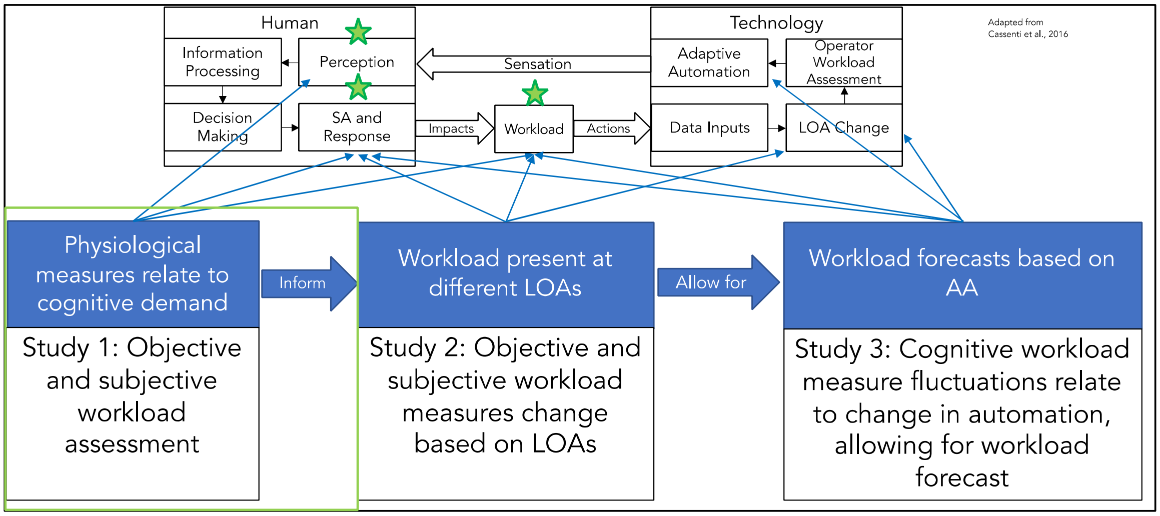


Figure 17. Study 1 mapping to the adapted MLCC framework.

The study consisted of an initial MATB-II training period on one day for all participants, with the novice participants conducting their MATB-II 20-minute test trial immediately following their training. The experienced participants conducted four additional training scenarios on their training day and then returned within 72 hours of their training to complete their 20-minute test trial.

The supported research questions and hypotheses for this study are as follows:

Research Question 1. How do operator cognitive workload predictions, psychophysiological measures, and subjective workload measures correlate in a manually completed multi-tasking simulation?

Ha1: Cognitive workload predictions are directly correlated with objective and subjective workload measures.

Ha2: Cognitive workload measures differ between the low and high workload conditions.

Ha3: Performance and workload measures differ between novice and experienced participants.

Research Question 2. What is the relationship between cognitive workload and situation awareness during a manually completed multi-tasking simulation?

Ha4: Cognitive workload and situation awareness are inversely related.

B. PILOT STUDY

A pilot study involving 4 participants was conducted to determine the training time it would take for a user to achieve a MATB-II FOM of 90%. Participants watched the MATB-II training video, completed part task training, and then completed two 5-minute MATB-II scenario trials at medium and high workload conditions, respectively. Based on the results of the pilot trials, we determined that the training and relative straightforward nature of MATB-II did not warrant having participants conduct multiple training sessions to achieve a 90% FOM. To delineate between novice and experienced users, the researchers replaced the FOM threshold of 90% with number of training exposures in a high workload condition prior to testing on the 20-minute scenario. Experienced users would instead train on four different high workload scenarios to differentiate them from the novice group, which was exposed to only one low workload training scenario. The purpose of this grouping of participants in these two categories was to highlight differences in experiences of pilots who will enter FVL aircraft. Some pilots will have only their flight school training, while others will have hundreds of flight hours to give them more experience. Additionally, some pilots will have a significant level of experience but may have to transition to a new airframe. This transition can be aided or hindered based on the pilots' ability to efficiently integrate into the new system.

C. PARTICIPANTS

1. Selection

The Naval Postgraduate School (NPS) Institutional Review Board (IRB) reviewed and approved the research methods used in this study. Participants were treated in accordance with the Department of the Navy's Human Research Protection Program standards. All participants were informed of their rights as participants in the study and signed consent forms. Participants were recruited through personal communication, email, and campus-wide announcements on the student personnel accountability website.

The inclusion criteria for participation in the study applied to any students or employees assigned to NPS, which included the following:

- Personnel in all branches of service and any specialty branch.
- No specific age, gender, or service required.
- Minimum age of subjects is 18 years old.
- Visual acuity within service standards.

The exclusion criteria for all three studies applied to the same participant sample pool. These criteria were developed to account for anomalies that would interfere with oculometric data collection and to account for confounding learning effects. The exclusion criteria for the study were as follows:

- Personnel with glasses that are bifocal, trifocal, or beyond.
- Personnel with corrective lenses that have near infrared blocking coating.
- Personnel who are red-green colorblind.
- Personnel with previous experience using NASA's Multi-Attribute Task Battery.
- Personnel can only complete one of the three studies.

2. Demographics

From the 44 participants enrolled in the study, four participants were not included in the analysis due to incomplete or faulty data collection files. Forty participants completed the study (mean age = 35.53, standard deviation [SD] = 6.53). Participants included 34 males and 6 females. Of the 40 participants, 36 were in the military (13 in the U.S. Army, 12 in the U.S. Navy, 4 in the U.S. Marine Corps, 5 in the U.S. Air Force, and 4 Department of the Navy civilians). The military participants' occupational specialties within their respective services included operations, operations support, and force sustainment. All participants were graduate students or employees at NPS. The rank breakdown of the participants is depicted in Table 2. The participants' time in service ranged from 5 months to 25 years (mean years in service = 12.79, SD = 5.48).

Table 2. Study 1 participants' military rank.

Participant Rank	Number
E-6	1
O-1	1
O-2	14
O-3	14
O-4	4
O-5	4
Civilian	2
Total	40

D. MATERIALS

The experiment used multiple displays and the equipment to interact with MATB-II, i.e., one flat-panel color monitor, a joystick, a keyboard, and a computer mouse. Psychophysiological measurement devices included an eye tracker, an fNIRS system, and a heart rate monitor. The experimenter sat behind and out of frontal view of the participants. The workstation configuration and psychophysiological measurement devices are seen in Figure 18. Detailed descriptions of all the materials used in the study are provided in Appendix A.

Figure 18. Participant and experimenter workstations.

E. VARIABLES

1. Independent Variables

The two independent variables manipulated in this study were workload and experience levels. Presentation of the workload levels was counterbalanced to account for order effects. The overview of Study 1 is shown in Figure 19.

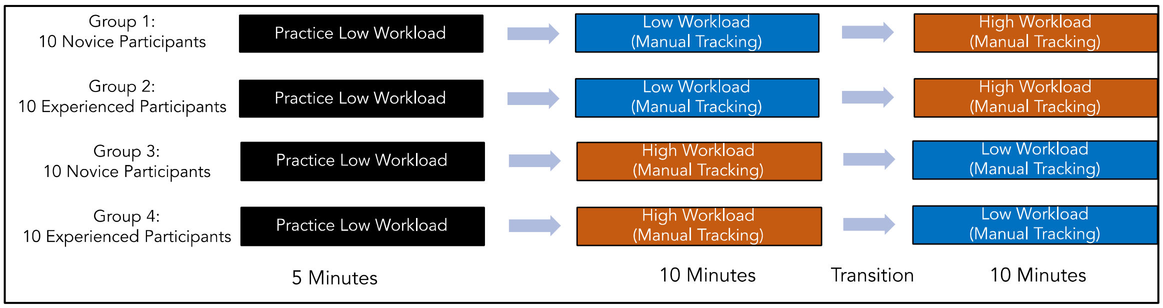


Figure 19. Study 1 overview.

a. *Workload.*

All participants experienced a low and high workload condition based on previous MATB-II studies (McCurry et al., 2022). Based on their experience level, participants were randomly assigned to a group in which they received either the low or high workload condition first. The number of tasks associated with each condition are shown in Table 3. Tasks were randomly assigned throughout the scenarios by the researcher. Of note, the communications tasks in the high workload condition were limited to those audio files that were eight seconds or less in duration to allow for execution of the audio file and the 15 second response window for the operator within the 10-minute trial period. This meant that 59 out of the 80 available audio files were available for random assignment in the high workload condition. These restrictions were not present in the low workload condition since there was more time available between communications. Thus, the full complement of 80 audio files was available for random selection in the low workload condition.

Table 3. MATB-II system settings for each Study 1's conditions.

	SYSMON	TRACK	COMM	RESMAN
Low Workload	11 Events	Low Joystick Response High Update Rate	3 Events	1 Pump Failure 1 Pump Shutoff
High Workload	20 Events	Low Joystick Response High Update Rate	12 Events	10 Pump Failures 10 Pump Shutoffs

b. Experience.

Participants were assigned to either a novice or experienced group based on scheduling availability. These groups were developed to assess differences between novice and experienced operators. Participants in both groups received the same baseline training on MATB-II. The baseline training included an orientation to the input devices, a NASA instructional video, part-task training, and a 5-minute MATB-II practice session in a low workload condition. Experienced participants conducted four additional 5-minute MATB-II practice sessions in a high workload condition. All the MATB-II sessions were written such that participants did not receive the same scenario more than once.

2. Dependent Variables

There were multiple dependent variables collected in this study. The MATB-II FOM score was used to assess differences in performance between novice and experienced participants. CSWAG and NASA-TLX ratings were collected to determine self-reported workload measures during and after the low and high workload sessions. SART ratings were collected post hoc to determine operator SA. Eye tracking data, heart rate data, and pre-frontal cortex (PFC) blood oxygenation levels were collected in real time as surrogate measurements for workload.

F. PROCEDURE

1. Performing the MATB-II

Participants concurrently performed four primary tasks in MATB-II, i.e., System Monitoring (SYSMON), Tracking (TRACK), Communications (COMM), and Resource Management (RESMAN). The MATB-II user interface is shown in Figure 20. Participants interacted with the SYSMON, COMM, and RESMAN tasks with a mouse. They used a joystick to accomplish the TRACK task. Participants were required to notice system status changes in the SYSMON task. They had to keep the circle within the square reticle during the TRACK task. Participants were also required to change radio frequencies in the COMM task and manage fuel levels in the RESMAN task. Specific requirements about the operation of MATB-II are listed in Appendix A.

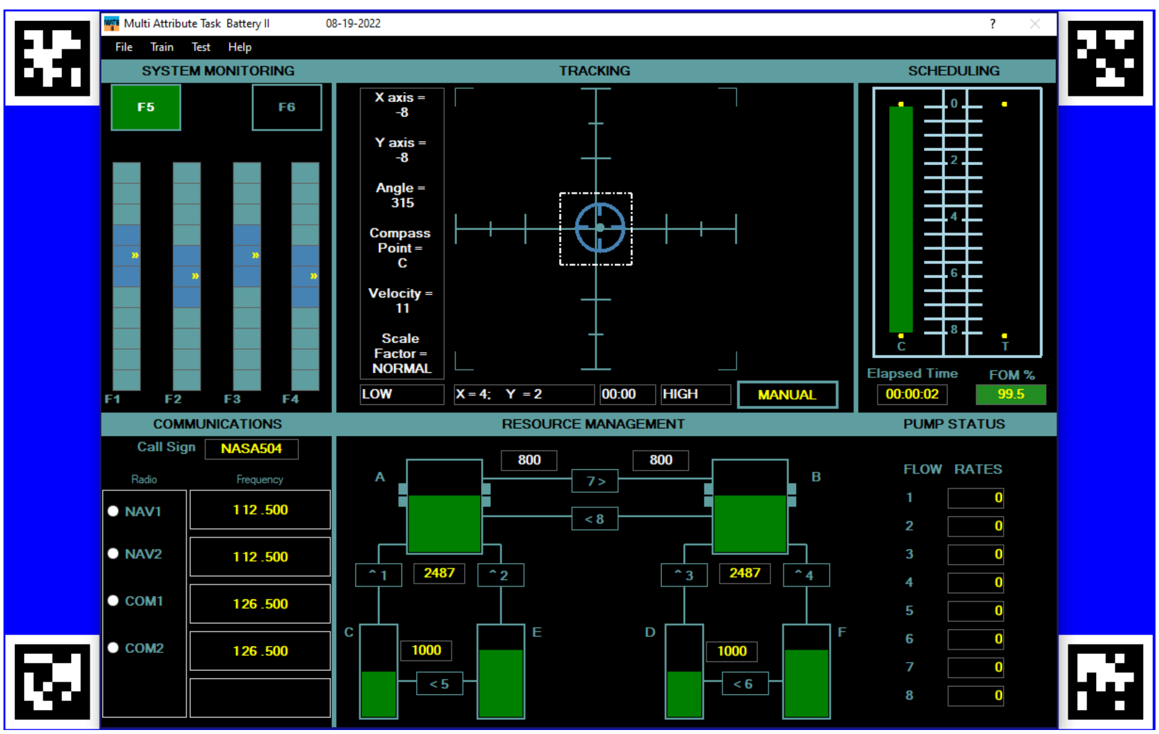


Figure 20. MATB-II user display.

2. Novice Participants.

Novice participants coordinated for a participation time and reported to the Human Systems Integration Laboratory (HSIL). The sequence of the novice group's participation is depicted in Figure 21. Detailed experimental procedures are provided in Appendix A.

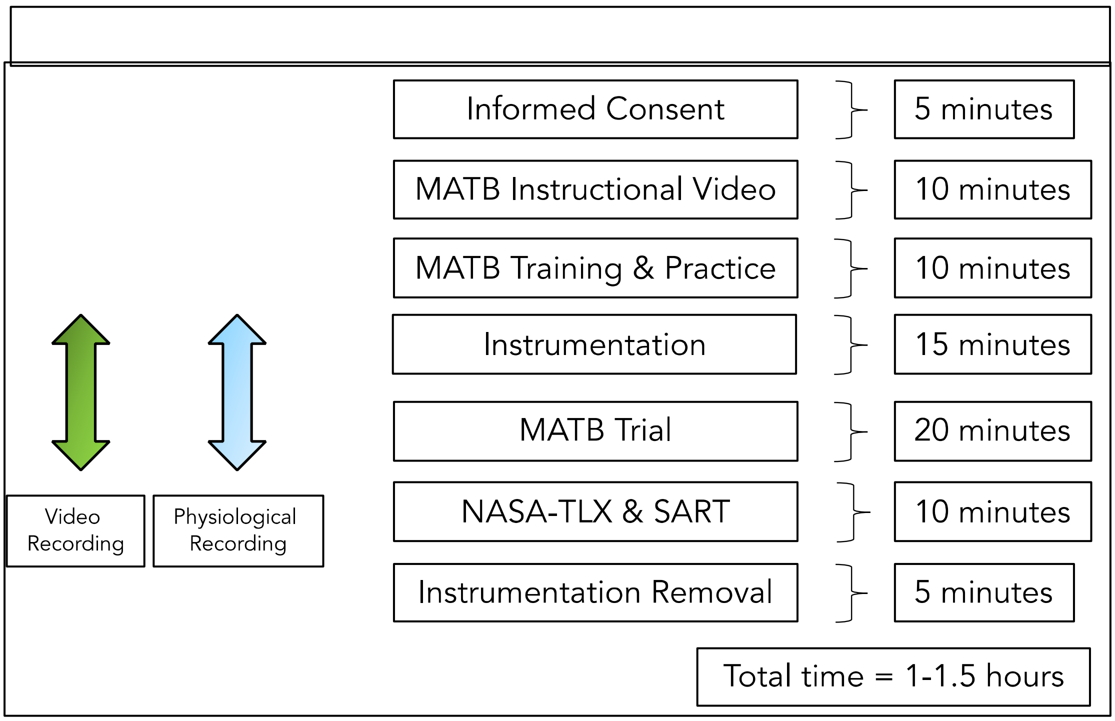


Figure 21. Novice participants' Study 1 experimental sequence.

3. Experienced Participants

Experienced participants completed the study with the same training approach as the novice participants. However, upon completion of the baseline training progression, the experienced participants then completed four MATB-II trials at a high workload level as shown in Figure 22. These trials were five minutes in duration, with a less than 30 second break between trials to allow for loading of the next trial. Experienced participants then confirmed their experimental data collection time that was within 72 hours of their training. The purpose of the training progression and time between studies was to minimize learning effects that have been associated with MATB-II (Kong et al., 2022).

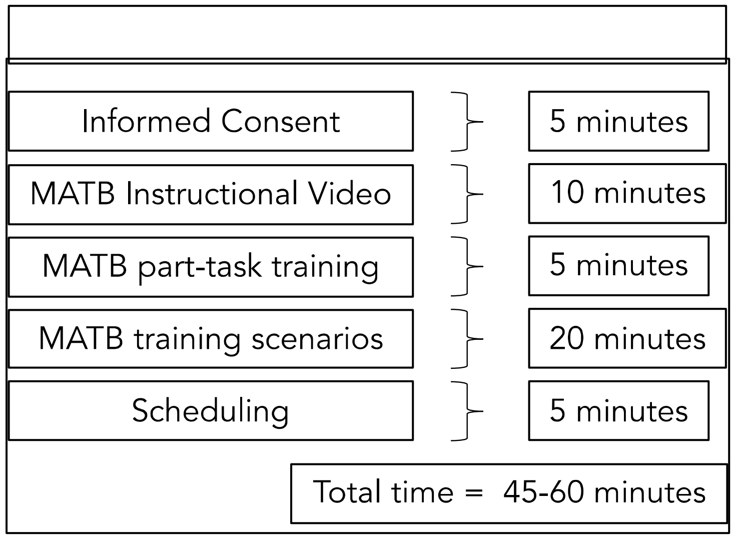


Figure 22. Experienced participants' Study 1 training sequence.

Within 72 hours of their initial training session, the experienced participants returned to the HSIL for their experimental session. The sequence that the experienced group followed is shown in Figure 23. The participants were refamiliarized with their consent forms and asked if they had any questions regarding the study. Participants then conducted a 5-minute low workload practice scenario that was different from their previous training session. Following the practice session, participants were instrumented with the psychophysiological measurement devices in the same manner as the novice group. Experienced participants were randomly assigned a workload condition such that half the participants completed the low workload condition first and vice versa. Upon completion of the experimental trial, participants followed the same sequence as the novice group by removing the instrumentation and completing the NASA-TLX and SART. Participants were given the same debriefing and instructions as the novice group to conclude their involvement in the study.

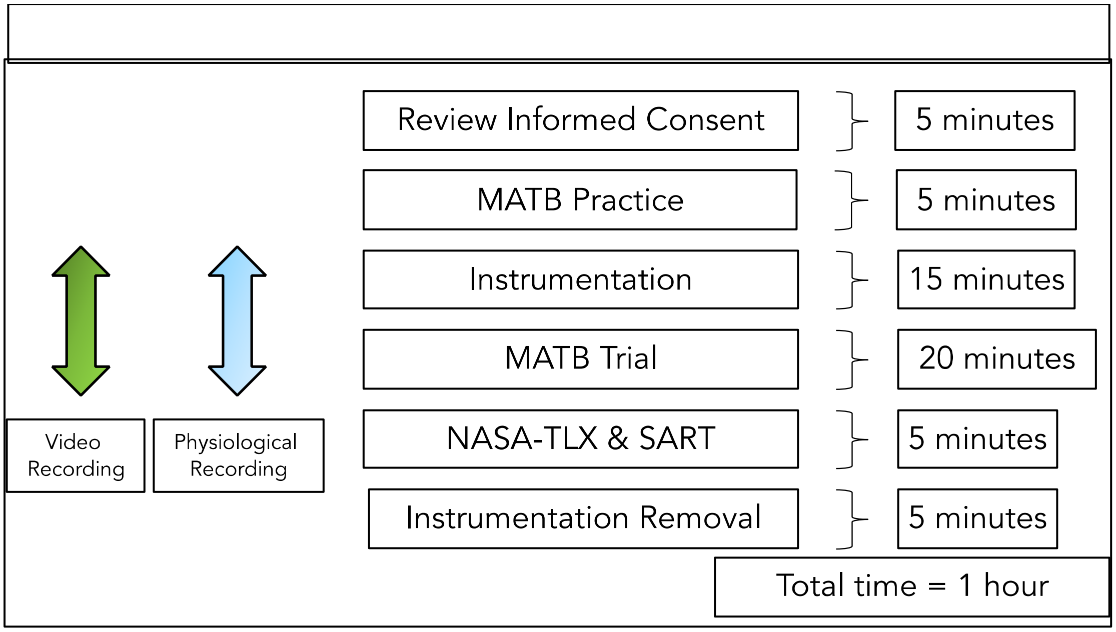


Figure 23. Experienced participants' Study 1 trial sequence.

4. IMPRINT Modeling

The researcher constructed an IMPRINT model of both workload conditions used in Study 1. These models provided the basis for comparative analysis between the cognitive workload predictions in IMPRINT and the collected workload measures in Study 1. The task network diagrams were the same for the low and high workload conditions is seen in Figure 24. The external events that triggered the communications and manual tracking tasks are shown in Figure 25. The IMPRINT models were developed after conducting a task analysis that determined the cognitive and physical requirements necessary to complete the MATB-II tasks scripted in Study 1's scenarios. The four main MATB-II tasks are depicted in the models. The TRACK, SYSMON, and RESMAN tasks were modeled such that they occur continuously but with user engagement distributed among them. The researcher modeled these tasks with the assistance of an IMPRINT subject matter expert to ensure that the task network flowed properly and that logic within the model allowed for proper execution.

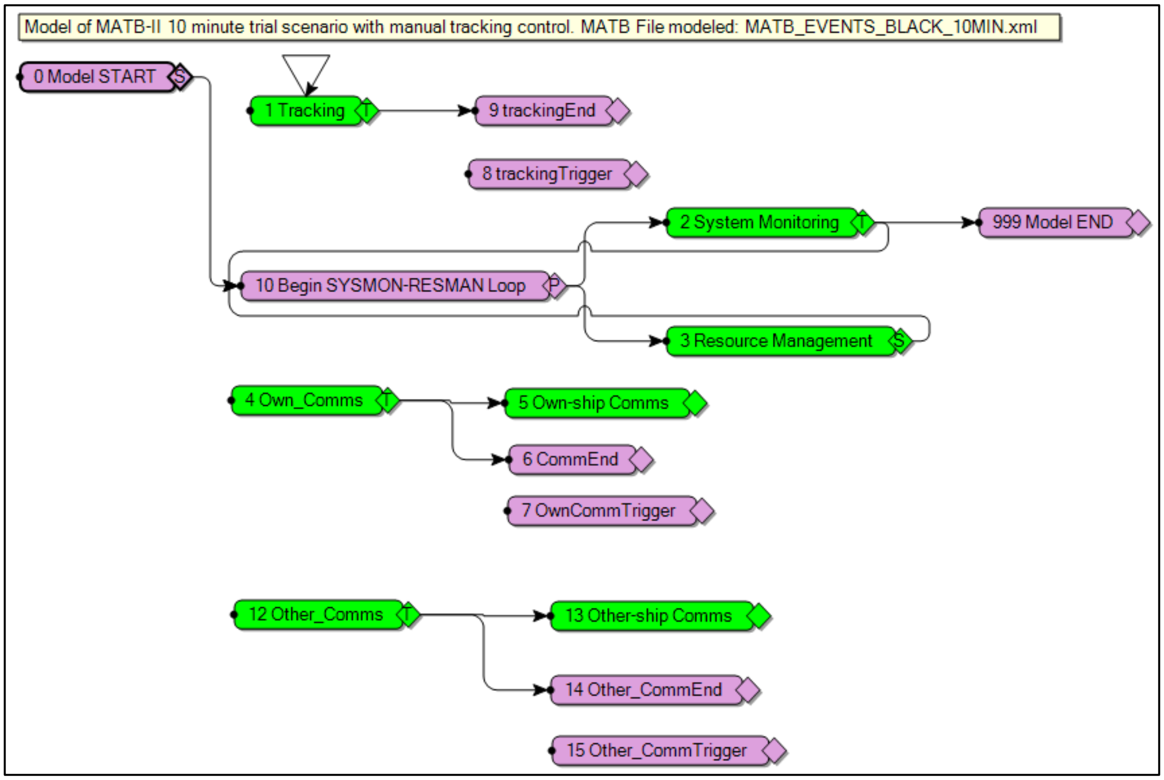


Figure 24. Study 1 task network diagram.

Mission Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Own		Value: 00:00:25.00	Root (Root)	4 Own_Comms
Comm02_Own		Value: 00:00:25.00	Root (Root)	4 Own_Comms
Comm03_Other		Value: 00:04:01.00	Root (Root)	12 Other_Comms
Comm04_Own		Value: 00:06:05.00	Root (Root)	4 Own_Comms
Comm05_Other		Value: 00:07:47.00	Root (Root)	12 Other_Comms
Comm06_Own		Value: 00:07:47.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:00:00.50	Root (Root)	1 Tracking

Mission Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Other		Value: 00:00:05.00	Root (Root)	12 Other_Comms
Comm02_Own		Value: 00:00:29.00	Root (Root)	4 Own_Comms
Comm03_Own		Value: 00:00:58.00	Root (Root)	4 Own_Comms
Comm04_Other		Value: 00:01:23.00	Root (Root)	12 Other_Comms
Comm05_Own		Value: 00:01:46.00	Root (Root)	4 Own_Comms
Comm06_Other		Value: 00:02:09.00	Root (Root)	12 Other_Comms
Comm07_Own		Value: 00:02:32.00	Root (Root)	4 Own_Comms
Comm08_Own		Value: 00:02:56.00	Root (Root)	4 Own_Comms
Comm09_Other		Value: 00:03:20.00	Root (Root)	12 Other_Comms
Comm10_Own		Value: 00:03:45.00	Root (Root)	4 Own_Comms
Comm11_Own		Value: 00:04:10.00	Root (Root)	4 Own_Comms
Comm12_Own		Value: 00:04:34.00	Root (Root)	4 Own_Comms
Comm13_Own		Value: 00:05:06.00	Root (Root)	4 Own_Comms
Comm14_Own		Value: 00:05:35.00	Root (Root)	4 Own_Comms
Comm16_Own		Value: 00:06:00.00	Root (Root)	4 Own_Comms
Comm17_Other		Value: 00:06:25.00	Root (Root)	12 Other_Comms
Comm17_Other1		Value: 00:06:52.00	Root (Root)	12 Other_Comms
Comm18_Own		Value: 00:07:15.00	Root (Root)	4 Own_Comms
Comm19_Own		Value: 00:07:39.00	Root (Root)	4 Own_Comms
Comm20_Other		Value: 00:08:02.00	Root (Root)	12 Other_Comms
Comm21_Own		Value: 00:08:27.00	Root (Root)	4 Own_Comms
Comm22_Own		Value: 00:08:52.00	Root (Root)	4 Own_Comms
Comm23_Other		Value: 00:09:15.00	Root (Root)	12 Other_Comms
Comm24_Own		Value: 00:09:37.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:00:00.50	Root (Root)	1 Tracking

Figure 25. Study 1 low (above) and high (below) workload external event matrices.

The resource workload demand values for each of the MATB-II tasks were chosen based on the options provided by IMPRINT. These cognitive workload value benchmarks provided for use in IMPRINT were calculated in a previous study that developed a model of cognitive workload in system design (Aldrich, Szabo, & Bierbaum, 1989). The researcher modeled the devices used to operate MATB-II by creating joystick, mouse, and speaker interfaces and assigning resource-interface demand relationships between them. The associated demand values available for selection with auditory, cognitive, fine motor, and visual resources in IMPRINT are depicted in Figure 26. Other demand values exist for gross motor control, speech, and tactile resources, but they were not applicable in modeling MATB-II.

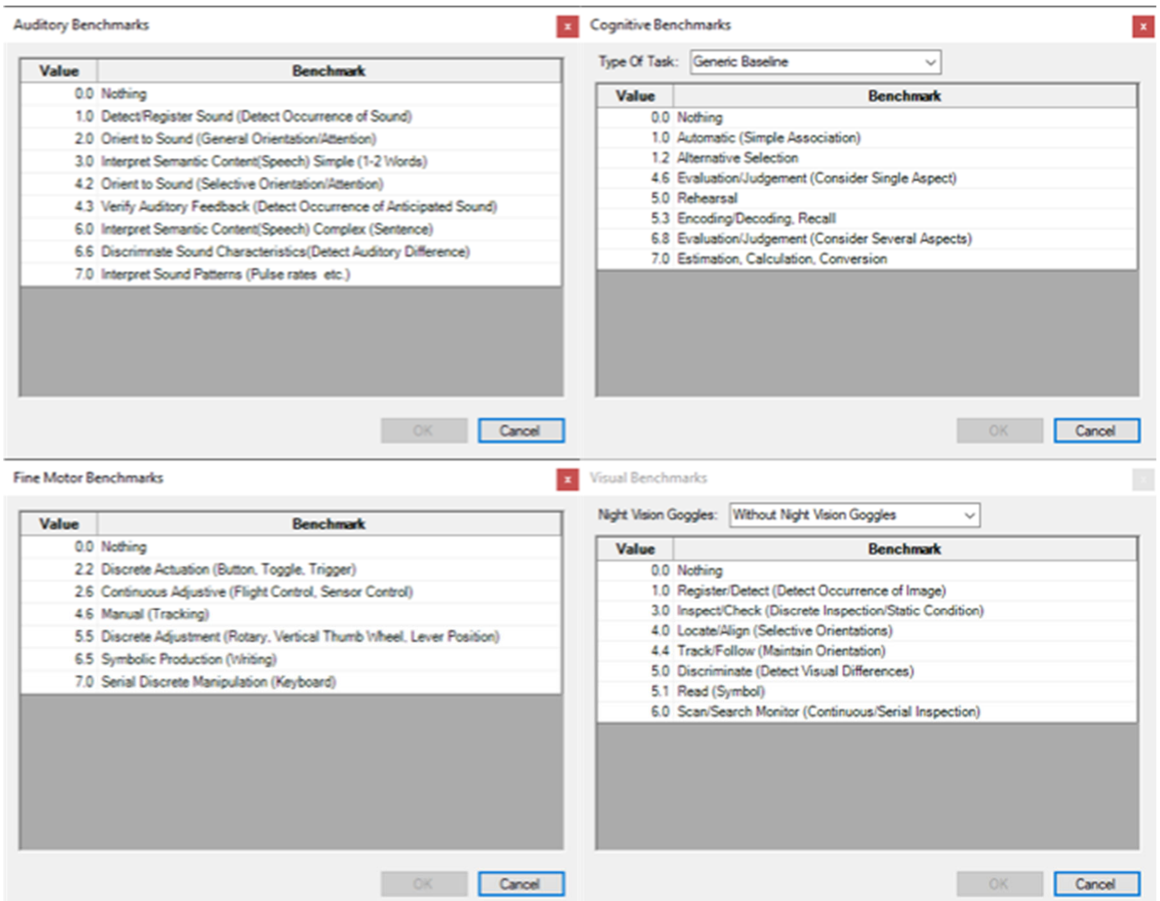


Figure 26. Selected IMPRINT workload demand benchmarks. Source: Alion Science (2018).

To assign workload levels for each of the MATB-II tasks, the researcher constructed multiple models using two approaches. In the first approach, the researcher assigned workload values to the MATB-II tasks using the default anchor values provided in IMPRINT that most closely fit the description of each task. These values are depicted in Table 4.

Table 4. Study 1 researcher-derived IMPRINT workload demand values.

Task: Tracking	**RI Pair Demand Values**						
Total Task Demand 10.00	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Joystick		1.00	4.60				4.40
Task: System Monitoring	**RI Pair Demand Values**						
Total Task Demand 8.40	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		1.20	2.20				5.00
Task: Resource Management	**RI Pair Demand Values**						
Total Task Demand 8.40	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		1.20	2.20				5.00
Task: Own Comms	**RI Pair Demand Values**						
Total Task Demand: 10.50	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		1.00	2.20				1.00
Interface: Speaker	4.30	1.00					1.00
Task: Other Comms	**RI Pair Demand Values**						
Total Task Demand: 6.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse							
Interface: Speaker	4.30	1.00					1.00

The researcher gathered workload rating feedback from three MATB expert operators through a cognitive walk-through in the second approach. These operators had logged 20–100 hours using an adapted version of MATB-II. Based on the results of their feedback, the researcher inputted the mean values for each of the MATB-II tasks. These values were used in both the low and high workload conditions as they are representative of the demand of each task. Of note, these expert users used an adapted version MATB that was based on NASA's instance. The experts used buttons on the joystick to control the SYSMON task, had only own ship calls in the COMM task, and only operated the TRACK task in manual mode. Additionally, the expert user group did not receive a standardized MATB training progression as was used in the current studies. Instead, they accomplished the tasks through individual strategy development based on their understanding of the task as they gained experience. Because of the differences in training and prioritization of tasks, expert feedback was used to assess task difficulty rather than sequencing and task duration. This approach allowed for default anchor values derived by the researcher and expert provided values to be input into the task network diagram that reflected an equal weighting to each of the four tasks. A consolidation measure was calculated for each resource-interface pair by taking the mean of the expert ratings. The consolidated expert-derived values are provided in Table 5.

Table 5. Study 1 expert-derived IMPRINT workload demand values.

Task: Tracking	**RI Pair Demand Values**						
Total Task Demand 11.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Joystick		3.60	3.90				3.80
Task: System Monitoring	**RI Pair Demand Values**						
Total Task Demand 7.87	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		2.40	3.30				2.17
Task: Resource Management	**RI Pair Demand Values**						
Total Task Demand 13.37	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		5.80	3.57				4.00
Task: Own Comms	**RI Pair Demand Values**						
Total Task Demand: 16.87	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		2.07	4.07				1.83
Interface: Speaker	5.00	2.07					1.83
Task: Other Comms	**RI Pair Demand Values**						
Total Task Demand: 6.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse							
Interface: Speaker	4.30	1.00					1.00

To differentiate between low and high workload IMPRINT models, the researcher created event triggers in IMPRINT to initiate the “Own_Comms” and “Other_Comms” communications loops. This meant that there were six communications events in the low workload condition and 24 in the high workload condition. This approach was utilized to support the overall effort’s aim to assess IMPRINT’s ability to provide cognitive workload predictions when using adaptive automation.

The IMPRINT-predicted time average workload was calculated by multiplying the operator’s predicted workload by the time elapsed between the discrete interval model events. This calculation yielded a time weighted workload value for each discrete interval. The time weighted workload values were then summed for the whole model and then divided by the total time of the simulation. The total researcher-derived IMPRINT predicted mean workload for the low workload condition was 38.48 and 41.05 for the high workload condition. The expert-derived IMPRINT predicted mean workload values for the low and high workload conditions were 34.13 and 41.15, respectively. Additionally, both the researcher and expert-derived high workload condition models had increased numbers of workload spikes as seen in Figures 27 and 28. These spikes in the high workload trial model are intuitive as there were more communications tasks introduced. The increase in communications tasks created more resource demand as audio cues were coupled with visual attention requirements. These requirements created a time-sharing conflict of the visual processing channel explained by Wickens' (1981) MRT whereby visual attention for the other tasks were interrupted by having to attend to the communications task. The drops in workload represent communications tasks for other callsigns. These dips in workload again highlight MRT in that resources can be shared across modalities if they are used to process information separately.

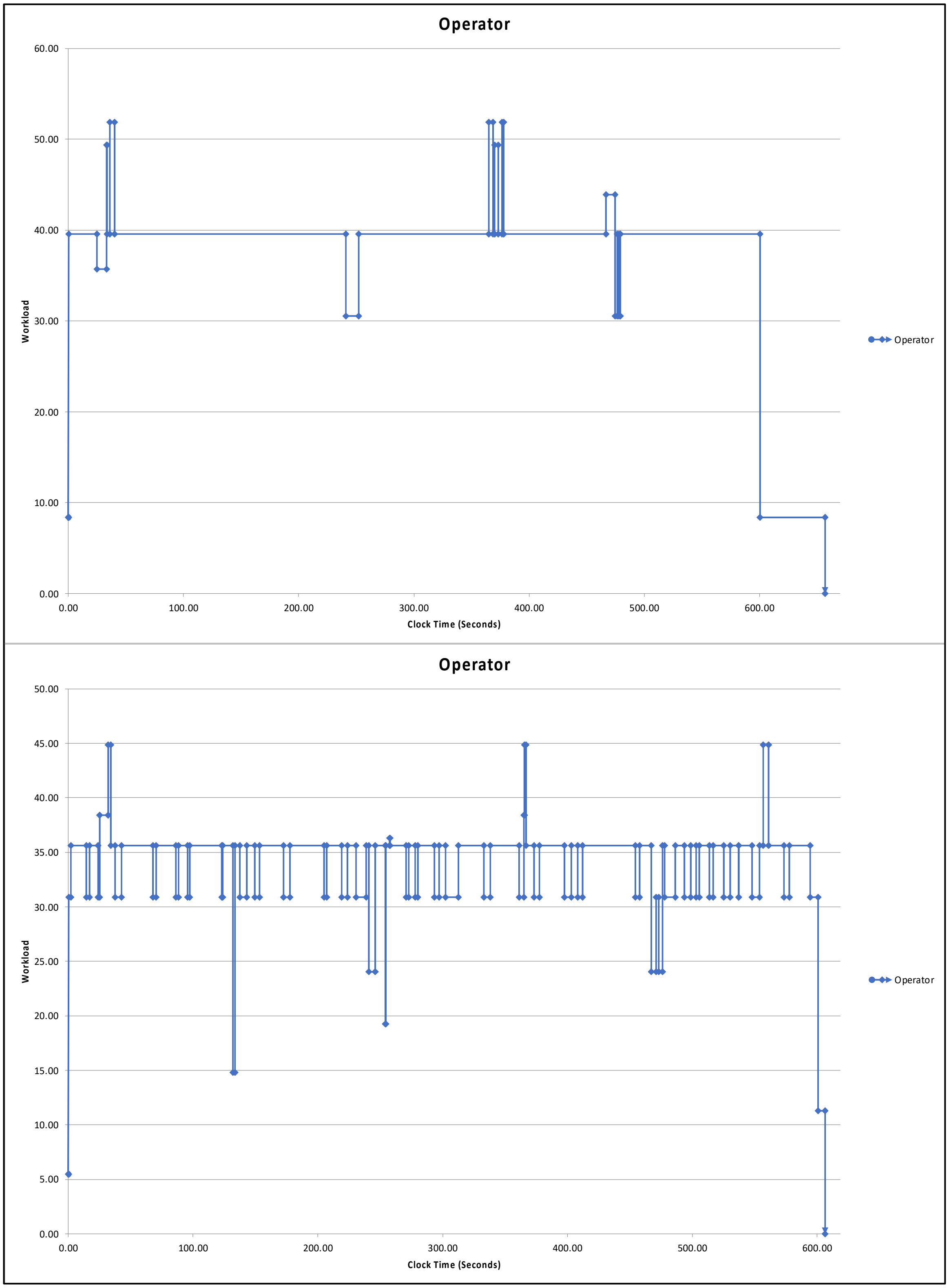
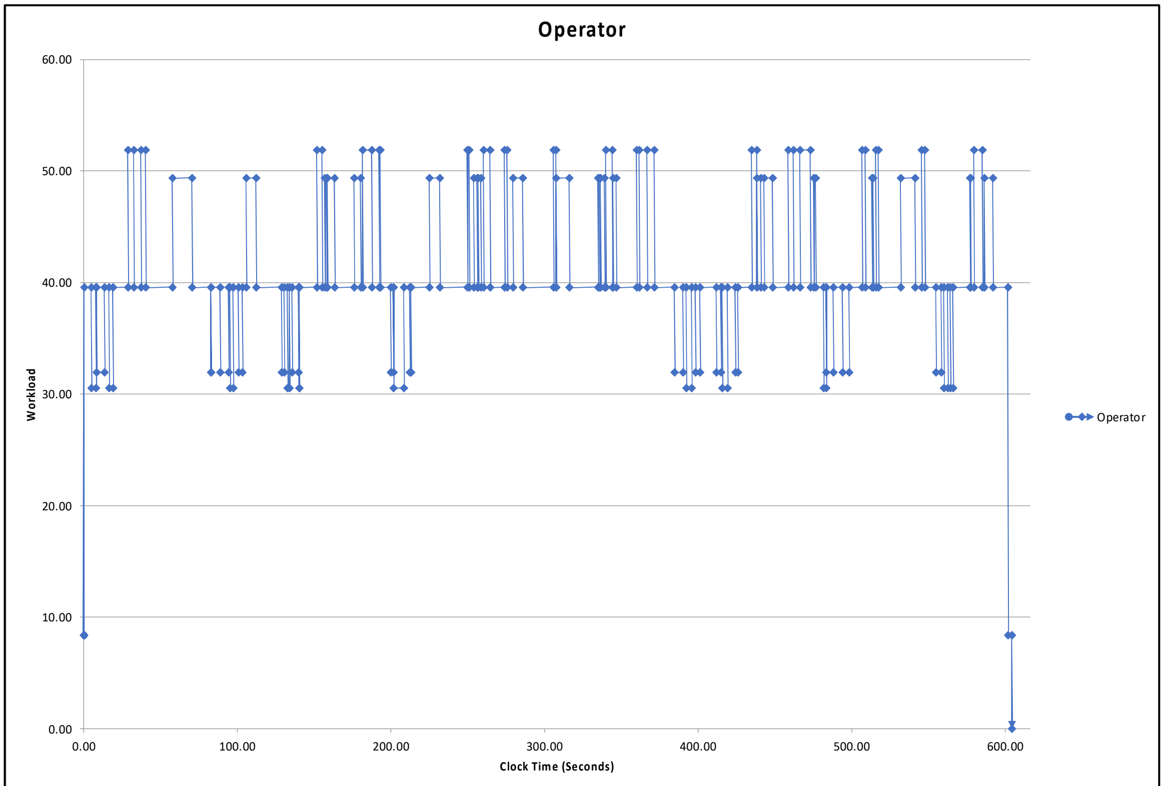


Figure 27. Study 1 Researcher (above) and expert (below) derived IMPRINT model graph for the low workload condition.



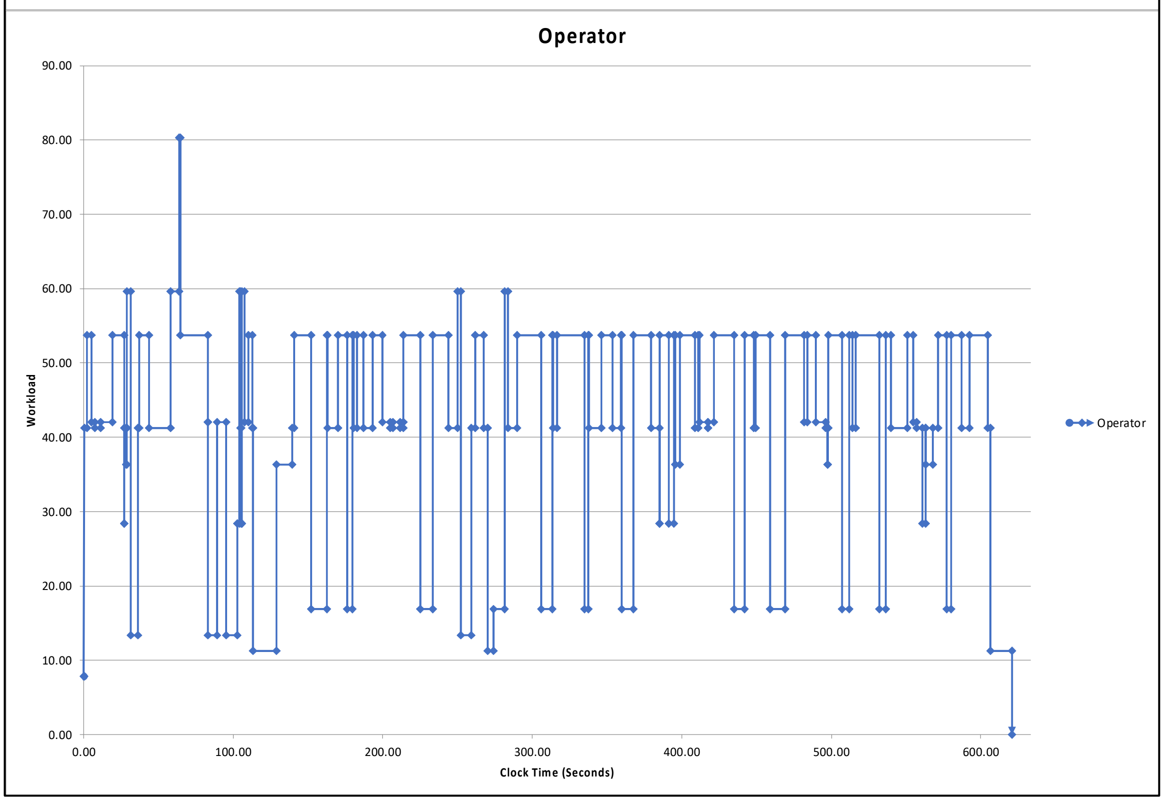


Figure 28. Study 1 Researcher (above) and expert (below) derived IMPRINT model graph for the high workload condition.

G. ANALYTICAL APPROACH

Psychophysiological data collected during the study were consolidated in LSL's Extensible Data Format (.xdf) files. This synchronized format allowed for data extracting, processing, and analysis using a collaboratively developed MATLAB tool (Vogl, 2022). An additional Neurokit2 script was used and collaboratively adapted for HRV data processing (Makowski et al., 2021; O'Brien, 2022). Timestamps from the Neurokit2 script provided the basis for pupil diameter data processing from Pupil Labs' Pupil Player software. A separate process for processing the fNIRS data was conducted with NIRX's nirsLab analysis software. Subjective measures were consolidated from the CSWAG, NASA-TLX, and SART instruments. Additionally, data from MATB-II were available for analysis, with the composite FOM score used as an overall assessment of participants' performance during their trials.

Pupil data were processed using Pupil Player version 3.5.7. Participant data files were loaded into Pupil Player and were processed for the duration of each trial run. The minimum data confidence was set to the default value of 0.60. The use of Pupil Player's data processing was leveraged to export refined pupil data, specifically pupil diameter, for further analysis. The resulting data provided pupil diameter for each eye along with confidence levels in those measurements.

The fNIRS data were processed and filtered using nirsLAB version 2019.04 (Xu et al., 2017). Participant data files were loaded with probe settings automatically applied from the initial NIRSTAR setup and calibration from each session. Event markers were applied to note the start, stop, and communications events during the trial runs. Data pre-processing procedures included truncating all participants sessions to the 60 seconds before and after the trial runs. Detector saturation intervals were interpolated to the maximum of four frames or 0.51 seconds to account for noise artifacts in the collected data channels. Data channels that exceeded the four-frame saturation interval were excluded from analysis. Data qualities were checked using the default gain setting of eight. This default gain setting provided a factor to amplify the light-produced photocurrent of the NIRSport system. Channels that exceeded this gain setting were excluded from analysis. The data from the excluded channels were not included in the analysis of PFC blood oxygenation. Data that fell outside of five standard deviations were removed as part of pre-processing as well. Spike artifacts that represented data loss during trial runs were removed and interpolated by nirsLAB. These artifacts were calculated using the data immediately before and after the signal loss. Finally, a frequency filter was applied using a low cutoff frequency of .01 Hz and high cutoff frequency of 0.2. The current study used Gratzer's spectrum for application of the Beer-Lambert Law to generate the HbO and Hb levels (Prahl, 1998). Finally, hemodynamic states were computed with these given parameters and yielded oxygenated, de-oxygenated, and total oxygenated hemoglobin levels in the PFC for each participant.

Following data processing procedures, data were analyzed using statistical methods. A mixed-effects model analysis approach was used to analyze the collected measures. For both workload conditions, dependent measures were analyzed against the fixed effects of experience level and workload level. Random effects were modeled using each participant with their experience level nested. All statistical tests were conducted using JMP version 16.0.0.

H. RESULTS

A summary of the results for Study 1 is shown in Table 6 (with presentation order included in the analytical model). Because presentation order was not found to be statistically different across measures, it was removed from the mixed-effects model to determine the relationships across measures and conditions. The resulting differences followed the same patterns of statistical significance across variables as seen in Table 7.

Participant data were excluded due to extreme values after analyzing residual plots for each modeled measure. These exclusions are listed in the notes below Tables 6 and 7. Of note, the removal of these extreme values did not change the pattern of statistical significance in the results of Study 1.

Table 6. Study 1 summary results with presentation order included.

Measure Category	Measure Type	Dependent Measure	Novice vs. Experienced	High vs. Low Workload	Presentation Order
Performance	**MATB-II Figure of Merit**	**Composite FOM****	Experienced group -> higher FOM $p = .002$*	***High workload -> lower FOM*** $\boldsymbol{p < .001}$*	$p = .934$
Psychophysiological	**HRV**	**Mean HRV*****	$p = .928$	***High workload -> lower HRV*** $\boldsymbol{p = .008}$*	$p = .184$
	fNIRS	**Mean HbO*****	$p = .057$	$p = .208$	$p = .343$
		Mean Hb****	$p = .373$	$p = .411$	$p = .215$
		Mean Total Hb***	$p = .325$	$p = .814$	$p = .122$
	Pupil***	**Mean Right Pupil Diameter**	$p = .477$	***High workload -> larger pupil diameter*** $\boldsymbol{p = .024}$*	***First condition -> larger pupil diameter*** $\boldsymbol{p < .001}$*
		Mean Left Pupil Diameter	$p = .158$	***High workload -> larger pupil diameter*** $\boldsymbol{p < .001}$*	***First condition -> larger pupil diameter*** $\boldsymbol{p < .001}$*
Subjective Workload	**Continuous Subjective Workload Assessment**	**Mean CSWAG**	$p = .859$	***High workload -> higher CSWAG***	$p = .416$

Measure Category	**Measure Type**	**Dependent Measure**	**Novice vs. Experienced**	**High vs. Low Workload**	**Presentation Order**
	Graph			$\boldsymbol{p < .001*}$	
	NASA-TLX	**NASA-TLX Rating**	$p = .967$	$p = .143$	N/A
Situation Awareness	**Situation Awareness Rating Technique**	**SART Rating**	$p = .799$	$p = .311$	N/A

Table 7. Study 1 summary results with presentation order omitted.

Measure Category	**Measure Type**	**Dependent Measure**	**Novice vs. Experienced**	**High vs. Low Workload**
Performance	**MATB-II Figure of Merit**	**Composite FOM****	***Experienced group -> higher FOM*** $\boldsymbol{p = .002*}$	***High workload -> lower FOM*** $\boldsymbol{p < .001*}$
Psychophysiological	**HRV**	**Mean HRV**	$p = .928$	***High workload -> lower HRV*** $\boldsymbol{p = .009*}$
	fNIRS	**Mean HbO*****	$p = .057$	$p = .199$
		Mean Hb****	$p = .820$	$p = .535$
		Mean Total Hb***	$p = .325$	$p = .845$
	Pupil***	**Mean Right Pupil Diameter**	$p = .477$	$p = .106$
		Mean Left Pupil Diameter	$p = .158$	***High workload -> larger pupil***

Measure Category	Measure Type	Dependent Measure	Novice vs. Experienced	High vs. Low Workload
				diameter $\boldsymbol{p = .002*}$
Subjective Workload	**Continuous Subjective Workload Assessment Graph**	**Mean CSWAG**	$p = .859$	***High workload -> higher CSWAG*** $\boldsymbol{p < .001*}$
	NASA-TLX	**NASA-TLX Rating**	$p = .967$	$p = .143$
Situation Awareness	**Situation Awareness Rating Technique**	**SART Rating**	$p = .799$	$p = .311$

Tables 5 and 6 Notes:

* p<.05

** 3 Novice participants excluded due to extreme results

*** 1 Novice participant excluded due to extreme results

**** 2 Novice and 1 Experienced participant excluded due to extreme results

Figure 29 shows that experienced participants scored higher FOMs than the novice participants (M=94.46, SD=2.46, SE=0.39 vs. M=92.33, SD=2.92, SE=0.50). Participants' mean FOMs were higher in the low workload condition than the high workload condition (M=94.46, SD=2.55, SE=0.42 vs. M=92.50, SD=2.86, SE=0.47).

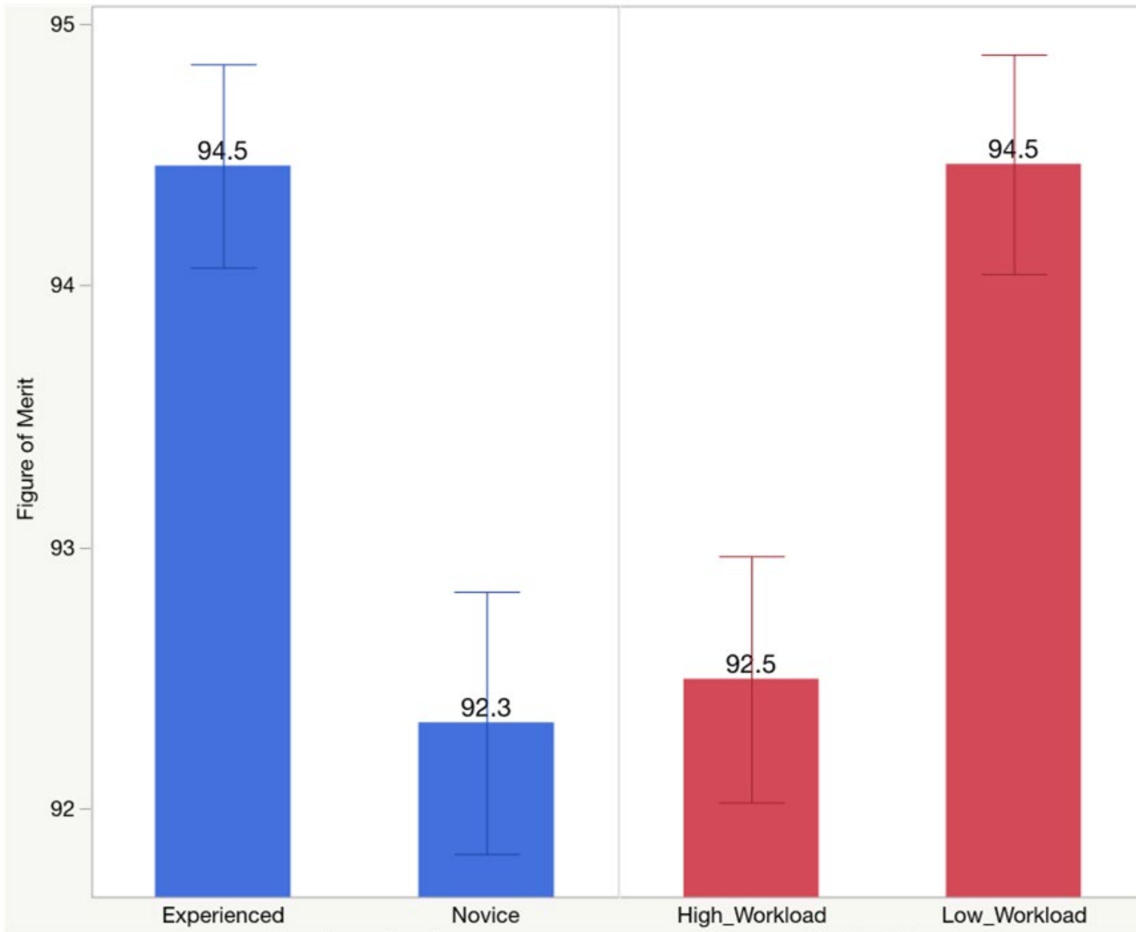


Figure 29. Study 1 Mean FOM vs. experience level and workload level. Error bars denote the standard error.

Figure 30 shows that novice and experienced participants had lower HRV in the high workload condition than the low workload condition (M=808.99 ms, SD=134.47, SE=21.58 vs. M=822.45ms, SD=136.22, SE=21.54). There were no statistically significant differences between novice and experienced participants' mean HRV.

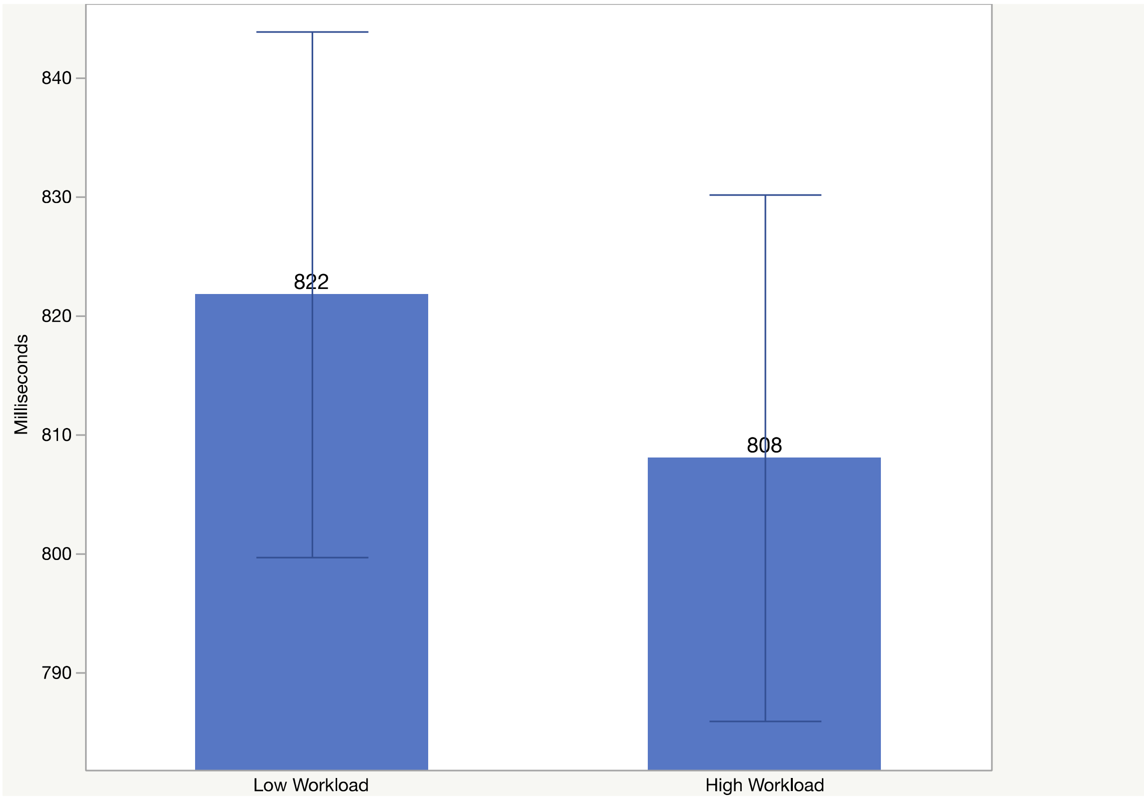


Figure 30. Mean HRV in milliseconds for Study 1 by workload condition. Error bars denote the standard error.

Participants had larger left eye pupil diameters in the high workload condition than the low workload condition $F(1, 115) = 8.41,\ p < .01$ (M=3.45mm, SD=0.96, SE=0.11 vs. M=3.40mm, SD=0.95, SE=0.11). There were no statistically significant differences in right eye pupil diameter between workload conditions. These results are depicted in Figure 31. There were no statistically significant differences in pupil diameter between experience groups in either pupil for any condition.

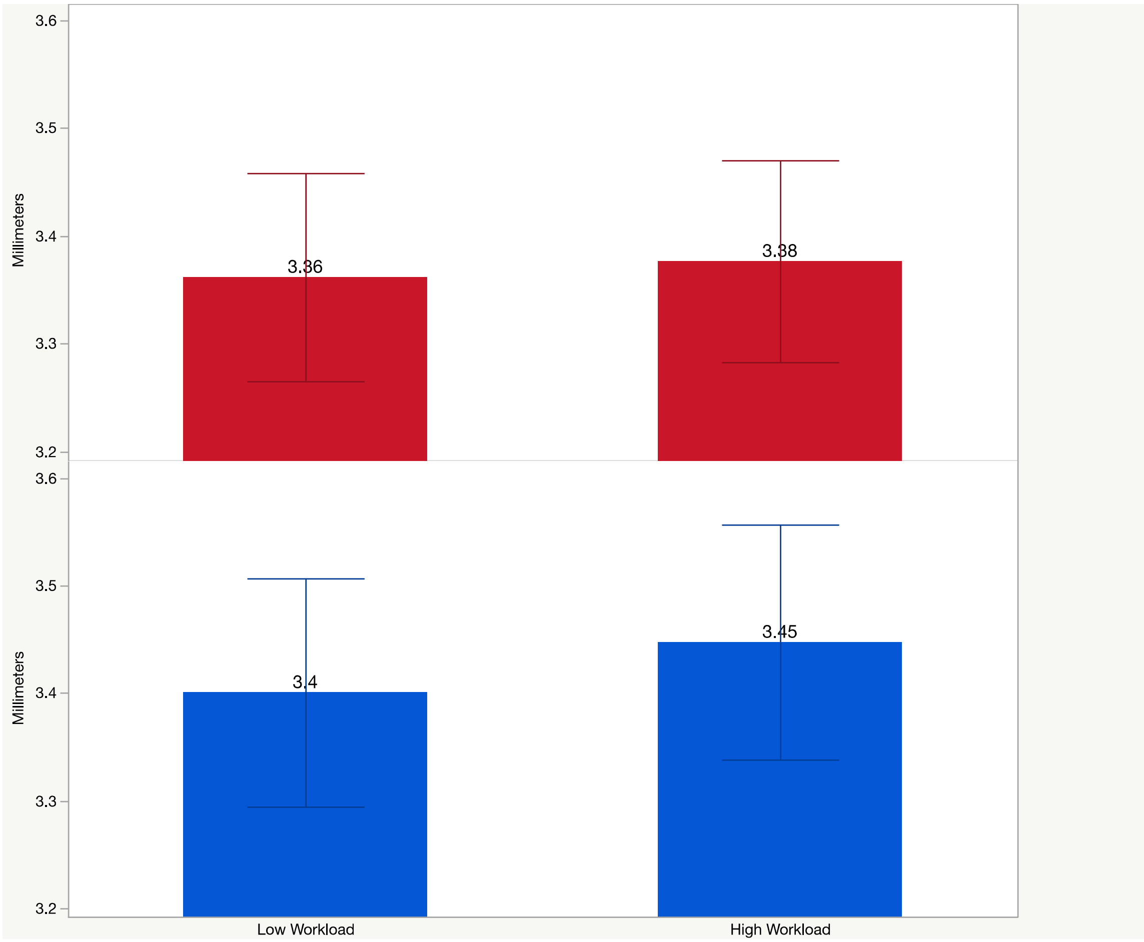


Figure 31. Study 1 mean right (top) and left (bottom) pupil diameters in millimeters by workload condition. Error bars denote the standard error.

However, there was a statistically significant difference in all participants' right pupil diameters five seconds after a communications event compared to the diameters five seconds preceding the communications events, $F(1, 115) = 16.46,\ p < .001$ (M=3.40, SD=0.85, SE=0.10 vs. M=3.34, SD=0.83, SE=0.09). The same pattern was also seen in all participants' left pupils, $F(1, 115) = 19.31,\ p < .001$ (M=3.46, SD=0.97, SE=0.11 vs. M=3.38, SD=0.93, SE=0.11). These results are shown in Figure 32.

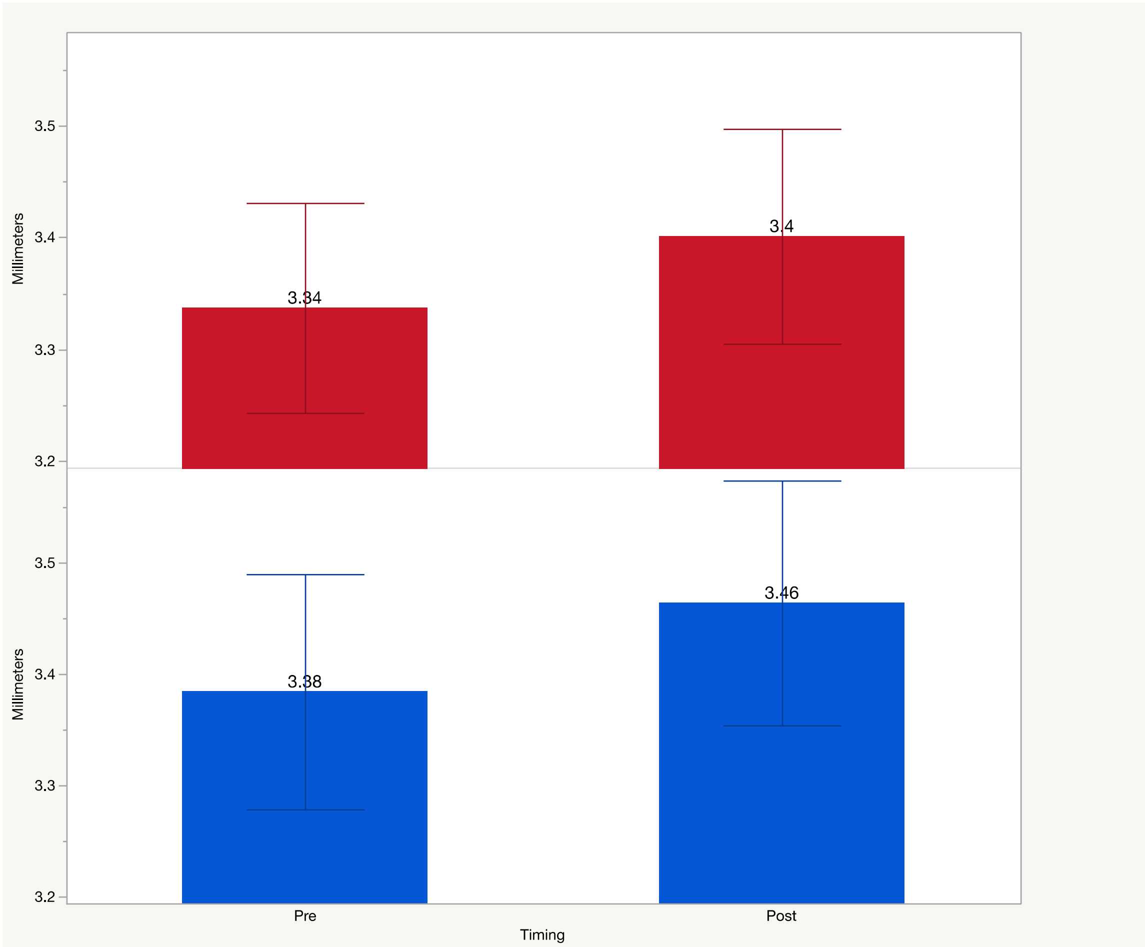


Figure 32. Study 1 mean right (top) and left (bottom) pupil diameters by all communications event timings. Error bars denote the standard error.

Further, participants' pupils were larger five seconds after a communications event directed at their ship (NASA 504) and other ships. Own ship communications resulted in statistically significant differences in right pupil diameter $F(1, 115) = 19.37$, $p < .001$ (M=3.41, SD=0.85, SE=0.10 vs. M=3.35, SD=0.84, SE=0.10) and left pupil diameter $F(1, 115) = 16.19$, $p < .001$ (M=3.45, SD=0.95, SE=0.11 vs. M=3.38, SD=0.93, SE=0.11). Other ship communications followed the same pattern of statistical significant differences with the right pupil diameter $(1, 115) = 5.67$, $p < .05$ (M=3.37, SD=0.82, SE=0.09 vs. M=3.31, SD=0.81, SE=0.09) and left pupil diameter $F(1, 115) = 17.03$, $p < .001$ (M=3.44, SD=0.95, SE=0.11 vs. M=3.37, SD=0.93, SE=0.11). These results are displayed in Figure 33.

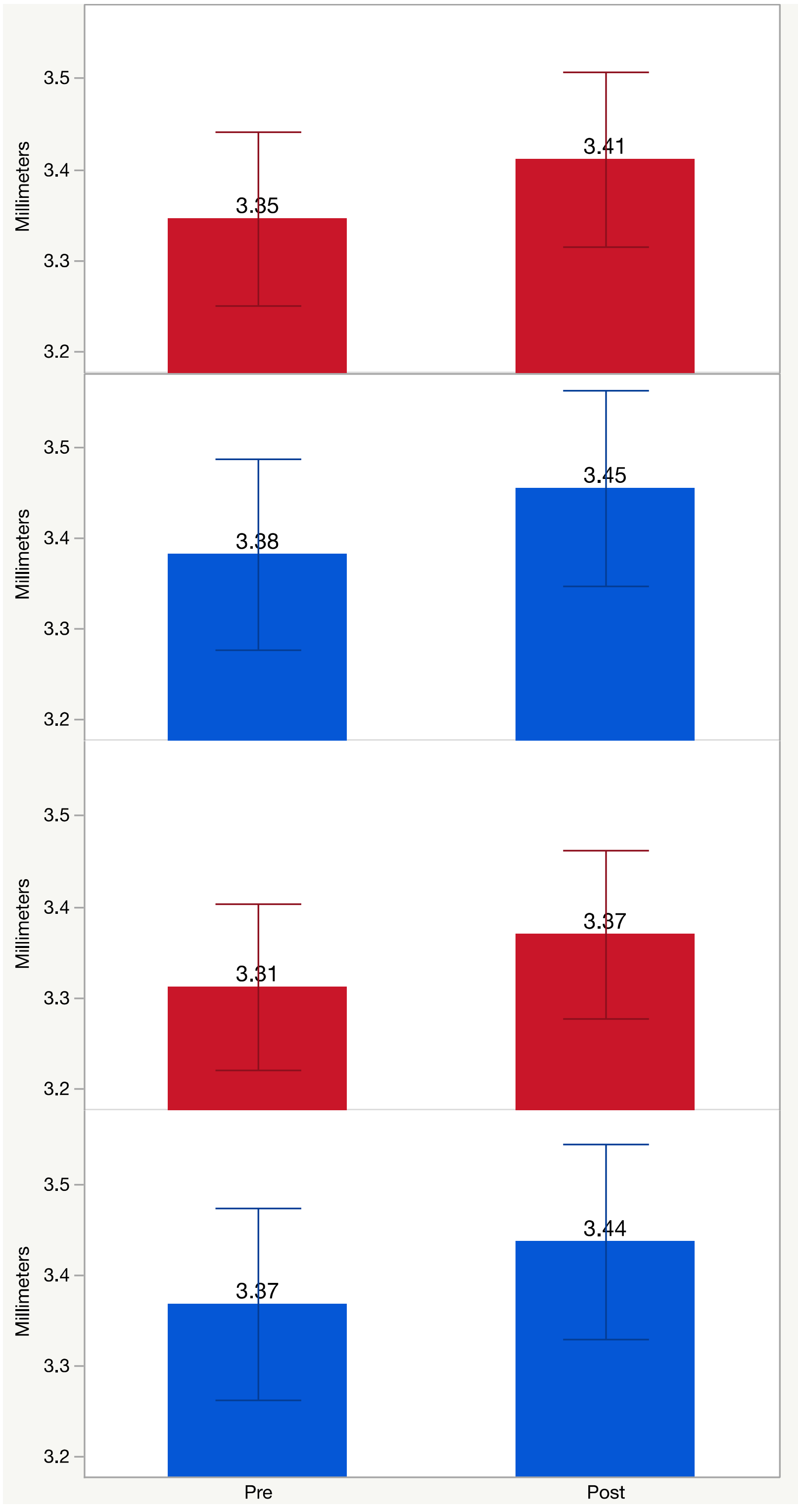


Figure 33. Study 1 mean right (red) and left (blue) pupil diameter in millimeters by all communications event timings. Error bars denote the standard error.

Differences between own and other ship communications were not statistically significant as shown in Figure 34. Two additional novice participants were excluded from this portion of the analysis due to extreme results as analyzed in the data's residual plots.

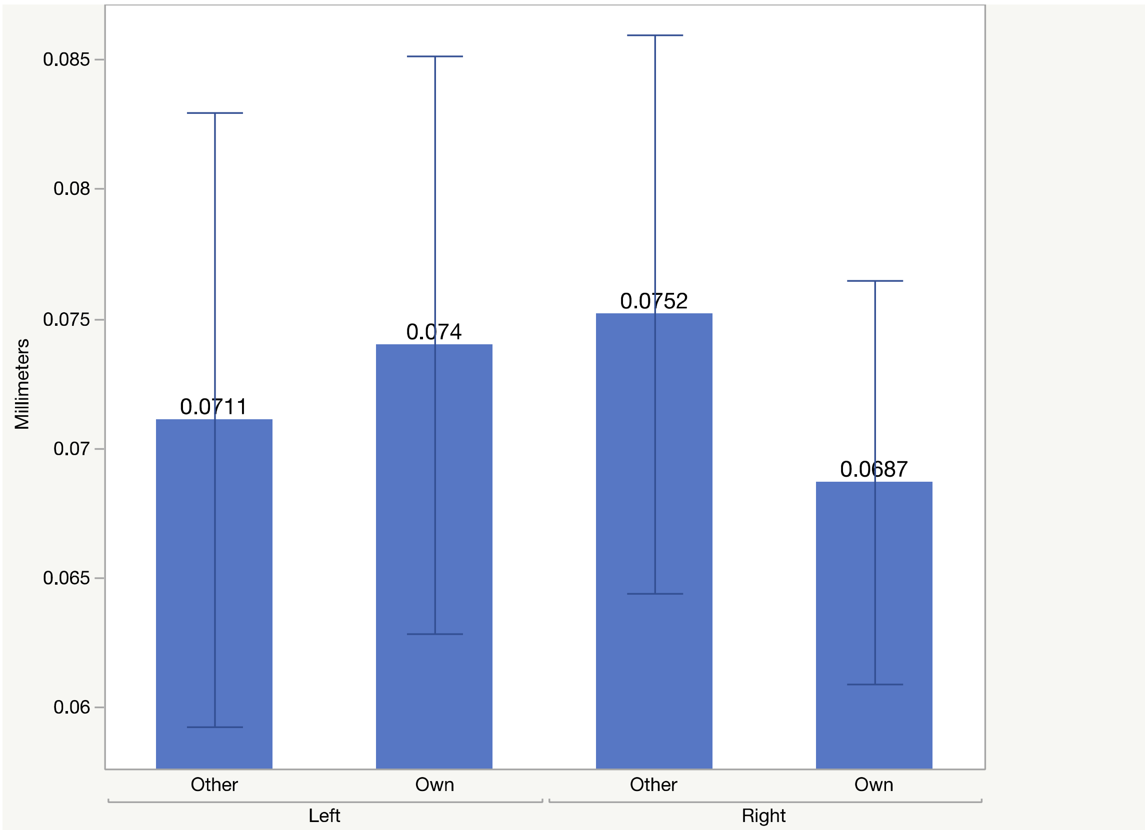


Figure 34. Study 1 mean pupil diameter differences in. millimeters between eyes and communications target. Error bars denote the standard error.

The relationship between subjectively assessed cognitive workload and SA was conducted using post-trial questionnaire assessment. Correlation analysis was conducted to identify associations between NASA-TLX and SART ratings. There were no statistically significant correlations between those two ratings (p=0.57 for Spearman's ρ).

I. DISCUSSION

The following discussion pertains to the results from Study 1. The hypotheses for Study 1 are reviewed, followed by analysis on their impacts on Studies 2 and 3. This section also discusses how Study 1 addressed the aspects of the MLCC that were identified in Figure 17.

1. Research Questions and Hypotheses

Research Question 1: IMPRINT workload predictions correlate to two of the three main psychophysiological tools used in the study, HRV and pupil diameter. Of the two subjective workload measures collected, the measure that correlated with the workload predictions was the CSWAG. This finding suggests that more continuous gauging of an operator's subjective workload is more sensitive to real-time workload changes. While the NASA-TLX has been used effectively in numerous studies, its application to this design was not associated with cognitive workload.

Ha1: Cognitive workload predictions are directly correlated with objective and subjective workload measures.

The first hypothesis that cognitive workload predictions were directly correlated with objective and subjective workload measures was partially supported. Statistical tests that examined the differences between the two workload conditions indicated that higher levels of workload were experienced in the high workload condition. This finding followed the IMPRINT cognitive workload prediction models that yielded higher predictive values in the high workload condition than the low workload condition.

Previous studies have investigated mean pupil diameter within a 12-second task window during an arithmetic and search task (Chen & Epps, 2014). The current study used a 10-second task window for the communications task to allow for analysis. Additionally, Kruger, Hefer, & Matthew (2013) suggested investigating pupil diameter at least two seconds after the initiation of a stimulus. Both pupil diameters were analyzed in the five seconds before and after the initiation of a communications to allow for setting a baseline period and measurement period in accordance with these previous studies. The larger pupil sizes present five seconds after the initiation of any radio communication served to validate the workload spikes seen in the IMPRINT models. These spikes were hypothesized to be caused by resource demand conflicts present with the introduction of communications tasks. There was no difference between pupil diameters during own and other ship radio communications, however. This non-finding suggests that participants experienced the auditory resource conflict in similar fashion for both own and other ship communications. Further, participants strategies to handle the different communications may have led to differences in pupil diameter being seen at another time interval outside of 5 seconds before and after an event.

Objective cognitive workload surrogate measures that yielded statistically significant results between workload conditions were mean HRV and left eye pupil diameter. Participants' mean HRV data were lower for the high workload condition than the low workload condition. Additionally, their mean left pupil diameters were larger for the higher workload condition. These findings follow previous research that suggest lower HRV and larger pupil diameters are indicators of higher cognitive workload (Aura et al., 2021; Hughes et al., 2019; Steinhauer et al., 2022; Vogl et al., 2020).

The NASA-TLX was administered after the trial. The timing of its administration may have potentially confounded participants' responses. Participants had to recall two separate trials as one. Additionally, participants ended their trials in different conditions, potentially creating a recency effect on their responses (e.g., lower reported NASA-TLX ratings when ending in the low workload condition versus ending in the high workload condition). These results support the approach of using IMPRINT models and multiple measures to validate the cognitive workload predictions. These findings suggest that certain measures might be more sensitive to cognitive workload changes than others in similar multi-attribute tasks.

Pre-frontal cortex blood oxygenation from fNIRS was not associated with increased workload. While fNIRS measures changes to blood oxygen, it is slower in measuring changes than other measures of brain activity such as EEG. The MATB-II scenarios may have introduced changes that were too fast to identify changes in PFC blood oxygenation using the fNIRS capturing approach (Girouard et al., 2010).

Ha2: There is a significant difference between cognitive workload measures in the low and high workload conditions.

The second hypothesis that there was a significant difference between cognitive workload measures in the low and high workload conditions was partially supported through mean HRV, left pupil diameter, and CSWAG scores. However, fNIRS and NASA-TLX results did not show any statistically significant differences between the two workload conditions. These non-findings can again be explained by the fast-moving pace of the trials that are not conducive to measurement through fNIRS. Additionally, the post-trial rating of the NASA-TLX had participants rely on their recollection of two separate trials while trying to assess their time in the experiment as one trial. This potentially introduced an issue in the administration of the NASA-TLX. However, this approach was used to try to gain understanding of cognitive workload from the NASA-TLX ratings. The NASA-TLX was not administered during the break for eye recalibration due to the time it would have taken to complete the survey. This time would have created a longer gap in the trial run and not allowed for any analysis on potential transition effects.

This study assessed cognitive workload surrogate measures in two workload conditions across two experience levels. Results from the mixed model supported the assertion that cognitive workload measures differed between the workload conditions. This finding allowed for building Study 2 using low- and high-level classifications of workload in the same manner as Study 1. The difference in FOMs between workload conditions helps validate the initial framework used to determine the number of tasks to be included in the experimental design of the MATB-II tasks (McCurry et al., 2022). This finding also supports previous research that found increases in cognitive workload when conflicts arose in resources demanded through multiple information processing channels (Longo et al., 2022; Wickens, 1981, 2002, 2008a).

Ha3: There is a significant difference between the novice and experienced participant groups' performance and workload measures.

The third hypothesis that there was a significant difference between the novice and experienced participant groups' performance and workload measures was supported with performance scores only. In particular, the FOMs between the novice and experienced groups showed differences such that experienced participants scored higher than their novice counterparts. There were no statistically significant differences between novice and experienced participants for any of the cognitive workload measures collected. This non-finding suggests that the training design with MATB-II in the current study was not able to train different experience groups. However, this training approach was important to try to mirror real-world training approaches to determine the effects of experience on cognitive workload. While this design did not find any significant differences between experience groups' cognitive workload measures, this line of investigation is worthy in future studies to assess cognitive workload impacts at different levels of training. For instance, the effects of experience on cognitive workload throughout a training progression from novice to expert have demonstrated that novices put forth more effort than experts (Fairclough et al., 2005).

The results from Study 1 indicated that performance was significantly different between both novices and experienced participants, as well as between the high and low workload conditions. Experienced participants achieved higher composite FOMs than the novice participants. This finding indicates that the differences between the groups was significant and that the training approaches yielded different levels of performance. This difference would serve to inform the design of Study 3 by leveraging the experienced group training approach as the baseline for all participants. The FOMs were also higher for the low workload condition than the high workload condition, suggesting that lower workload allowed for higher performance across participants.

Research Question 2. Cognitive workload and SA were assessed using results from the NASA-TLX and the SART questionnaires. There was no significant relationship between the two ratings. While both instruments have been used in previous studies, the experimental design in this study was not sensitive to these measures.

Ha4: Cognitive workload and situation awareness are inversely related.

We failed to reject the null hypothesis that cognitive workload and situation awareness are related. Research has demonstrated multiple possible relationships between cognitive workload and SA (Endsley, 2021; Kaber & Endsley, 2004; Wickens, 2008a). The psychophysiological data showed significant differences between the workload conditions. Subjective workload assessments differed in their results. The results from the CSWAG were different between workload conditions, whereas the results from the NASA-TLX were not. This difference may be explained by the CSWAG being collected every minute, providing a more real-time assessment of the participant's workload. The researchers administered the NASA-TLX at the conclusion of the whole experimental trial to mitigate disruptions to the data collection and minimize confounding time increases between scenarios.

Situation awareness was measured using the SART. However, the SART data did not show any significant differences between groups or workload conditions. The SART was administered in the same manner as the NASA-TLX, possibly not allowing participants to properly gauge their SA during each of their two trial runs. Participants may not have been able to accurately gauge their SA beyond Level 1 given the nature of the MATB-II tasks. Additionally, participants' perceptions of what was new and understood varied when completing the SART. These factors highlight some of the considerations when using the SART for similar tasks. SART results provide self-assessed SA ratings. Self-assessed SA approaches might provide a participant's confidence in their SA rather than actual SA (Endsley, 1995). The basis to gauge SA was also difficult in this design. Participants were not able to ascertain a ground truth to base their SA assessment, further highlighting issues associated with self-assessed SA approaches.

2. Findings Informing Study 2

Study 2's design was informed partially based on the findings of Study 1. Study 1 provided validation that the design of the two MATB-II scenarios yielded two different workload conditions as determined by the differences in performance. This finding served as the basis for building Study 2's MATB-II scenarios with automation being introduced for an additional level of analysis. The use of objective and subjective measures in Study 1 allowed for analysis of their associations in two different workload conditions. The same cognitive workload surrogate measures were used in Study 2 to further examine the impact of automation on those measures. The training progression and delineation of experience groups addressed concerns with learning effects that had been known to exist with multi-tasking batteries in previous research (Kong et al., 2022). Across all participants and workload conditions, the CSWAG appeared sensitive to subjective cognitive workload assessment.

While fNIRS and post-trial subjective assessments did not yield any statistically significant results between experience groups or workload conditions, their use was continued in Study 2. Their inclusion in Study 2 served to investigate the potential impacts on cognitive workload of introducing automation in the TRACK task. The design of Study 1 informed Study 2's design by replicating its data collection strategy to investigate cognitive workload with different LOAs present in the two 10-minute trials.

3. MLCC Review

Study 1 sought to investigate three areas of the adapted MLCC framework: perception, SA and response, and workload. Results from Study 1 indicate that the human operator was able to perceive differences in experienced workload both objectively and subjectively. This finding informed an operator's responses as indicated by their performance metrics. Cognitive workload measures were shown to be different between workload conditions, suggesting that the cognitive workload aspect of the adapted MLCC continuously impacted the remaining elements in the framework in an iterative fashion.

These results paved the way to examine other portions of the MLCC framework in Study 2. Given that cognitive workload and performance data indicated patterns and results for a manually completed scenario, the examination of the MLCC framework could then be expanded to investigate the impact of introducing a higher level of automation. The inclusion of a high and low level of automation would facilitate the exploration of the unintended negative consequences of adaptive automation in the human-system loop as seen in the adapted MLCC framework. Additionally, the use of IMPRINT modelling provided a basis of comparison for the different experimental conditions that were validated by the performance and workload measurements collected.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. STUDY 2

A. OVERVIEW

The second experiment focused on investigating the effect of different levels of automation on objective and subjective measures of workload. Additional analysis of SA and performance was conducted as well. Study 2 leveraged MATB-II again. However, participants conducted trial runs with automated tracking assistance. Figure 35 maps the areas investigated in this study to the adapted MLCC framework.

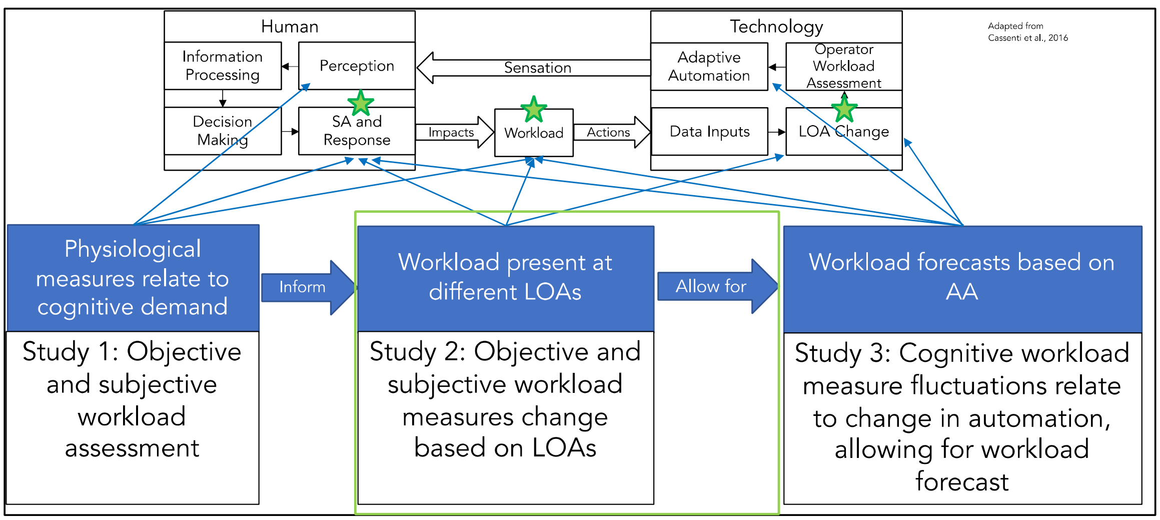


Figure 35. Study 2 mapping to the adapted MLCC framework.

The purpose of this study was to determine levels of workload present at a low (or highly manual) and high LOA (or highly automated level). The interval of introducing changes to the automation were time-based, occurring in two 10-minute increments. This interval was used in both Studies 1 and 2 because it was a factor of the five minute MATB-II task-to-workload mapping used in previous studies (McCurry et al., 2022). Further, the 10-minute interval allowed for collection of more psychophysiological data points. The approach used in Study 2 enabled analysis on the effects of automation changes on objectively measured and self-reported workload. Investigation into SA and performance was conducted again in the same manner as Study 1. An analysis to determine which measures are predictive of workload as LOAs change resulted from this study.

The supported research questions and hypotheses for this study are as follows:

Research Question 1. How do IMPRINT operator cognitive workload predictions, psychophysiological measures, and subjective workload measures correlate at different levels of automation?

Ha1: Cognitive workload predictions are directly correlated with objective and subjective workload measures.

Ha2: There is a significant difference between cognitive workload measures in low and high levels of automation within the same level of workload demand.

Research Question 2. What is the relationship between cognitive workload and situation awareness during different levels of automation when completing a multi-tasking simulation?

Ha3: Cognitive workload and situation awareness are inversely related.

B. PILOT DATA

A pilot study was not originally planned for Study 2 because the only important difference from the previous study was the introduction of automation for one of the 10-minute scenarios. Additionally, participants in Study 1 stated that they were expecting to see some sort of automation in the TRACK task since it was explicitly covered during the instructional video. Due to these considerations, the researchers proceeded with Study 2 with the assumption that no significant changes would be necessary to the study's execution. However, initial observations from the first seven participants necessitated a modification in the study's execution.

Six of the first seven participants (three in the novice and three in the experienced groups) did not realize that one of their trial runs had the TRACK task in automatic mode. While the TRACK task is automated, there is no requirement for the user to provide any inputs to the joystick. The six participants completed their training progressions in the same manner as the participants in Study 1. There were four visual indicators that the TRACK task was in an automated mode as seen in Figure 36. Additionally, the color of the circular reticle changed from a dark shade of blue to a lighter shade of blue as mentioned in the instructional video. A different audio tone at the start of the MATB-II scenario also indicated that the TRACK task was in automated mode. Finally, participant inputs via the joystick had no impact on the track task, yet the six participants who did not realize that the task was in automated mode continued to make inputs via the joystick for the duration of their experimental trials.

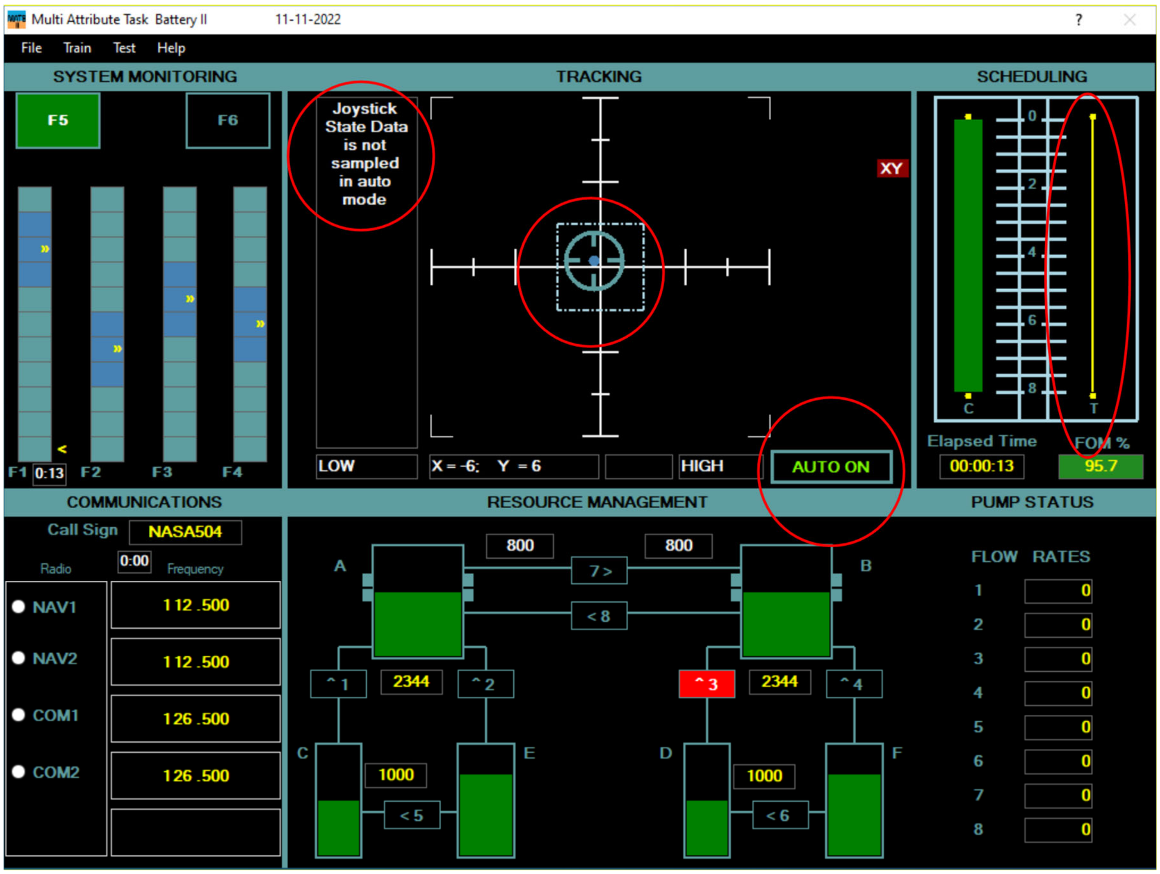


Figure 36. MATB-II with TRACK in AUTO ON mode and indicators highlighted in red.

As a result of these unexpected participant behaviors, instructions were modified prior to beginning the trial runs. Participants were told that they were going to complete two trials runs: one with the TRACK task in “Auto On” mode and one with the TRACK task in “Manual” mode. The participants were instructed to recall their training and refer to the MATB-II Reference Sheet for assistance in determining which mode the simulation was running. After this addition to the instruction script, there were no further issues with participants mistaking the state of the TRACK task during their trial runs.

While this emergent behavior caused the study to begin with a slight delay, it was ultimately informative. A key aspect of this research is to investigate unintended negative consequences of adaptive automation. In this case, an unintended consequence manifested in operators not being aware of the system status. This phenomenon has been investigated over the years in human computer interaction research. Operators can experience mode error whereby they are not aware of the status of the system that they are operating or monitoring (Sarter & Woods, 1995). The participants were so engrossed with the TRACK task that they were unaware the system was in an automated mode. The continued engagement with the TRACK task consumed participants’ attentional resources that could have been allocated to the other tasks and caused inadvertent increases in reported cognitive workload. This type of phenomenon is not uncommon in many accidents that involved a user and a miscalibration of task responsibility with an automated system. In essence, the operator did not realize the system was in a state that was supposed to alleviate workload demands. The operator’s unawareness of the system’s state created additional workload instead (Sarter & Woods, 1995).

C. PARTICIPANTS

1. Selection

The NPS IRB approved the research methods used for this study. There were no changes to the inclusion and exclusion criteria in this study. All participants were informed of their rights as participants in the study and signed consent forms. Participants were recruited through personal communication, email, and campus-wide announcements on the student personnel accountability website. Participants used the same consent form from Study 1. No participants from Study 1 were eligible to participate in this study due to the exclusion criteria.

2. Demographics

Forty participants completed the study (mean age in years = 36.04, standard deviation [SD] = 8.34). Participants included 33 males and 7 females. Of the 40 participants, 39 were in the military (15 in the U.S. Army, 6 in the U.S. Navy, 13 in the U.S. Marine Corps, 1 in the U.S. Air Force, 3 in foreign militaries, and 2 Department of the Navy civilians). The military participants' occupational specialties within their respective services included operations, operations support, and force sustainment. All participants were graduate students or employees at NPS. The rank breakdown of the participants is depicted in Table 8. The participants' time in service ranged from 5 months to 20 years (mean years in service = 11.71, SD = 4.33). All participants met the screening criteria listed in the inclusion and exclusion criteria.

Table 8. Study 2 participants' military rank.

Participant Rank	Number
O-3	18
O-4	17
O-5	3
Civilian	2
Total	40

D. MATERIALS

In Study 2, we used the same configuration, materials, and workstation as in Study 1.

E. VARIABLES

1. Independent Variables

The three independent variables manipulated in this study were experience, levels of automation, and workload levels. Presentation of the levels of automation was counterbalanced to account for order effects. The four conditions used in Study 2 are depicted in Figure 37.

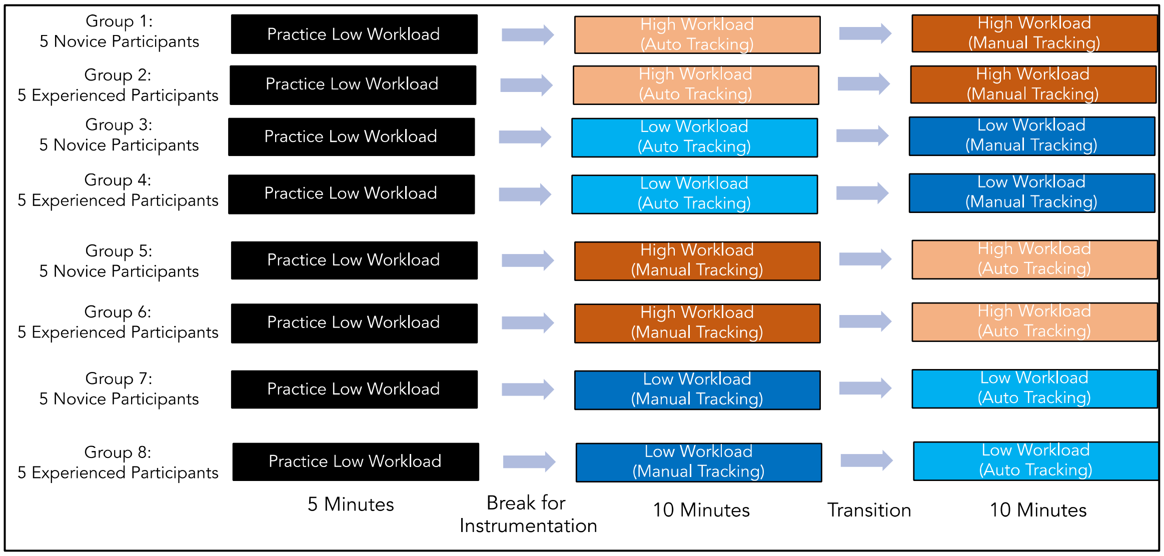


Figure 37. Study 2 conditions.

a. Workload

All participants were presented either a low or high workload condition based on previous MATB-II studies (McCurry et al., 2022). Participants were randomly assigned to a condition where they would receive either the low automation or high automation condition first. The number of tasks associated with each manual condition were the same used in Study 1 (see Table 8). The researcher assigned tasks throughout the scenarios in accordance with the parameters in Table 9.

Table 9. MATB-II system settings for Study 2's conditions.

	System Monitoring	**Tracking**	**Communications**	**Resource Management**
Low Workload (Manual Tracking)	11 Events	Low Joystick Response High Update Rate	3 Events	1 Pump Failure 1 Pump Shutoff
Low Workload (Auto Tracking)	11 Events	Automatic	3 Events	1 Pump Failure 1 Pump Shutoff

	System Monitoring	Tracking	Communications	Resource Management
High Workload **(Manual Tracking)**	20 Events	Low Joystick Response High Update Rate	12 Events	10 Pump Failures 10 Pump Shutoffs
High Workload **(Auto Tracking)**	20 Events	Automatic	12 Events	10 Pump Failures 10 Pump Shutoffs

b. Experience

Participants were assigned to either a novice or experienced group based on scheduling availability. These groups were developed to assess differences between novice and experienced operators. All participants in both groups received the same baseline training on MATB-II. The baseline training included an orientation to the input devices, a NASA instructional video, part-task training, and a 5-minute MATB-II practice session in a low workload condition. Experienced participants conducted four additional 5-minute MATB-II practice sessions in a high workload condition. All the MATB-II sessions were written such that participants did not receive the same scenario more than once.

2. Dependent Variables

Study 2 collected the same dependent variables as Study 1. These collected measures included the MATB-II FOM score; CSWAG, NASA TLX, and SART ratings; and eye tracking, heart rate, and PFC blood oxygenation level data.

F. PROCEDURE

1. Participants

Participants in Study 2 completed the same procedures as participants in Study 1. Additional instructions were provided to the participants based on the emergent behavior seen in the early Study 2 trials. Prior to beginning their experimental runs, participants were informed that they would be presented with two scenarios. One scenario would have the TRACK task in manual mode while the other would have it in auto mode. Participants were instructed to recall their training and refer to the MATB-II reference guide on the desk to determine what the system status of the TRACK task was. Novice participants completed their training and experimental runs in one session, while experienced participants completed the study over two sessions within 72 hours of each other. The key difference between the procedures in Study 1 and Study 2 was that novices and experienced users were assigned to conditions that presented low and high levels of automation in different orders that were counterbalanced across all participants. Additionally, each user only experienced one workload condition. For instance, a user would conduct two trial runs, with one run being at a high level of automation in a low workload condition, followed by a second run at a low level of automation in a low workload condition.

2. IMPRINT Modeling

The researcher constructed IMPRINT models using the same approach used in developing Study 1's models. However, two additional IMPRINT models were produced to capture workload predictions when completing each workload condition with either a low or high level of automation. The baseline models used for the manual conditions in Study 1 did not change. The task network diagram of the low and high conditions is seen in Figures 38. Figure 39 depicts the external event triggers used to model the initiation of communications events in Study 2. Of note, the manual tracking trigger was omitted for the two automated tracking conditions in Study 2. Additionally, automatic tracking was not included in the task network diagram. This omission was due to participants being told in their training that they had no responsibility when the TRACK mode was set to "AUTO ON."

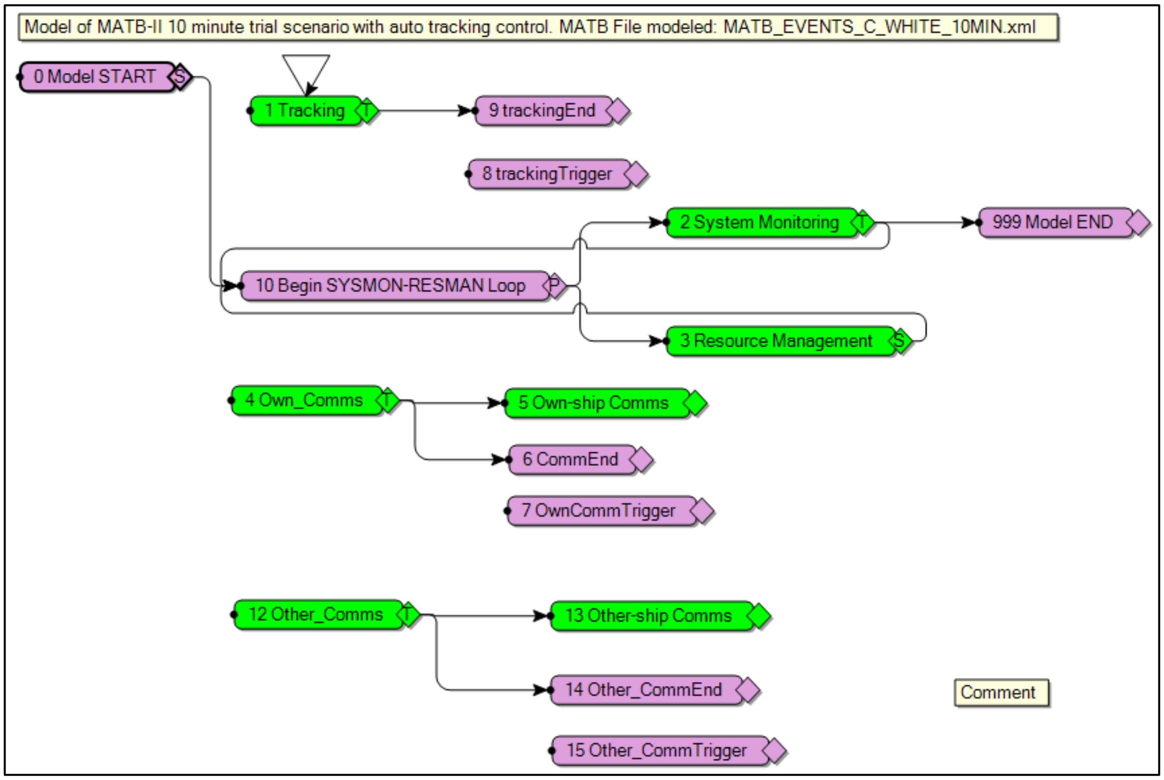
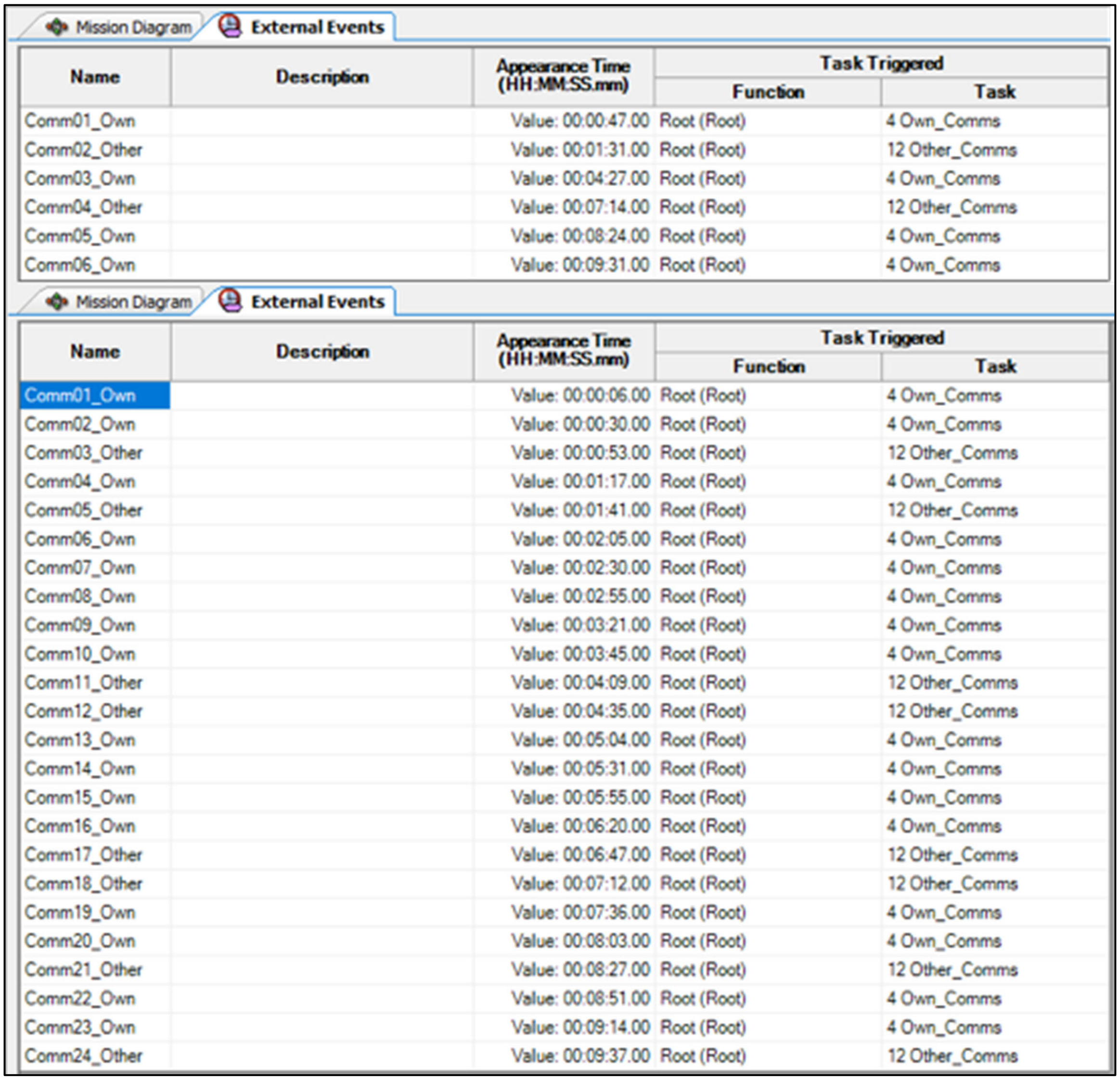


Figure 38. Study 2 researcher-derived task network diagram IMPRINT model.



Mission Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Own		Value: 00:00:47.00	Root (Root)	4 Own_Comms
Comm02_Other		Value: 00:01:31.00	Root (Root)	12 Other_Comms
Comm03_Own		Value: 00:04:27.00	Root (Root)	4 Own_Comms
Comm04_Other		Value: 00:07:14.00	Root (Root)	12 Other_Comms
Comm05_Own		Value: 00:08:24.00	Root (Root)	4 Own_Comms
Comm06_Own		Value: 00:09:31.00	Root (Root)	4 Own_Comms

Mission Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Own		Value: 00:00:06.00	Root (Root)	4 Own_Comms
Comm02_Own		Value: 00:00:30.00	Root (Root)	4 Own_Comms
Comm03_Other		Value: 00:00:53.00	Root (Root)	12 Other_Comms
Comm04_Own		Value: 00:01:17.00	Root (Root)	4 Own_Comms
Comm05_Other		Value: 00:01:41.00	Root (Root)	12 Other_Comms
Comm06_Own		Value: 00:02:05.00	Root (Root)	4 Own_Comms
Comm07_Own		Value: 00:02:30.00	Root (Root)	4 Own_Comms
Comm08_Own		Value: 00:02:55.00	Root (Root)	4 Own_Comms
Comm09_Own		Value: 00:03:21.00	Root (Root)	4 Own_Comms
Comm10_Own		Value: 00:03:45.00	Root (Root)	4 Own_Comms
Comm11_Other		Value: 00:04:09.00	Root (Root)	12 Other_Comms
Comm12_Other		Value: 00:04:35.00	Root (Root)	12 Other_Comms
Comm13_Own		Value: 00:05:04.00	Root (Root)	4 Own_Comms
Comm14_Own		Value: 00:05:31.00	Root (Root)	4 Own_Comms
Comm15_Own		Value: 00:05:55.00	Root (Root)	4 Own_Comms
Comm16_Own		Value: 00:06:20.00	Root (Root)	4 Own_Comms
Comm17_Other		Value: 00:06:47.00	Root (Root)	12 Other_Comms
Comm18_Other		Value: 00:07:12.00	Root (Root)	12 Other_Comms
Comm19_Own		Value: 00:07:36.00	Root (Root)	4 Own_Comms
Comm20_Own		Value: 00:08:03.00	Root (Root)	4 Own_Comms
Comm21_Other		Value: 00:08:27.00	Root (Root)	12 Other_Comms
Comm22_Own		Value: 00:08:51.00	Root (Root)	4 Own_Comms
Comm23_Own		Value: 00:09:14.00	Root (Root)	4 Own_Comms
Comm24_Other		Value: 00:09:37.00	Root (Root)	12 Other_Comms

Figure 39. Study 2 external event triggers for the low (above) and high (below) workload conditions with automatic TRACK.

The MATB-II researcher-derived cognitive workload demand ratings are provided in Table 10. These ratings were again based on the default anchors provided in IMPRINT.

Table 10. Study 2 researcher-derived IMPRINT workload demand values.

Task: Tracking	**RI Pair Demand Values**						
Total Task Demand 10.00	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Joystick		1.00	4.60				4.40
Task: System Monitoring	**RI Pair Demand Values**						
Total Task Demand 8.40	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		1.20	2.20				5.00
Task: Resource Management	**RI Pair Demand Values**						
Total Task Demand 8.40	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		1.20	2.20				5.00
Task: Own Comms	**RI Pair Demand Values**						
Total Task Demand: 10.50	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		1.00	2.20				1.00
Interface: Speaker	4.30	1.00					1.00
Task: Other Comms	**RI Pair Demand Values**						
Total Task Demand: 6.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse							
Interface: Speaker	4.30	1.00					1.00

A consolidation measure was calculated for each resource-interface pair by taking the mean of the expert ratings as was done in Study 1. However, the experts never experienced the TRACK task in "AUTO ON" mode in their interactions with their version of MATB. Therefore, they were asked to provide their best prediction as to the impacts on the resource-demand values if the TRACK task was automated. The consolidated expert-derived values are provided in Table 11.

Table 11. Study 2 expert-derived IMPRINT workload demand values.

Task: Tracking	**RI Pair Demand Values**						
Total Task Demand 11.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Joystick		3.60	3.90				3.80

Task: Tracking (AUTO)	**RI Pair Demand Values**						
Total Task Demand 1.67	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Joystick	0.00	0.67	0.00				1.00

Task: System Monitoring	**RI Pair Demand Values**						
Total Task Demand 7.87	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		2.40	3.30				2.17

Task: Resource Management	**RI Pair Demand Values**						
Total Task Demand 13.37	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		5.80	3.57				4.00

Task: Own Comms	**RI Pair Demand Values**						
Total Task Demand: 16.87	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		2.07	4.07				1.83
Interface: Speaker	5.00	2.07					1.83

Task: Other Comms	**RI Pair Demand Values**						
Total Task Demand: 6.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse							
Interface: Speaker	4.30	1.00					1.00

The researcher-derived IMPRINT-predicted time average workload value for the whole low workload condition with high automation was 11.65. The total high workload with high automation condition time average workload value was 19.20. Expert predicted values for the low and high LOA conditions were and 13.97 and 16.54, respectively. There were significant differences between the average workload seen between the low and high levels of automation for the TRACK task. These workload differences were intuitive as participants had more attentional resources available to execute three tasks instead of four. However, both the low and high workload conditions saw workload spikes as seen in Figures 40 and 41. These workload spikes were associated with the initiation of communications tasks in both studies. These types of tasks create additional demand on the user while also creating workload resource conflicts with the other tasks.

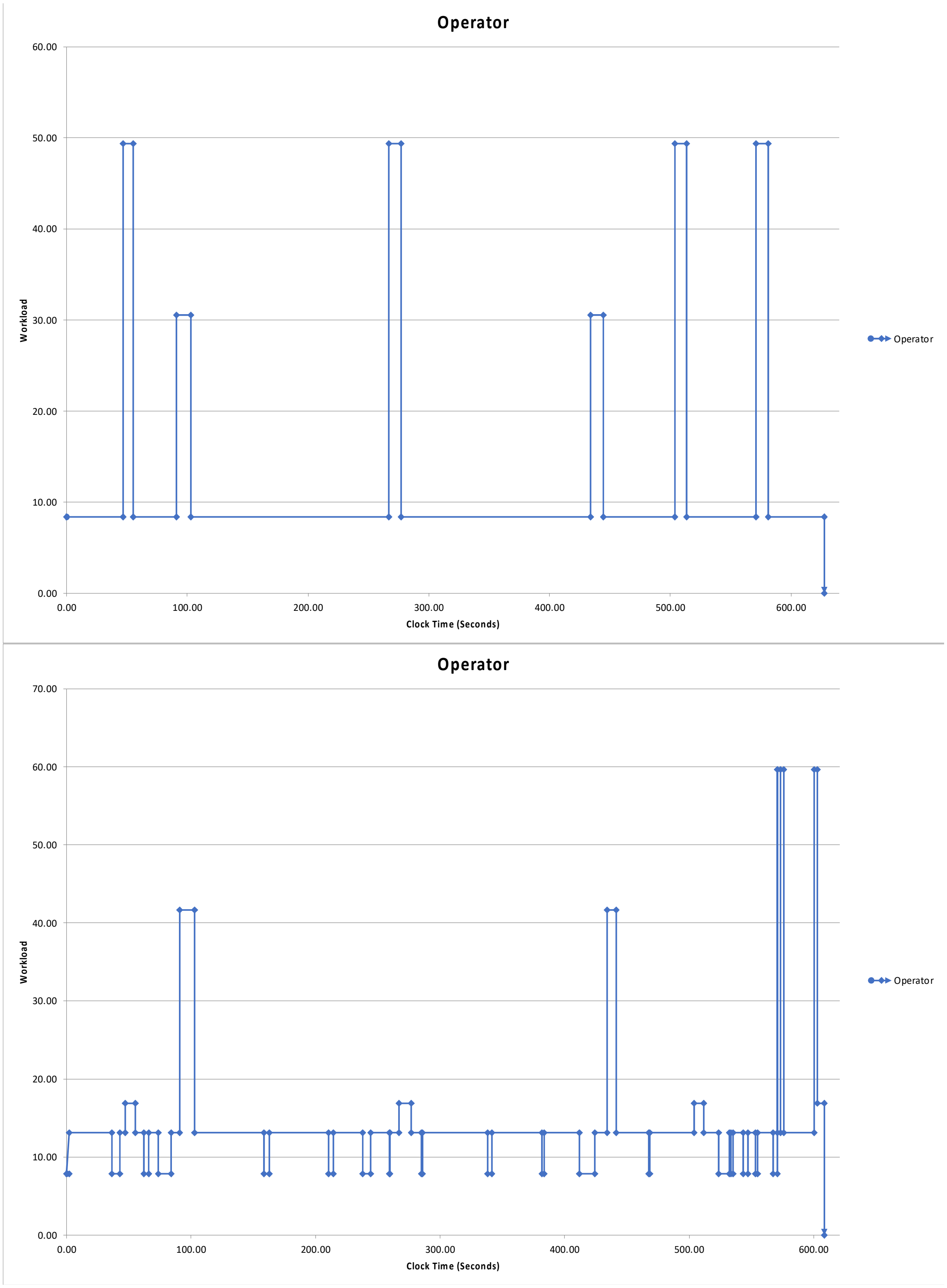


Figure 40. Study 2 researcher-derived (top) and expert-derived IMPRINT (below) model graphs for the low workload condition with high level of automation.

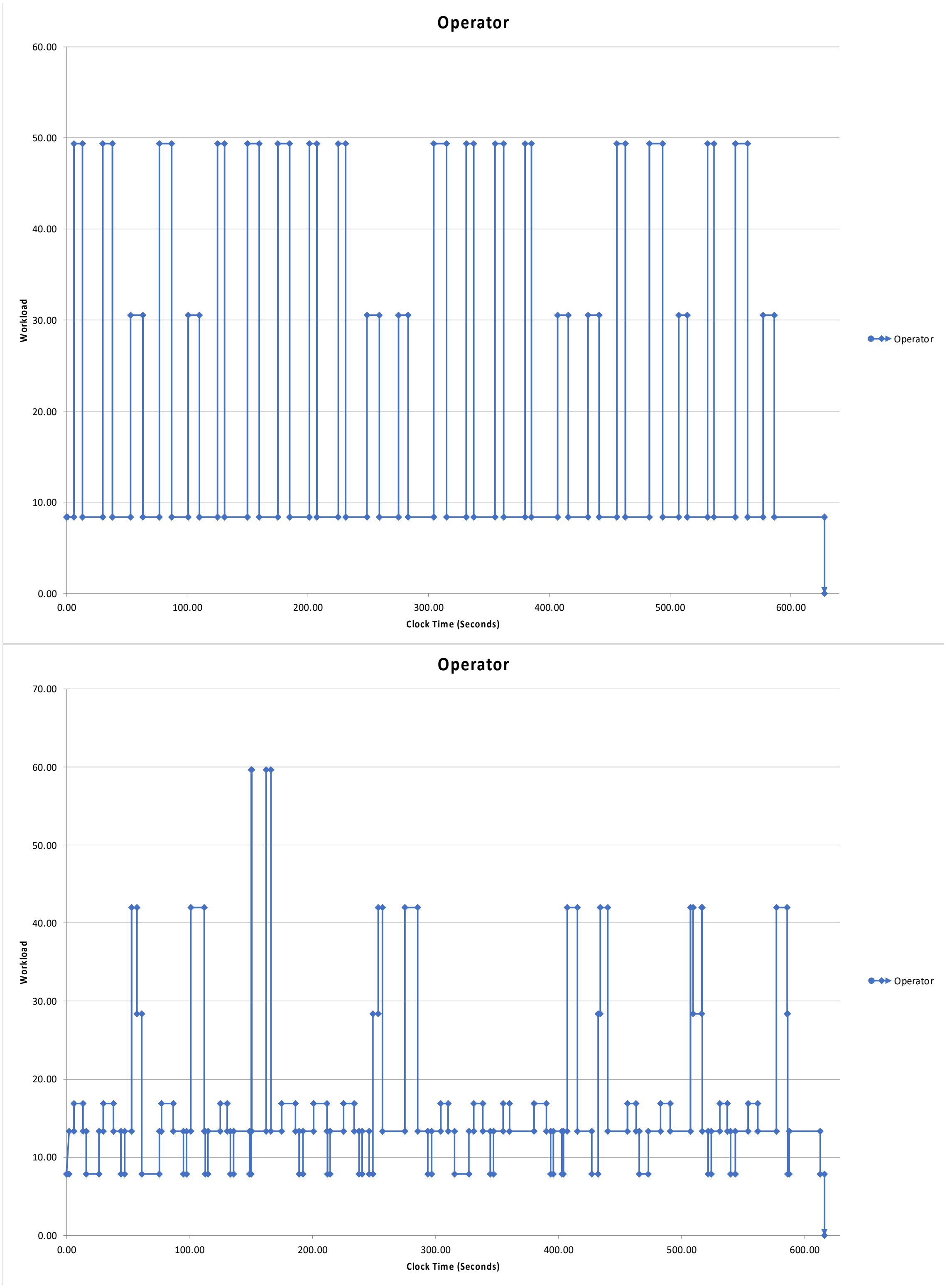


Figure 41. Study 2 researcher-derived (top) and expert-derived IMPRINT model graphs for the high workload condition with high level of automation.

G. RESULTS

Data results were gathered in the same manner as Study 1 by leveraging the participants' .xdf files and subjective ratings. A mixed model analysis approach was again used to analyze the collected measures in JMP version 16.0.0. Fixed effects included experience level, workload level, and automation condition in the TRACK task. Random effects were modeled using each participant with their experience and workload levels nested to account for the differences between experimental condition exposure for all participants. A summary of the results for Study 2 are listed in Tables 10 (with presentation order included in the analytical model) and 11 (without presentation order included in the analytical model). Participant data were excluded due to extreme values after analyzing residual plots for each modeled measure. These exclusions are listed in the notes below Tables 5 and 6.

Table 12. Study 2 summary results table with presentation order included.

Measure Category	Measure Type	Dependent Measure	Novice vs. Experienced	High vs. Low Workload	Tracking Mode	Presentation Order
Performance	**MATB-II Figure of Merit**	**Composite FOM****	$p = .105$	***High workload -> lower FOM*** $\boldsymbol{p = .039*}$	$p = .308$	***First condition -> lower FOM*** $\boldsymbol{p = .017*}$
Psycho-physiological	**HRV**	**Mean HRV*****	$p = .679$	$p = .179$	***Manual Tracking-> lower HRV*** $\boldsymbol{p < .001*}$	***First condition -> lower HRV*** $\boldsymbol{p = .04*}$
	fNIRS	**Mean HbO****	$p = .307$	$p = .526$	$p = .092$	$p = .841$
		Mean Hb	$p = .965$	$p = .338$	$p = .087$	$p = .343$
		Mean Total Hb**	$p = .635$	$p = .923$	$p = .131$	$p = .701$
	Pupil	**Mean Right Pupil Diameter****	$p = .098$	$p = .361$	***Auto Tracking -> smaller diameter***	***First condition -> larger pupil diameter***

Measure Category	Measure Type	Dependent Measure	Novice vs. Experienced	High vs. Low Workload	Tracking Mode	Presentation Order
					$p < .001$*	$p = .004$*
		Mean Left Pupil Diameter****	$p = .337$	$p = .443$	***Auto Tracking -> smaller diameter*** $p < .001$*	***First condition -> larger pupil diameter*** $p = .003$*
Subjective Workload	Continuous Subjective Workload Assessment Graph	Mean CSWAG	$p = .772$	***High workload -> higher CSWAG rating*** $p = .023$ *	***Participants reported lower CSWAG in Auto Track*** $p < .001$*	$p = .810$
	NASA-TLX	NASA-TLX Rating	$p = .244$	$p = .250$	$p = .815$	N/A
Situation Awareness	Situation Awareness Rating Technique	SART Rating	$p = .429$	$p = .166$	$p = .608$	N/A

Table 13. Study 2 summary results table with presentation order omitted.

Measure Category	Measure Type	Dependent Measure	Novice vs. Experienced	High vs. Low Workload	Tracking Mode
Performance	MATB-II Figure of Merit	Composite FOM**	$p = .105$	***High workload -> lower FOM*** $p = .039$*	$p = .369$
Psycho-physiological	HRV	Mean HRV***	$p = .679$	$p = .179$	***Manual Tracking -> lower HRV*** $p < .001$*

Measure Category	Measure Type	Dependent Measure	Novice vs. Experienced	High vs. Low Workload	Tracking Mode
	fNIRS	Mean HbO**	$p = .307$	$p = .526$	$p = .086$
		Mean Hb	$p = .965$	$p = .338$	$p = .086$
		Mean Total Hb**	$p = .635$	$p = .923$	$p = .128$
	Pupil	Mean Right Pupil Diameter**	$p = .098$	$p = .361$	*Auto Tracking -> smaller diameter* $\boldsymbol{p < .001}$*
		Mean Left Pupil Diameter****	$p = .541$	$p = .921$	*Auto Tracking -> smaller diameter* $\boldsymbol{p \leq .005}$*
Subjective Workload	Continuous Subjective Workload Assessment Graph	Mean CSWAG	$p = .772$	*High workload -> higher CSWAG rating* $\boldsymbol{p = .002}$*	*Auto Tracking-> Lower reported workload* $\boldsymbol{p < .001}$*
	NASA-TLX	NASA-TLX Rating	$p = .244$	$p = .250$	$p = .815$
Situation	Situation Awareness	SART Rating			

Measure Category	Measure Type	Dependent Measure	Novice vs. Experienced	High vs. Low Workload	Tracking Mode
Awareness	**Rating Technique**		$p = .429$	$p = .166$	$p = .608$

Tables 10 and 11 Notes:

* p<.05

** 1 Novice participant excluded due to extreme results

*** 2 Experienced participants excluded due to extreme results

**** 2 Novice participants excluded due to extreme results

There were no statistically significant differences between novice and experienced participants' FOMs in Study 2. However, Figure 42 shows that participants' mean FOMs were higher in the low workload condition than the high workload condition (M=93.61, SD=4.27, SE=0.67 vs. M=91.25, SD=4.16, SE=0.67)

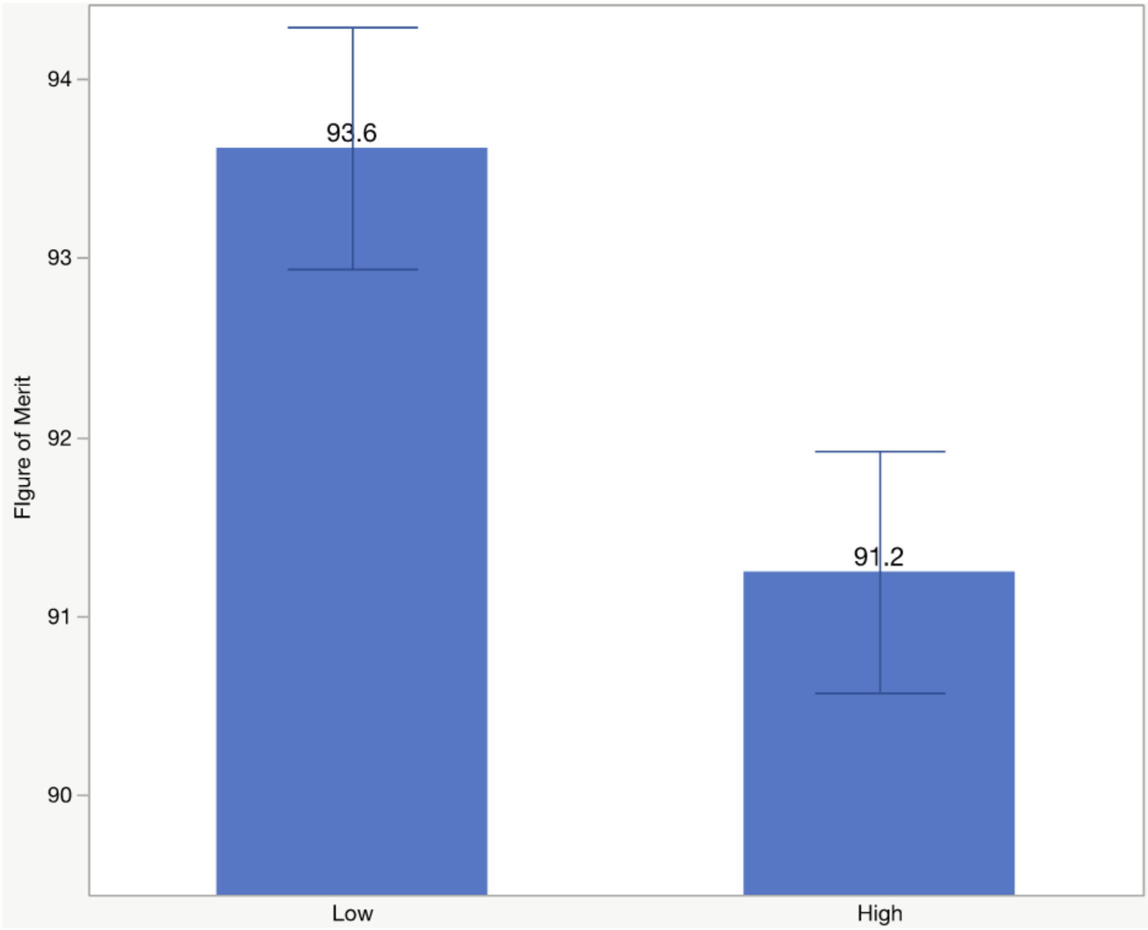


Figure 42. Study 2 mean FOMs by workload level. Error bars represent standard errors.

There were no statistically significant differences between novice or experienced participants' mean HRV. There were also no differences in mean HRV between workload levels in Study 2. However, there were statistically significant differences in mean HRV between tracking modes. Auto tracking had a higher mean HRV in milliseconds (M=860.69, SD=107.34, SE=17.41) than manual tracking (M=839.05, SD=106.51, SE=17.28) as depicted in Figure 43.

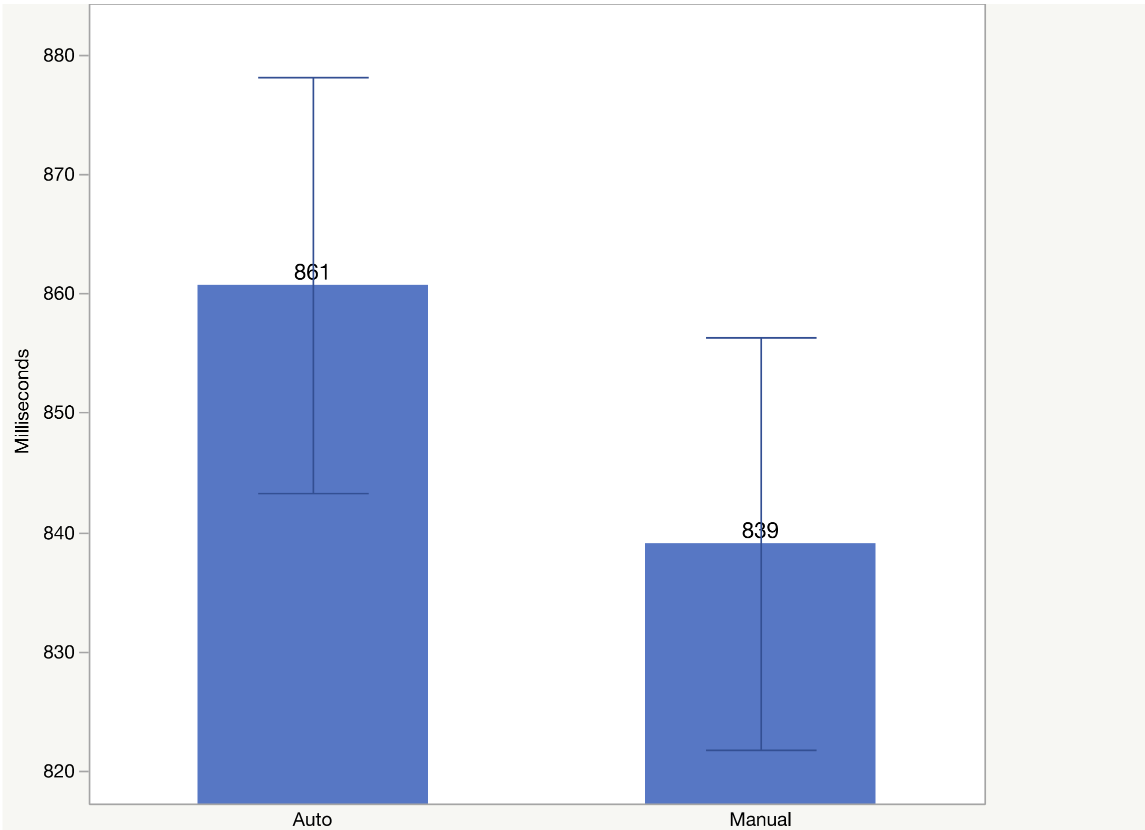


Figure 43. Mean HRV in milliseconds for Study 2 by tracking condition. Error bars represent standard errors.

The mean left and right pupil diameter differences in Study 2 were statistically significant for the tracking condition. Figure 44 illustrates the mean right pupil diameter being larger (M=3.23mm, SD=0.70, SE=0.11) for the manual tracking condition than the auto tracking condition (M=3.14mm, SD=0.66, SE=0.11). Figure 45 shows the differences for the left pupil in the manual tracking condition (M=3.38mm, SD=0.70, SE=0.11) and the auto tracking condition (M=3.28mm, SD=0.64, SE=0.10).

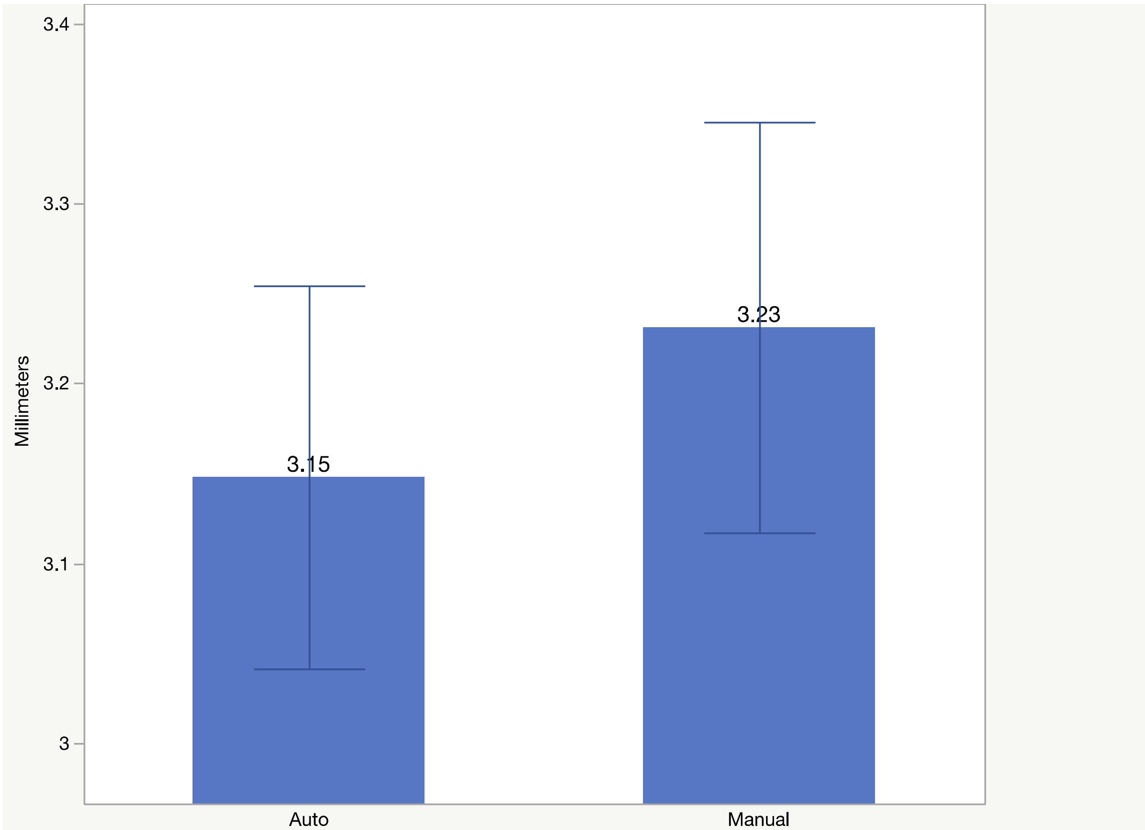


Figure 44. Mean right pupil diameter in millimeters by tracking condition for Study 2. Error bars represent standard errors.

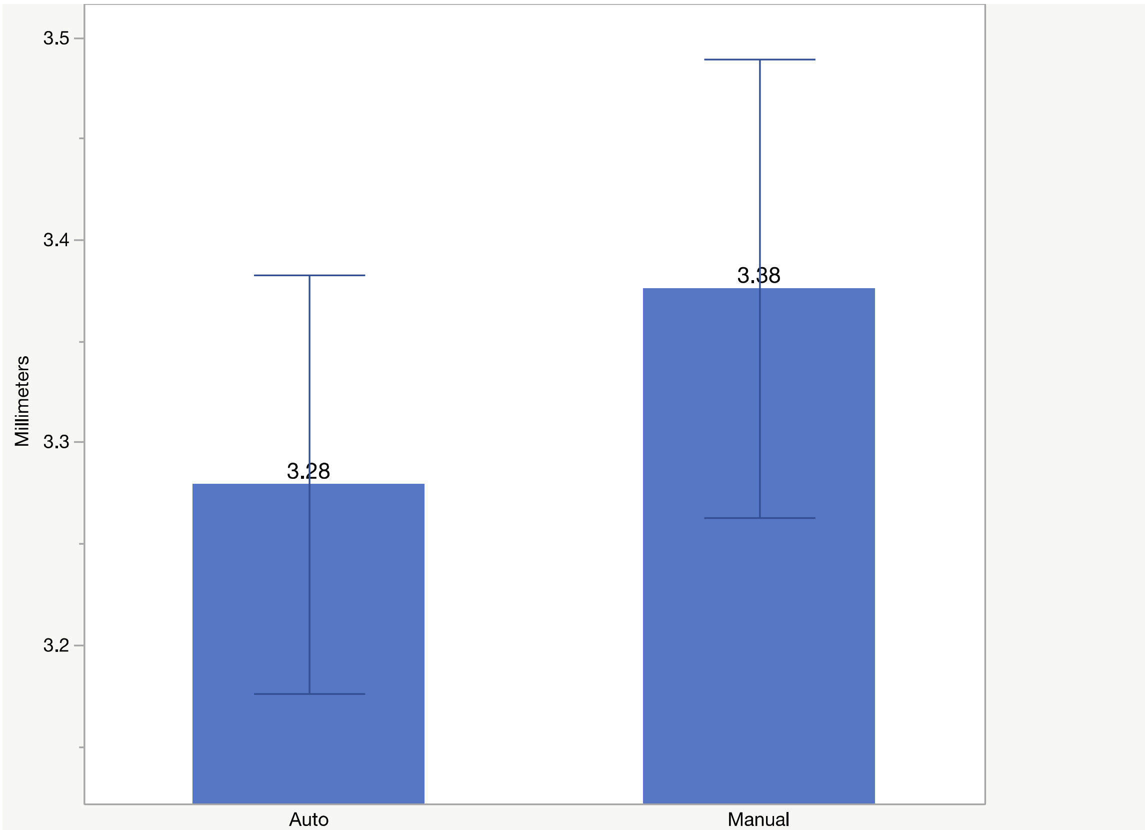


Figure 45. Mean left pupil diameter by tracking condition for Study 2. Error bars represent standard errors.

Figure 46 depicts progressively larger right pupil diameters from pre-communications event to post-communications event for Study 2 (M=3.19mm, SD=0.67, SE=0.08 vs. M=3.23mm, SD=0.70, SE=0.08). Left pupil diameters also followed the same progression as shown in Figure 47 (M=3.28mm, SD=0.73, SE=0.08 vs. M=3.39, SD=0.74, SE=0.09). There were statistically significant differences between participants' left pupil diameters in the five second period before and after communications events, $F(1, 112) = 20.17, p < .001$. Participants' right pupil diameters were also different between the same time periods ($F(1, 112) = 7.76, p < .01$.

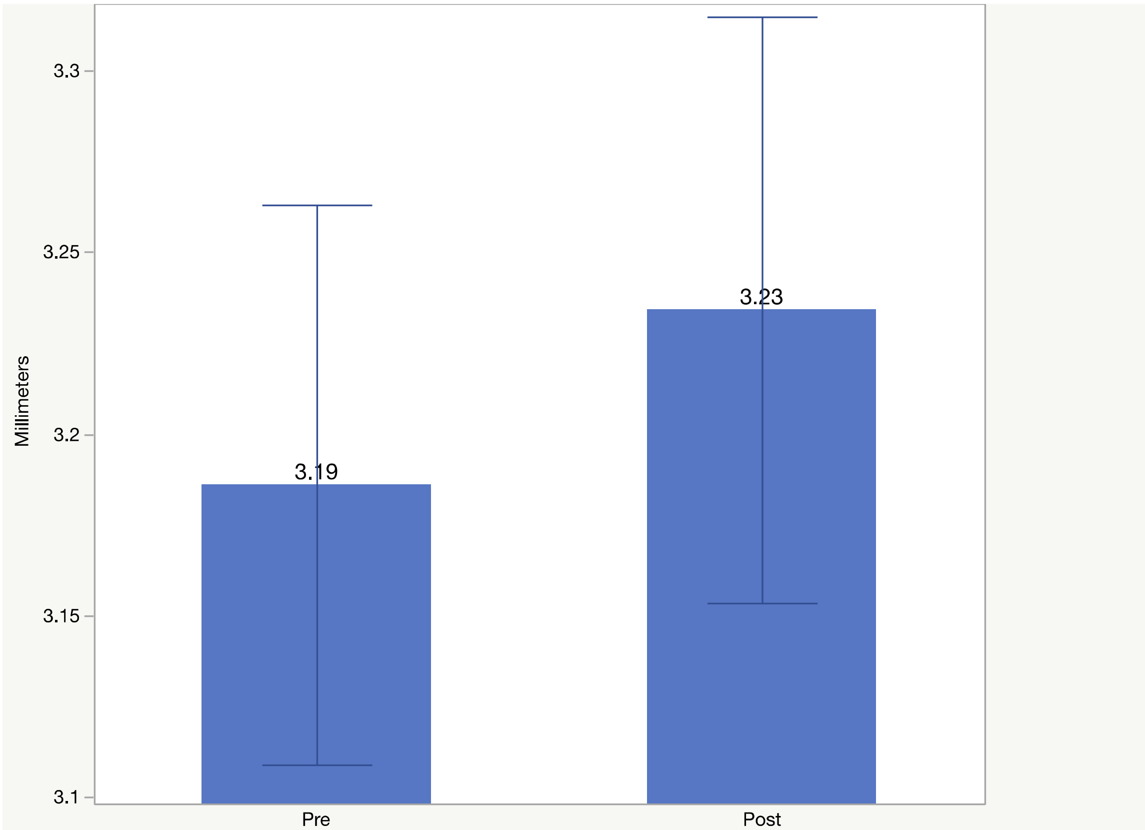


Figure 46. Study 2 mean right diameter in millimeters by communications event timing. Error bars represent standard errors.

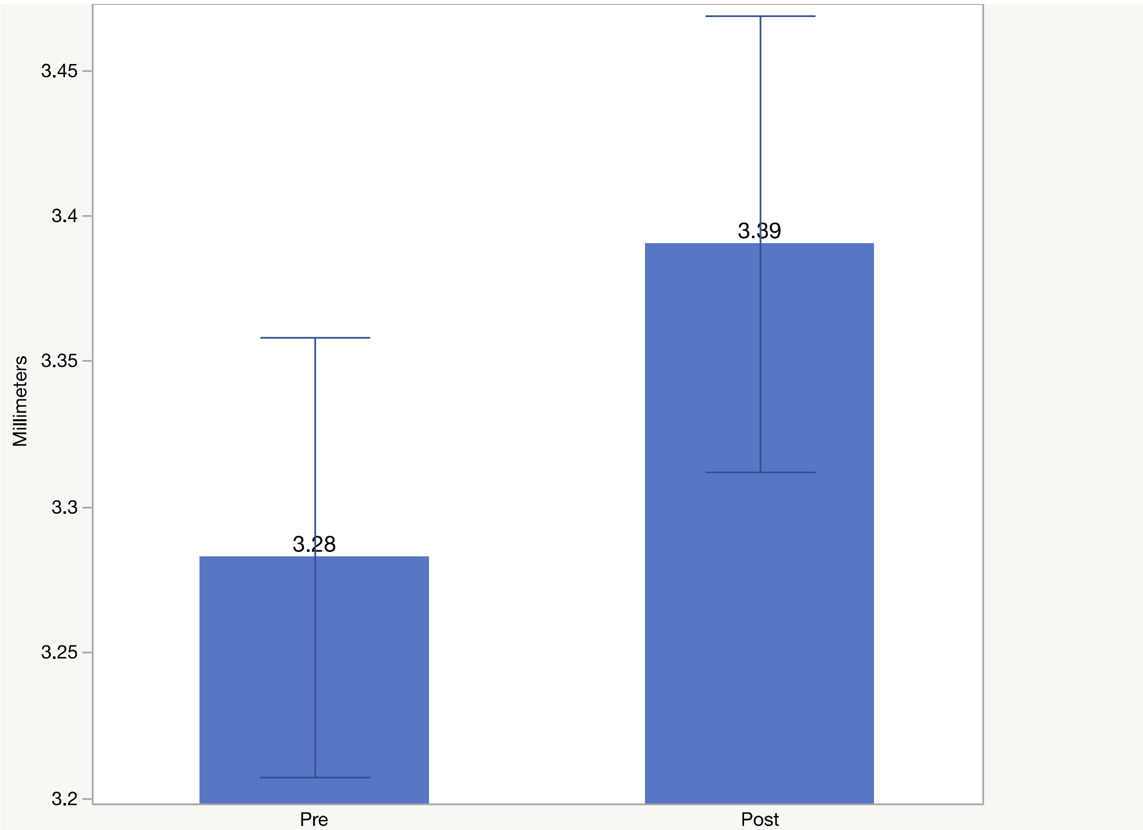


Figure 47. Study 2 mean left pupil diameter in millimeters by communications event timing. Error bars represent standard errors.

Participants' pupils were larger five seconds after a communications event directed at their ship (NASA 504) and other ships, following the pattern seen in Study 1. Own ship communications resulted in statistically significant differences before and after communications events in right pupil diameter $F(1, 112) = 5.70,\ p < .05$ (M=3.19, SD=0.67, SE=0.08 vs. M=3.23, SD=0.70, SE=0.08) and left pupil diameter $F(1, 112) = 20.02,\ p < .001$ (M=3.29, SD=0.66, SE=0.08 vs. M=3.40, SD=0.69, SE=0.08). Other ship communications followed the same pattern of statistical significant differences with the right pupil diameter $F(1, 112) = 10.02,\ p < .01$ (M=3.18, SD=0.68, SE=0.08 vs. M=3.31, SD=0.81, SE=0.09) and left pupil diameter $F(1, 112) = 16.38,\ p < .001$ (M=3.26 SD=0.67, SE=0.08 vs. M=3.37, SD=0.68, SE=0.08). These results are displayed in Figures 48 and 49.

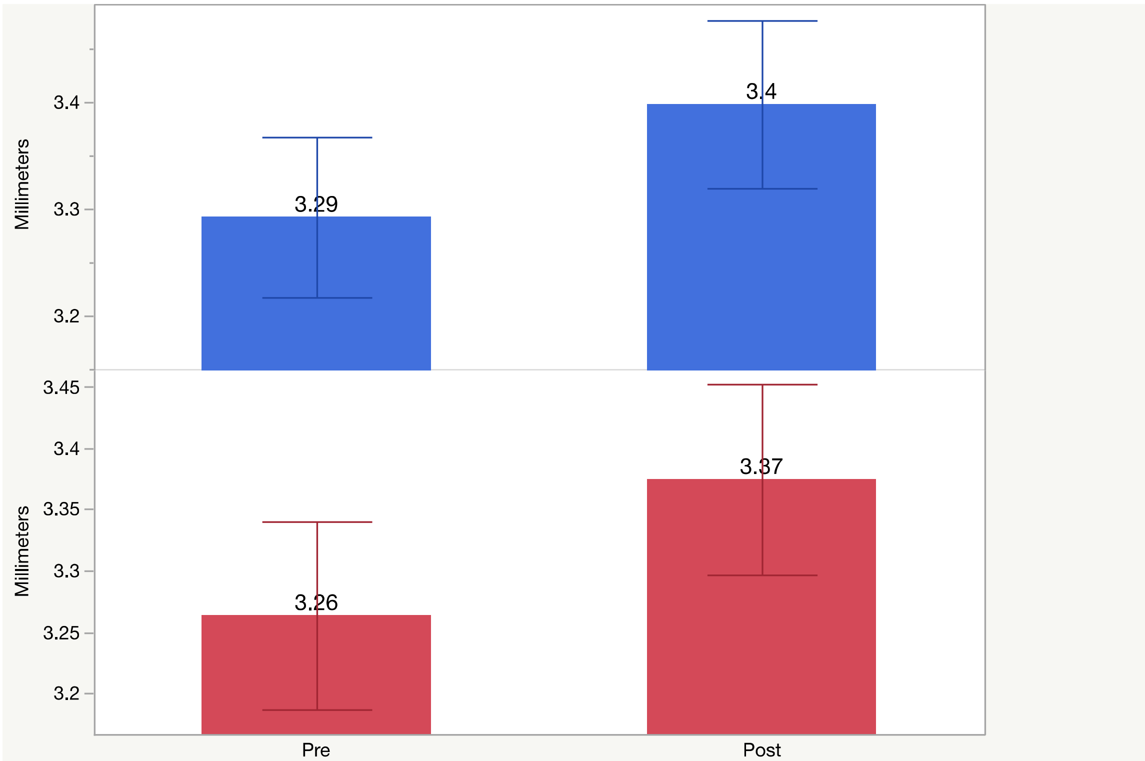


Figure 48. Study 2 left pupil diameter in millimeters by own (top) and other (bottom) event communications timing. Error bars represent standard errors.

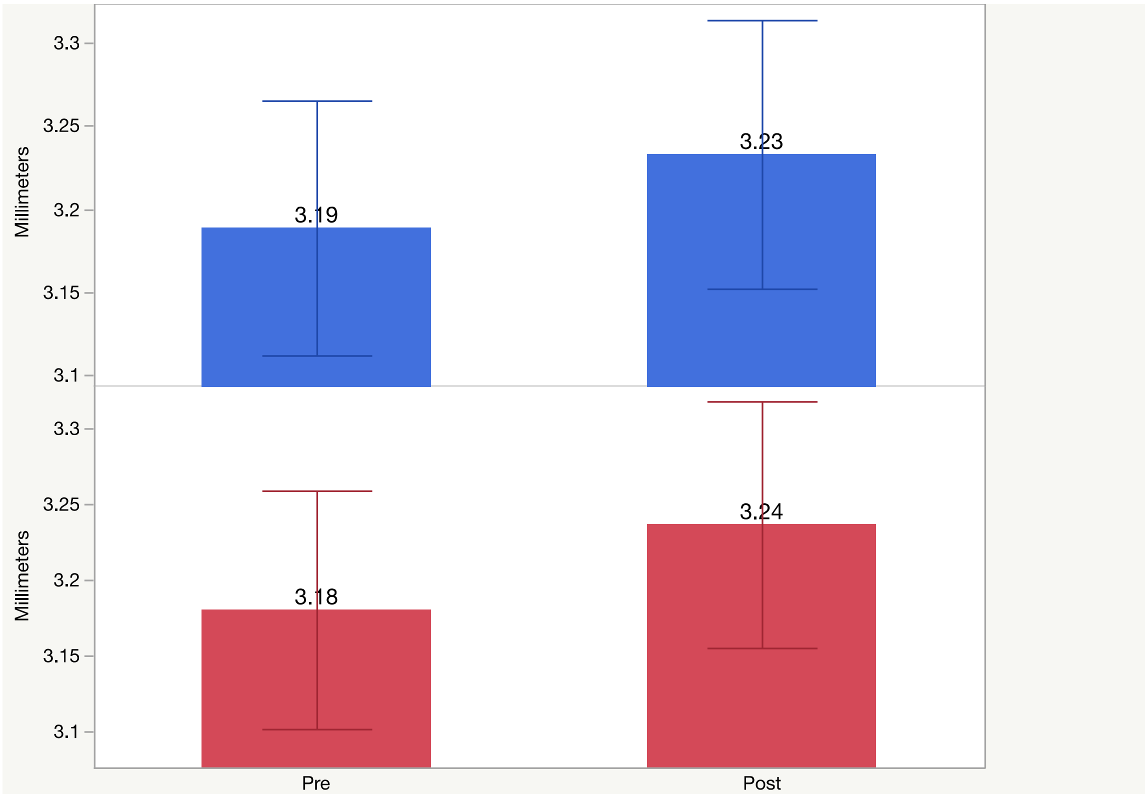


Figure 49. Study 2 right diameter in by own (top) and other (bottom) event communications timing. Error bars represent standard errors.

Differences between own and other ship communications were not statistically significant as shown in Figure 50. One experienced participant was excluded from this portion of the analysis due to extreme results as analyzed in the data's residual plots.

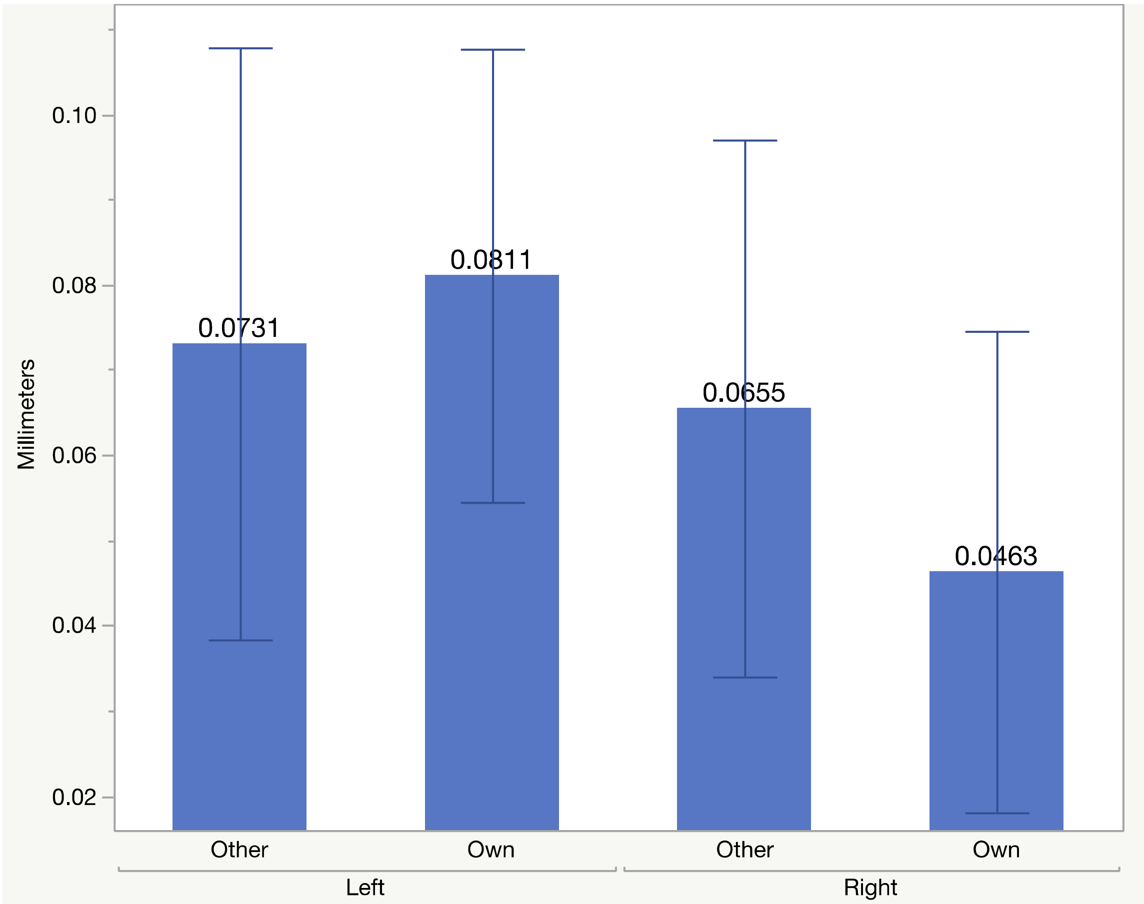


Figure 50. Study 1 mean pupil diameter differences in millimeters between eyes and communications target. Error bars represent standard errors.

Differences in participants' CSWAG percentages were statistically significant between tracking conditions as shown in Figure 51. Manual tracking resulted in higher reported CSWAG (M=55.64, SD=12.24, SE=1.94) than the auto tracking condition (M=37.95, SD=12.03, SE=1.90). Additionally, the high workload condition resulted in higher reported CSWAG than the low workload condition (M=50.86, SD=13.39, SE=2.11 vs. M=42.73, SD=15.156, SE=2.46).

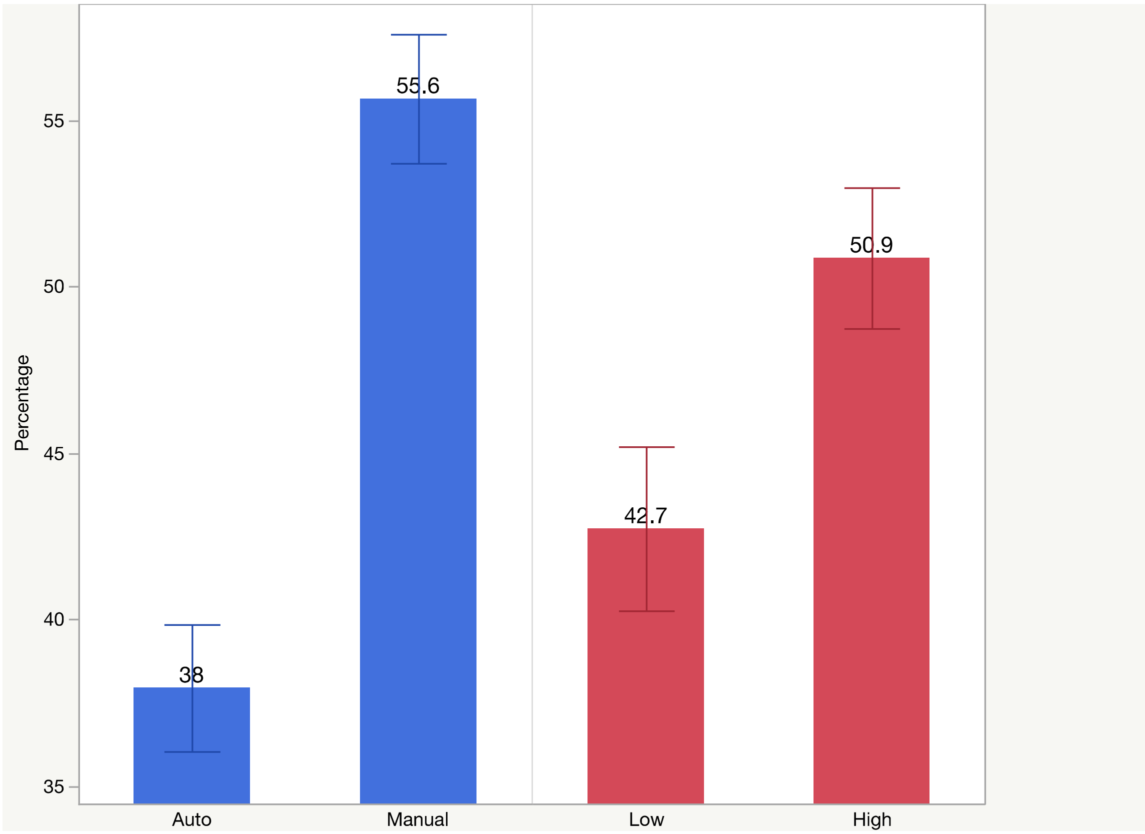


Figure 51. Reported CSWAG percentage means by tracking condition and workload levels in Study 2. Error bars represent standard error.

Post-trial situation awareness and cognitive workload ratings were collected using the SART and NASA-TLX, respectively. There was a statistically significant inverse correlation between SART ratings and NASA-TLX ratings, $\rho = -0.33, p = .04$. This relationship is illustrated in Figure 52.

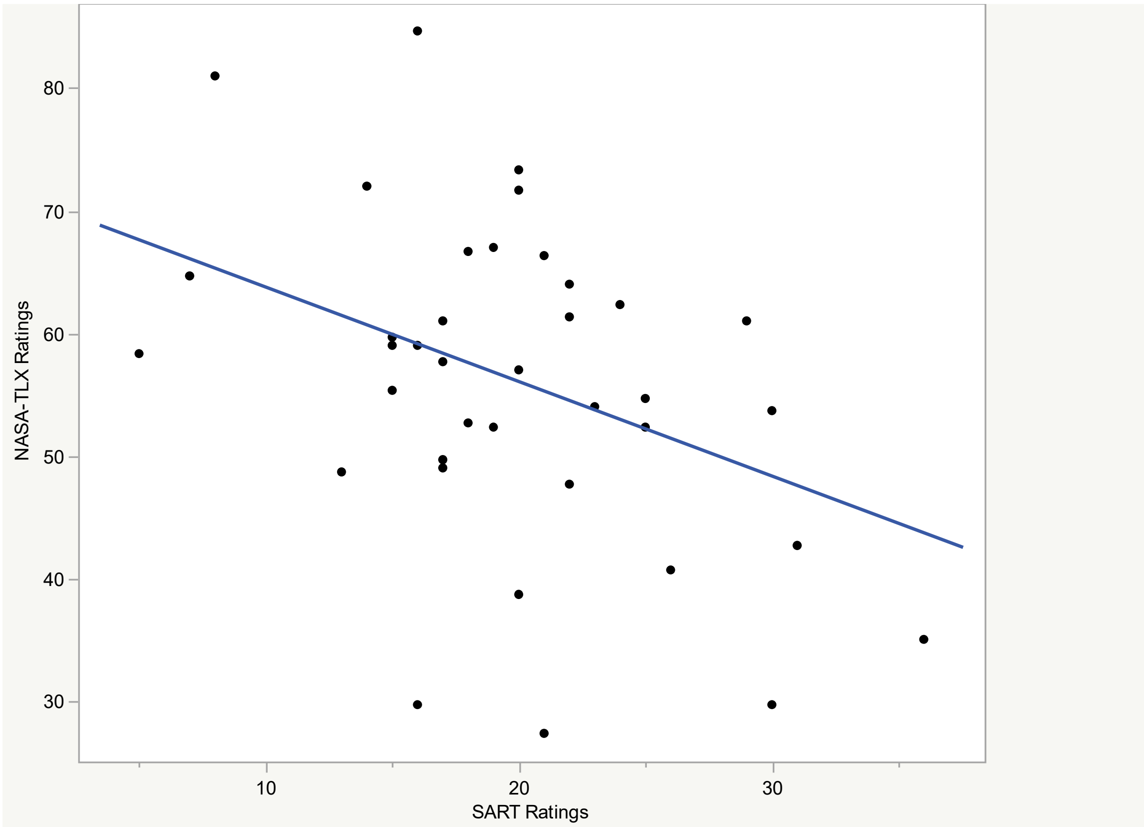


Figure 52. Pairwise correlation graph for Study 2's NASA-TLX and SART ratings.

H. DISCUSSION

The results from Study 2 are discussed in the proceeding section. Study 2's research questions and hypotheses are reviewed, followed by analysis of their associated statistical results. These results were analyzed to provide recommendations for the refinement of Study 3's design. This discussion also addresses the specific aspects of the adapted MLCC framework that were investigated in Study 2.

1. Research Questions and Hypotheses

Research Question 1. Operator cognitive workload predictions correlated to two psychophysiological measures, HRV and pupil diameter. The subjective measure that correlated to the different LOAs was CSWAG. These findings follow the results seen in Study 1 and provide additional support for more continuous assessment of an operators' workload using these approaches. NASA-TLX and fNIRS data did not yield significant results in Study 2.

Higher levels of automation yielded smaller mean pupil sizes and larger HRV. The values follow previous research that suggest that smaller pupil diameters and increased HRV are analogous to lower workload (Aura et al., 2021; Beatty & Lucero-Wagoner, 2000; Duchowski et al., 2018; Fairclough et al., 2005; Pfleging et al., 2016; Tao et al., 2019). Participants also reported lower CSWAG values in the high level of automation condition. These findings followed the IMPRINT workload value predictions that the higher LOA condition would yield lower workload values than the lower LOA condition. The pattern of the high workload condition having higher predicted workload values than the low workload condition regardless of automation level remained consistent between Studies 1 and 2.

Ha1: Cognitive workload predictions are directly correlated with objective and subjective workload measures.

Hypothesis 1 was partially supported with HRV, pupil diameter, and CSWAG ratings being directly associated with cognitive workload predictions. Further, the pattern of pupil diameter being larger after a communications event than the five seconds preceding it remained intact in Study 2, serving to validate the workload spikes in the IMPRINT models again. We fail to reject the null hypothesis that cognitive workload predictions are directly correlated with cognitive workload measures when using fNIRS data and NASA-TLX ratings. These findings suggest that certain measures might be more sensitive to cognitive workload changes than others in similar multi-attribute tasks.

The non-findings with the fNIRS data follow the results from Study 1 in that the fast-paced changes of task may not be conducive to its use as a surrogate measure to workload (Girouard et al., 2010). The same approach that limited NASA-TLX's assessment in Study 1 appeared to have the same result in Study 2. It was difficult for participants to accurately combine their cognitive workload assessment over two separate trials.

Presentation order appears to have had a significant impact on performance, HRV, and pupil diameter. The first condition yielded lower performance results and increased objective cognitive workload surrogate measures. This finding indicates that there is a cost and period of gaining in-the-loop familiarity with a system's operation at the beginning of a task regardless of the LOA. Kaber and Endsley (2004) suggested that operators may need time to orient to their system's status in order to understand how to engage appropriately. Applying Kaber and Endsley's proposal to Study 2, participants were told that they would conduct the trial runs in either manual or auto TRACK mode. This instruction would introduce criteria that the participants would need to process to interact with MATB-II appropriately, and thus help account for the first conditions they saw as being more cognitively demanding. The period of gaining in-the-loop familiarity is also nested with the concept of operators working toward achieving an optimal level of workload as seen in the inverted U-curve of cognitive workload (Ernst et al., 2020; Yerkes & Dodson, 1908; Zhang et al., 2021). Additionally, this finding also suggests that the introduction of higher LOAs might warrant investigation to determine the impacts of LOA transitions on operators. This assessment could help determine the specific impacts of establishing in-the-loop familiarity on cognitive workload.

For subjective workload measures, hypothesis 1 was partially supported through the CSWAG results. The low automation condition had a higher predicted workload score than the high automation condition. When assessed with the CSWAG results, automatic tracking yielded lower CSWAG percentages across participants. This result mirrored the outcome of Study 1 where cognitive workload predictions and subjective workload assessments showed higher results in the more difficult conditions. For Study 2, the manual TRACK condition resulted in higher CSWAG percentages than in the auto TRACK condition. These results suggest that cognitive workload is lower with the introduction of automation and follows previous literature that supports this assertion. Additionally, an aim of adaptive automation is to help manage an operator's workload to optimal levels. The findings in Study 2 suggest that this workload management level is possible given the MATB-II task that participants completed.

Ha2: There is a significant difference between cognitive workload measures in low and high levels of automation within the same level of workload demand.

For Study 2, participants completed their trials with one level of workload and two levels of automation in the TRACK task. Within this framework, hypothesis 2 was partially supported with similar objective and subjective workload surrogate measures as seen in Study 1. Higher levels of automation yielded smaller pupil sizes and larger HRV. These results follow IMPRINT workload predictions that higher LOAs will yield lower overall cognitive workload. Like Study 1, there were no significant differences with fNIRS PFC blood oxygenation levels data across conditions. Hypothesis 2 is also partially supported with CSWAG results as they indicate that workload differences were more sensitive to continuous assessment than whole-trial assessment with the NASA-TLX. Additionally, CSWAG results suggest lower self-assessed workload during the trials when the participant had automated tracking.

Research Question 2. Cognitive workload and SA were inversely related for Study 2 when analyzed through SART and NASA-TLX ratings. This finding may be explained by analyzing the participants' perceived workload in the different conditions. The introduction of automation was a key difference between Study 1 and Study 2 and can help explain the significant findings seen across participant groups Study 2. Participants were not aware that they were in a high or low workload condition during their time operating MATB-II. However, they were aware when automation was introduced, potentially providing a lower perceived workload. Previous research using MATB-II has used automatic and manual tracking to differentiate between low and high workload conditions (Heard & Adams, 2019). Therefore, the significant correlation between SART ratings and NASA-TLX ratings in Study 2 could be explained through the differences in the conditions.

Ha3: Cognitive workload and situation awareness are inversely related.

There are multiple possibilities that can describe the relationship between cognitive workload and SA when assessed in an automated environment (Wickens, 2008b). One relationship investigated in this current effort was that they were inversely related. The statistical results indicate that this relationship does exist between the surrogate measures of SART and NASA-TLX ratings in Study 2. Therefore, we reject null hypothesis 3 that there was no relationship between cognitive workload and SA.

Participants had to gauge their subjective ratings of SA and cognitive workload after their time in the study was complete in the same manner seen in Study 1. The main difference between the two studies was the use of automation in the TRACK task. This task condition change meant that participants had an overt indicator of a difference in the simulation. However, it was difficult to ascertain what their SA was being rated against because the SART did not ask them for specific aspects of the trial. For instance, the SART questionnaire asked participants the value of the information they gained in the situation. This information could have varied greatly between participants, and there was no way of standardizing that across all participants with this instrument. Additionally, participants had fewer tasks to monitor when the automation was active. Participants were informed that they had no responsibility with the TRACK task when it was in "AUTO ON" mode. Therefore, it is possible that participants had additional information processing resources that might have contributed to a sense of increased SA.

2. Analysis Informing Study 3

Results from Studies 1 and 2 informed the design of Study 3. Study 2 provided insights into the effects of introducing automation while completing MATB-II tasks. Automating the TRACK task yielded results that indicate participants experienced lower workload when using automation. Participants in Study 2 reported higher subjective workload ratings using the CSWAG for the high workload and the low automation conditions. Combined with the results of Study 1, Study 2's results provided validation that the MATB-II scenarios again yielded different workload levels. These results were important to build on to Study 3, where cognitive workload would be investigated at dynamically changing levels of automation and workload demand conditions.

The same objective and subjective measure categories that yielded statistically significant differences in Study 1 continued in the same pattern in Study 2 with HRV, pupil diameter, and CSWAG. Based on these patterns, Study 3's collected cognitive workload measures were modified. HRV, pupil diameter, and CSWAG were chosen as measures to gauge participants' workload in Study 3. The use of HRV as a cognitive workload surrogate has shown to be effective in tasks where workload levels are continuously altered (Hughes et al., 2019). Additionally, pupil diameter has shown to be sensitive to rapid manipulations in cognitive workload (Aura et al., 2020). CSWAG was sensitive to different levels of cognitive workload and tracking mode presented in MATB-II. Because of the statistically significant differences in these collected measures, they were chosen to serve as the primary cognitive workload surrogates in Study 3.

However, fNIRS and SART were again not reliable surrogate measures of workload or SA, respectively. The use of fNIRS was eliminated for Study 3 due to non-significant results in the first two studies. While NASA-TLX and SART indicated a significant inverse correlation in Study 2, these results may not have been the most representative measure of SA. The proposed design for Study 3 would eliminate the mid-trial eye tracking calibration and loading of the second scenario. To investigate the effects of dynamically changing levels of automation, the final study would leverage a continuous 20-minute trial instead of two separate 10-minute trials. This difference in the administration of the MATB-II scenarios could potentially lead to different results and analysis based on the participants' groups that followed different condition progressions. The NASA-TLX and SART were once again included as post-trial questionnaires in Study 3 to determine if their relationship persisted with the introduction of dynamically changing levels of automation.

3. MLCC Review

The purpose of Study 2 was to investigate the following areas of the adapted MLCC framework: SA and response, workload, and LOA change. Study 2's results suggest that the participants were once again able to perceive differences in their experienced workload through their collected cognitive workload surrogate measures. Further, participants' performance scores indicated that higher levels of workload had an effect that produced lower levels of performance. These performance results follow the same pattern seen in Study 1. Cognitive workload measures were different between workload conditions and LOAs. These results followed IMPRINT cognitive modeling predictions as they did in Study 1. Further, the increasing progression of pupil diameters before, during, and after a communications event highlighted the resource demand conflict that was seen in the IMRPINT models. Building on the results of Study 1, these findings led to the assertion that the changes in the cognitive cybernetic loop were sensitive to changes in multiple conditions as seen in the cognitive workload measures and performance data.

The data analysis of Study 2 allowed for recommended investigation areas of the adapted MLCC framework in Study 3. In particular, the impacts of adaptive automation could now be assessed to determine changes in operators' performance and cognitive workload measures as LOAs dynamically changed. Cognitive workload and performance measures were sensitive to manipulations to task difficulty and automated assistance in the first two studies. Therefore, utilizing the MLCC framework would allow for investigation into how cognitive workload fluctuations relate to dynamically changing levels of automation. These cognitive workload changes could be modeled to allow for cognitive workload forecasting, which would serve as a key element leading into Study 3.

THIS PAGE INTENTIONALLY LEFT BLANK

V. STUDY 3

A. OVERVIEW

The final study in this research effort sought to assess workload forecasts based on changes in levels of automation. Study 3 continued with the use of MATB-II. However, participants conducted trial runs with dynamic changes in LOA during a continuous 20-minute trial. The mapping of the areas investigated in relation to the adapted MLCC framework is depicted in Figure 53.

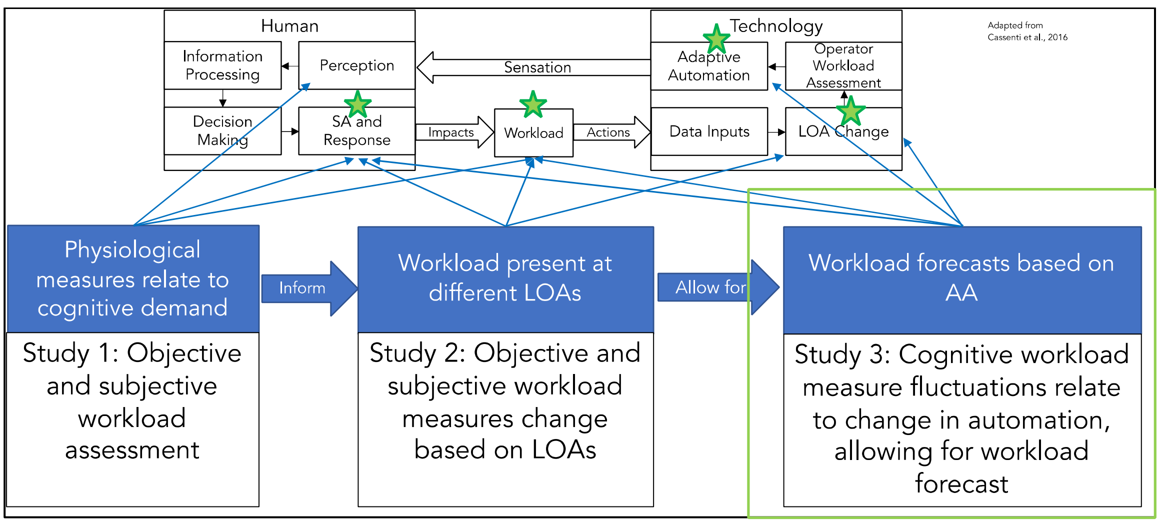


Figure 53. Study 3 mapping to the adapted MLCC framework.

The purpose of this study was to determine levels of workload present during dynamic changes in LOAs and workload. The interval of introducing changes to the automation were time-based, occurring in 5-minute increments during the 20-minute trial run. This interval was used in both Studies 1 and 2 because it followed the 5-minute MATB-II task-to-workload mapping and allowed for HRV R-R interval analysis (Delliaux et al., 2019; Malik et al., 1996). Additionally, the 20-minute interval allowed for the uninterrupted collection of more physiological data points. Investigation into SA and performance was conducted in the same manner as Studies 1 and 2. The approach used in Study 3 enabled analysis of the effects of dynamically changing LOAs and workload conditions on cognitive workload.

Study 3 was informed and designed with insights gained from Studies 1 and 2. MATB-II performance measures differed significantly between experience and workload levels in both studies. Heart rate variability and pupil diameter were sensitive to changes in workload conditions in the first two studies. Pupil diameter and HRV data did not appear to suffer from any data loss or quality issues over the two 10-minute periods in the first two studies. Therefore, consolidating the trial into one continuous 20-minute period was selected as the collection interval for Study 3. This 20-minute period would allow for insight into transition periods and more fidelity on post-trial assessments. The use of CSWAG during the trial runs was used again. However, the timing was shifted to begin at the 30 second mark and every 60 seconds thereafter. This shift in timing would allow for collection of CSWAG percentages 30 seconds before and after a system state transition. Further, IMPRINT workload value predictions followed the results of the first two studies in that higher workload conditions were associated with higher objective and subjective cognitive workload surrogate measures.

The results of Studies 1 and 2 did not yield significant differences between novice and experienced participants' objective and subjective workload measures. Therefore, the decision was made by the researchers to eliminate the novice group of participants in Study 3. While it is essential to investigate the impacts of cognitive workload when using AA with different experience levels, the results from the current effort did not support further investigation into this relationship. To ensure further accounting for any order or learning effects, the researchers decided to leverage the experienced group training progression in Study 3. This meant that participants received a one-hour training session before returning for their experimental data collection on a second day that was within 72 hours of their training day.

The supported research questions and hypotheses for Study 3 sought to answer the main objectives of the dissertation and are reviewed below.

Research Question 1: Can cognitive workload modeling inform design decisions in AA systems?

Ha_1: Effective cognitive workload modeling will reflect changes that occur as LOAs vary within AA systems.

Research Question 2: Do cognitive workload predictions forecast future performance in AA systems?

Ha_2: Cognitive workload measures can be used to predict future performance in AA systems.

Research Question 2a: Will unintended (or unanticipated) design consequences of AA systems emerge in the form of changes in performance?

Ha_{2a}: Unintended (or unanticipated) design decisions of AA lead to performance changes.

B. PILOT DATA

A pilot study was not conducted for Study 3. Changes to the execution of Study 3 included the elimination of fNIRS data collection and the mid-point break for eye tracking calibration. Because the execution of the rest of the study remained consistent with the first two studies, the researchers decided that a pilot study was not necessary for any data analysis ahead of Study 3.

C. PARTICIPANTS

1. Selection

The NPS IRB approved the research methods used for this study. There were no changes to the inclusion and exclusion criteria for this study. Participants were recruited across the NPS campus as in Studies 1 and 2. Consent forms were updated to reflect the changes in the conditions that were being investigated in Study 3. No participants that were enrolled in Studies 1 or 2 were eligible to participate in this study due to the exclusion criteria.

2. Demographics

Forty-three participants were enrolled in Study 3, with 40 participants completing the study (mean age in years=34.18, SD= 4.90). Participants included 29 males and 11 females. Of the 40 participants, 39 were in the military (8 in the U.S. Army, 10 in the U.S. Navy, 13 in the U.S. Marine Corps, 5 in the U.S. Air Force, 3 in foreign militaries, and 1 Department of the Navy civilian). The military participants' military occupational specialties again ranged from operations to operational sustainment. All participants were graduate students or employees at NPS. The rank breakdown of military participants is depicted in Table 14. The participants' time in service ranged from 5 to 25 years of service (M=11.99 years, SD=5.52). All participants met the screening criteria listed in the inclusion and exclusion criteria.

Table 14. Study 3 participants' military rank.

Participant Rank	Number
E-7	1
O-3	18
O-4	16
O-5	4
Civilian	1
Total	40

D. MATERIALS

The only major change in the materials used for this study was the elimination of the NIRSport system. Because no statistically significant differences were found between experience groups, workload conditions, or levels of automation in the first two studies, fNIRS analysis was omitted for Study 3. The remaining configuration, materials, and workstation were the same used in Studies 1 and 2.

variables

1. Independent Variables

The two independent variables manipulated in this study were workload and tracking condition. Presentation of the workload and tracking levels were counterbalanced to account for order effects. The conditions used in Study 3 are shown in Figure 54. All participants were presented with both low and high workload conditions using McCurry et al.'s (2022) proposed task distribution. Participants were randomly assigned to one of the four groups shown in Figure 54.

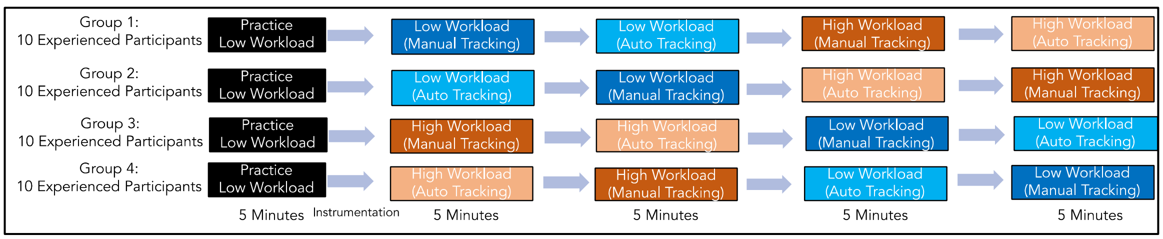


Figure 54. Study 3 conditions.

The researcher again assigned tasks throughout the scenarios in accordance with the parameters in Table 15.

Table 15. Study 3 MATB-II system settings for each condition.

	System Monitoring	Tracking	Communications	Resource Management
Low Workload (Manual Tracking)	11 Events	Low Joystick Response High Update Rate	3 Events	1 Pump Failure 1 Pump Shutoff
Low Workload (Auto Tracking)	11 Events	Automatic	3 Events	1 Pump Failure 1 Pump Shutoff
High Workload (Manual Tracking)	20 Events	Low Joystick Response High Update Rate	12 Events	10 Pump Failures 10 Pump Shutoffs
High Workload (Auto Tracking)	20 Events	Automatic	12 Events	10 Pump Failures 10 Pump Shutoffs

2. Dependent Variables

Study 3 used the same dependent measures as Studies 1 and 2 except that fNIRS data were not collected. The measures collected included the following.

- Performance metric: MATB-II FOM score.
- Subjective workload: CSWAG and NASA-TLX.
- SA rating: SART.
- Physiological metrics: Eye tracking and heart rate.

E. PROCEDURE

1. Participants

Participants enrolled in Study 3 completed the same training progression as the experienced participants in Studies 1 and 2. Prior to beginning their experimental trial runs, participants were given the same instructions as participants in Study 2 that reminded them that they would experience the TRACK task in both "MANUAL" and "AUTO ON" modes. The researcher then instructed the participants to recall their training to determine when the TRACK task was in those different modes. Participants also were reminded that the MATB-II reference sheet was available to them to assist with remembering how to determine the TRACK mode. The key difference in Study 3 was that participants conducted the experimental run in one 20-minute trial instead of two 10-minute trials. This approach was used to investigate the effects of dynamically changing levels of automation without interruption from any recalibration or changing simulation scenarios. All participants experienced low and high workload conditions with manual and auto tracking for each condition.

2. IMPRINT Modeling

The researcher constructed four IMPRINT models that reflected the scenarios to be presented in Study 3. The models for Study 3 were built from the models used in Studies 1 and 2, because they used the same MATB-II event files. These models were updated to reflect the accurate timing execution of the COMM and TRACK task. The researcher developed IMPRINT models using default anchors provided by the system to generate one set of models. The researcher-derived task network diagram for Study 3 is shown in Figure 55.

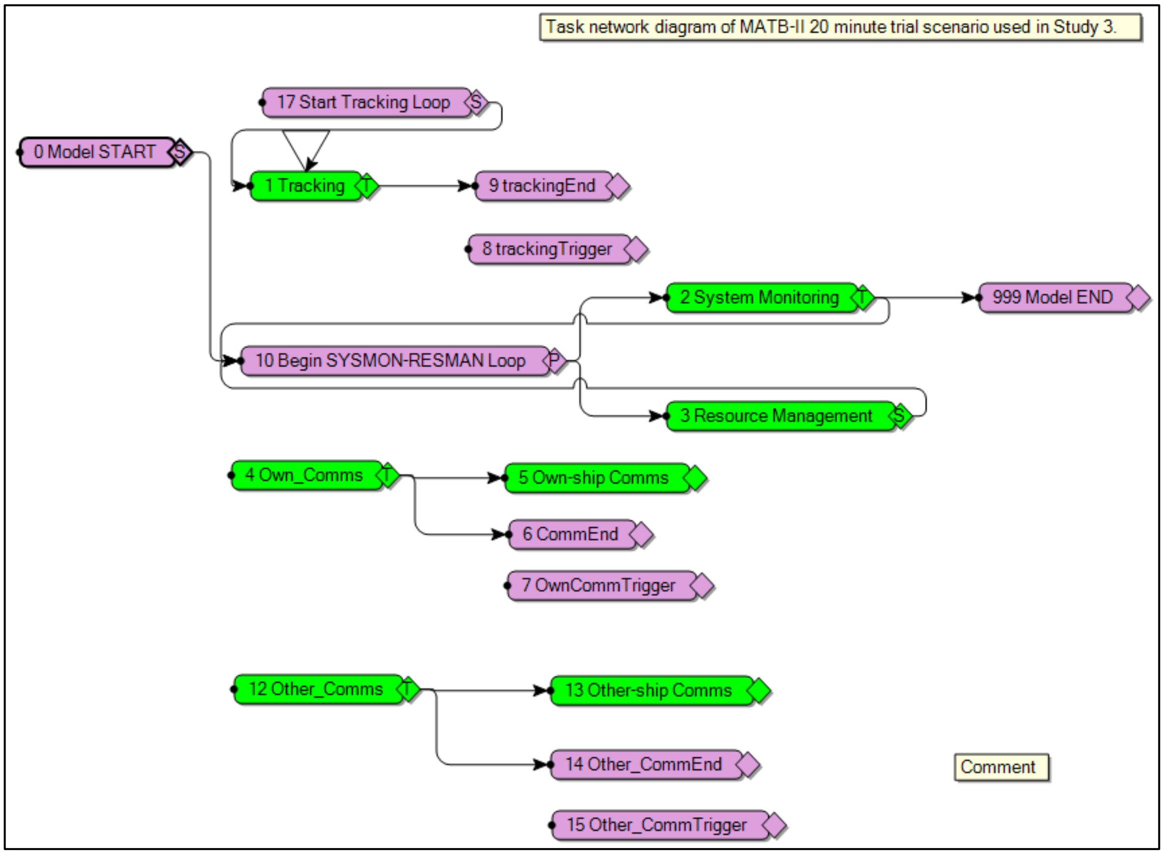


Figure 55. Study 3 IMPRINT researcher-derived task network diagram.

The same resource-interface demand values were used in Study 3 as were used in Studies 1 and 2. The research-derived IMPRINT-predicted time average workload values for each group's segments and total trials are listed in Table 16. The researcher-derived IMPRINT model workload graphs for Group 1–4 and their corresponding external event triggers are contained in Appendix I.

Table 16. Study 3 researcher-derived time-weighted predicted workload values.

Group	Low Manual	Low Auto	High Manual	High Auto	Total
1	40.55	11.70	41.54	25.25	26.51
2	40.31	14.34	43.48	21.12	27.41
3	40.20	11.81	42.35	24.29	28.37
4	39.95	14.92	41.29	20.61	28.15

An additional set of models was developed using the expert feedback received for Studies 1 and 2. The key difference between the researcher-derived models and the expert-derived models was the inclusion of auto TRACK task with its associated workload requirements based on the expert feedback. The task network diagram reflecting the experts' inputs with the auto TRACK task is shown in Figure 56. The expert-derived IMPRINT model graphs for Groups 1–4 and their associated external event triggers are contained in Appendix I.

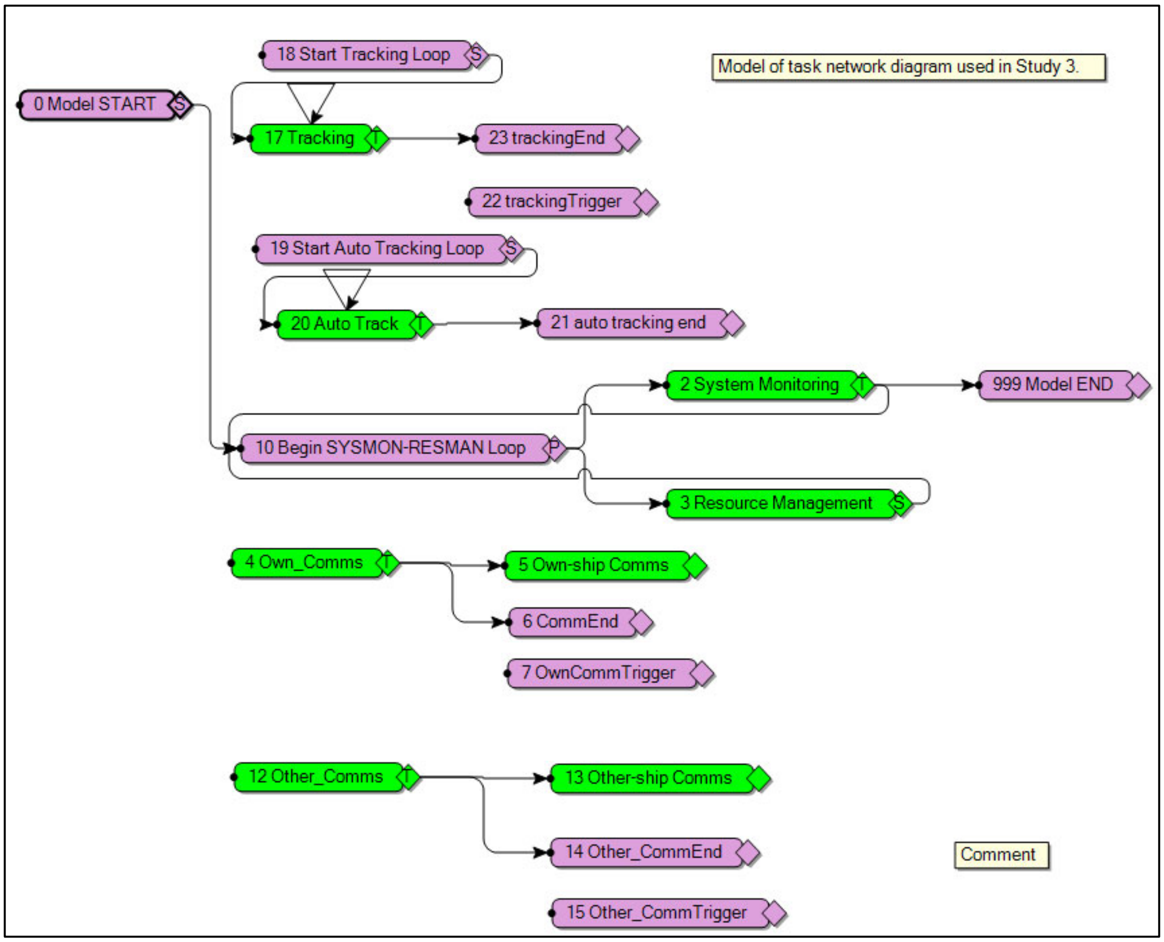


Figure 56. Study 3 IMPRINT expert-derived task network diagram.

The expert-derived IMPRINT-predicted time average workload values for each group's segments and total trials are listed in Table 17.

Table 17. Study 3 expert-derived time-weighted predicted workload values.

Group	Low Manual	Low Auto	High Manual	High Auto	Total
1	40.00	12.89	41.58	23.79	28.72
2	40.33	13.35	42.05	21.95	29.03
3	40.03	13.41	42.38	22.86	29.02
4	40.71	14.31	41.70	22.14	28.71

The predicted workload values were similar for both the researcher-derived and expert-derived models. In both instances, workload predictions followed the patterns seen in Studies 1 and 2 with higher workload conditions being associated with higher forecasted levels of workload. Additionally, the expert-derived workload predictions were slightly higher due to the inclusion of monitoring demands present during the auto TRACK task.

F. RESULTS

Data results were gathered in the same manner as in the first two studies using .xdf files created using LSL. Subjective and performance metrics were also included in the analysis. A mixed-effects model approach was used to analyze the data with JMP version 16.0.0. Fixed effects included workload level and automation condition. Analysis of transition periods was conducted with the addition of communications event timing to the model. Included as a random effect, participants were nested within groups. There were no statistically significant differences between group conditions for any of the collected measures. These results indicate that there was no effect of presentation order between the groups, and therefore order is not included in the summary results listed in Table 18. No participant data were excluded due to extreme values after analyzing residual plots for each modeled measure.

Table 18. Study 3 summary results table.

Measure Category	Measure Type	Dependent Measure	High vs. Low Workload	Tracking Mode
Performance	MATB-II FOM	Composite FOM	*High workload -> lower FOM* $p < .001$*	*Manual Tracking -> lower FOM* $p < .001$*
Psycho-physiological	HRV	Mean HRV	$p = .087$	*Manual Tracking -> lower HRV* $p < .001$*
	Pupil	Mean Right Pupil Diameter	$p = .467$	$p = .410$
		Mean Left Pupil Diameter	*Low Workload-> smaller diameter* $p = .011$*	$p = .627$
Subjective Workload	CSWAG	Mean CSWAG	*Low Workload -> lower CSWAG* $p < .001$*	*Lower CSWAG in Auto Track* $p < .001$*
	NASA-TLX	NASA-TLX Rating	*Low Workload-> lower rating* $p < .001$*	*Manual Mode -> lower rating* $p = .012$*
Situation Awareness	SART	SART Rating	*Low Workload-> higher rating* $p < .001$*	*Manual Mode -> higher rating* $p = .012$*

Participants had lower FOMs in the high workload condition than the low workload condition (M=94.02, SD=2.88, SE=0.32 vs. M=95.25, SD=2.76, SE=0.30). This pattern was also present in the manual TRACK condition compared to the auto TRACK condition (M=94.09, SD=2.71, SE=0.30 vs. M=95.19, SD=2.96, SE=0.33). These results are shown in Figure 57. There were no statistically significant differences between FOMs in the 30 seconds before and after a system state transition.

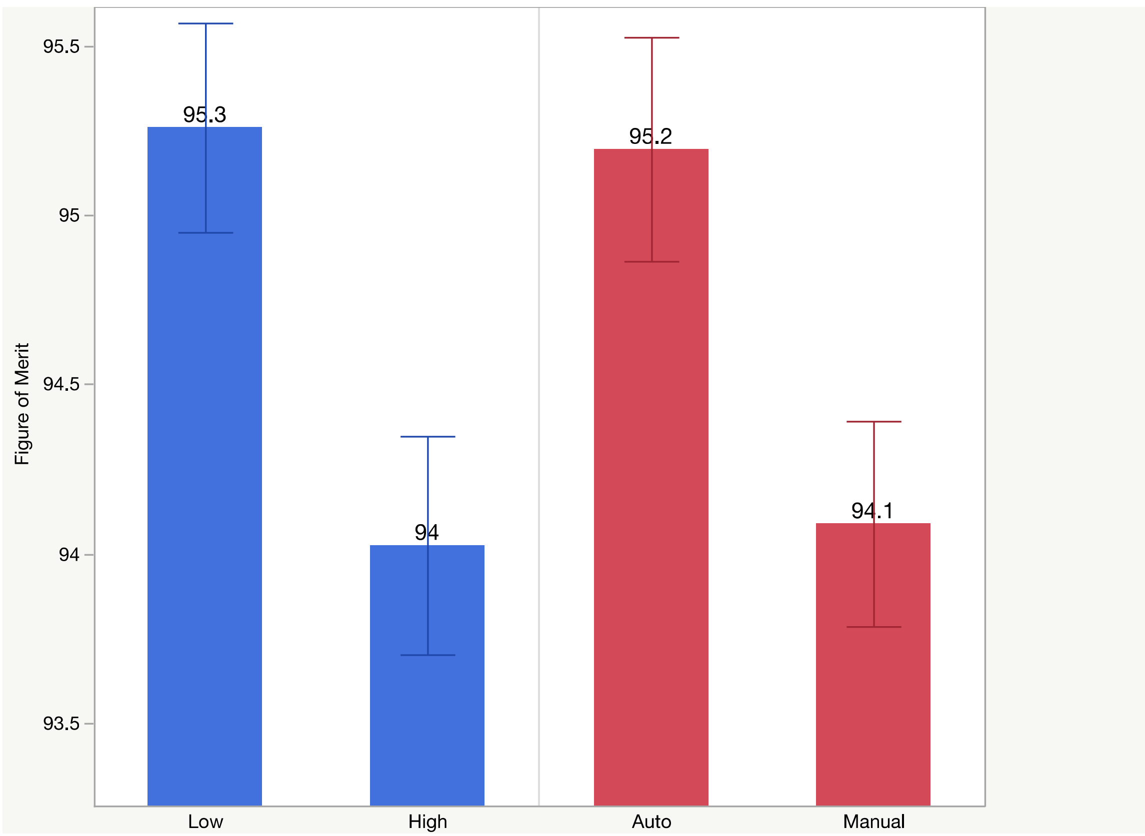


Figure 57. Study 3 FOMs by workload conditions (left) and tracking conditions (right). Error bars denote the standard error.

There were no statistically significant differences in mean HRV R-R intervals between workload levels in Study 3. However, Figure 58 shows the difference between mean HRV R-R intervals in the auto TRACK condition compared to the manual TRACK condition was statistically significant (M=845.23 milliseconds, SD=120.96, SE=13.52 vs. M=833 milliseconds, SD=121.09, SE=13.53).

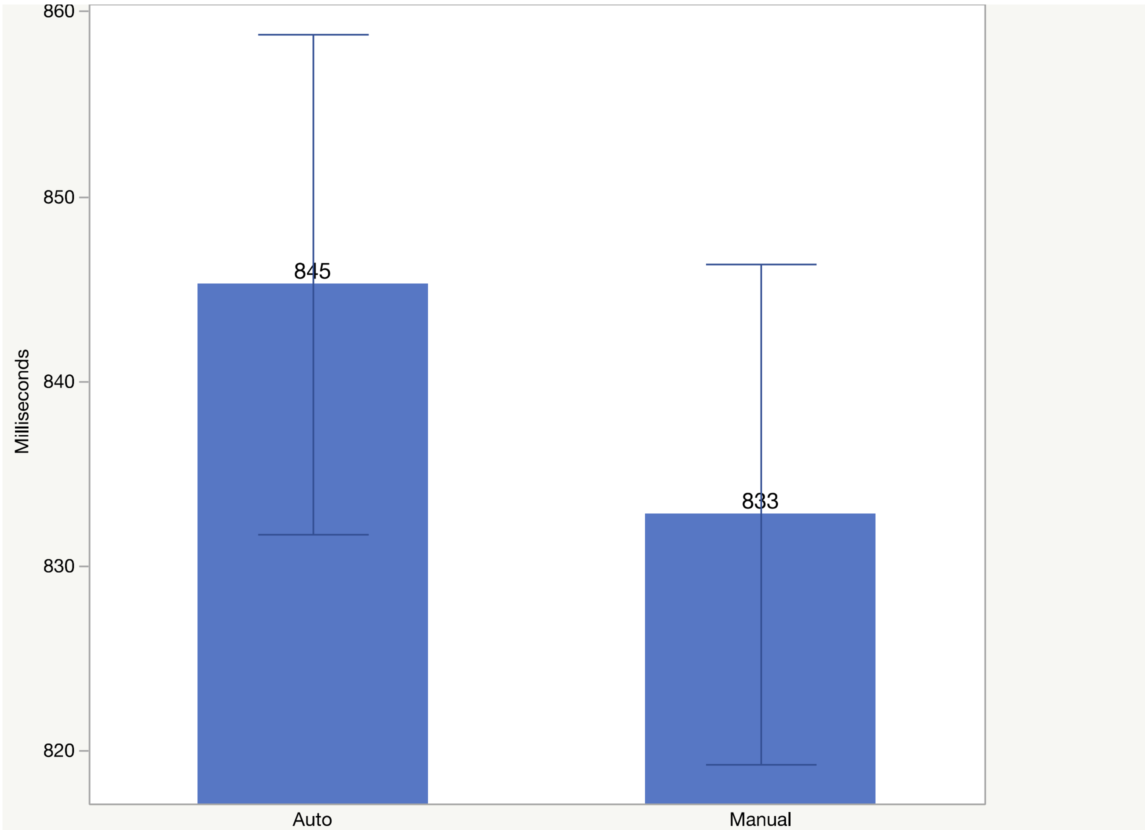


Figure 58. Study 3 mean HRV R-R interval by TRACK condition. Error bars denote the standard error.

There were no statistically significant differences in mean pupil diameters between TRACK conditions. However, there was a statistically significant difference between mean left pupil diameters in the low and high workload conditions as seen in Figure 59 (M=3.45mm, SD=0.93, SE=0.10 vs. M=3.40mm, SD=0.89, SE=0.09).

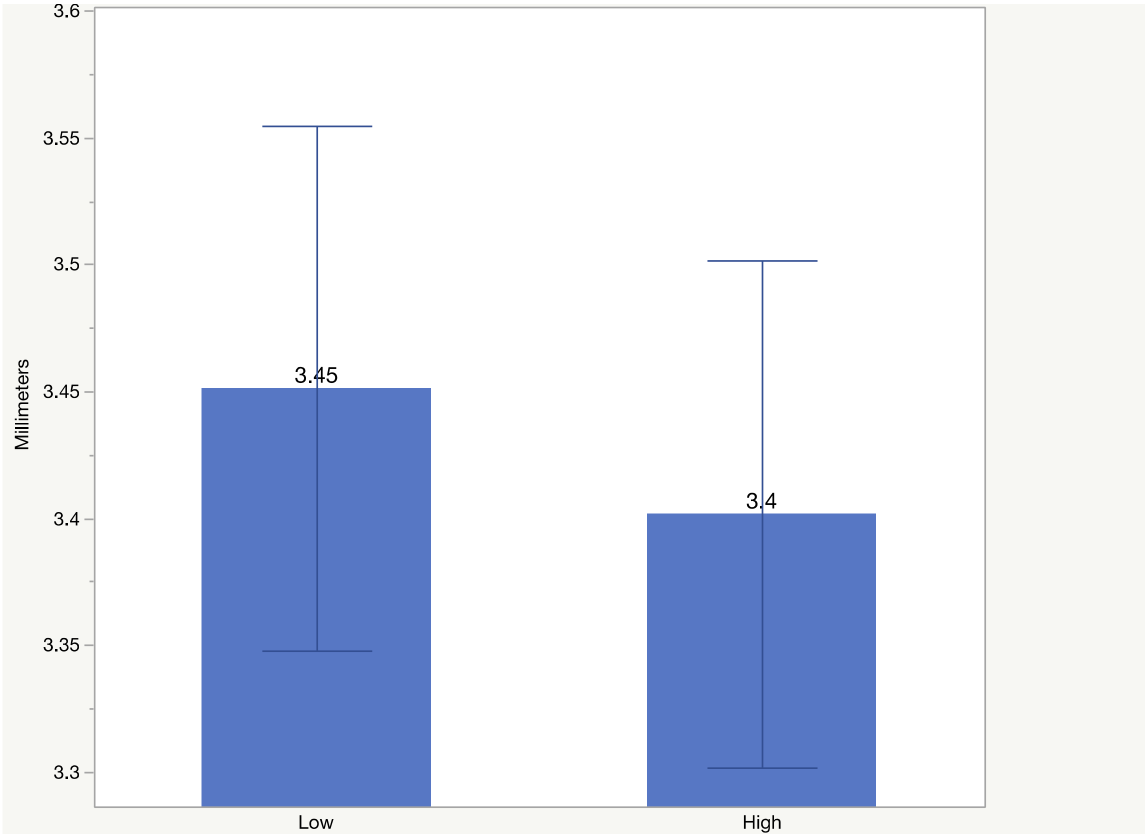


Figure 59. Study 3 mean left pupil diameter in millimeters by workload condition. Error bars denote the standard error.

Figures 60 depicts progressively larger pupil diameters in the five seconds preceding and following a communications event for left pupil $F(1, 277) = 5.70, p = 0.02$ (M=3.31mm, SD=0.86, SE=0.07 vs. M=3.36mm, SD=0.88, SE=0.07) and the right pupil $F(1, 277) = 8.70, p = .003$ (M=3.46mm, 0.94, 0.07 vs. M=3.51mm, SD=0.96 ,SE=0.08).

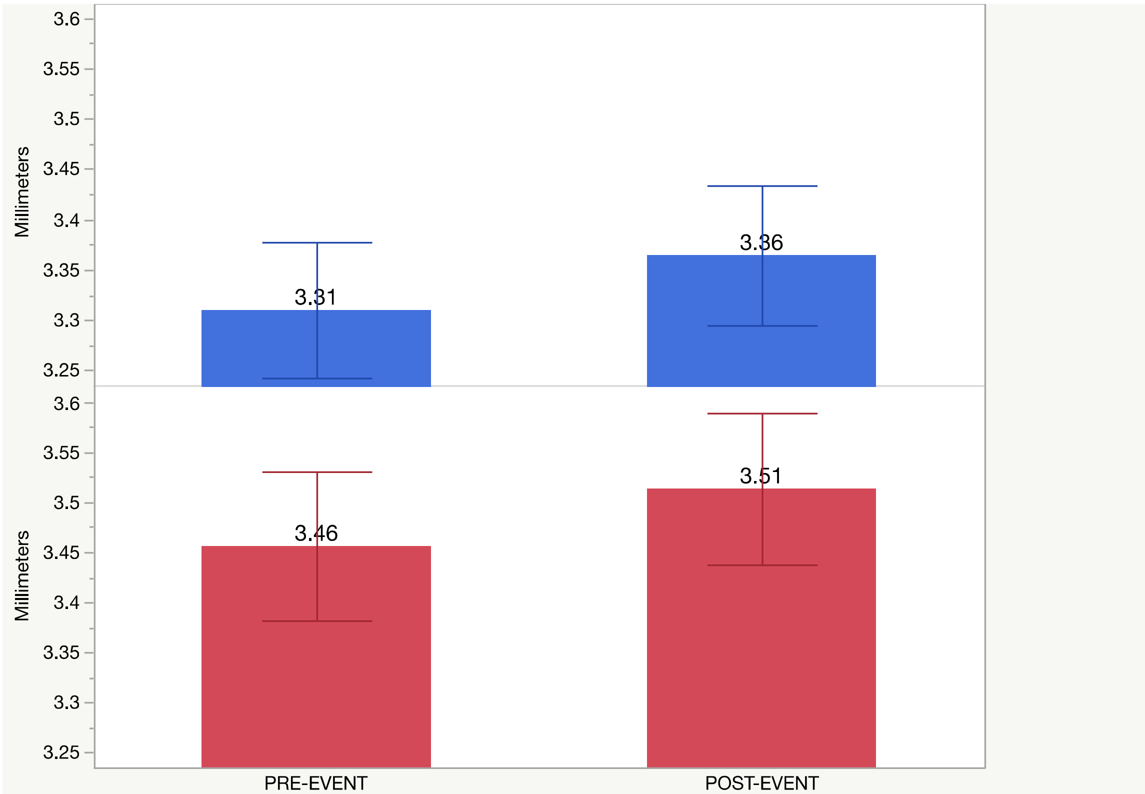


Figure 60. Study 3 mean left (top) and right (bottom) pupil diameters in millimeters by communications event timing. Error bars denote the standard errors.

Left pupil diameter differences in the same five second window were statistically significant between own and other ship communications as seen in Figure 61. This relationship was not statistically significant with right pupil diameters.

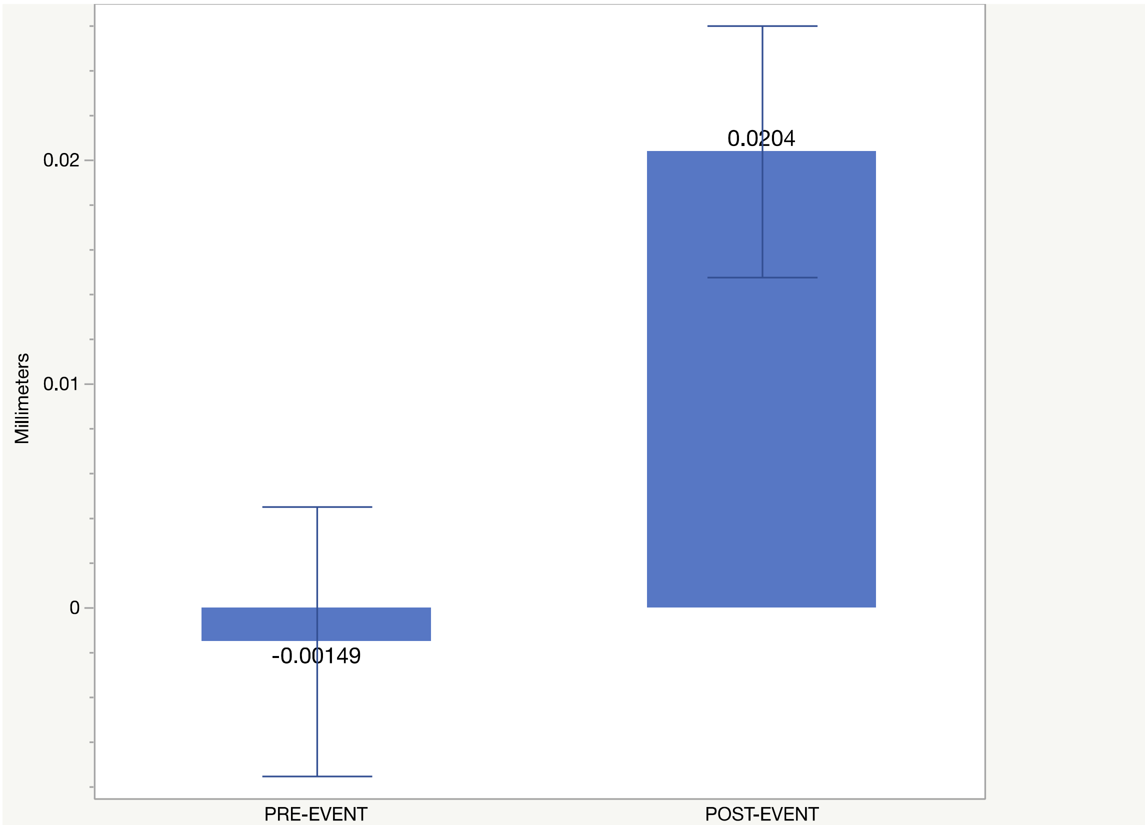


Figure 61. Study 3 mean left pupil diameter differences in millimeters between own and other radio transmissions in the 5 second window before and after a communications event. Error bars denote the standard errors.

There were statistically significant differences in both pupil diameters at the 5, 15, and 30 second windows surrounding a system transition. Study 3's mixed-effects model statistical results are listed in Table 19.

Table 19. Study 3 pupil diameters before and after a system state transition.

Transition Window	Left Pupil			Right Pupil		
	TRACK Mode	Pre-Post Transition Event Duration	Workload Condition	TRACK Mode	Pre-Post Transition Event Duration	Workload Condition
5	0.02*	0.0001*	0.26	0.13	0.02*	0.12
15	0.002*	0.001*	0.35	0.07	0.0004*	0.99
30	0.01*	0.01*	0.56	0.05	0.006*	0.84
60	0.04*	0.10	0.78	0.001*	0.08	0.80
90	0.26	0.42	0.58	0.001*	0.37	0.43
120	0.16	0.57	0.56	0.002*	0.73	0.57

Mean CSWAG differences were statistically significant between both workload and tracking conditions as seen in Figure 62. Low workload resulted in lower reported CSWAG percentages than in the high workload condition (M=38.79, SD=14.13, SE=1.58 vs. M=48.75, SD=13.30, SE=1.49). Automatic tracking also resulted in lower reported CSWAG percentages than manual tracking (M=50.51, SD=12.37, SE=1.38 vs. M=37.04, SD=13.49, SE=1.51).

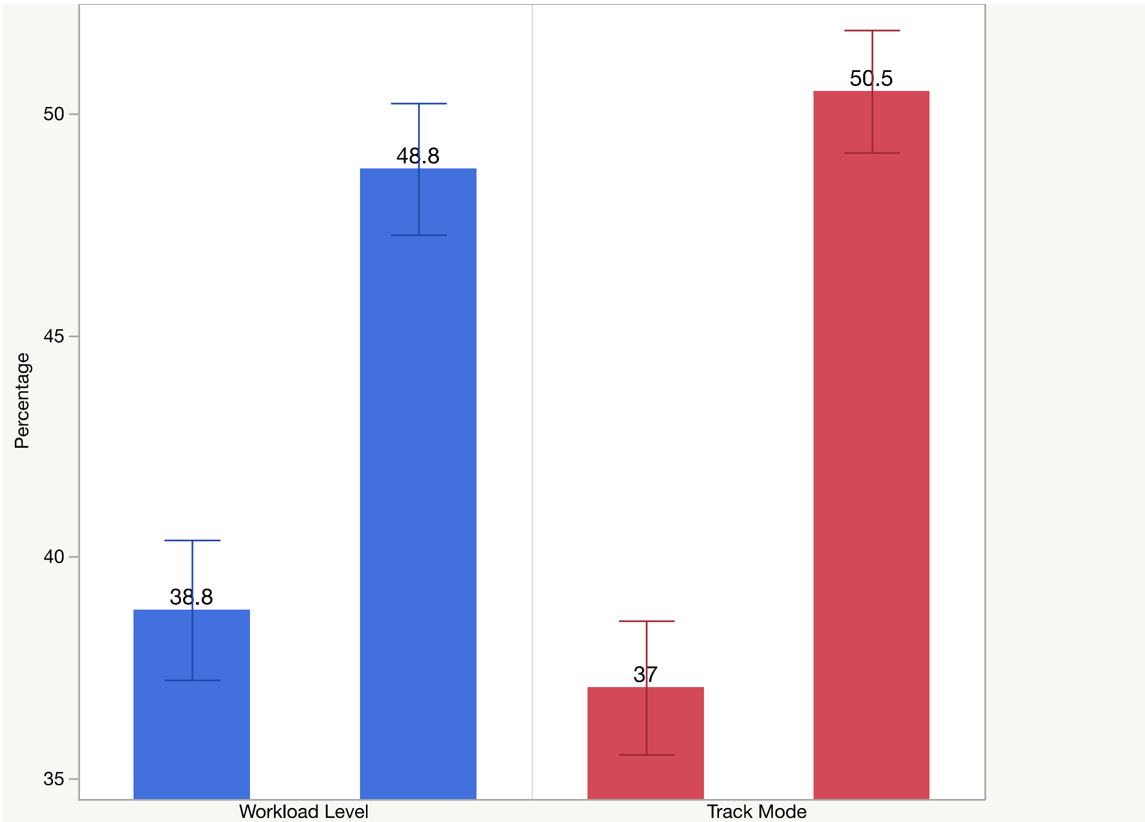


Figure 62. Study 3 CSWAG differences between workload (left) and tracking (right) conditions. Error bars denote the standard error.

Additionally, the pattern of workload differences between workload and tracking conditions followed in the 30 second window before and after a system state transition. Mean reported CSWAG was lower in the low workload condition than the high workload condition, $F(1,198) = 59.30, p < .001$ (M=40.13, SD=15.06, SE=1.37 vs. M=49.06, SD=14.91, SE=1.36). Mean reported CSWAG was also lower in the auto TRACK condition (M=38.73, SD=14.18, SE=1.29) than the manual TRACK condition, (M=50.46, SD=14.80, SE=1.35; $F(1,198) = 102.31, p < .001$). These differences are shown in Figure 63.

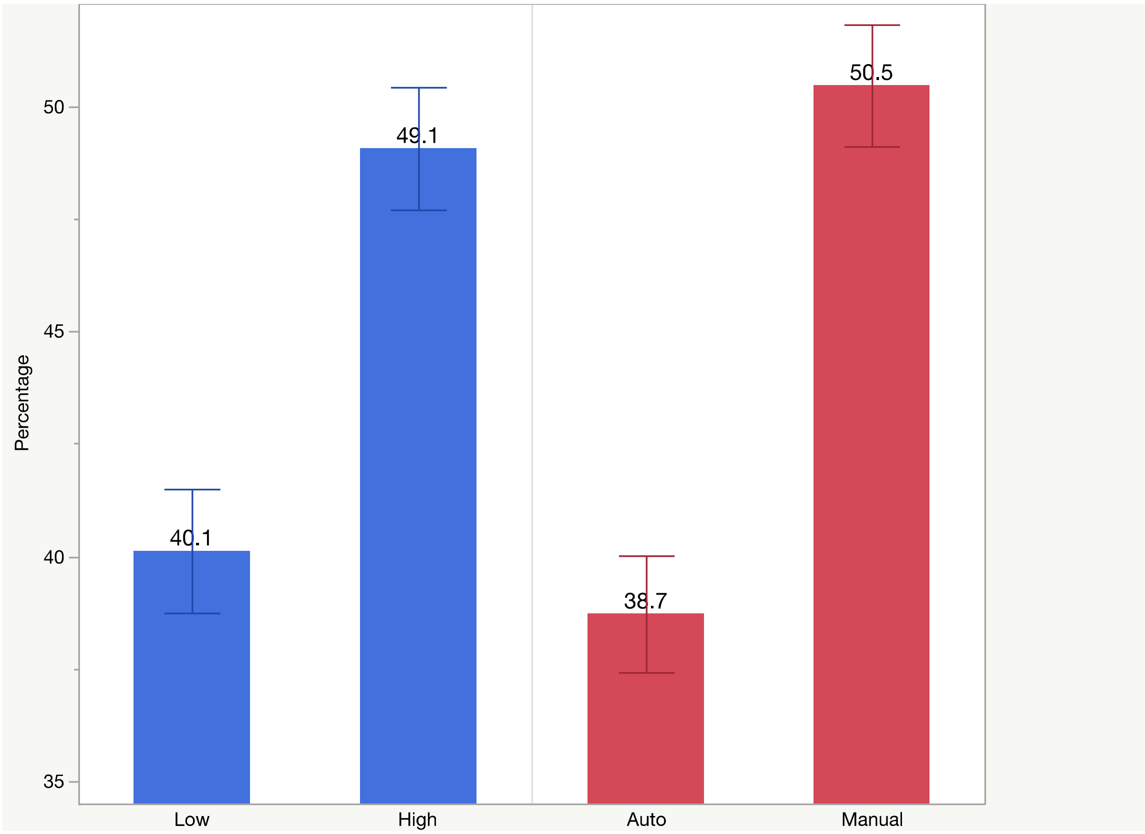


Figure 63. Study 3 CSWAG differences between workload (left) and tracking (right) conditions in the 30 second system state transition window. Error bars denote the standard error.

Post-trial situation awareness and cognitive workload ratings were collected using the SART and NASA-TLX, respectively. There was a statistically significant inverse correlation between SART ratings and NASA-TLX ratings, $\rho = -0.65, p < .001$. This relationship is illustrated in Figure 64.

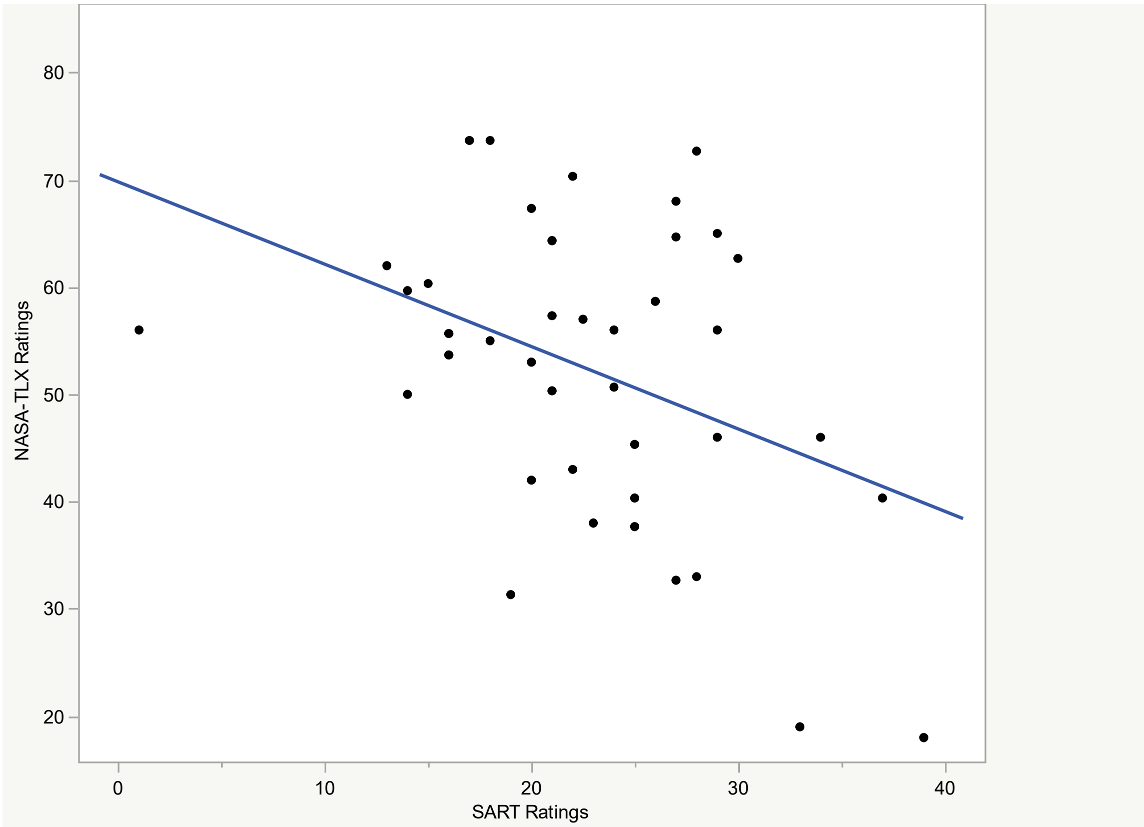


Figure 64. Pairwise correlation graph for NASA-TLX and SART ratings in Study 3.

G. DISCUSSION

The following section discusses the results of Study 3. The study's research questions and hypotheses are reviewed, followed by analysis of the study's associated statistical results. The analysis seeks to assess the cognitive workload forecasts made in the cognitive workload prediction models developed using IMPRINT. Finally, the discussion addresses specific aspects of the adapted MLCC framework that were investigated in Study 3.

1. Research Questions and Hypotheses

Research Question 1. Cognitive workload modeling followed the results of two physiological measures: HRV and pupil diameter. Subjective cognitive workload measures were also related to the IMPRINT predictions through the CSWAG and NASA-

TLX results. These patterns continue the trends seen throughout this dissertation and support the use of these measures to validate cognitive workload prediction models. Validated models using this approach can then substantively inform design decisions in AA systems to give additional understanding of the impacts on operators' states.

One of the more salient findings from Study 3 was the impact of system state transitions throughout the trial runs. The cognitive workload prediction models provided workload values without consideration for the impacts of transitions on workload. As dynamic shifts in automation occur in AA systems, it became evident that these changes should be modeled to provide a more accurate representation of changes in operator cognitive workload.

Ha_1: Effective cognitive workload modeling will reflect changes that occur as LOAs vary within AA systems.

Hypothesis 1 was supported through assessment of the time-weighted workload values for each LOA shift. The changes in cognitive workload modeling prediction values were supported in a more continuous manner with pupil diameter in the 30 second window surrounding a transition. Participants' left and right pupils were smaller in the 5, 15, and 30 seconds preceding a system state transition than they were in the same time window following a transition. At the 60, 90, and 120 second windows, there were no statistically significant differences for either pupil before and after a system state transition. However, there was a statistically significant difference in manual vs. automatic tracking at the 60 second window for both pupils. This result indicates that transitions in the system state in the 30 second window before and after the transition event contributed to increased workload. At the 60 second mark, manual tracking seemed to contribute to higher workload instead of the system state transition. This result highlights the time that it took for participants to gain in-the-loop familiarity and the workload resource demand cost associated with the system state transitions.

Participants were required to use both hands when using MATB-II. Mean right pupil diameters were larger during the system state transition windows than the left pupils (M=3.52mm vs. M=3.37mm). Figure 65 depicts the differences in p-values throughout the transition windows in both pupils. The differences in levels of statistical significance with pupil diameters at the transition windows can possibly be explained due to the variability in each pupils' mean diameters. These results follow previous research that found that right pupil diameters were larger than left pupil diameters in surgical residents completing tasks that required both hands (Cagiltay & Menekse Dalveren, 2012). Additionally, left pupil diameter has been associated with sympathetic activity in the autonomic nervous system (ANS), whereas right pupil diameter has been associated with parasympathetic activity (Burtis et al., 2014). The significant differences in pupil diameters in the left eye in the 5 to 60 second window may indicate sympathetic nervous system activation to address the increased demands during the tracking mode change. The differences in the right pupil from the 60 to 120 second window could indicate the parasympathetic response became dominant in participants when they perceived workload to demand fewer cognitive resources.

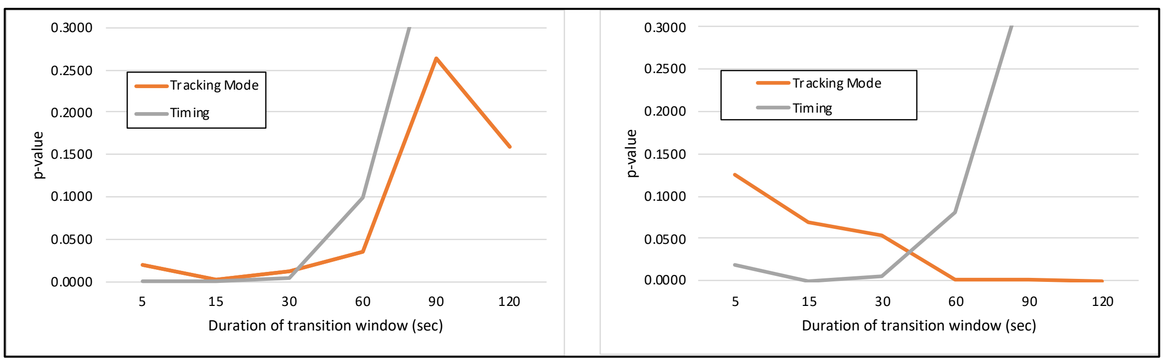


Figure 65. Study 3 p-value differences in left and right pupils during system state transitions.

Further, HRV was higher in the low workload and high automation conditions, suggesting that time-based measures of HRV are sensitive to changes in conditions when assessed over the appropriate time interval (five minutes in the case of each condition segment in Study 3). These results indicate that gaining in-the-loop familiarity is a measurable process when assessing AA systems. While AA can ultimately reduce workload, an unintended consequence of its introduction is a temporary increase in cognitive workload as seen with the physiological and subjective results in Study 3 (Kaber et al., 2001; Scerbo, 2008; Woods et al., 2021).

There was no significant difference in the presentation order of the workload and automation conditions in Study 3. This non-finding accounted for potential order effects in the research design. Further, this non-finding allowed for analysis of each condition's segment in the 20-minute trials. While there were differences between workload and automation conditions, the impacts of the transitions between them were also significant. Cognitive workload models must account for automation state transitions as drivers of workload to provide a complete picture of the cost of invoking and revoking AA (Goodman, Miller, & Rusnock, 2015; Rusnock & Geiger, 2017).

Subjective cognitive workload measures used in Study 3 supported Hypothesis 1. Mean CSWAG values were significantly different in the 30 seconds before and after a condition transition. These differences indicate that participants were able to subjectively assess the impacts of the systems' transitioning states. These results showed that experienced workload was significantly different between workload and automation conditions. Therefore, cognitive workload predictions must address these changes in cognitive workload during the transitions to give a more accurate prediction for the entirety of a task's model. Additionally, NASA-TLX ratings differed with statistical significance in Study 3 based on the group condition. This finding indicated a potential recency effect as participants might have rated their experiences according to the final condition that they experienced (Guastello et al., 2015). Participants who ended in higher workload conditions reported higher cognitive workload through the NASA-TLX and vice versa. The use of the TLX through a continuous study, as opposed to the segmented approach in the first two studies appears to have addressed concerns in the timing of its administration for Study 3.

Research Question 2. The MATB-II FOM results from Study 3 follow the pattern seen in the first two studies that show differences in workload and tracking conditions. However, this performance data did not differ significantly when assessed at the transition periods in Study 3. One potential explanation for this was that the version of MATB-II used in this study collected performance data every five seconds and used the mean FOM over that period to provide performance scores at the end of those five second periods. This configuration was a limitation of the system that did not allow for a more in-depth analysis of the impacts on performance as LOAs changed.

Ha_2: Cognitive workload measures can be used to predict future performance in AA systems.

Hypothesis 2 was partially supported with performance data in Study 3. Even though changes in system state conditions resulted in differences in workload measures at different LOAs, performance measures did not follow this pattern. Additionally, cognitive workload predictions were generally associated with performance results. As workload values increased, performance scores decreased. However, this trend was not present around the system state transition windows. This finding suggests that performance does not always directly associate with cognitive workload. For instance, participants could be achieving the same levels of performance, but their associated workload may be different based on individual factors such as experience (Guastello et al., 2015; Patten, Kircher, Ostlun, Nilsson, & Svenson, 2006).

Research Question 2a. Changes in performance were not significantly different when analyzing the system state transition windows in Study 3. Therefore, it was difficult to ascertain if any changes in performance manifested as unintended consequences when the system state was transitioning. However, changes in objective and subjective surrogate measures of cognitive workload suggested times when unanticipated consequences of LOA shifts manifested around these transition windows. For instance, right pupil diameters were significantly different in the 30 seconds surrounding a system transition. Then, differences in the right pupil diameter were significantly different depending on the tracking mode. The 30–60 second transition window differences in right pupil diameter indicate that participants were actively addressing the system state change for at least 30 seconds after the transition. After this 30 second period, the system state, not the transition, was the main driver of experienced workload. This finding is important for the current effort as it highlights an unanticipated consequence of the dynamic shift in LOA: there will be a time cost as operators deal with the transition of system states that must be accounted for in predictive workload models and ultimately, system design.

Ha_{2a}: Unintended (or unanticipated) design decisions of AA lead to performance changes.

Hypothesis 2a was supported with MATB-II performance data. Participants performed better in the lower workload and higher automation conditions than in the higher workload and lower automation conditions. These results are intuitive and follow the pattern of statistically significant differences seen throughout the studies in this dissertation. These findings suggest that operators' performance is sensitive to changes in LOAs. Therefore, it is important to consider LOAs to provide to an operator when assessed against other requirements that happen concurrently in a completion of a task.

2. Cognitive Workload Models with AA Transitions

The results of Study 3's experimental analysis indicated that there were transition periods between AA states that needed to be modeled. These modeling efforts would help refine workload predictions that occur around the AA transitions using workload demand values provided by the default anchors in IMPRINT. Figure 66 shows an updated IMPRINT task network diagram for Study 3 that adds a state transition task to the model.

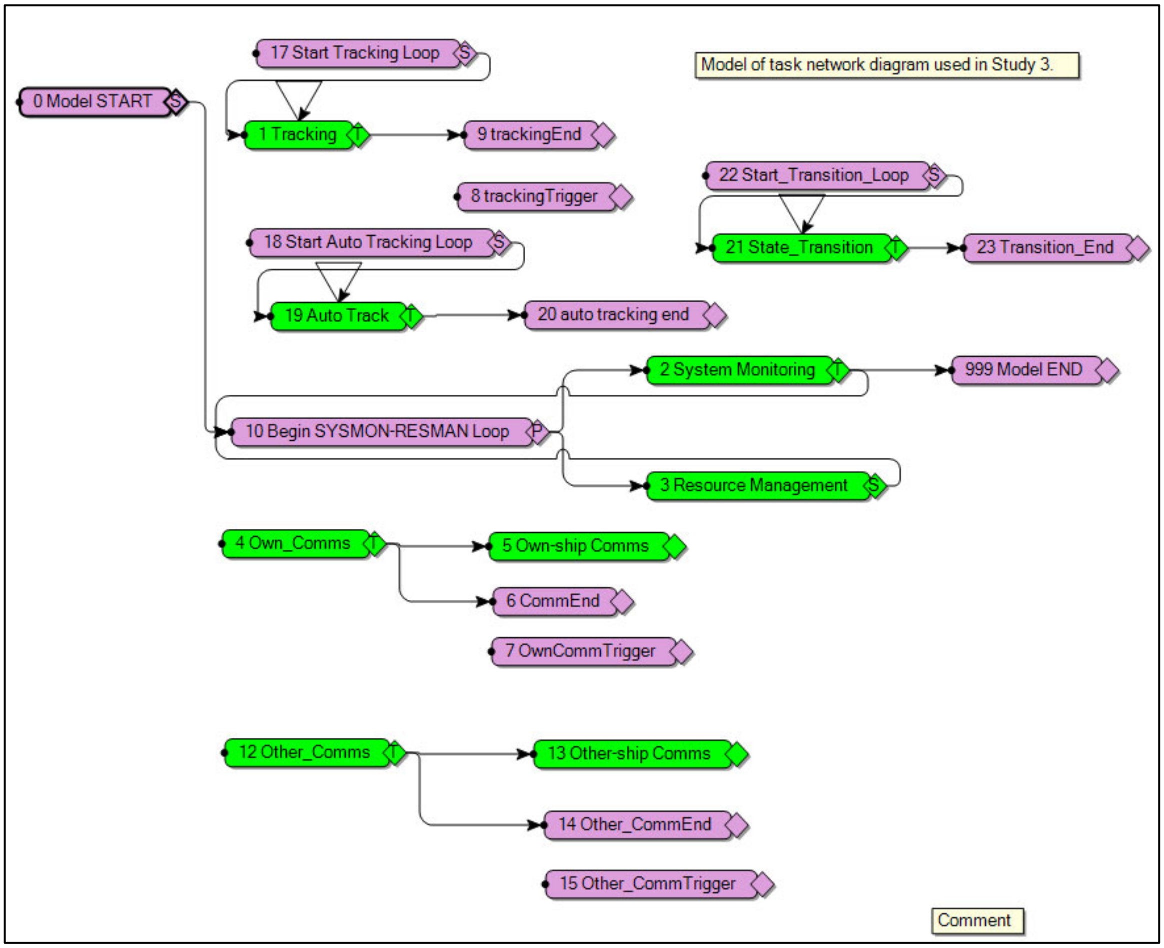


Figure 66. Refined Study 3 IMPRINT task network diagram with the state transition task added to the model.

The updated MATB-II resource-interface pair cognitive workload demand ratings are provided in Table 20. The workload demand ratings for the system monitoring, tracking, communications, and resource management tasks are based on the consolidated expert ratings used in Study 2. These ratings were used since they included expert-derived values for the automated tracking condition. The key difference between this the ratings used in Study 2 was the addition of research-derived default anchor ratings for the transition task. The analysis from Study 3 indicated that additional workload needed to be modeled at these transition points. Because the MATB SMEs interviewed for this dissertation did not experience transitions such as those experienced in the current effort, the researcher leveraged the default anchors in IMPRINT to update the model.

The interface interaction selected as the cognitive workload driver for the transition task was the operator and the crew station. The crew station was modeled as the whole of the interfaces the participants interacted with in MATB-II. The researcher selected this interface because an assumption of the updated model was that multiple interfaces of the MATB-II system demanded resources during the transition at the same time rather than one or two (i.e., not just the mouse or joystick). This demand was then assumed to create increased workload overall across all the interfaces. Therefore, the system state was modeled as an additional task that is listed at the bottom of Table 20.

Table 20. Study 3 workload demand ratings with state transition demands added.

Task: Tracking	**RI Pair Demand Values**						
Total Task Demand 11.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Joystick		3.60	3.90				3.80
Task: Tracking (AUTO)	**RI Pair Demand Values**						
Total Task Demand 1.67	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Joystick	0.00	0.67	0.00				1.00
Task: System Monitoring	**RI Pair Demand Values**						
Total Task Demand 7.87	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		2.40	3.30				2.17
Task: Resource Management	**RI Pair Demand Values**						
Total Task Demand 13.37	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		5.80	3.57				4.00
Task: Own Comms	**RI Pair Demand Values**						
Total Task Demand: 16.87	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse		2.07	4.07				1.83
Interface: Speaker	5.00	2.07					1.83
Task: Other Comms	**RI Pair Demand Values**						
Total Task Demand: 6.30	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Mouse							
Interface: Speaker	4.30	1.00					1.00
Task: State_Transition	**RI Pair Demand Values**						
Total Task Demand: 7.60	Auditory	Cognitive	Fine Motor	Gross Motor	Speech	Tacticle	Visual
Interface: Crew Station	1.00	1.00	2.60				3.00

RI Pair = resource-interface pair

The only external event trigger modifications in this model from the one used in Study 2 was the addition of the three times a system state transition occurred with the tracking task. Transition state event triggers occurred at the 5-, 10-, and 15-minute marks in accordance with the MATB-II event files. The updated IMPRINT-predicted time average workload values for each groups' segments and total trials with transitions included are listed in Table 21. These workload values reflect increases in predicted workload across all conditions.

Table 21. Study 3 updated time-weighed predicted workload values with differences from Study 3's predicted values in parentheses.

Group	Low Manual	Low Auto	High Manual	High Auto	Total
1	40.02 (+0.02)	13.69 (+0.80)	42.70 (+1.12)	28.15 (+4.36)	31.00 (+2.28)
2	40.89 (+0.56)	13.18 (-0.17)	45.16 (+3.11)	23.89 (+1.94)	30.63 (+1.60)
3	41.35 (+1.32)	16.40 (+2.99)	42.49 (+0.11)	24.13 (+1.27)	30.81 (+1.79)
4	41.29 (+0.58)	14.12 (-0.19)	42.11 (+0.41)	22.17 (+0.03)	29.78 (+1.07)

The results of the refined IMPRINT workload models are graphically depicted in Figures 67–70. The 30-second state transition periods are highlighted on each graph. Increases in predicted workload are seen throughout most but not all the trial runs at the transition periods.

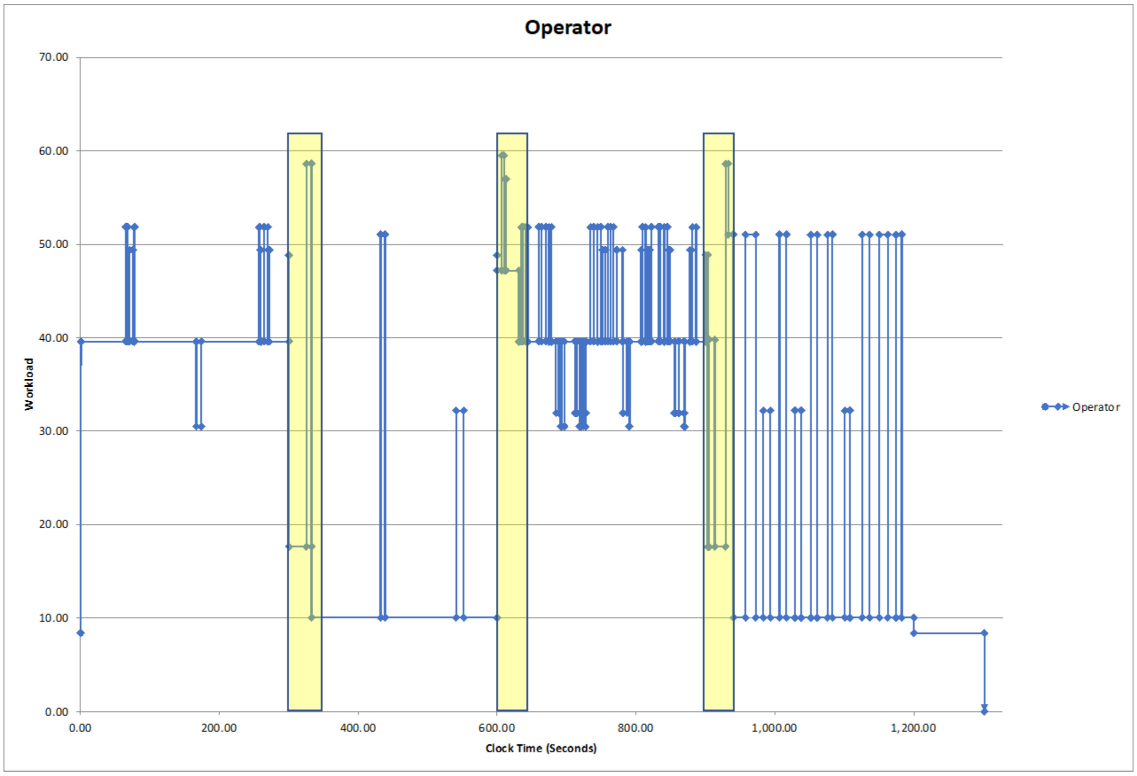


Figure 67. Study 3 group 1 IMPRINT workload graph with state transition highlighted.

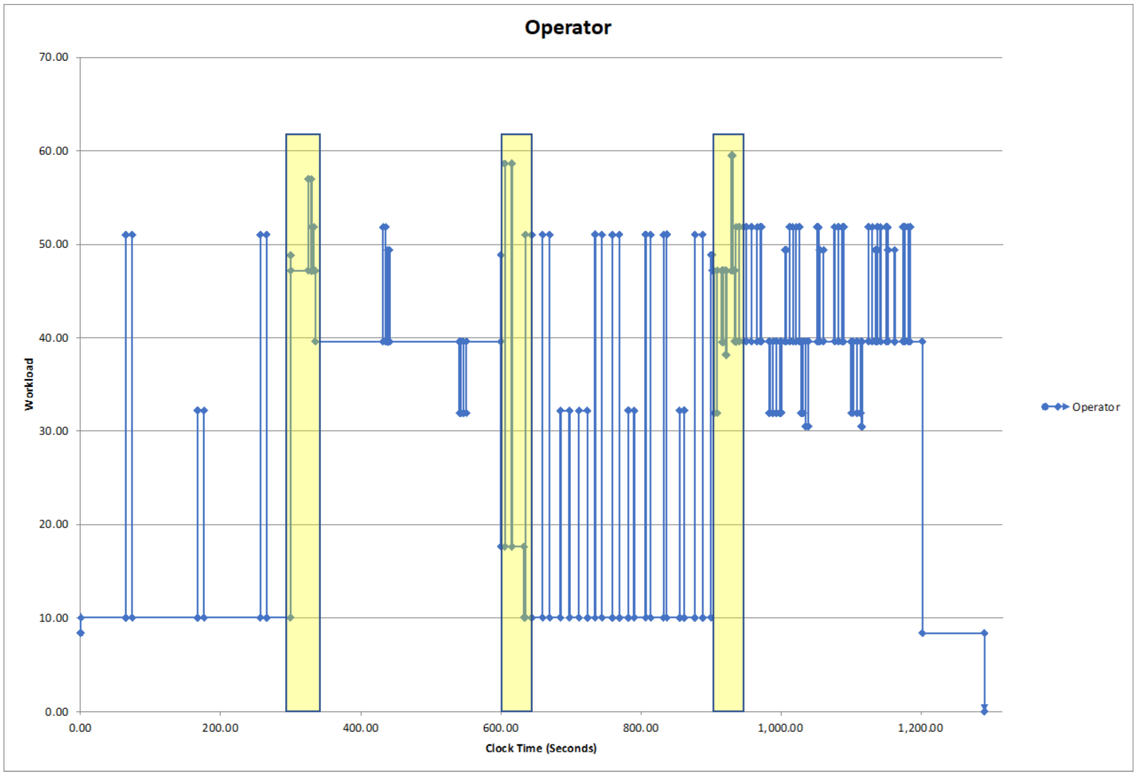


Figure 68. Study 3 group 2 IMPRINT workload graph with state transition highlighted.

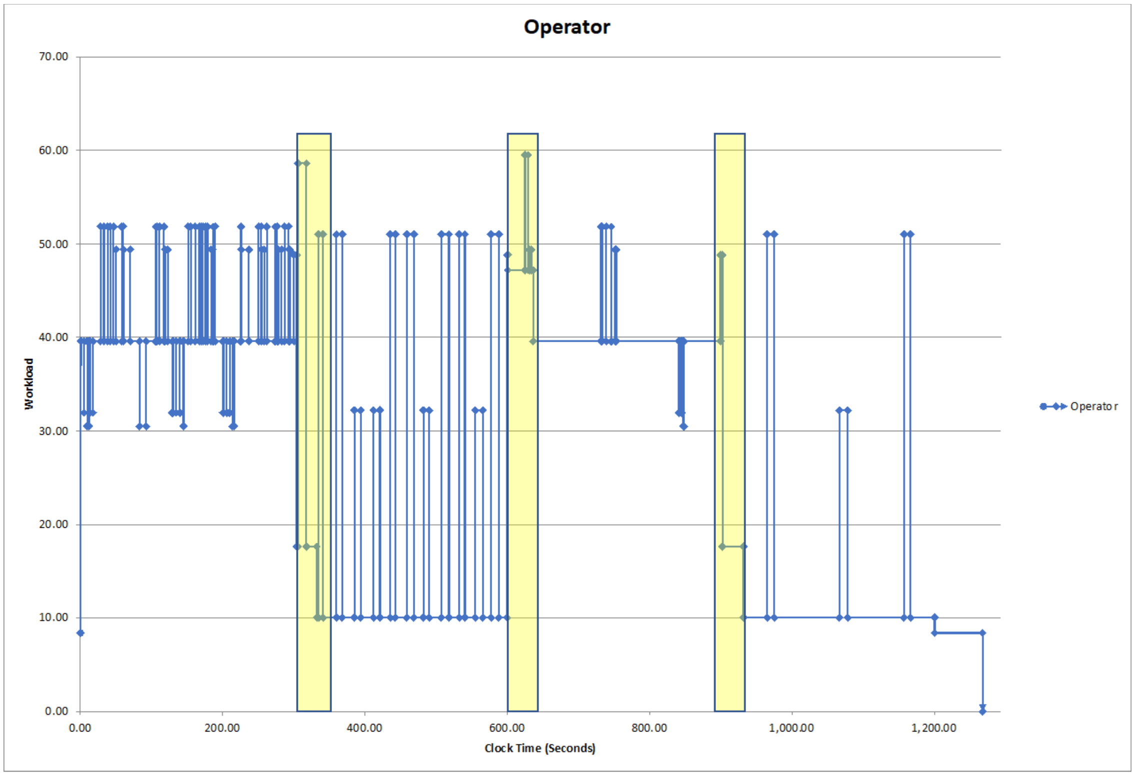


Figure 69. Study 3 group 3 IMPRINT workload graph with state transition highlighted.

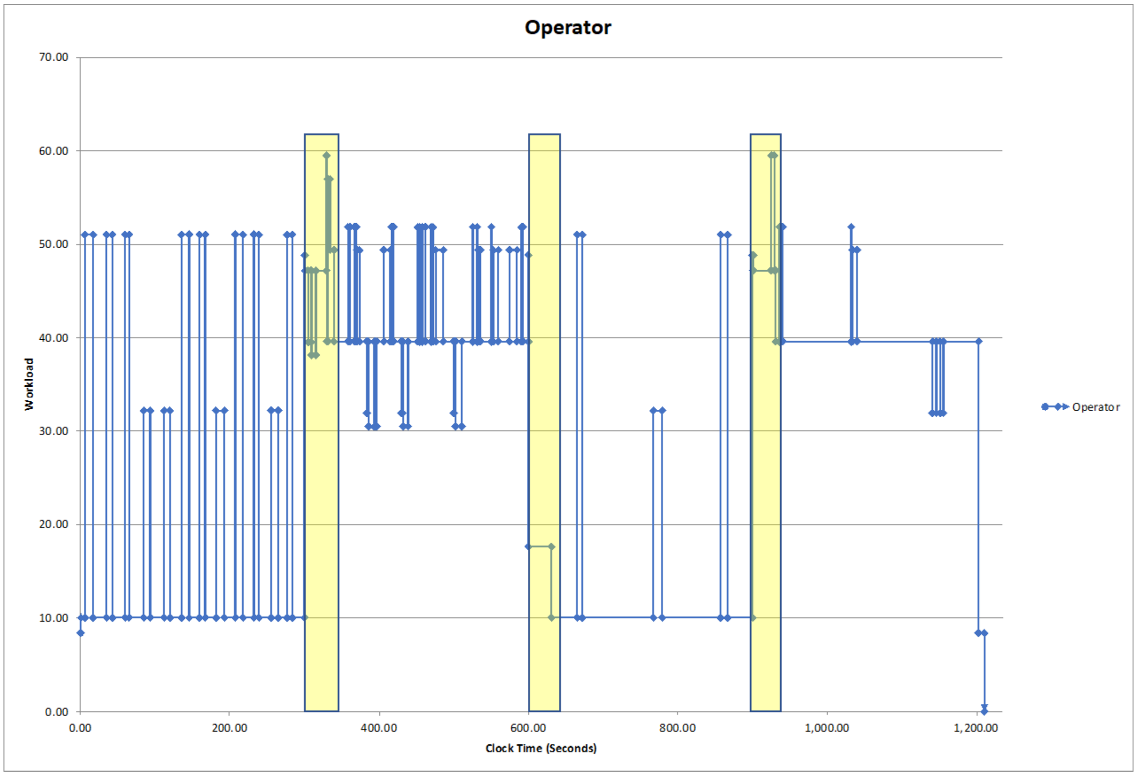


Figure 70. Study 3 group 4 IMPRINT workload graph with state transition highlighted.

Participants' CSWAG ratings were found to be sensitive to different workload and automated tracking conditions in Study 3. The composite results of all participants' CSWAG ratings are overlayed on the group 1 workload prediction model in Figure 71. The results depict the differences in conditions and show general pattern alignment with the workload prediction values provided by IMPRINT. The CSWAG ratings began 30 seconds after initiating the experimental trial, and subsequent ratings were elicited every minute thereafter. Additionally, CSWAG elicitations were delayed when there was a communications task ongoing to mitigate any confounding audio demands on the participant. Therefore, direct mappings to subjective workload experienced during communications tasks was difficult to ascertain. However, the increased pupil diameters seen during communications task windows highlight the workload demand spikes in the IMPRINT models at those specific intervals.

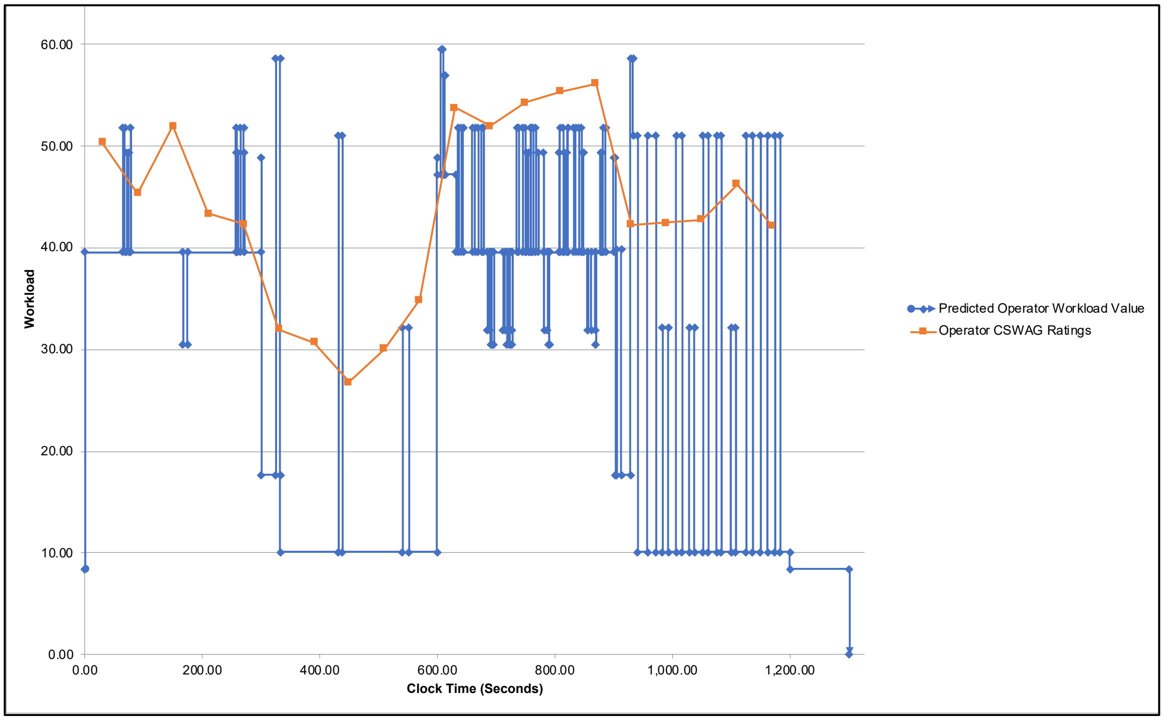


Figure 71. Study 3 group 1 IMPRINT prediction model with composite CSWAG values overlayed.

3. MLCC Review

The purpose of Study 3 was to investigate cognitive workload fluctuations in relation to dynamically changing LOAs. Study 3 addressed SA and response, workload, LOA change, and AA as parts of the adapted MLCC framework. This investigation allowed for assessment of cognitive workload forecasts made based on the results of the first two studies. Study 3's results were consistent with the previous two studies' outcomes. Participants' performance and surrogate workload measures again differed between workload and tracking conditions.

Situation awareness and participant response were inversely related to cognitive workload when assessed through the post-trial SART and NASA-TLX questionnaires. Participants reported higher levels of SA when their workload was lower and vice versa. These results follow previous work that highlights the dynamic relationship between SA and cognitive workload (Endsley, 2021; Kaber & Endsley, 2004; Wickens, 2008b). While increased SA may decrease experienced workload, this relationship is only truly beneficial when SA is sufficiently grounded in truth to mitigate any unanticipated changes in regaining in-the-loop familiarity.

Study 3 continued with the emphasis of using multiple cognitive workload measures to gain an understanding of an operator's state. Time interval analysis from HRV data was leveraged again in Study 3. The R-R interval HRV data followed previous studies that showed sensitivity to this metric in a five-minute window (Delliaux et al., 2019; Malik et al., 1996). Higher workload conditions resulted in lower HRV again in Study 3. These results support the continued use of HRV when assessing longer duration tasks within a cognitive cybernetic framework.

Dynamic changes in LOAs lead to changes in operator state that then lead to AA shifts according to the adapted MLCC framework. The results from Study 3 indicate that these LOA changes impact an operator's state when looking at objective and subjective data around the times of the automation transitions. This finding can help inform not only the recommended level to adjust the AA, but also the way those transitions occur. The transition in system state should not be done in a simple on or off manner, but rather with consideration for task responsibility handoffs (Rusnock & Geiger, 2017).

The use of the adapted MLCC framework provided a bounded guide for this dissertation research. The inextricable relationships between key concepts such as SA, workload, dynamically shifting LOAs, and the delivery of AA are accounted for in this framework. Further, being able to assess impacts of each element on the cybernetic loop with continuous feedback is an effective method to generate future analysis of the unintended consequences of AA systems.

VI. DISCUSSION

The use of adaptive automation can help decrease cognitive workload, but it also can unintentionally contribute to increased cognitive workload. To complicate matters, cognitive workload is difficult to model and measure. This is not a new issue and is in fact well-documented (Cain, 2007; Hancock & Matthews, 2019; Lohani et al., 2019; O'Donnell & Eggemeier, 1986; Vogl et al., 2020). The current effort sought to assess cognitive workload through predictive modeling and experimental validation. The multi-modal approach to this research attempted to provide a more complete assessment of cognitive workload in AA systems. This chapter discusses the findings of this investigation into the emergent negative unintended consequences of AA on cognitive workload.

Main findings indicate that there is a system state transition cost on an operator's workload in AA systems that must be accounted for in human performance prediction models and system design. Additionally, human operators can suffer from negative unintended consequences of AA systems with improper salience as different system state modes are presented to them. The failure to properly detect mode changes will impact operator SA. While an aim of AA is to manage cognitive workload within optimal levels, the effect of these workload adjustments may be delayed as control transitions between humans and the technological system. Operators appear to suffer decreases in SA as their workload increases in AA systems.

Additional findings emerged from the analysis of the three experimental studies and their corresponding cognitive workload models. Some of these additional research contributions included an analysis of different workload measurement techniques, an illustration of cognitive workload modeling techniques to predict future performance, and the role of both training and human factors considerations in system design. Research limitations and future work recommendations conclude the discussion.

A. PRIMARY CONTRIBUTIONS: ASSESSING THE UNINTENDED CONSEQUENCES OF ADAPTIVE AUTOMATION

This dissertation saw multiple unintended consequences of AA emerge in Studies 2 and 3. Adaptive automation can lead to unintended consequences (Kaber et al., 2001; P. Smith & Baumann, 2020; Tonn & Stiefel, 2019; Woods, 2016). This investigation specifically into AA's unintended negative consequences sought to investigate some of these consequences as the technological landscape continues to rapidly evolve, especially in the aeronautical field. This dissertation effort has yielded insights to facilitate future research and design in this area and beyond. While AA has shown the potential to alleviate excessive cognitive demands on humans, the results from this work have highlighted considerations for modeling and implementing AA in human-machine systems.

1. AA Transition Cost

The inadvertent increase in workload when automation was introduced follows previous findings that more demands are placed on an operator as they manage multiple components of a system (Bainbridge, 1982b; Endsley, 2017a). This research highlighted that transitioning a system state between manual and automatic tracking modes came at a cost to an operator's cognitive workload. The results from Studies 2 and 3 suggest that introducing higher levels of automation can increase cognitive workload for up to 30 seconds. This phenomenon is seen in the increased mean pupil diameter and CSWAG ratings around the time of system state transitions.

The introduction of adaptive automation may yield appropriate shifts in task allocation to facilitate manageable levels of workload. However, the act of introducing the automation in this manner comes with an unintended consequence of temporarily increasing workload and introducing unnecessary risk (P. Smith & Baumann, 2020). These transition-induced increases in workload can be explained through resource conflicts that arise in accordance with Multiple Resource Theory (Wickens, 1981, 2008a). These workload increases need to be accounted for in both modeling and system design to allow for proper handoff between the operator and machine.

2. System Indication Salience and Mode Determination Issues with AA Systems

An intended consequence of introducing AA in this research was to lower cognitive workload. This intent was realized through smaller pupil diameters, higher HRV, and decreased CSWAG ratings. While the intent of the AA was to lower cognitive workload, six participants in Study 2 did not experience this reduction due to incorrect interpretation of the system state. This emergent behavior was an indication of an unintended negative consequence of introducing AA. Increasing the number of system modes led to an increased number of mode errors. Even though training and cues were provided to operators to detect mode changes, they failed to recognize these changes on several occasions.

Increased training in the form of additional instructions mitigated this behavior for the rest of the participants in Study 2. However, even after training was modified to include automation transitions in Study 3, 15 participants did not immediately notice the start of, or transition to, an automated tracking condition. While there were six overt cues that alerted participants to the automation's state and additional training that allowed them to see and hear those cues, participants still had difficulty in noticing them. Resource conflict in accordance with MRT can help account for the apparent lack of resources to notice the change in system state (Wickens, 1981). Even though there were multiple overt cues, these cues may not have had sufficient salience to divert attentional resources away from the tracking task. This finding highlights a competition between salience of visual stimuli and different cues seen in previous research (Dowd & Mitroff, 2013). Therefore, future AA systems should be designed with salience of informational cues as a priority to mitigate the number of system state modes that may be available to an operator.

3. Delayed Cognitive Workload Management

This dissertation highlighted an aspect of the human-technology relationship by identifying potential areas where AA assisted and hindered the operator. Adaptative automation has been described in terms of managing operator workload to an optimal level. In other words, AA can increase or decrease operator workload to facilitate optimal system performance (de Greef & Arciszewski, 2007; Endsley, 2017a; Endsley & Kiris, 1995; Inagaki, 2003). When operators had automation assistance, their surrogate workload measures were lower overall in different workload conditions.

However, operator workload levels were not immediately reduced, but increased until the operator reached a level of equilibrium. This pattern suggests the importance of gaining in-the-loop familiarity. Further, this process represents a delay in achieving the intended outcome of cognitive workload management that AA systems are designed to provide. These delays were evident in the increased pupil sizes that were present in the 30 seconds following a system state transition. The increased pupil sizes were possibly related to ANS responses. This relationship further highlights the need to provide appropriate time for ANS responses to complete before transition steady state operational control to a human or machine operator.

4. Situation Awareness and Cognitive Workload in AA Systems

An intended outcome of AA is to increase SA (Kaber & Endsley, 2004). Results from the Studies 2 and 3 indicate that participants did experience increased SA when their cognitive workload was perceived to be lower. However, participants reported decreased SA when their cognitive workload was increased. These results follow previous studies that demonstrated these relationships (Endsley, 2021; Wickens, 2008b).

One unintended negative outcome with the introduction of higher LOAs is that SA decreases when it should increase. Results from this research indicated that participants reported higher SA when they were had most recently completed the tracking task in manual mode. This result suggests that participants were able to achieve higher levels of SA because they were in the system's operational loop. This finding highlights a complex relationship between providing cognitive workload management to the human operator at higher LOAs with the potential cost of reducing SA.

The inverse SA and cognitive workload relationship seen in this dissertation is particularly important when analyzing the system state transition periods. Results from Study 3 suggest that participants experienced increased cognitive workload for 30 seconds following a system state transition. So, it can be assumed that they experienced decreased SA around these transition windows. Therefore, AA systems should provide salient information at times of system state transitions until an operator can properly perceive and comprehend those cues to facilitate regaining SA.

B. ADDITIONAL CONTRIBUTIONS

1. Cognitive Workload Modeling Considerations

The cognitive workload models seen in this dissertation were developed using the discrete event modeling approach provided in IMPRINT. IMPRINT has been previously used in numerous studies to investigate the relationship between an operator and different task conditions (Ernst et al., 2020; Goodman et al., 2015; Lebiere et al., 2005; Militello et al., 2019; Samms, 2010; Wojciechowski, 2004). This research effort yielded insights into using IMPRINT task network models to investigate the unintended impacts of AA on cognitive workload. Some of these findings included the utility of IMPRINT models, the process of modeling and measuring unintended negative consequences of AA systems, and potential improvements for future attempts to use discrete-event modeling systems to understand AA systems.

a. IMPRINT Utility

Downes and Weisberg describe a two-level approach to assessing a model's representational capacity. The extent to which a model is accurate can be used to assess its utility (Downes, 2020). Models can be further related to their real-world target through investigating various components and manipulating their assigned value within different analyses (Weisberg, 2013). Both considerations were investigated in this research.

First, the accuracy of the models showed representational capacity to the real-world target when the predicted workload values were assessed against the results of the three experimental studies. Additionally, task workload values were manipulated in the model because of expert feedback, subjective cognitive workload ratings via CSWAG, and statistically significant differences in pupil diameter. The IMPRINT models showed utility by being able to incorporate multiple workload value updates. This dissertation demonstrated that representative AA system task models can be developed using IMPRINT to provide design recommendations.

b. AA Unintended Consequence Modeling and Measurement Process

This dissertation used a novel modeling and simulation approach to investigate AA's unintended negative consequences. The adapted MLCC framework guided the research efforts to investigate the impacts of how specific manipulations of MLCC variables affected cognitive workload. This approach was used to assess current workload prediction capabilities and to refine them based on validation measures gathered throughout the HITL studies. The results indicate that this methodology can help inform design decisions when appropriate modeling and measurement techniques are used.

An important aspect of this research's approach was the inclusion of subject matter expert knowledge elicitation through cognitive walk-throughs. The insights gained from these interviews allowed for refinement of the cognitive workload models that were derived from the default anchors provided in IMPRINT. The cognitive walk-throughs facilitated the process of verifying and validating the IMPRINT models. The experimental results served to validate the models and led to the discovery of transition effects that were not captured in the models previously. This framework highlights the importance of knowledge elicitation in refining the task analyses that serve as inputs to workload prediction modeling tools such as IMPRINT.

The multiple iterations of cognitive workload measurement throughout the three experimental studies revealed metrics that were sensitive to changes in workload. In the end, the current research attempted to surround the truth of what cognitive workload looks like using objective and subjective surrogate measures. The findings support previous work that recommended multi-modal approaches to investigating cognitive workload (Aricò et al., 2016; Cain, 2007; Hancock & Matthews, 2019). The method used here of combining predictive modeling tools and multi-modal cognitive workload measurement could be applied to different domains.

c. Future IMPRINT Considerations

Future research and design efforts can continue to benefit from the use of IMPRINT and similar tools to investigate impacts of various environments on system performance. While using these types of models can be informative, it is important to note that there are corresponding preliminary requirements that would help ensure successful modeling efforts. IMPRINT provides a user interface that modelers can use to graphically depict the flow of a task. They can define the operators and interfaces that will interact to accomplish the task. Further, modelers can input workload demand values for each interaction between the operator and a corresponding interface, while also determining impacts based on system reliability and success probabilities. These customizable parameters are numerous and can help ensure that models are developed with high levels of fidelity.

One key finding from using IMPRINT in this research effort was the importance of ensuring proper inputs into the model's workload demand values that resulted from the expert users' cognitive walk-throughs. A future consideration would be to include insights from novice operators to provide insights into the spectrum of cognitive workload that may be experienced during a task. While there were no validating cognitive workload measures found when assessing experience levels in this research, these non-findings may have been due to the nature of the MATB-II design used. However, future AA system models could stand to benefit from investigation of experience-related workload impacts.

IMPRINT was chosen in this research effort because its algorithm allowed for the investigation of the operator and system interactions together. The researcher was able to develop the progression of the tasks with their associated workload values. This approach allowed for analyzing cognitive workload through the human-system interactions that occurred with each event in MATB-II. Future modeling work could take the IMPRINT demand values and incorporate them into the processes seen in cognitive architectures. Used together, IMPRINT and cognitive architectures could inform how human operators might accomplish tasks using AA and provide resulting cognitive workload values based on the insights gleaned from the cognitive architectures (Lebiere et al., 2005). Both human performance models (i.e., IMPRINT) and cognitive architectures (i.e., ACT-R) rely on events as foundational concepts. Therefore, the modeled events can be used to integrate IMPRINT and ACT-R, for example (Lebiere et al., 2005). This framework begins to mirror the process of federating distributed simulations with a communications architecture that supports passing data in an integrated manner. The results of such an integration could provide insights into different approaches of accomplishing a task and the corresponding predicted workload values.

Modeling and simulation provide approaches and standards to integrating simulations through common architectures (Strickland, 2011). Applied to investigating the unintended consequences of AA, IMPRINT could provide information as to which tasks are being executed by the operator and ACT-R could list the actions being taken by the operator at the different interfaces. The two simulations could communicate over an architecture provided by a standard M&S protocol. The resulting information could be displayed to show the system state and provide an additional level of model verification and validation. This integration description follows Lebiere and colleagues' proposed modeling and simulation integration approach for task network models and cognitive architectures.

Because IMPRINT and ACT-R have shown potential for integration using M&S standards, future design analysis could include further integration of live, virtual, or constructive simulations to further assess cognitive workload impacts in AA systems. While the FVL platforms are still in development, their system requirements are established (Lacdan, 2022). Therefore, future modeling efforts using IMPRINT could be refined with real-time inputs from operators in a virtual simulation. These operator inputs could also inform the cognitive decision paths that exist in cognitive architectures. There exists great potential for M&S to be used in this complementary manner. The inclusion of multiple M&S tools could provide more robust insight into such systems as FVL where requirements are known, but the final materiel solution is not.

2. Towards Real-time Operator State Monitoring

AA will need to rely on real-time operator state monitoring. Therefore, it is important to investigate multiple candidate measures and validate them for future design consideration and to use insights from these measures in workload prediction models to facilitate early design decisions and later system modifications. Pupil diameter was sensitive to changes in workload and tracking conditions throughout this dissertation. Additionally, mean R-R intervals for HRV were sensitive to the scenario conditions in the five-minute increments that were assessed. Both physiological measures could be candidates for inclusion in operator state monitoring (OSM) efforts for future Army aircraft (Feltman, Kelley, Bernhardt, Bass, & Morabito, 2021). However, more analysis is necessary to determine feasibility in operational environments (Wilkins, Feltman, & Aura, 2022).

The use of fNIRS has been shown to be sensitive to changes in cognitive workload under certain conditions (Harrison et al., 2014; Herff et al., 2014; Scerbo, 2008). However, results from the current research did not follow this trend. This may be explained by the pace of the MATB-II tasks and the lag in the PFC blood oxygenation level response. Other studies also have not found PFC blood oxygenation to be an effective workload measure (Girouard et al., 2010). While fNIRS was not sensitive to changes in cognitive workload conditions in the current research, there may be applications of AA systems that would benefit from further investigation into changes of blood oxygenation in different parts of the brain.

Additionally, the results from this dissertation can help inform design considerations in future flight controls. If both hands will be required for operation, pupil metrics might be different in each eye due to ANS activation and deactivation. This finding could provide clarifying information to understand a pilot's cognitive workload through surrogate eye measures when accomplishing tasks that require both hands being engaged by the operator. Further investigation into different areas of the brain with fNIRS or EEG that are associated with ANS responses could help provide more understanding of the pupil size differences in each eye around the system state transition windows.

The results from the current effort support the inclusion of certain metrics to gauge operator cognitive workload. It is important to note that these measures can be leveraged to detect both increases and decreases in task demands to keep operators in the band of optimal workload. In other words, reducing workload should not be the only aim of introducing AA into a system.

3. Training and Human Factors Considerations

This research effort sought to investigate the effect of experience on MATB-II performance. Two training progressions yielded novice and experienced participants in the first two studies. While there were no statistically significant differences in cognitive workload surrogate measures between the two experience groups, experienced participants performed better on the MATB-II. Future work should include this type of analysis to account for the varied levels of experience that might interact with any system (Dreyfus, 2004; Hutton & Klein, 1999).

Human factors engineering considerations derived from this research included the impacts of intended and unintended consequences of AA on cognitive workload and performance. Study 1 demonstrated that there were increased pupil diameters, decreased HRV, and increased CSWAG ratings in a manually completed MATB-II task. Additionally, performance scores were lower in a manually completed MATB-II scenarios. The introduction of automation in Studies 2 and 3 represented improved HFE. Pupil sizes were smaller, HRV increased, and CSWAG ratings were lower when participants had automated assistance in the tracking task. While those measures all indicated a lower cognitive workload in the current design, they should not be used to make design decisions independent of other variables such as performance, SA, decision-making, or fatigue to name a few.

C. LIMITATIONS

The experimental studies in this dissertation each had a few threats to validity and limitations that should be considered. The experimental studies were limited to a laboratory setting where many potential confounding effects were controlled. While this setting allowed for the collection of a rich data set, this approach should be investigated in more operational settings using the same psychophysiological measures. Additionally, this dissertation leveraged a multi-attribute task battery that replicated aspects of flight to investigate cognitive workload rather than an actual flight task. The design used in this dissertation could be applied to more high-fidelity flight tasks to support different levels of analysis into AA.

The study samples used in this effort included predominantly mid-career military officers in graduate school. This population demographic is certainly unique and different than those of other settings that rely on either specific population sampling or more varied participation (i.e., military aviators or civilian undergraduate students). However, being able to investigate AA's unintended consequences with this population helped provide insight for future military system development. Although the study's scope was limited in this regard, the modeling and validation methodology used could be beneficial in different populations.

The modeling of the task was limited due to the lack of experts on the MATB-II task used in this study. Additionally, the small number of experts that provided feedback for the MATB-II task analysis threatened the validity of the model predictions and thus the models' outputs. While the expert-derived workload demand values were close to the default anchors provided in IMPRINT, further elicitation of workload demand values would be necessary to provide more refined models.

MATB-II afforded investigation of dynamically changing LOAs. However, this investigation was limited and did not allow for the full exploration of all LOAs. Further, the FOM data were aggregated at longer time intervals than all the physiological measurement devices. This time difference did not allow for higher fidelity comparison of performance and objective surrogate measures of cognitive workload.

D. DESIGNING FOR THE FUTURE

Technology will continue to evolve at a fast pace to meet growing requirements in changing environments, many of which are not yet known (Hollnagel & Woods, 2005). Therefore, it is critical to include human considerations when modeling and predicting what these future operational environments might be. These considerations are perhaps some of the only opportunities to evaluate system design in systems and environments that have not been realized yet. Early modeling and simulation efforts can help inform system acquisition processes. Therefore, it is important to have informed models that attempt to address potential issues.

Future Vertical Lift platforms are being developed with candidate technologies that will support the management of operator cognitive workload. Two of these technologies include heart rate and eye tracking. While there will be limitations on the number of technologies that can be included into a system due to competing resource demands, eye tracking and heart rate monitoring are two strong candidates to assess cognitive workload. The results from this dissertation support the inclusion of these capabilities to gather operator state data to inform the use of AA. Both heart rate and eye tracking metrics can be gathered in minimally intrusive ways and potentially through existing or soon to be fielded systems (Roth, Klein, Sushereba, Ernst, & Militello, 2022).

E. NEXT STEPS

The results of this dissertation highlighted numerous areas that would benefit from follow on studies. Future work should include further analysis of differences in physiological measures as objective means to assess cognitive workload in context. The use of fNIRS analysis in this effort did not yield significant results. However, future studies should continue to explore the use of this technology in different environments to assess the conditions in which these data might provide substantive information. Additional fNIRS analysis on the current data set could investigate any differences in specific channels to determine if specific areas of the PFC are sensitive to the rapid changes seen in the MATB-II scenarios.

HRV and pupil diameter appeared to be the most sensitive physiological measures in this research. Pupil diameter showed to be a consistent measure of cognitive workload, and future studies should continue to develop methods to capture this metric in various environmental conditions. Much of the future of adaptive automation rests with being able to have physiological inputs from the human operator be analyzed for action by a larger system (Ayaz et al., 2012; Parasuraman, 2003; Rusnock & Geiger, 2017). It is, therefore, imperative that the use of these measures be refined to provide quality signal from noise to properly advise systems of the future. Further analysis of saccades and fixations in relation to the current study could yield additional insights into unanticipated consequences of adaptive automation.

The investigation of the relationship with SA and cognitive workload warrants more investigation. While individual SA and cognitive workload were key considerations in this study and many others, the relationship between those two variables in a team setting warrant study as well. Smith and Baumann (2020) posit that team SA can suffer as an unintended negative consequences of AA. Therefore, applying the modeling and experimentation approach used in this research effort to a team setting (human-human and human-autonomy teams) could lead to an informative area of inquiry.

This research supported growing efforts to examine the effects of adaptive automation on operator workload and system performance using modeling and simulation tools. The resulting output of this effort yielded a model of cognitive workload that incorporates psychophysiological measurements, subjective measures, and performance data when operators execute tasks using adaptive automation. Combining data derived from IMPRINT's visual, audio, cognitive, and psychomotor (VACP) workload values and then comparing those data with performance results from an HITL study using physiological and subjective measures can assist in making operator workload modeling more robust. Performance metrics also served as a point of analysis to determine the relationship between objective and subjective workload measurements. This research effort also contributed to the Army's HSA-DM efforts in support of FVL to determine cognitive workload drivers and analyze their impacts on system performance. The results from this effort will support design considerations for adaptive automation across domains to address cognitive workload throughout a system's operation.

F. CONCLUSION

This dissertation sought to investigate the negative unintended consequences of AA on cognitive workload through leveraging modeling and simulations. Adaptive automation can introduce inadvertent drivers of cognitive workload. Human performance modeling can forecast these increases in cognitive workload when appropriate tools and complementary methods are used. These modeling predictions can be strengthened when assessed against objective and subjective measures. Further, human performance modeling can help inform early AA system design to ensure task demands are accounted for properly. These considerations can help facilitate appropriate total systems integration as new technologies are realized.

There exists a great opportunity to include the considerations of AA on operator workload in the Army's FVL program and beyond. While the nature of work will continue to transition to more automation-assisted activity, there will remain the impetus to ensure that human capabilities and limitations are identified, addressed, and accounted for through dynamic levels of technology integration. Introducing AA cannot be a simple on or off mechanism. Rather, the relationship that must exist between the human and the machine require seamless transitions that will develop over time rather than discretely. Let us then endeavor to ensure that those transitions take us further and faster into the future as intended.

APPENDIX A. EXPERIMENTAL MATERIALS AND METHODS

A. EQUIPMENT SPECIFICATIONS

The specifications for the equipment used in the studies were:

- 1 x MSI GE66 Raider gaming laptop computer, running Windows 10 Enterprise, with an Intel 11th Generation Core i9 processor at 3.30Ghz and 32.0 GB RAM.
- 1 x Acer H227HU 27-inch LCD flat panel display, connected to the main laptop computer via an HDMI cable. This external display was kept at a resolution of 1400x900 pixels to provide the largest possible display of MATB-II in the screen.
- 1 x Logitech G-Extreme 3D Pro USB Joystick.
- 1 x Dell KB813 USB Keyboard.
- 1 x Dell N231 USB Optical Mouse.
- 1 x NIRx NIRSport wearable, multi-channel neuroimaging system with an 8-light source/8-light detector configuration with 3cm source-detector separation, using a preconfigured NIRx prefrontal cortex headband. This configuration consisted of 21 data channels across the anterior prefrontal cortex (Giles et al., 2017). The NIRSport system collected data at a 7.81Hz sampling rate.
- 1 x Pupil Labs Core head-worn eye tracking system connected via USB-C, with a 200Hz eye camera sampling frequency and up to 120hz@480p scene camera sampling frequency.
- 1 x Polar H10 chest-worn heart rate monitor, with a 130Hz sampling frequency.

- 1 x Apple iPad, 9.7," Model Mp2H2LL/A, running iOS version 10.3.3.
- Lab Streaming Layer (LSL), version 1.14.
- Lab Streaming Layer Keyboard/Mouse connector application or Windows, version 1.15.0.
- Pupil Labs Capture, version 3.5.1.
- NIRx NIRStar, version 15.3.
- NASA MATB-II, version 3.5, presented in color inverse mode (black background) to allow for analysis using the Pupil Labs Core.
- NASA-TLX mobile application, version 1.0.2.
- PolarBand2LSL Bluetooth to LSL software connection.
- GNU Image Manipulation Program (GIMP), version 2.10.32.
- Apriltags visual fiduciary system markers, Tag 36h11. The researcher created a solid blue background with four Apriltags placed at the corners of the MATB-II window using the GIMP image editor. MATB-II was opened on the same centered spot on the monitor every time, with the Apriltags appearing on the corners of the MATB-II window as seen in Figure 20. This setup allowed for redundant eye tracking calibration and post hoc analysis.
- The Improved Performance Research Integration Tool (IMPRINT), version 4.6.60.0.

B. MATB-II REQUIREMENTS

Participants were instructed to ensure that the F5 button remained illuminated as green by clicking on the F5 should the square turn black. Similarly, participants were to ensure that the F6 square stayed black by clicking on the F6 square when it turned red.

For the F1-F4 sliding scales, users must ensure that the three dark blue rectangles do not deviate in their movement past the center point. An indicator arrow appeared next to the F1-F4 text, indicating that the scale moved out of tolerance. Participants had to click on the F1-F4 column (matching to the slider that is out of tolerance). Participants had to identify deviations in the F1-F4 columns within 15 seconds and light changes in F5-F6 within 10 seconds. If participants did not click on these indicators, it was counted as a failed task.

Participants interacted with the tracking task through a joystick when the TRACK task was set to manual mode. The TRACK task can also be set to automated where the operator is only responsible for monitoring the status of the tracking status (automated or manual). For Study 1, the TRACK task remained in manual mode. The center reticle illuminated to dark blue when in manual mode and light blue when in automated mode. There was also an audio tone alert that signaled any changes to the state of the TRACK task's automation level. Additionally, the text data in the TRACK task window provided data of position, velocity, and refresh rates when in manual mode. These data were reduced in automatic tracking mode. Participants were responsible for keeping the reticle in the center of the TRACK display. Participants could determine when they will be required to manually control the TRACK by identifying when the green bar was present on the scheduling window.

The COMM task can use up to 80 different audio clips to simulate radio traffic. Operators had to monitor for their callsign (NASA 504) and use their mouse to change the frequency of one of the four directed radio channels. Different voices are used, and time intervals for radio calls were configured ahead of time. Like the TRACK task, operators could determine when they would be required to respond to the COMM task by looking at the scheduling task. However, the green bar represents a period when radio traffic will come through, not necessarily that all communications traffic will be directed at their aircraft.

The RESMAN task required participants to interact with a series of fuel tanks and pumps to keep two main fuel tanks (A and B) at volumes around 2500. Fuel level indicators directly below tanks A and B displayed red numbers if the fuel levels were outside of +/- 500. Pumps C and D could hold volumes of 2000 and tanks E and F could hold an unlimited amount of fuel. Pumps 1–8 have different flow rates that the researcher set ahead of time using the default configuration file. The window directly to the right of the RESMAN task showed the fuel flow rates of each pump. When a pump was activated by a participant clicking on it, it would turn green. When a pump turned red, it was inoperable and could only be automatically fixed by MATB-II. These inoperable and fixed statuses were randomly pre-programmed by the researcher. Participants had to wait for the pump to turn black, signifying it had been fixed, before they could re-engage it. The RESMAN task window provides levels of fuel and their deviations from the 2500 fuel volume level. Additionally, participants were able to view the flow rates of the activated pumps to guide their strategy in accomplishing the RESMAN task.

C. EXPERIMENTAL PROCEDURES

The researcher welcomed the participants to the lab and oriented them to the experimental site. Participants were then given a consent form, and the researcher answered any resulting questions. Once participants signed the consent form, they completed the demographic questionnaire. This sequence served to identify any participants who met exclusion criteria. The researcher then discussed the subjective assessments (CSWAG and NASA-TLX) that were going to be collected. Participants received instructions on reporting their workload values every minute during the trial runs using CSWAG. They were also given a quick reference guide to help them report their workload assessments appropriately. Each of the psychophysiological measurement devices were displayed, and the researcher described them and their uses to the participants.

Next, participants then watched a NASA-produced training video on MATB-II and were provided a corresponding quick reference guide that was available for the duration of their time in the study room. Upon completion of the training video, participants began part-task training using MATB-II. Participants trained on each task (SYSMON, TRACK, COMM, and RESMAN) separately, and each training scenario lasted one minute. Once the part-task training was complete, participants completed a 5-

minute MATB-II practice session in a low workload condition (see Table 3) with all four tasks active. This training progression served as the training baseline for both novice and experienced operators in the study.

Upon completion of MATB-II training, the participants were instrumented with the three psychophysiological measurement devices. Participants received the Polar H10 heart rate monitor and were given instructions on how to moisten the leads on the chest strap. Further, participants had access to an instructional diagram that was posted in the private changing room in the HSIL. Participants returned to the main study site in the HSIL where the researchers ensured that the Polar2LSL application was connected to the heart rate monitor and sending data to the computer via Bluetooth in a Python environment.

Upon verification of the Polar H10 communicating with the collection system, members of the research team placed the fNIRS headband on the participant. The NIRx headbands were placed and centered on the participants' foreheads between the left and right pre-auricular regions in accordance with international 10–20 standards (Pinti et al., 2015). The researcher then calibrated the fNIRS device using NIRSTAR version 15.3 to validate proper setup of the fNIRS headband. Default values provided by NIRx in their software were used throughout the study to maintain consistent collection. If the calibration resulted in acceptable or better results (as provided by the NIRSTAR software), then the next system calibration steps commenced. If the calibration was not successful, then the researcher restarted the software and adjusted the NIRSport headband prior to running another calibration sequence to satisfactory levels.

The participants then placed the Pupil Labs Core system on their heads with the assistance of the research team to ensure proper fit and comfort. Participants were then invited to move up to a comfortable position in their chair so that they would be able to have the joystick, keyboard, and mouse within reach and have the screen in view. The researcher then instructed the participants to maintain a stationary position as the eye tracker calibration would happen next. Participants then completed the eye tracking calibration using the Pupil Capture software by moving their eye to five points on the screen. The center dot within the calibration circle that they fixated on turned from red to green in each position when the eye was properly registered. If participants did not successfully have their eyes calibrated to the screen, the researcher would examine the placement of the glasses on the participant and adjust the eye cameras so that each eye was recognized by Pupil Capture. Participants would then complete the calibration again, with this process being repeated until the software confirmed the participants' eye location. Once eye calibration was complete, the researcher locked the eye model toggle for each participant in Pupil Capture to account for potential drift or movement of the physical eye tracking system. The resolution of the eye cameras was set at 400 x 400 pixels based on the Pupil Labs Core User Guide (2022) and at the recommendation of a U.S. Army Aeromedical Research Laboratory (USAARL) vision scientist with previous experience using Pupil Labs (C. Aura, personal communication, August 18, 2022). The process of calibrating the system, locking the eye model, and setting the resolution at the recommended level was completed to account for differences in right and left pupil diameters found in pilot testing. While the pupil sizes fluctuated at the same rate, they were not equal diameters. Lighting conditions were held constant with the use of one lamp positioned behind the experimental station to provide ambient light and mitigate light artifacts from being directly in front of the participants. This setup was established to reduce detected light noise by the eye tracker and the fNIRS headband.

Once all three of the psychophysiological measurement devices were set up, the researcher opened Lab Streaming Layer and selected each of the modalities that were to be collected during the trial run. These modalities included the NIRSport, Pupil Labs Core, Polar H10, mouse position, mouse click, and keyboard entries. The keyboard entries also served as the definitive time start marker when analyzing the synchronized data. The researcher began recording on LSL and then instructed the participants to relax for one minute while pre-trial psychophysiological baselines were collected. Once the baseline period was complete, the researcher reminded the participants of the CSWAG instructions.

The researcher opened the MATB-II application, and then selected the trial run for the participant based on their assigned condition. Participants were instructed that to begin their trial, they would use the keystroke of Control + S. This keystroke entry would display a prompt for the participant asking if they were ready to begin MATB-II. When they were ready, participants pressed the Enter key on the keyboard. The release of the Enter key served as the marker for when the trial started and allowed for a common starting point for data analysis. They were also instructed that when the scenario was over, they would need to press, and release Enter once more to close out MATB-II. Participants then executed the instructions for MATB-II and completed their first ten-minute trial run in either the low or high workload condition.

Once the first trial was complete, participants conducted an additional eye tracking calibration. Then, the researcher loaded the second MATB-II trial in accordance with the participant's assigned trial condition. Participants were reminded that their CSWAG data would be collected again and that the keystrokes used in the first trial were the same. Barring any questions, the researcher initiated a new LSL recording in the same sequence as the first trial run. The participants were instructed to begin the second trial with the same keystrokes they used previously. Once participants were complete with their second trial, the researcher instructed the participants to rest for one minute. The researcher then saved and stopped the data collection. Participants were asked to remove the fNIRS headband and eye tracker and hand the devices back to the researcher.

Next, the participants completed the NASA-TLX and the SART. Participants were given an iPad with the NASA-TLX application. They first conducted pairwise comparisons to establish workload rating weights. Participants then completed the assessment of their workload using the six scales listed in the NASA-TLX. When the participants were complete with providing their ratings in the NASA-TLX application, the researcher took the iPad and handed them a paper copy of the SART along with a pen. Participants were instructed to refer to the instructions on the SART and to complete the questionnaire as it pertained to their experience completing the MATB-II trials.

The participants were then directed to the private changing room to remove the Polar H10 device. Once they returned to the study room, the researcher informed the participants that they had completed the study and thanked them for their participation. The researcher asked the participants to not discuss the study with others and that results would be available to them later upon publication of this research effort.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. RECRUITING FLYER

Research Volunteers Needed!

The Naval Postgraduate School's Modeling, Virtual Environments, and Simulation (MOVES) Institute is seeking volunteers to participate in a research study to observe the impacts of cognitive workload when interacting with adaptive automation. The study aims to assess the impacts of adaptive automation's unintended consequences through a series of studies using a simulated flight task. No flight experience required.

You will participate in two sessions within 72 hours of each other. Each session will last approximately one hour in order to capture objective and subjective measures of cognitive workload.

We invite and welcome all NPS Staff, Faculty, and Students to participate in this study. The end state is to assess ways of measuring cognitive workload in real-time when using adaptive automation. Additionally, the research seeks to provide design recommendations to one of the US Army's Future Vertical Lift Program subsidiary efforts, the Holistic Situational Awareness and Decision Making (HSA-DM) Program.

All volunteers must meet the following criteria:

- **At least 18 years old**
- **Visual acuity within service standards (20/20 corrected)**
- **Not wear bifocal, trifocal, or beyond glasses**
- **Not wear corrective lenses that have near infrared blocking coating**
- **Not be red-green colorblind**
- **Not have experience using NASA's Multi-Attribute Task Battery**

To sign up or for more information, please contact LTC Charles Rowan at charles.rowan@nps.edu.

Points of Contact: If you have any questions or comments about this research, or you experience an injury or have questions about any discomforts that you experience while taking part in this study, please contact:

- Principal Investigator: Dr. Lawrence Shattuck, lgshattu@nps.edu, 831.656.2473.
- IRB Vice-Chair, Bryan Hudgens at bryan.hudgens@nps.edu, 831.656.2039.
- IRB Vice-Chair, LT Aditya Prasad at aditya.prasad@nps.edu, 831.656.7675.

This is an NPS Institutional Review Board Approved research protocol.

Figure 72. Recruiting flyer.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. MATB-II REFERENCE SHEET

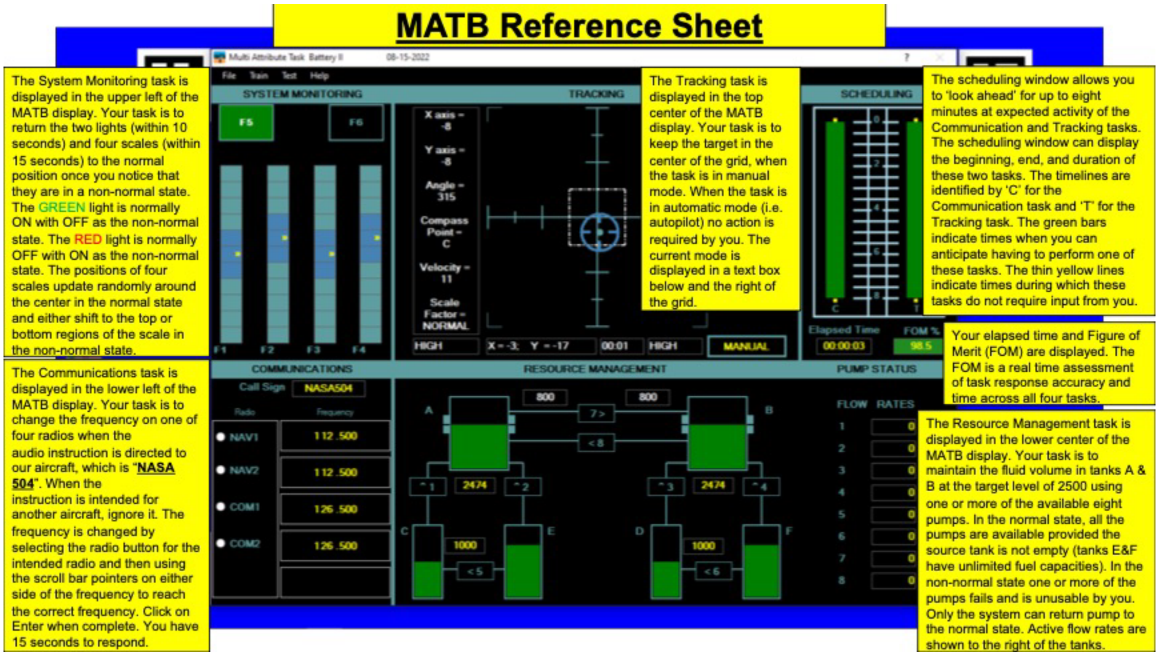


Figure 73. MATB-II participant reference sheet.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. CONTINUOUS SUBJECTIVE WORKLOAD ASSESSMENT GRAPH (CSWAG) REFERENCE GUIDES

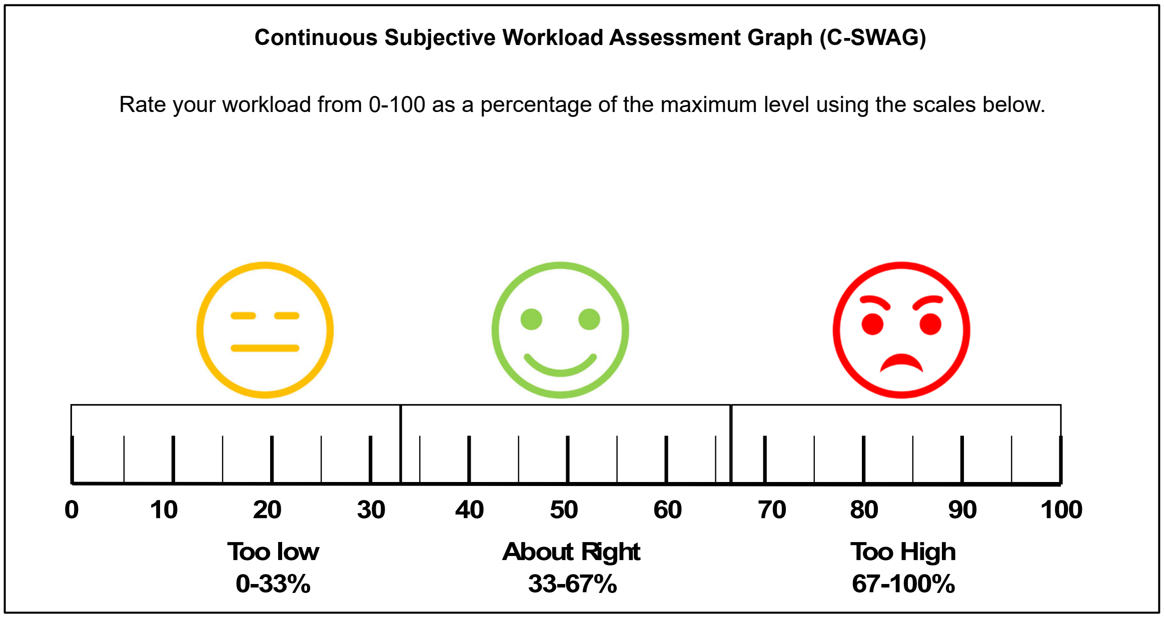


Figure 74. CSWAG participant reference visual.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. DEMOGRAPHIC SURVEYS

Demographic Questionnaire
Participant #: ____________

1. Do you wear bifocal, trifocal, or beyond corrective lenses?

   ____

2. Do you wear corrective lenses that have near infrared blocking coating?

   ____

3. Have you been diagnosed as being red-green colorblind?

   ____

4. Do you have previous experience using NASA's Multi-Attribute Task Battery?

   ____

5. What is your gender?

   ____

6. What is your age?

   ____

7. What is your visual acuity and dominant eye?

   ____

-------------------Military Participants-------------------

8. What is your branch of service?

   ____

9. What is your military occupational specialty?

   ____

10. What is your rank?

    ____

11. How many years in service do you have?

    ____

Figure 75. Studies 1 and 2 demographic survey.

Demographic Questionnaire
Participant #: ____________

1. Do you wear bifocal, trifocal, or beyond corrective lenses?

   ____

2. Do you wear corrective lenses that have near infrared blocking coating?

   ____

3. Have you been diagnosed as being red-green colorblind?

   ____

4. Do you have previous experience using NASA's Multi-Attribute Task Battery?

   ____

5. What is your gender?

   ____

6. What is your age?

   ____

7. What is your dominant hand?

   ____

8. What is your visual acuity and dominant eye?

   ____

-------------------Military Participants-------------------

9. What is your branch of service?

   ____

10. What is your military occupational specialty?

    ____

11. What is your rank?

    ____

12. How many years in service do you have?

    ____

Figure 76. Study 3 demographic survey.

APPENDIX F. SITUATION AWARENESS RATING TECHNIQUE FORM

Situation Awareness Rating Technique **Date: _________ Participant ID: _________**

The following is a Situation Awareness Rating Technique (SART) questionnaire (Taylor, 1990). Please answer each of the 10 questions by selecting a number between 1-7 on each scale as they relate to your experience performing the simulation.

1. How changeable is the situation? [Instability of Situation]

Stable and straightforward	1	2	3	4	5	6	7	Changing suddenly

2. How complicated is the situation [Complexity of Situation]

Simple and straightforward	1	2	3	4	5	6	7	Complex with many interrelated components

3. How many variable are changing with the situation [Variability of Situation]

Very few variables changing	1	2	3	4	5	6	7	A large number of factors changing

4. How aroused are you in the situation [Arousal]

A low degree of alertness	1	2	3	4	5	6	7	Alert and ready for activity

5. How much are you concentrating on the situation? [Concentration of Attention]

Focusing on only one	1	2	3	4	5	6	7	Concentrating on many aspects of the situation

6. How much is your attention divided in the situation? [Division of Attention]

Focusing on only one	1	2	3	4	5	6	7	Concentrating on many aspects of the situation

7. How much mental capacity do you have to spare in the situation [Spare Mental Capacity]

Nothing to spare at all	1	2	3	4	5	6	7	Sufficient to attend to many variables

8. How much information have you gained about the situation [Information Quantity]

Very little	1	2	3	4	5	6	7	A great deal of knowledge

9. How valuable or accurate was information you gained about the situation [Information Quality]

Very little	1	2	3	4	5	6	7	A great deal of value

10. How familiar are you with the situation? [Familiarity with Situation]

New situation	1	2	3	4	5	6	7	Great deal of relevant experience

Figure 77. Situation awareness technique questionnaire.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX G. NASA-TLX

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 78. NASA-TLX.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX H. TASK ANALYSIS QUESTIONNAIRE

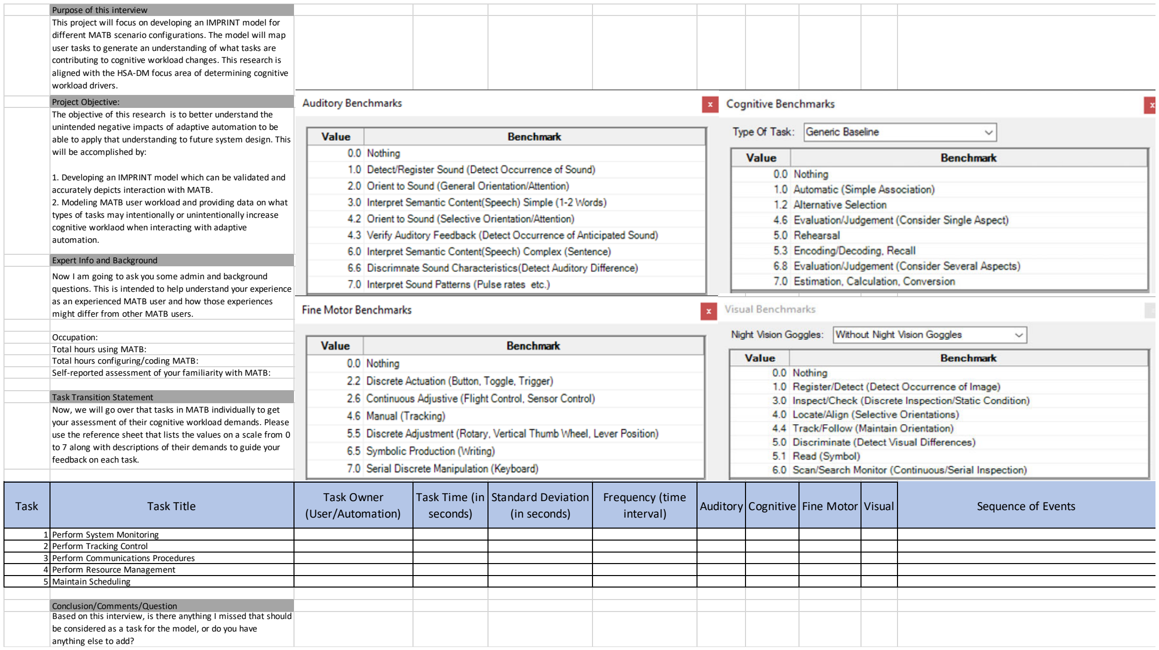


Figure 79. MATB-II task analysis questionnaire.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX I. STUDY 3 IMPRINT DIAGRAMS

S3_Group1_...HA Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Own		Value: 00:00:25.00	Root (Root)	4 Own_Comms
Comm02_Own		Value: 00:02:12.00	Root (Root)	4 Own_Comms
Comm03_Other		Value: 00:04:01.00	Root (Root)	12 Other_Comms
Comm04_Own		Value: 00:06:05.00	Root (Root)	4 Own_Comms
Comm05_Other		Value: 00:07:47.00	Root (Root)	12 Other_Comms
Comm06_Own		Value: 00:09:17.00	Root (Root)	4 Own_Comms
Comm07_Other		Value: 00:10:05.00	Root (Root)	12 Other_Comms
Comm08_Own		Value: 00:10:29.00	Root (Root)	4 Own_Comms
Comm09_Own		Value: 00:10:58.00	Root (Root)	4 Own_Comms
Comm10_Other		Value: 00:11:23.00	Root (Root)	12 Other_Comms
Comm11_Own		Value: 00:11:46.00	Root (Root)	4 Own_Comms
Comm12_Other		Value: 00:12:09.00	Root (Root)	12 Other_Comms
Comm13_Own		Value: 00:12:32.00	Root (Root)	4 Own_Comms
Comm14_Own		Value: 00:12:56.00	Root (Root)	4 Own_Comms
Comm15_Other		Value: 00:13:20.00	Root (Root)	12 Other_Comms
Comm16_Own		Value: 00:13:45.00	Root (Root)	4 Own_Comms
Comm17_Own		Value: 00:14:10.00	Root (Root)	4 Own_Comms
Comm18_Own		Value: 00:14:34.00	Root (Root)	4 Own_Comms
Comm19_Own		Value: 00:15:06.00	Root (Root)	4 Own_Comms
Comm20_Own		Value: 00:15:35.00	Root (Root)	4 Own_Comms
Comm21_Own		Value: 00:16:00.00	Root (Root)	4 Own_Comms
Comm22_Other		Value: 00:16:25.00	Root (Root)	12 Other_Comms
Comm23_Other		Value: 00:16:52.00	Root (Root)	12 Other_Comms
Comm24_Own		Value: 00:17:15.00	Root (Root)	4 Own_Comms
Comm25_Own		Value: 00:17:39.00	Root (Root)	4 Own_Comms
Comm26_Other		Value: 00:18:02.00	Root (Root)	12 Other_Comms
Comm27_Own		Value: 00:18:27.00	Root (Root)	4 Own_Comms
Comm28_Own		Value: 00:18:52.00	Root (Root)	4 Own_Comms
Comm29_Other		Value: 00:19:15.00	Root (Root)	12 Other_Comms
Comm30_Own		Value: 00:19:37.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:00:00.50	Root (Root)	17 Start Tracking Loop
ManualTracking2		Value: 00:10:00.00	Root (Root)	17 Start Tracking Loop

Figure 80. Group 1 researcher-derived model external event triggers.

S3_Group2_...HM Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Own		Value: 00:01:05.00	Root (Root)	4 Own_Comms
Comm02_Other		Value: 00:02:47.00	Root (Root)	12 Other_Comms
Comm03_Own		Value: 00:04:17.00	Root (Root)	4 Own_Comms
Comm04_Own		Value: 00:05:25.00	Root (Root)	4 Own_Comms
Comm05_Own		Value: 00:07:12.00	Root (Root)	4 Own_Comms
Comm06_Other		Value: 00:09:01.00	Root (Root)	12 Other_Comms
Comm07_Own		Value: 00:10:06.00	Root (Root)	4 Own_Comms
Comm08_Own		Value: 00:10:35.00	Root (Root)	4 Own_Comms
Comm09_Own		Value: 00:11:00.00	Root (Root)	4 Own_Comms
Comm10_Other		Value: 00:11:25.00	Root (Root)	12 Other_Comms
Comm11_Other		Value: 00:11:52.00	Root (Root)	12 Other_Comms
Comm12_Own		Value: 00:12:15.00	Root (Root)	4 Own_Comms
Comm13_Own		Value: 00:12:39.00	Root (Root)	4 Own_Comms
Comm14_Other		Value: 00:13:02.00	Root (Root)	12 Other_Comms
Comm15_Own		Value: 00:13:27.00	Root (Root)	4 Own_Comms
Comm16_Own		Value: 00:13:52.00	Root (Root)	4 Own_Comms
Comm17_Other		Value: 00:14:15.00	Root (Root)	12 Other_Comms
Comm18_Own		Value: 00:14:37.00	Root (Root)	4 Own_Comms
Comm19_Other		Value: 00:15:05.00	Root (Root)	12 Other_Comms
Comm20_Own		Value: 00:15:29.00	Root (Root)	4 Own_Comms
Comm21_Own		Value: 00:15:58.00	Root (Root)	4 Own_Comms
Comm22_Other		Value: 00:16:23.00	Root (Root)	12 Other_Comms
Comm23_Own		Value: 00:16:46.00	Root (Root)	4 Own_Comms
Comm24_Other		Value: 00:17:09.00	Root (Root)	12 Other_Comms
Comm25_Own		Value: 00:17:32.00	Root (Root)	4 Own_Comms
Comm26_Own		Value: 00:17:56.00	Root (Root)	4 Own_Comms
Comm27_Other		Value: 00:18:20.00	Root (Root)	12 Other_Comms
Comm28_Own		Value: 00:18:45.00	Root (Root)	4 Own_Comms
Comm29_Own		Value: 00:19:10.00	Root (Root)	4 Own_Comms
Comm30_Own		Value: 00:19:34.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:05:00.00	Root (Root)	17 Start Tracking Loop
ManualTracking2		Value: 00:15:00.00	Root (Root)	17 Start Tracking Loop

Figure 81. Group 2 researcher-derived model external event triggers.

S3_Group3_...LA Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Other		Value: 00:00:05.00	Root (Root)	12 Other_Comms
Comm02_Own		Value: 00:00:29.00	Root (Root)	4 Own_Comms
Comm03_Own		Value: 00:00:58.00	Root (Root)	4 Own_Comms
Comm04_Other		Value: 00:01:23.00	Root (Root)	12 Other_Comms
Comm05_Own		Value: 00:01:46.00	Root (Root)	4 Own_Comms
Comm06_Other		Value: 00:02:09.00	Root (Root)	12 Other_Comms
Comm07_Own		Value: 00:02:32.00	Root (Root)	4 Own_Comms
Comm08_Own		Value: 00:02:56.00	Root (Root)	4 Own_Comms
Comm09_Other		Value: 00:03:20.00	Root (Root)	12 Other_Comms
Comm10_Own		Value: 00:03:45.00	Root (Root)	4 Own_Comms
Comm11_Own		Value: 00:04:10.00	Root (Root)	4 Own_Comms
Comm12_Own		Value: 00:04:34.00	Root (Root)	4 Own_Comms
Comm13_Own		Value: 00:05:06.00	Root (Root)	4 Own_Comms
Comm14_Own		Value: 00:05:35.00	Root (Root)	4 Own_Comms
Comm15_Own		Value: 00:06:00.00	Root (Root)	4 Own_Comms
Comm16_Other		Value: 00:06:25.00	Root (Root)	12 Other_Comms
Comm17_Other		Value: 00:06:52.00	Root (Root)	12 Other_Comms
Comm18_Own		Value: 00:07:15.00	Root (Root)	4 Own_Comms
Comm19_Own		Value: 00:07:39.00	Root (Root)	4 Own_Comms
Comm20_Other		Value: 00:08:02.00	Root (Root)	12 Other_Comms
Comm21_Own		Value: 00:08:27.00	Root (Root)	4 Own_Comms
Comm22_Own		Value: 00:08:52.00	Root (Root)	4 Own_Comms
Comm23_Other		Value: 00:09:15.00	Root (Root)	12 Other_Comms
Comm24_Own		Value: 00:09:37.00	Root (Root)	4 Own_Comms
Comm25_Own		Value: 00:10:25.00	Root (Root)	4 Own_Comms
Comm26_Own		Value: 00:12:12.00	Root (Root)	4 Own_Comms
Comm27_Other		Value: 00:14:01.00	Root (Root)	12 Other_Comms
Comm28_Own		Value: 00:16:05.00	Root (Root)	4 Own_Comms
Comm29_Other		Value: 00:17:47.00	Root (Root)	12 Other_Comms
Comm30_Own		Value: 00:19:17.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:00:00.50	Root (Root)	17 Start Tracking Loop
ManualTracking2		Value: 00:10:00.00	Root (Root)	17 Start Tracking Loop

Figure 82. Group 3 researcher-derived model external event triggers.

S3_Group4_...LM Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
Comm01_Own		Value: 00:00:06.00	Root (Root)	4 Own_Comms
Comm02_Own		Value: 00:00:35.00	Root (Root)	4 Own_Comms
Comm03_Own		Value: 00:01:00.00	Root (Root)	4 Own_Comms
Comm04_Other		Value: 00:01:25.00	Root (Root)	12 Other_Comms
Comm05_Other		Value: 00:01:52.00	Root (Root)	12 Other_Comms
Comm06_Own		Value: 00:02:15.00	Root (Root)	4 Own_Comms
Comm07_Own		Value: 00:02:39.00	Root (Root)	4 Own_Comms
Comm08_Other		Value: 00:03:02.00	Root (Root)	12 Other_Comms
Comm09_Own		Value: 00:03:27.00	Root (Root)	4 Own_Comms
Comm10_Own		Value: 00:03:52.00	Root (Root)	4 Own_Comms
Comm11_Other		Value: 00:04:15.00	Root (Root)	12 Other_Comms
Comm12_Own		Value: 00:04:37.00	Root (Root)	4 Own_Comms
Comm13_Other		Value: 00:05:05.00	Root (Root)	12 Other_Comms
Comm14_Own		Value: 00:05:29.00	Root (Root)	4 Own_Comms
Comm15_Own		Value: 00:05:58.00	Root (Root)	4 Own_Comms
Comm16_Other		Value: 00:06:23.00	Root (Root)	12 Other_Comms
Comm17_Own		Value: 00:06:46.00	Root (Root)	4 Own_Comms
Comm18_Other		Value: 00:07:09.00	Root (Root)	12 Other_Comms
Comm19_Own		Value: 00:07:32.00	Root (Root)	4 Own_Comms
Comm20_Own		Value: 00:07:56.00	Root (Root)	4 Own_Comms
Comm21_Other		Value: 00:08:20.00	Root (Root)	12 Other_Comms
Comm22_Own		Value: 00:08:45.00	Root (Root)	4 Own_Comms
Comm23_Own		Value: 00:09:10.00	Root (Root)	4 Own_Comms
Comm24_Own		Value: 00:09:34.00	Root (Root)	4 Own_Comms
Comm25_Own		Value: 00:11:05.00	Root (Root)	4 Own_Comms
Comm26_Other		Value: 00:12:47.00	Root (Root)	12 Other_Comms
Comm27_Own		Value: 00:14:17.00	Root (Root)	4 Own_Comms
Comm28_Own		Value: 00:15:25.00	Root (Root)	4 Own_Comms
Comm29_Own		Value: 00:17:12.00	Root (Root)	4 Own_Comms
Comm30_Other		Value: 00:19:01.00	Root (Root)	12 Other_Comms
ManualTracking1		Value: 00:05:00.00	Root (Root)	17 Start Tracking Loop
ManualTracking2		Value: 00:15:00.00	Root (Root)	17 Start Tracking Loop

Figure 83. Group 4 researcher-derived model external event triggers.

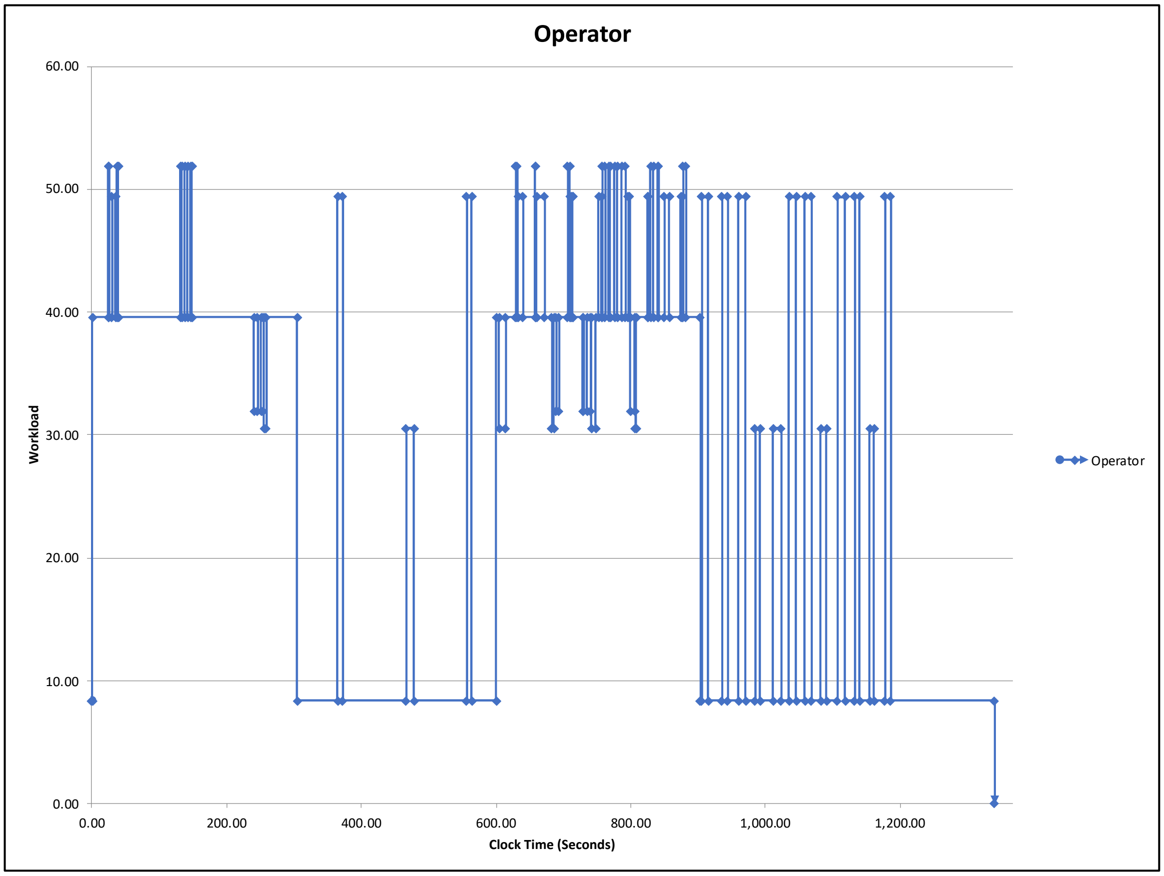


Figure 84. Study 3 researcher-derived model group 1 IMPRINT workload graph.

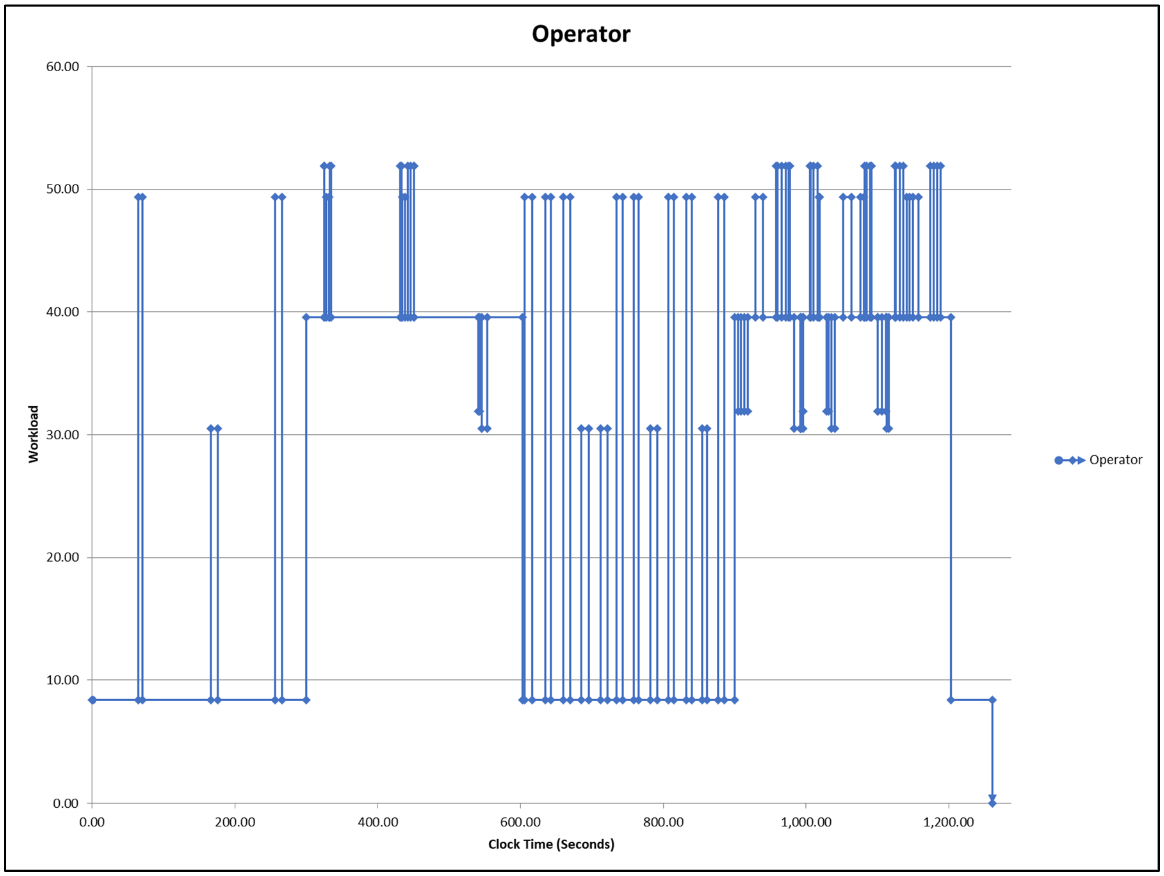


Figure 85. Study 3 researcher-derived model group 2 IMPRINT workload graph.

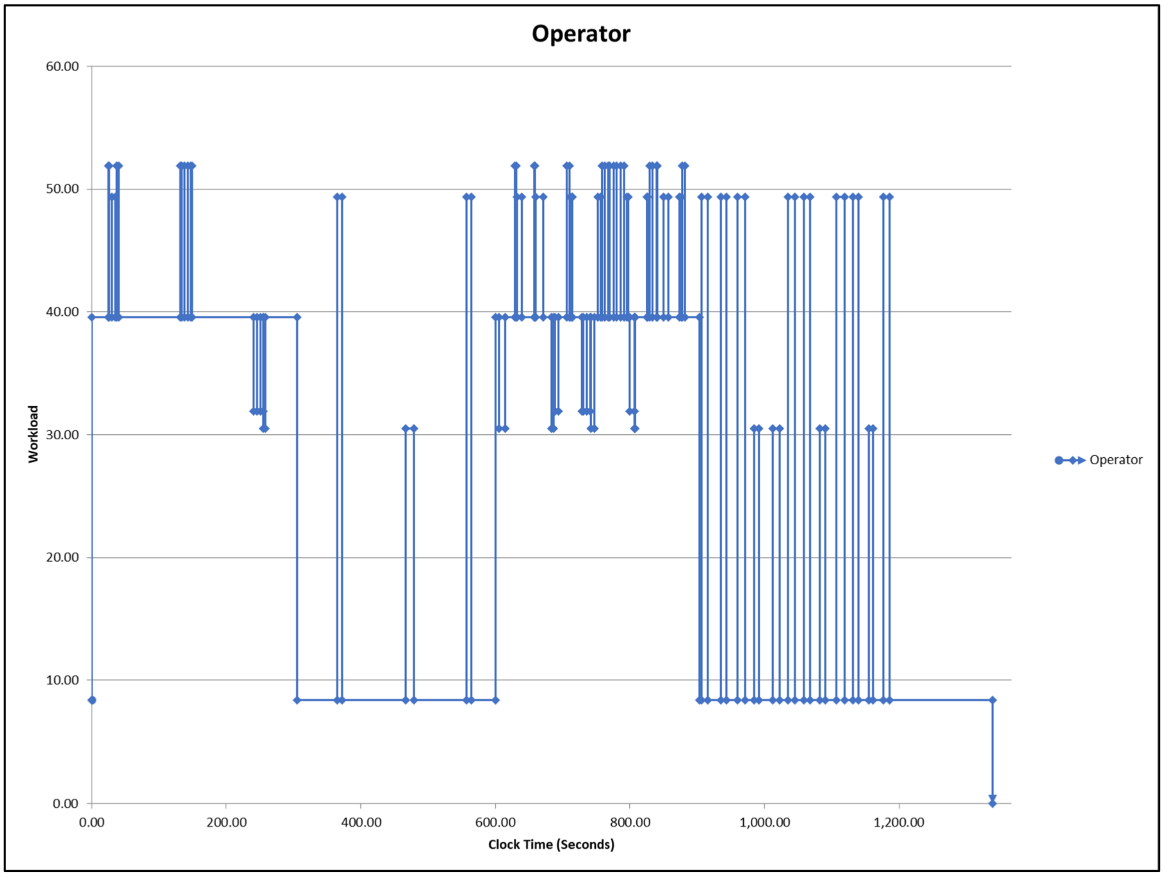


Figure 86. Study 3 researcher-derived model group 3 IMPRINT workload graph.

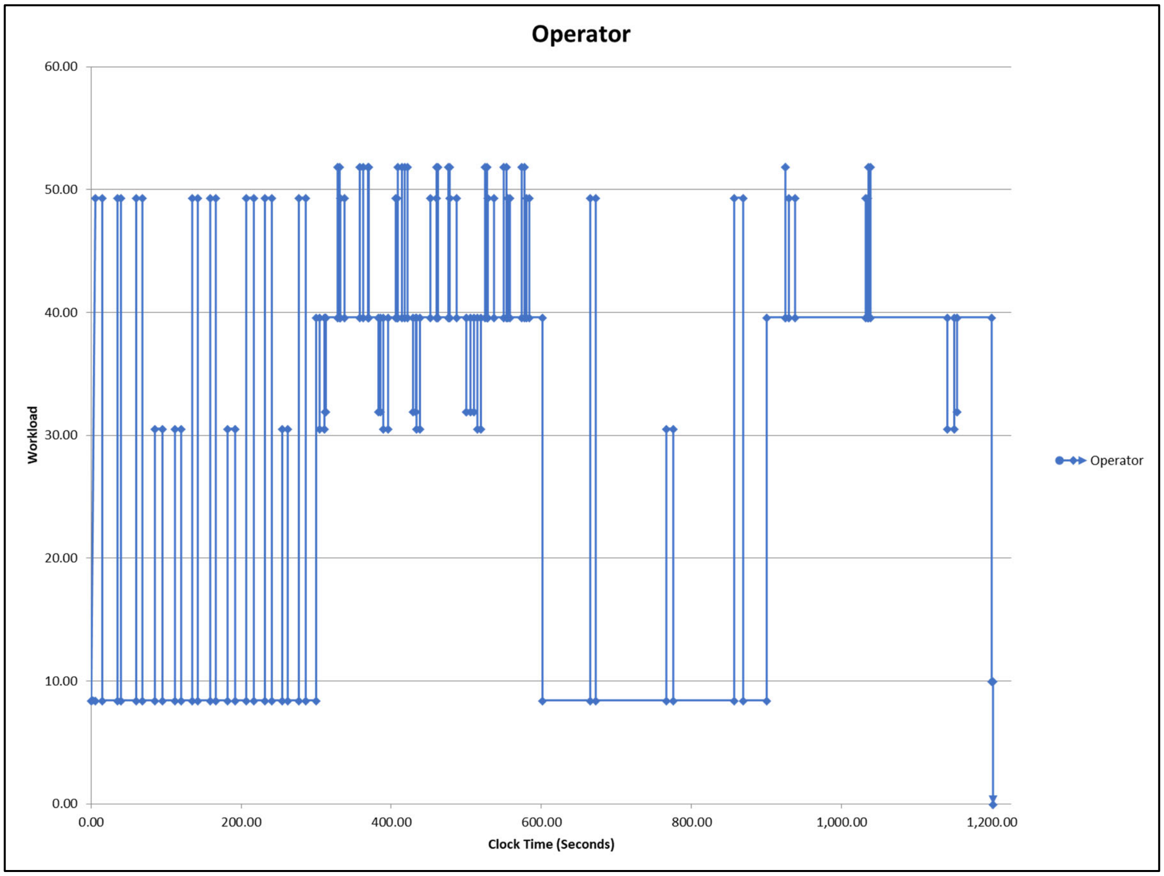


Figure 87. Study 3 researcher-derived model group 4 IMPRINT workload graph.

S3_Group1_...HA Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
AutoTracking1		Value: 00:05:00.00	Root (Root)	20 Start Auto Tracking Loop
AutoTracking2		Value: 00:15:00.00	Root (Root)	20 Start Auto Tracking Loop
Comm01_Own		Value: 00:00:25.00	Root (Root)	4 Own_Comms
Comm02_Own		Value: 00:02:12.00	Root (Root)	4 Own_Comms
Comm03_Other		Value: 00:04:01.00	Root (Root)	12 Other_Comms
Comm04_Own		Value: 00:06:05.00	Root (Root)	4 Own_Comms
Comm05_Other		Value: 00:07:47.00	Root (Root)	12 Other_Comms
Comm06_Own		Value: 00:09:17.00	Root (Root)	4 Own_Comms
Comm07_Other		Value: 00:10:05.00	Root (Root)	12 Other_Comms
Comm08_Own		Value: 00:10:29.00	Root (Root)	4 Own_Comms
Comm09_Own		Value: 00:10:58.00	Root (Root)	4 Own_Comms
Comm10_Other		Value: 00:11:23.00	Root (Root)	12 Other_Comms
Comm11_Own		Value: 00:11:46.00	Root (Root)	4 Own_Comms
Comm12_Other		Value: 00:12:09.00	Root (Root)	12 Other_Comms
Comm13_Own		Value: 00:12:32.00	Root (Root)	4 Own_Comms
Comm14_Own		Value: 00:12:56.00	Root (Root)	4 Own_Comms
Comm15_Other		Value: 00:13:20.00	Root (Root)	12 Other_Comms
Comm16_Own		Value: 00:13:45.00	Root (Root)	4 Own_Comms
Comm17_Own		Value: 00:14:10.00	Root (Root)	4 Own_Comms
Comm18_Own		Value: 00:14:34.00	Root (Root)	4 Own_Comms
Comm19_Own		Value: 00:15:06.00	Root (Root)	4 Own_Comms
Comm20_Own		Value: 00:15:35.00	Root (Root)	4 Own_Comms
Comm21_Own		Value: 00:16:00.00	Root (Root)	4 Own_Comms
Comm22_Other		Value: 00:16:25.00	Root (Root)	12 Other_Comms
Comm23_Other		Value: 00:16:52.00	Root (Root)	12 Other_Comms
Comm24_Own		Value: 00:17:15.00	Root (Root)	4 Own_Comms
Comm25_Own		Value: 00:17:39.00	Root (Root)	4 Own_Comms
Comm26_Other		Value: 00:18:02.00	Root (Root)	12 Other_Comms
Comm27_Own		Value: 00:18:27.00	Root (Root)	4 Own_Comms
Comm28_Own		Value: 00:18:52.00	Root (Root)	4 Own_Comms
Comm29_Other		Value: 00:19:15.00	Root (Root)	12 Other_Comms
Comm30_Own		Value: 00:19:37.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:00:00.50	Root (Root)	17 Start Tracking Loop
ManualTracking2		Value: 00:10:00.00	Root (Root)	17 Start Tracking Loop

Figure 88. Group 1 expert-derived model external event triggers.

S3_Group2_...HM Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
AutoTracking1		Value: 00:00:00.50	Root (Root)	19 Start Auto Tracking Loop
AutoTracking2		Value: 00:10:00.00	Root (Root)	19 Start Auto Tracking Loop
Comm01_Own		Value: 00:01:05.00	Root (Root)	4 Own_Comms
Comm02_Other		Value: 00:02:47.00	Root (Root)	12 Other_Comms
Comm03_Own		Value: 00:04:17.00	Root (Root)	4 Own_Comms
Comm04_Own		Value: 00:05:25.00	Root (Root)	4 Own_Comms
Comm05_Own		Value: 00:07:12.00	Root (Root)	4 Own_Comms
Comm06_Other		Value: 00:09:01.00	Root (Root)	12 Other_Comms
Comm07_Own		Value: 00:10:06.00	Root (Root)	4 Own_Comms
Comm08_Own		Value: 00:10:35.00	Root (Root)	4 Own_Comms
Comm09_Own		Value: 00:11:00.00	Root (Root)	4 Own_Comms
Comm10_Other		Value: 00:11:25.00	Root (Root)	12 Other_Comms
Comm11_Other		Value: 00:11:52.00	Root (Root)	12 Other_Comms
Comm12_Own		Value: 00:12:15.00	Root (Root)	4 Own_Comms
Comm13_Own		Value: 00:12:39.00	Root (Root)	4 Own_Comms
Comm14_Other		Value: 00:13:02.00	Root (Root)	12 Other_Comms
Comm15_Own		Value: 00:13:27.00	Root (Root)	4 Own_Comms
Comm16_Own		Value: 00:13:52.00	Root (Root)	4 Own_Comms
Comm17_Other		Value: 00:14:15.00	Root (Root)	12 Other_Comms
Comm18_Own		Value: 00:14:37.00	Root (Root)	4 Own_Comms
Comm19_Other		Value: 00:15:05.00	Root (Root)	12 Other_Comms
Comm20_Own		Value: 00:15:29.00	Root (Root)	4 Own_Comms
Comm21_Own		Value: 00:15:58.00	Root (Root)	4 Own_Comms
Comm22_Other		Value: 00:16:23.00	Root (Root)	12 Other_Comms
Comm23_Own		Value: 00:16:46.00	Root (Root)	4 Own_Comms
Comm24_Other		Value: 00:17:09.00	Root (Root)	12 Other_Comms
Comm25_Own		Value: 00:17:32.00	Root (Root)	4 Own_Comms
Comm26_Own		Value: 00:17:56.00	Root (Root)	4 Own_Comms
Comm27_Other		Value: 00:18:20.00	Root (Root)	12 Other_Comms
Comm28_Own		Value: 00:18:45.00	Root (Root)	4 Own_Comms
Comm29_Own		Value: 00:19:10.00	Root (Root)	4 Own_Comms
Comm30_Own		Value: 00:19:34.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:05:00.00	Root (Root)	18 Start Tracking Loop
ManualTracking2		Value: 00:15:00.00	Root (Root)	18 Start Tracking Loop

Figure 89. Group 2 expert-derived model external event triggers.

S3_Group3_...LA Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
AutoTracking1		Value: 00:05:00.00	Root (Root)	19 Start Auto Tracking Loop
AutoTracking2		Value: 00:15:00.00	Root (Root)	19 Start Auto Tracking Loop
Comm01_Other		Value: 00:00:05.00	Root (Root)	12 Other_Comms
Comm02_Own		Value: 00:00:29.00	Root (Root)	4 Own_Comms
Comm03_Own		Value: 00:00:58.00	Root (Root)	4 Own_Comms
Comm04_Other		Value: 00:01:23.00	Root (Root)	12 Other_Comms
Comm05_Own		Value: 00:01:46.00	Root (Root)	4 Own_Comms
Comm06_Other		Value: 00:02:09.00	Root (Root)	12 Other_Comms
Comm07_Own		Value: 00:02:32.00	Root (Root)	4 Own_Comms
Comm08_Own		Value: 00:02:56.00	Root (Root)	4 Own_Comms
Comm09_Other		Value: 00:03:20.00	Root (Root)	12 Other_Comms
Comm10_Own		Value: 00:03:45.00	Root (Root)	4 Own_Comms
Comm11_Own		Value: 00:04:10.00	Root (Root)	4 Own_Comms
Comm12_Own		Value: 00:04:34.00	Root (Root)	4 Own_Comms
Comm13_Own		Value: 00:05:06.00	Root (Root)	4 Own_Comms
Comm14_Own		Value: 00:05:35.00	Root (Root)	4 Own_Comms
Comm15_Own		Value: 00:06:00.00	Root (Root)	4 Own_Comms
Comm16_Other		Value: 00:06:25.00	Root (Root)	12 Other_Comms
Comm17_Other		Value: 00:06:52.00	Root (Root)	12 Other_Comms
Comm18_Own		Value: 00:07:15.00	Root (Root)	4 Own_Comms
Comm19_Own		Value: 00:07:39.00	Root (Root)	4 Own_Comms
Comm20_Other		Value: 00:08:02.00	Root (Root)	12 Other_Comms
Comm21_Own		Value: 00:08:27.00	Root (Root)	4 Own_Comms
Comm22_Own		Value: 00:08:52.00	Root (Root)	4 Own_Comms
Comm23_Other		Value: 00:09:15.00	Root (Root)	12 Other_Comms
Comm24_Own		Value: 00:09:37.00	Root (Root)	4 Own_Comms
Comm25_Own		Value: 00:10:25.00	Root (Root)	4 Own_Comms
Comm26_Own		Value: 00:12:12.00	Root (Root)	4 Own_Comms
Comm27_Other		Value: 00:14:01.00	Root (Root)	12 Other_Comms
Comm28_Own		Value: 00:16:05.00	Root (Root)	4 Own_Comms
Comm29_Other		Value: 00:17:47.00	Root (Root)	12 Other_Comms
Comm30_Own		Value: 00:19:17.00	Root (Root)	4 Own_Comms
ManualTracking1		Value: 00:00:00.50	Root (Root)	18 Start Tracking Loop
ManualTracking2		Value: 00:10:00.00	Root (Root)	18 Start Tracking Loop

Figure 90. Group 3 expert-derived model external event triggers.

S3_Group4_...LM Diagram | External Events

Name	Description	Appearance Time (HH:MM:SS.mm)	Task Triggered	
			Function	Task
AutoTracking1		Value: 00:00:00.50	Root (Root)	18 Start Auto Tracking Loop
AutoTracking2		Value: 00:10:00.00	Root (Root)	18 Start Auto Tracking Loop
Comm01_Own		Value: 00:00:06.00	Root (Root)	4 Own_Comms
Comm02_Own		Value: 00:00:35.00	Root (Root)	4 Own_Comms
Comm03_Own		Value: 00:01:00.00	Root (Root)	4 Own_Comms
Comm04_Other		Value: 00:01:25.00	Root (Root)	12 Other_Comms
Comm05_Other		Value: 00:01:52.00	Root (Root)	12 Other_Comms
Comm06_Own		Value: 00:02:15.00	Root (Root)	4 Own_Comms
Comm07_Own		Value: 00:02:39.00	Root (Root)	4 Own_Comms
Comm08_Other		Value: 00:03:02.00	Root (Root)	12 Other_Comms
Comm09_Own		Value: 00:03:27.00	Root (Root)	4 Own_Comms
Comm10_Own		Value: 00:03:52.00	Root (Root)	4 Own_Comms
Comm11_Other		Value: 00:04:15.00	Root (Root)	12 Other_Comms
Comm12_Own		Value: 00:04:37.00	Root (Root)	4 Own_Comms
Comm13_Other		Value: 00:05:05.00	Root (Root)	12 Other_Comms
Comm14_Own		Value: 00:05:29.00	Root (Root)	4 Own_Comms
Comm15_Own		Value: 00:05:58.00	Root (Root)	4 Own_Comms
Comm16_Other		Value: 00:06:23.00	Root (Root)	12 Other_Comms
Comm17_Own		Value: 00:06:46.00	Root (Root)	4 Own_Comms
Comm18_Other		Value: 00:07:09.00	Root (Root)	12 Other_Comms
Comm19_Own		Value: 00:07:32.00	Root (Root)	4 Own_Comms
Comm20_Own		Value: 00:07:56.00	Root (Root)	4 Own_Comms
Comm21_Other		Value: 00:08:20.00	Root (Root)	12 Other_Comms
Comm22_Own		Value: 00:08:45.00	Root (Root)	4 Own_Comms
Comm23_Own		Value: 00:09:10.00	Root (Root)	4 Own_Comms
Comm24_Own		Value: 00:09:34.00	Root (Root)	4 Own_Comms
Comm25_Own		Value: 00:11:05.00	Root (Root)	4 Own_Comms
Comm26_Other		Value: 00:12:47.00	Root (Root)	12 Other_Comms
Comm27_Own		Value: 00:14:17.00	Root (Root)	4 Own_Comms
Comm28_Own		Value: 00:15:25.00	Root (Root)	4 Own_Comms
Comm29_Own		Value: 00:17:12.00	Root (Root)	4 Own_Comms
Comm30_Other		Value: 00:19:01.00	Root (Root)	12 Other_Comms
ManualTracking1		Value: 00:05:00.00	Root (Root)	17 Start Tracking Loop
ManualTracking2		Value: 00:15:00.00	Root (Root)	17 Start Tracking Loop

Figure 91. Group 4 expert-derived model external event triggers.

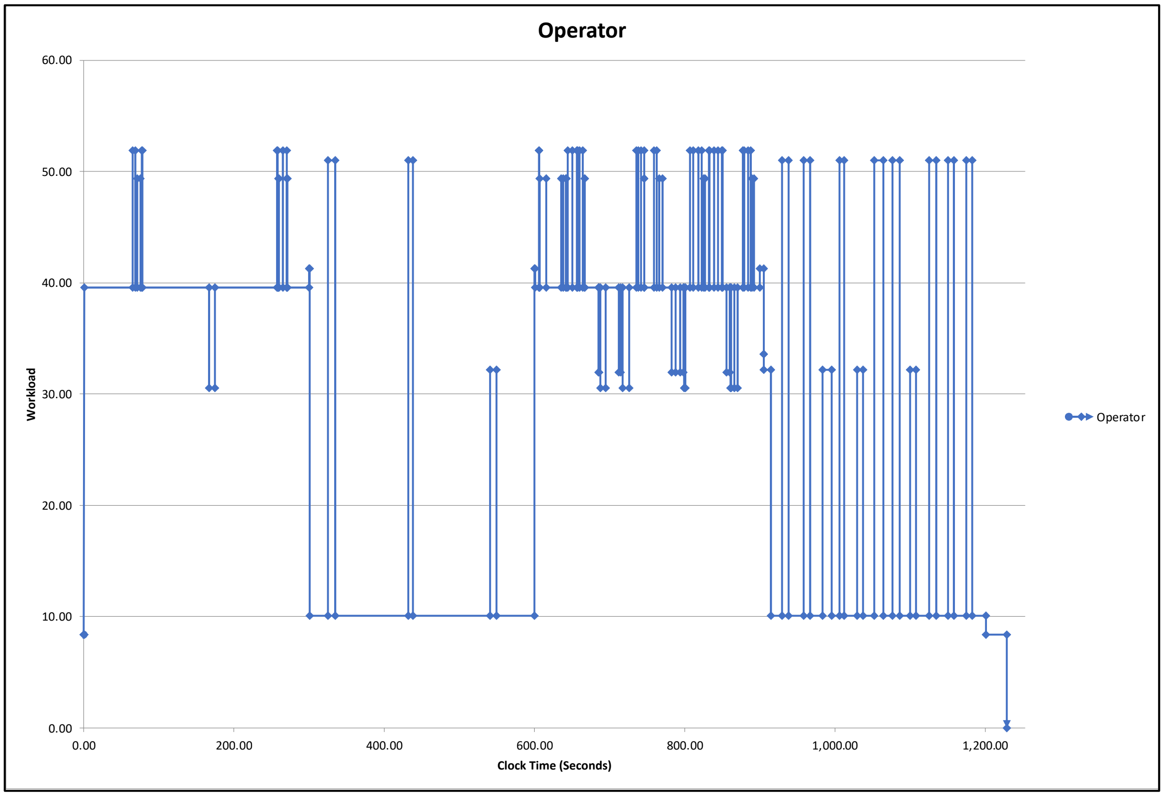


Figure 92. Study 3 expert-derived model group 1 IMPRINT workload graph.

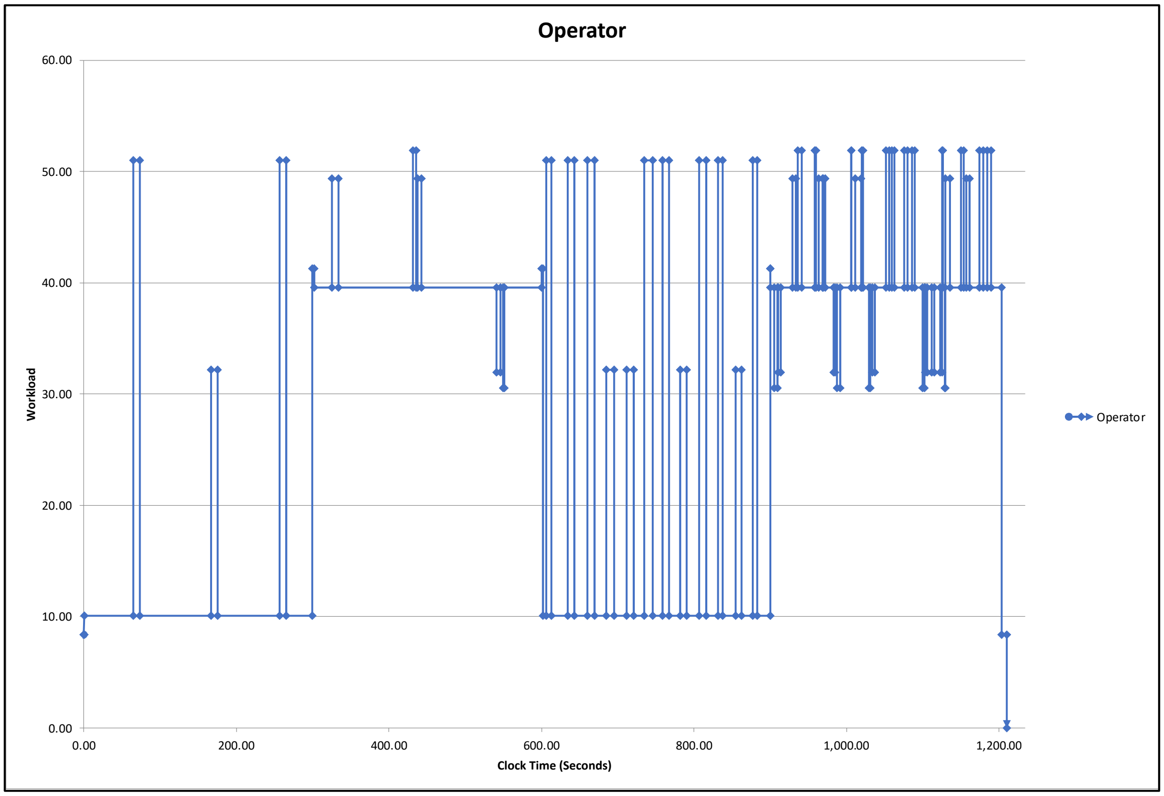


Figure 93. Study 3 expert-derived model group 2 IMPRINT workload graph.

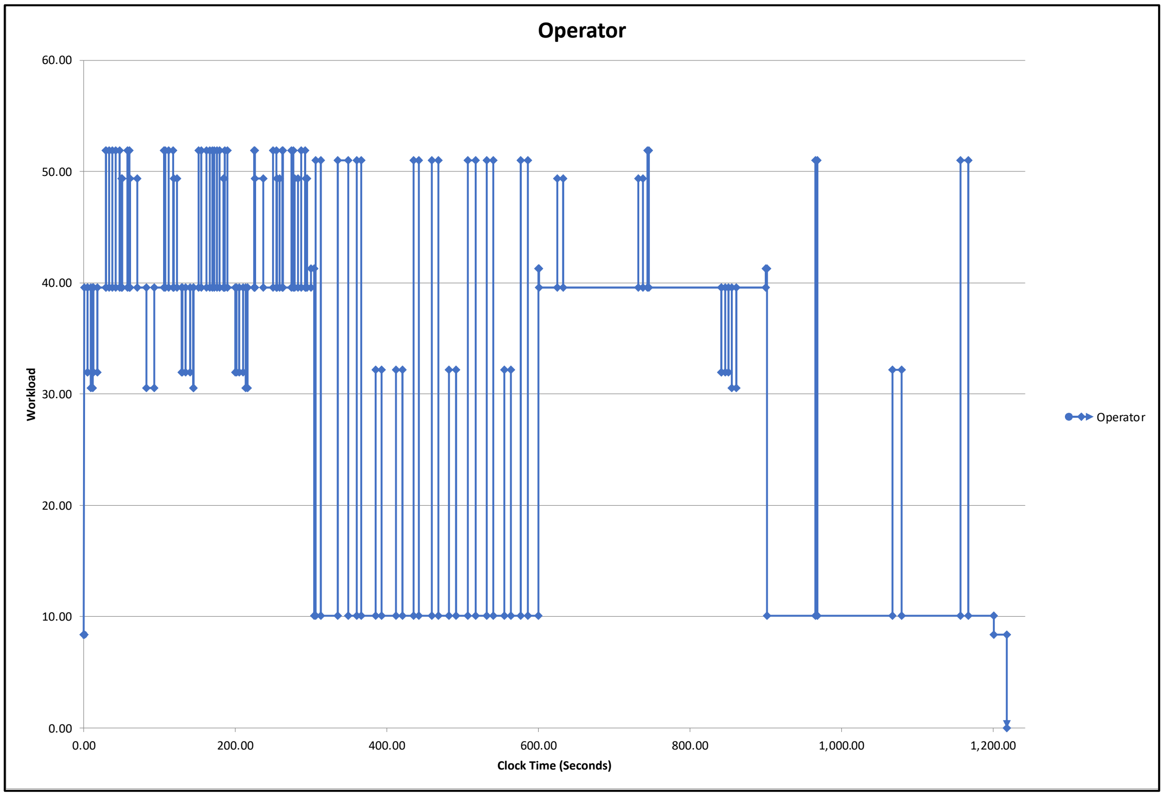


Figure 94. Study 3 expert-derived model group 3 IMPRINT workload graph

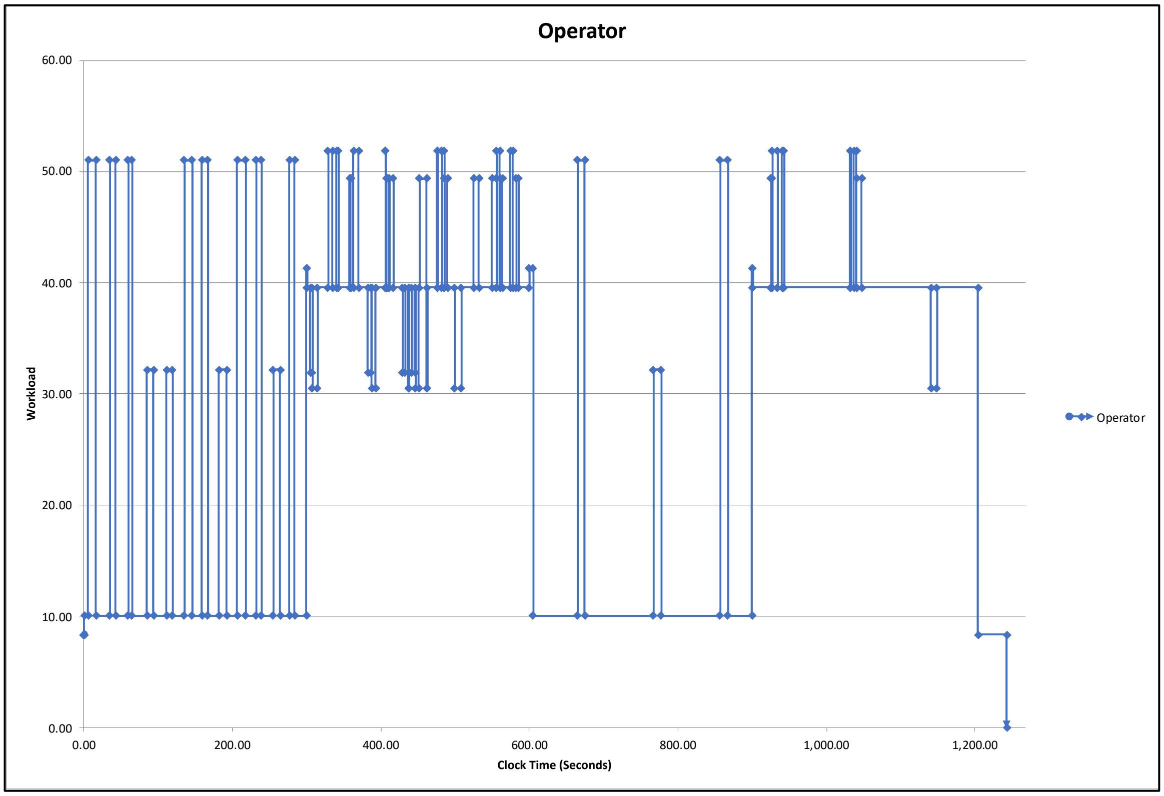


Figure 95. Study 3 expert-derived model group 4 IMPRINT workload graph

LIST OF REFERENCES

Adams, A. E., Rogers, W. A., & Fisk, A. D. (2012). Choosing the right task analysis tool. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *20*(1), 4–10. https://doi.org/10.1177/1064804611428925

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623–636.

Aldrich, T. B., Szabo, S. M., & Bierbaum, C. R. (1989). The development and application of models to predict operator workload during system design. In *Applications of Human Performance Models to System Design* (pp. 65–80). Springer.

Alion Science. (2018). *IMPRINT Pro User Guide*.

Alion Science. (2021). *IMPRINT Overview*.

Allender, L., Kelley, T., Lockett, J., & Archer, S. (2005). Human performance modeling in the Army: A long and winding road. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(12), 1191–1195. https://doi.org/10.1177/154193120504901220

Ames, L. L., & George, E. J. (1993). *Revision and Verification of a Seven-Point Workload Estimate Scale*. Air Force Test Center Edwards AFB CA.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., & Babiloni, F. (2016). Adaptive automation triggered by EEG-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Frontiers in Human Neuroscience*, *10*, 539. https://doi.org/10.3389/fnhum.2016.00539

Aura, C., Temme, L., Onge, P. S., McAtee, A., Wilson, M., Bowers, B., Baugher, K., & Hayes, A. (2021). Operator state monitoring via pupillometry: Measuring mental workload under varying luminance conditions. *Smart Biomedical and Physiological Sensor Technology XVIII*, *11757*, 78–86. https://doi.org/10.1117/12.2585989

Aura, C., Temme, L., & St. Onge, P. (2020). A Hands-Free, Visual Search Task Featuring Scalable Cognitive Challenge Based Upon the U.S. Army Trail Making Test. https://apps.dtic.mil/sti/citations/AD1094485

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, *59*(1), 36–47. https://doi.org/10.1016/j.neuroimage.2011.06.023

Backs, R. W. (1995). Going beyond heart rate: Autonomic space and cardiovascular assessment of mental workload. *The International Journal of Aviation Psychology*, *5*(1), 25–48. https://doi.org/10.1207/s15327108ijap0501_3

Bainbridge, L. (1982). Ironies of automation. *IFAC Proceedings Volumes*, *15*(6), 129–135. https://doi.org/10.1016/S1474-6670(17)62897-0

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of Psychophysiology*, *2*(142–162).

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction*, *3*(2), 74–99. https://doi.org/10.5898/JHRI.3.2.Beer

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(3), 199–208. https://doi.org/10.1016/j.trf.2010.12.001

Benyon, D. (1993). Adaptive systems: A solution to usability problems. *User Modeling and User-Adapted Interaction*, *3*(1), 65–87. https://doi.org/10.1007/BF01099425

Bohnsack, R., Bidmon, C., & Pinkse, J. (2021). Sustainability in the digital age: Intended and unintended consequences of digital technologies for sustainable development. *Business Strategy and the Environment*, *31*. https://doi.org/10.1002/bse.2938

Brand, Y., & Schulte, A. (2017). Model-based prediction of workload for adaptive associate systems. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1722–1727. https://doi.org/10.1109/SMC.2017.8122864

Brennan, S. D., & Jordan, C. S. (1992). *Instantaneous Self-Assessment of Workload (ISA)*. https://ext.eurocontrol.int/ehp/?q=node/1585

Brown, A. J., Folger, J. R., Hardin, J. W., Moore, J. D., & Sica, Q. (2021). *Task Handoff Between Humans and Automation*. Naval Postgraduate School.

Burtis, D. B., Heilman, K. M., Mo, J., Wang, C., Lewis, G. F., Davilla, M. I., Ding, M., Porges, S. W., & Williamson, J. B. (2014). The effects of constrained left versus right monocular viewing on the autonomic nervous system. In *Biological Psychology*, 100, 79–85. https://doi.org/10.1016/j.biopsycho.2014.05.006

Byrne, M. D. (2009). Cognitive architecture. In A. Sears & J. A. Jacko (Eds.), *Human-computer interaction. Fundamentals*. CRC Press.

Cagiltay, N. E., & Menekse Dalveren, G. G. (2012). Are left- and right-eye pupil sizes always equal? *Journal of Eye Movement Research*, *12*(2), 10.16910/jemr.12.2.1. https://doi.org/10.16910/jemr.12.2.1

Cain, B. (2007). *A Review of the Mental Workload Literature*. Defence Research and Development Toronto (Canada). https://apps.dtic.mil/sti/citations/ADA474193

Calhoun, G. (2021). Adaptable (not adaptive) automation: The forefront of human–automation teaming. *Human Factors*, 00187208211037457. https://doi.org/10.1177/00187208211037457

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-computer Interaction*. L. Erlbaum Associates.

Carnegie Mellon University. (2013). *ACT-R*. http://act-r.psy.cmu.edu/

Cassenti, D., Cox, K., & Bakdash, J. (2017). Multi-level cognitive cybernetics in human factors. In. K Hale & K. M. Stanney (Eds.), *Advances in Neuroergonomics and Cognitive Engineering* (Vol. 488, pp. 315–326). Springer International Publishing. https://doi.org/10.1007/978-3-319-41691-5_27

Cassenti, D., & Veksler, V. D. (2018). Using cognitive modeling for adaptive automation triggering. In D. Cassenti (Ed.), *Advances in Human Factors in Simulation and Modeling* (Vol. 591, pp. 378–390). Springer International Publishing. https://doi.org/10.1007/978-3-319-60591-3_34

Causse, M., Chua, Z., Peysakhovich, V., Del Campo, N., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-05378-x

Chen, S., & Epps, J. (2014). Using task-induced pupil diameter and blink rate to infer cognitive load. *Human-Computer Interaction*, *29*(4), 390–413. https://doi.org/10.1080/07370024.2014.892428

Christoffersen, K., & Woods, D. D. (2002). 1. How to make automated systems team players. In *Advances in Human Performance and Cognitive Engineering Research* (Vol. 2, pp. 1–12). Emerald Group Publishing Limited. https://doi.org/10.1016/S1479-3601(02)02003-9

Comstock, J. R., & Arnegard, R. J. (1992). *The Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research* (NAS 1.15:104174). https://ntrs.nasa.gov/citations/19920007912

Cooper, G. E., & Harper, R. P. (1969). *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*. National Aeronautics and Space Administration.

Cummings, M. L., & Thornburg, K. M. (2011). Paying attention to the man behind the curtain. *IEEE Pervasive Computing*, *10*(1), 58–62. https://doi.org/10.1109/MPRV.2011.7

Dahn, D., & Laughery, K. R. (1997). The integrated performance modeling environment—simulating human-system performance. *Proceedings of the 29th Conference on Winter Simulation*, 1141–1145.

de Greef, T., & Arciszewski, H. (2009). Triggering adaptive automation in naval command and control. In S. Cong (Ed.), *Frontiers in Adaptive Control*. InTech. https://doi.org/10.5772/6429

de Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 209–231. https://doi.org/10.1177/1555343411410160

Dehais, F., Causse, M., & Tremblay, S. (2011). Mitigation of conflicts with automation: Use of cognitive countermeasures. *Human Factors*, *53*(5), 448–460. https://doi.org/10.1177/0018720811418635

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work*, *6*, 79–86. https://doi.org/10.1007/s10111-003-0136-9

Delliaux, S., Delaforge, A., Deharo, J.-C., & Chaumet, G. (2019). Mental workload alters heart rate variability, lowering non-linear dynamics. *Frontiers in Physiology*, *10*, 565. https://doi.org/10.3389/fphys.2019.00565

Department of Defense. (2018). Department of Defense Instruction 5000.61, DOD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500061p.pdf

Department of Defense. (2020a). *Revolutionary Technology and Strategies for the Holistic Situational Awareness—Decision Making (HSA-DM) Program Request for Information (RFI)*.

Department of Defense. (2020b, September 15). *Department of Defense, Modeling and Simulation Glossary*. https://www.msco.mil/MSReferences/Glossary/MSGlossary.aspx

Department of the Army. (2019, November 18). *Human Performance Modeling—IMPRINT*. The United States Army CCDC Data and Analysis Center. https://www.dac.ccdc.army.mil/HPM_IMPRINT.html

Dowd, E. W., & Mitroff, S. R. (2013). Attentional guidance by working memory overrides salience cues in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1786.

Downes, S.M. (2020). *Models and modeling in the sciences: A philosophical introduction*. Routledge.

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society*, *24*(3), 177–181. https://doi.org/10.1177/0270467604264992

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M., & Giannopoulos, I. (2018). The index of pupillary activity: measuring cognitive load vis-à-vis task difficulty with pupil oscillation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3173574.3173856

Durbin, D. B., & Hicks, J. S. (2009). *AH-64D Apache Longbow aircrew workload assessment for unmanned aerial system (UAS) employment:* Defense Technical Information Center. https://doi.org/10.21236/ADA494123

Durso, F. T., Hackworth, C. A., & Truitt, T. R. (1999). *Situation awareness as a predictor of performance in en route air traffic controllers.* (DOT/FAA/AM-99/3). Federal Aviation Administration.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. In *Multiple-Task Performance*. CRC Press.

Ellis, C. J. (1981). The pupillary light reflex in normal subjects. *The British Journal of Ophthalmology*, *65*(11), 754–759. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1039657/

Ellis, G. A., & Roscoe, A. H. (1982). *The airline pilot's view of flight deck workload: A preliminary study using a questionnaire.* Royal Aircraft Establishment Farnborough (England). https://apps.dtic.mil/sti/citations/ADA116314

Embrey, D., Blackett, C., Marsden, P., & Peachey, J. (2006). *Development of a human cognitive workload assessment tool* (MCA Final Report, p. 290).

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 4–32. https://doi.org/10.1177/1555343415572631

Endsley, M. R. (2017a). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350

Endsley, M. R. (2017b). Direct measurement of situation awareness: Validity and use of SAGAT. In E. Salas (Ed.), *Situational Awareness* (1st ed., pp. 129–156). Routledge. https://doi.org/10.4324/9781315087924-9

Endsley, M. R. (2021). Situation awareness. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (1st ed., pp. 434–455). Wiley. https://doi.org/10.1002/9781119636113.ch17

Endsley, M. R., & Jones, D. G. (2011). *Designing for situation awareness: An approach to user-centered design* (2nd ed). CRC Press.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Ernst, K., Militello, L., Roth, E., Sushereba, C., & Scheff, S. (2020). *Future Vertical Lift Cognitive Workload Risk Mitigation Study—Final Report*. Applied Decision Science, LLC, Roth Cognitive Engineering, and HF Designworks Inc.

Evans, D. C., & Fendley, M. (2017). A multi-measure approach for connecting cognitive workload and automation. *International Journal of Human-Computer Studies*, *97*, 182–189. https://doi.org/10.1016/j.ijhcs.2016.05.008

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, *56*(2), 171–184. https://doi.org/10.1016/j.ijpsycho.2004.11.003

Feltman, K., Kelley, A. M., Bernhardt, K., Basso, J., & Morabito, C. (2021). *Psychophysiological Indicators of Aviator Flight Performance for Operator State Monitoring* (Technical Report USAARL-TECH-FR--2021-17). U.S. Army Aeromedical Research Laboratory.

Fiore, S. M., & Wiltshire, T. J. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology*, *7*, 1531. https://doi.org/10.3389/fpsyg.2016.01531

Flach, J. M. (1995). Situation Awareness: Proceed with Caution. *Human Factors*, *37*(1), 149–157. https://doi.org/10.1518/001872095779049480

Giles, G. E., Cantelon, J. A., Eddy, M. D., Brunyé, T. T., Urry, H. L., Mahoney, C. R., & Kanarek, R. B. (2017). Habitual exercise is associated with cognitive control and cognitive reappraisal success. *Experimental Brain Research*, *235*(12), 3785–3797. https://doi.org/10.1007/s00221-017-5098-x

Girouard, A., Solovey, E. T., Hirshfield, L. M., Peck, E. M., Chauncey, K., Sassaroli, A., Fantini, S., & Jacob, R. J. K. (2010). From brain signals to adaptive interfaces: Using fNIRS in HCI. In D. S. Tan & A. Nijholt (Eds.), *Brain-Computer Interfaces* (pp. 221–237). Springer London. https://doi.org/10.1007/978-1-84996-272-8_13

Goillau, P., & Kelly, C. (1996, October 1). MAlvern Capacity Estimate (MACE)—A proposed cognitive measure for complex systems. In *Engineering Psychology and Cognitive Ergonomics* (pp. 219–225). Routledge.

Goodman, T. J., Miller, M. E., & Rusnock, C. F. (2015). Incorporating automation: Using modeling and simulation to enable task re-allocation. 2015 Winter Simulation Conference (WSC), 2388–2399. https://doi.org/10.1109/WSC.2015.7408350

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, *83*(2), 314–320. https://doi.org/10.1037/0033-2909.83.2.314

Guastello, S. J., Shircel, A., Malon, M., & Timm, P. (2015). Individual differences in the experience of cognitive workload. *Theoretical Issues in Ergonomics Science*, *16*(1), 20–52. https://doi.org/10.1080/1463922X.2013.869371

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2014). Workload overload modeling: An experiment with MATB II to inform a computational model of task management. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 849–853.

Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). Human-automation interaction research: Past, present, and future. *Ergonomics in Design*, *21*(2), 9–14. https://doi.org/10.1177/1064804613477099

Hancock, P. A., & Matthews, G. (2019). Workload and performance: Associations, insensitivities, and dissociations. *Human Factors*, *61*(3), 374–392. https://doi.org/10.1177/0018720818809590

Harrison, J., İzzetoğlu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Woo, H., Shewokis, P. A., Bunce, S. C., & Onaral, B. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *IEEE Transactions on Human-Machine Systems*, *44*(4), 429–440. https://doi.org/10.1109/THMS.2014.2319822

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Heard, J., & Adams, J. A. (2019). Multi-dimensional human workload assessment for supervisory human–machine teams. *Journal of Cognitive Engineering and Decision Making*, *13*(3), 146–170. https://doi.org/10.1177/1555343419847906

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2014). Mental workload during n-back task—Quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/article/10.3389/fnhum.2013.00935

Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). *Operator versus computer control of adaptive automation*. The Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program: Interim Research and Guidelines for the Application of Adaptive Automation, 31–36.

Hoffman, R. R., & Woods, D. D. (2011). Beyond Simon's slice: Five fundamental trade-offs that bound the performance of macrocognitive work systems. *IEEE Intelligent Systems*, *26*(6), 67–71. https://doi.org/10.1109/MIS.2011.97

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press. https://doi.org/10.1201/9781420038194

Hou, M., Banbury, S., & Burns, C. (2014). *Intelligent adaptive systems: An interaction-centered design perspective*. CRC Press.

Hughes, A. M., Hancock, G. M., Marlow, S. L., Stowers, K., & Salas, E. (2019). Cardiac measures of cognitive workload: A meta-analysis. *Human Factors*, *61*(3), 393–414. https://doi.org/10.1177/0018720819830553

Hutton, R. J. B., & Klein, G. (1999). Expert decision making. *Systems Engineering*, *2*(1), 32–45. https://doi.org/10.1002/(SICI)1520-6858(1999)2:1<32::AID-SYS3>3.0.CO;2-P

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. *Handbook of Cognitive Task Design*, *8*, 147–169.

Janisse, M. P. (1977). *Pupillometry: The psychology of pupillary response*. Hemisphere Pub. Corp.

Jo, S., Myung, R., & Yoon, D. (2012). Quantitative prediction of mental workload with the ACT-R cognitive architecture. *International Journal of Industrial Ergonomics*, *42*(4), 359–370. https://doi.org/10.1016/j.ergon.2012.03.004

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. https://doi.org/10.1080/1463922021000054335

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel III, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, *36*(5), 447–462.

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. https://doi.org/10.1207/S15327566IJCE0501_3

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kanaan, D., & Moacdieh, N. M. (2021). Eye tracking to evaluate the effects of interruptions and workload in a complex task. *Human Factors*, 0018720821990487. https://doi.org/10.1177/0018720821990487

Karwowski, W., Haas, M., & Salvendy, G. (2006). A Review and Reappraisal of Adaptive Human-Computer Interfaces in Complex Control Systems (Technical Report AFRL-HE-WP-TR-2006-0123). Air Force Research Laboratory.

Khawaja, M. A., Chen, F., & Marcus, N. (2010). Using language complexity to measure cognitive load for adaptive interaction design. *Proceedings of the 15th International Conference on Intelligent User Interfaces – IUI '10*, 333. https://doi.org/10.1145/1719970.1720024

Khawaja, M. A., Chen, F., Owen, C., & Hickey, G. (2009). Cognitive load measurement from user's linguistic speech features for adaptive interaction design. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2009* (Vol. 5726, pp. 485–489). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03655-2_54

Kirwan, B., & Ainsworth, L. K. (1992). *A guide to task analysis: The task analysis working group*. CRC Press.

Kong, Y., Posada-Quintero, H. F., Gever, D., Bonacci, L., Chon, K. H., & Bolkhovsky, J. (2022). Multi-Attribute Task Battery configuration to effectively assess pilot performance deterioration during prolonged wakefulness. *Informatics in Medicine Unlocked*, *28*, 100822. https://doi.org/10.1016/j.imu.2021.100822

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE*, *13*(9), e0203629. https://doi.org/10.1371/journal.pone.0203629

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts. *Proceedings of the 2013 Conference on Eye Tracking South Africa*, 62–66. https://doi.org/10.1145/2509315.2509331

Lacdan, Joe. (2022, December 14). *Army awards contract to develop future vertical lift capability.* https://www.army.mil/article/262755/army_awards_contract_to_develop_future_vertical_lift_capability

Lebiere, C., Archer, R., Warwick, W., & Schunk, D. (2005). Integrating modeling and simulation into a general-purpose tool. *Proceedings of the 11th International Conference on Human Computer Interaction*, 11.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). *Designing for People: An Introduction to Human Factors Engineering* (3rd edition). CreateSpace.

Li, W., Li, R., Xie, X., & Chang, Y. (2022). Evaluating mental workload during multitasking in simulated flight. *Brain and Behavior*, *12*(4), e2489. https://doi.org/10.1002/brb3.2489

Liu, S., Wadeson, A., & Nam, C. S. (2016). Toward quantitative modeling of user performance in multitasking environments. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 3–9). Springer International Publishing. https://doi.org/10.1007/978-3-319-40030-3_1

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in Human Neuroscience*, *13*. https://www.frontiersin.org/articles/10.3389/fnhum.2019.00057

Longo, L., Wickens, C. D., Hancock, G., & Hancock, P. A. (2022). Human mental workload: A survey and a novel inclusive definition. *Frontiers in Psychology*, *13*, 883321. https://doi.org/10.3389/fpsyg.2022.883321

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal, 17(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Marshall, S. P. (2002). The Index of Cognitive Activity: Measuring cognitive workload. *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, 7–7. https://doi.org/10.1109/HFPP.2002.1042860

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2014). The Psychometrics of mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(1), 125.

McCracken, J. H., & Aldrich, T. B. (1984). *Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals:* Defense Technical Information Center. https://doi.org/10.21236/ADA232330

McCurry, C. D., Thanoon, M. I., & Zein-Sabatto, M. S. (2022). *Tennessee State University Multi-Attribute Task Battery (TSU-MATB)*. Department of Electrical and Computer Engineering, Tennessee State University.

McKendrick, R., Feest, B., Harwood, A., & Falcone, B. (2019). Theories and Methods for Labeling Cognitive Workload: Classification and Transfer Learning. *Frontiers in Human Neuroscience*, *13*. https://www.frontiersin.org/articles/10.3389/fnhum.2019.00295

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record*, *2138*(1), 6–12. https://doi.org/10.3141/2138-02

Militello, L. G., Roth, E., Scheff, S., Ernst, K., Sushereba, C., Diiulio, J., & Klein, D. (2019). *Toward an Optimally Crewed Future Vertical Lift Vehicle: Crewing Strategy and Recommendations*. U.S. Army Combat Capabilities Development Command.

Miller, N. L., & Shattuck, L. G. (2004). *A process model of situated cognition in military command and control*. Naval Postgraduate School.

Mitchell, D. K. (2000). *Mental workload and ARL workload modeling tools* (Technical Report ARL-TN-161). Army Research Laboratory.

Mitchell, D. K. (2009). *Workload Analysis of the Crew of the Abrams V2 SEP: Phase I Baseline IMPRINT Model*. Army Research Laboratory. https://apps.dtic.mil/sti/citations/ADA508882

Moray, N. (2013). *Mental workload: Its theory and measurement*. Springer Science & Business Media.

Murai, K., Hayashi, Y., Maenaka, K., & Higuchi, K. (2015). Basic study on evaluation of navigator's mental workload by sticking plaster-type sensor. *2015 10th System of Systems Engineering Conference (SoSE)*, 83–87. https://doi.org/10.1109/SYSOSE.2015.7151931

Neubauer, C., Schaefer, K. E., Oiknine, A. H., Thurman, S., Files, B., Gordon, S., Bradford, J. C., Spangler, D., & Gremillion, G. (2020). *Multimodal Physiological and Behavioral Measures to Estimate Human States and Decisions for Improved Human–Autonomy Teaming*. Army Research Laboratory.

Nguyen, T., Lim, C. P., Nguyen, N. D., Gordon-Brown, L., & Nahavandi, S. (2019). A review of situation awareness assessment approaches in aviation environments. *IEEE Systems Journal*, *13*(3), 3590–3603. https://doi.org/10.1109/JSYST.2019.2918283

O'Brien, K. (2022). Neurokit2 HRV Visualization. U.S. Army Aeromedical Research Laboratory.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Cognitive processes and performance. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance: Vol. II*. John Wiley & Sons. https://apps.dtic.mil/sti/citations/ADA289587

Olson, E. (2010). *AprilTag*. https://april.eecs.umich.edu/software/apriltag

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2020). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors*, 0018720820960865. https://doi.org/10.1177/0018720820960865

Palma Fraga, R., Reddy, P., Kang, Z., & Izzetoglu, K. (2020). Multimodal analysis using neuroimaging and eye movements to assess cognitive workload. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Augmented Cognition. Theoretical and Technological Approaches* (Vol. 12196, pp. 50–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-50353-6_4

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, *29*(6), 695–706. https://doi.org/10.1177/001872088702900609

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 5–20. https://doi.org/10.1080/14639220210199753

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and Design of Adaptive Automation in Aviation Systems*. Naval Air Warfare Center. https://apps.dtic.mil/sti/citations/ADA254595

Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., & Rizzo, M. (Eds.). (2008). *Neuroergonomics: The Brain at Work*. Oxford University Press.

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. https://doi.org/10.1109/3468.844354

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Parsons, H. M. (1985). Automation and the individual: Comprehensive and comparative views. *Human Factors*, *27*(1), 99–111.

Patten, C. J. D., Kircher, A., Östlund, J., Nilsson, L., & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis & Prevention, 38*(5), 887–894. https://doi.org/10.1016/j.aap.2006.02.014

Perret-Guillaume, C., Joly, L., & Benetos, A. (2009). Heart rate as a risk factor for cardiovascular disease. *Progress in Cardiovascular Diseases*, *52*(1), 6–10. https://doi.org/10.1016/j.pcad.2009.05.003

Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016). A model relating pupil diameter to mental workload and lighting conditions. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5776–5788. https://doi.org/10.1145/2858036.2858117

Pinti, P., Aichelburg, C., Lind, F., Power, S., Swingler, E., Merla, A., Hamilton, A., Gilbert, S., Burgess, P., & Tachtsidis, I. (2015). Using fiberless, wearable fnirs to monitor brain activity in real-world cognitive tasks. *Journal of Visualized Experiments*, *106*, 53336. https://doi.org/10.3791/53336

Potter, S. S., & Bressler, J. R. (1989). *Subjective Workload Assessment Technique (SWAT): A User's Guide*. Systems Research Labs Inc.

Prahl, S. (1998). *Tabulated Molar Extinction Coefficient for Hemoglobin in Water*. https://omlc.org/spectra/hemoglobin/summary.html

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction*. Wiley.

Proctor, R. W., & Van Zandt, T. (2018). *Human factors in simple and complex systems* (Third edition). CRC Press.

Pupil Labs. (2022, August 15). *Core—Pupil capture*. Pupil Labs. https://docs.pupil-labs.com

Reddy, P., Shewokis, P. A., & Izzetoglu, K. (2022). Individual differences in skill acquisition and transfer assessed by dual task training performance and brain activity. *Brain Informatics*, *9*(1), 9. https://doi.org/10.1186/s40708-022-00157-5

Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of Surgery*, *252*(1), 177–182. https://doi.org/10.1097/SLA.0b013e3181e464fb

Ritter, F. E., & Young, R. M. (2001). Embodied models as simulated users: Introduction to this special issue on using cognitive models to improve interface design. *International Journal of Human-Computer Studies*, *55*(1), 1–14. https://doi.org/10.1006/ijhc.2001.0471

Rosala, M. (2020, September 20). *Task analysis: Support users in achieving their goals*. Nielsen Norman Group. https://www.nngroup.com/articles/task-analysis/

Roscoe, A. H. (1984). *Assessing pilot workload in flight*. Royal Aircraft Establishment Bedford (United Kingdom) Bedford United Kingdom.

Roth, E., Klein, D., Sushereba, C., Ernst, K., & Militello, L. (2022). *Methods and measures to evaluate technologies that influence aviator decision making and situation awareness*. Applied Decision Science.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors*, *30*(4), 431–443. https://doi.org/10.1177/001872088803000405

Rowe, D., Sibert, J., & Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. *CHI '98: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 480–487. https://doi.org/10.1145/274644.274709

Rusnock, C. F., & Geiger, C. D. (2017). Simulation-based evaluation of adaptive automation revoking strategies on cognitive workload and situation awareness. *IEEE Transactions on Human-Machine Systems*, *47*(6), 927–938. https://doi.org/10.1109/THMS.2016.2618004

Samms, C. (2010). Improved performance research integration tool (IMPRINT): Human performance modeling for improved system design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(7), 624–625. https://doi.org/10.1177/154193121005400701

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide* (Technical Report NASA/TM–2011-217164; p. 269). National Aeronautics and Space Administration.

Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation*, *7*(1), 12–24. https://doi.org/10.1057/jos.2012.20

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? mode error and awareness in supervisory control. *Human Factors*, *37*(1), 5–19. https://doi.org/10.1518/001872095779049516

Save, L., Feuerberg, B., & Avia, E. (2012). Designing human-automation interaction: A new level of automation taxonomy. *Proc. Human Factors of Systems and Technology*, *2012*.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In *Automation and Human Performance: Theory and Applications*. CRC Press.

Scerbo, M. W. (2008). Adaptive automation. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The Brain at Work* (pp. 239–252). Oxford University Press.

Scerbo, M. W., Freeman, F. G., Mikulka, P. J., Parasuraman, R., Nocero, F. D., & Prinzel, L. G. I. (2001). *The Efficacy of Psychophysiological Measures for Implementing Adaptive Technology*. NASA Langley Technical Report Server.

Shattuck, L. G. (2017). From cognitive systems engineering to human systems integration: A short but necessary journey. In *Cognitive Systems Engineering* (pp. 309–326). CRC Press.

Shattuck, L. G., & Miller, N. L. (2006). Extending naturalistic decision making to complex organizations: A dynamic model of situated cognition. *Organization Studies*, *27*(7), 989–1009. https://doi.org/10.1177/0170840606065706

Sheridan, T. B. (2011). Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, *41*(4), 662–667. https://doi.org/10.1109/TSMCA.2010.2093888

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, *37*(1), 137–148. https://doi.org/10.1518/001872095779049444

Smith, P., & Baumann, E. (2020). Human-automation teaming: Unintended consequences of automation on user performance. *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 1–9.

Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science: Situation awareness in individuals, teams and systems. *Ergonomics*, *60*(4), 449–466. https://doi.org/10.1080/00140139.2017.1278796

Steinhauer, S. R., Bradley, M. M., Siegle, G. J., Roecklein, K. A., & Dix, A. (2022). Publication guidelines and recommendations for pupillary measurement in psychophysiological studies. *Psychophysiology*, *59*(4). https://doi.org/10.1111/psyp.14035

Stilgoe, J. (2018). Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science*, *48*(1), 32.

Strickland, J. S. (2011). *Using math to defeat the enemy: Combat modeling for simulation* (1st Edition). Lulu, Inc.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of mental workload. *International Journal of Environmental Research and Public Health*, *16*(15), 2716. https://doi.org/10.3390/ijerph16152716

Taylor, R. M. (2011). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In E. Salas & A. S. Dietz (Eds.), *Situational Awareness* (p. 18). Routledge.

Tonn, B. E., & Stiefel, D. (2019). Anticipating the unanticipated-unintended consequences of scientific and technological purposive actions. *World Futures Review*, *11*(1), 19–50. https://doi.org/10.1177/1946756718789413

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, *39*(3), 358–381.

University of Michigan. (2021). *Soar*. Soar Cognitive Architecture. https://soar.eecs.umich.edu

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, *53*, 190–202. https://doi.org/10.1016/j.apergo.2015.09.013

Van Acker, B., Parmentier, D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: From a clarifying concept analysis toward an implementable framework. *Cognition, Technology & Work*, *20*. https://doi.org/10.1007/s10111-018-0481-3

Vidulich, M. A. (2000). The relationship between mental workload and situation awareness. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting*, *3*, 460. https://www.proquest.com/docview/235449236/citation/3165681E18C74582PQ/2

Vogl, J. (2022). XDF Data Visualization Script. U.S. Army Aeromedical Research Laboratory.

Vogl, J., Hayes, A., Aura, C., Milam, L., Temme, L., & St. Onge, P. (2020). *A Case for a Measurement of Incremental Workload*. U.S. Army Aeromedical Research Laboratory. https://apps.dtic.mil/sti/citations/AD1107452

von Lühmann, A., Herff, C., Heger, D., & Schultz, T. (2015). Toward a wireless open source instrument: Functional near-infrared spectroscopy in mobile neuroergonomics and BCI applications. *Frontiers in Human Neuroscience*, *9*. https://www.frontiersin.org/articles/10.3389/fnhum.2015.00617

Watson, A. B., & Yellott, J. I. (2012). A unified formula for light-adapted pupil size. *Journal of Vision*, *12*(10), 12. https://doi.org/10.1167/12.10.12

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199933662.001.0001

Welford, A. T. (1978). Mental work-load as a function of demand, capacity, strategy and skill. *Ergonomics*, *21*(3), 151–167. https://doi.org/10.1080/00140137808931710

Wickens, C. D. (1981). *Processing resources in attention, dual task performance, and workload assessment.* (Technical Report EPL-SI-3/ONR-81-3). Office of Naval Research. https://apps.dtic.mil/sti/citations/ADA102719

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D. (2008a). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wickens, C. D. (2008b). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *Human Factors*, *50*(3), 397–403. https://doi.org/10.1518/001872008X288420

Wickens, C. D., Boles, D., Tsang, P., & Carswell, M. (1984). *The limits of multiple resource theory in display formatting: Effects of task integration*. University of Illinois at Champagne-Urbana. https://apps.dtic.mil/sti/citations/ADP003321

Wiener, E. L. (1988). 13—Cockpit automation. In *Human factors in aviation* (pp. 433–461). Academic Press. https://doi.org/10.1016/B978-0-08-057090-7.50019-9

Wiener, E. L. (1989). *Human factors of advanced technology (glass cockpit) transport aircraft* (Technical Report No. 177528; NASA Ames). NASA.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, *23*(10), 995–1011. https://doi.org/10.1080/00140138008924809

Wiener, N. (1948). *Cybernetics or control and communication in the animal and the machine*. The MIT Press.

Wilkins, J. C., Feltman, K., & Aura, C. (2022). *The Feasibility of Operator State Monitoring in Army Aviation: A Survey of Subject Matter Experts* (Technical Report USAARL-TECH-TR–2023-01). U.S. Army Aeromedical Research Laboratory.

Wojciechowski, J. Q. (2004). *Validation of improved research integration tool (IMPRINT) driving model for workload analysis*. Army Research Lab Human Research and Engineering Directorate. https://apps.dtic.mil/sti/citations/ADA422389

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In *Automation and Human Performance: Theory and Applications*. CRC Press.

Woods, D. D. (2002). Steering the reverberations of technology change on fields of practice: Laws that govern cognitive work. *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*.

Woods, D. D. (2006). The law of stretched systems in action: Exploiting robots. HRI 2006: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robt Interaction.

Woods, D. D. (2016). The risks of autonomy: Doyle's catch. *Journal of Cognitive Engineering and Decision Making*, *10*(2), 131–133. https://doi.org/10.1177/1555343416653562

Woods, D. D., & Dekker, S. (2000). Anticipating the effects of technological change: A new era of dynamics for human factors. *Issues in Ergonomics Science*, *1*(3), 272–282. https://doi.org/10.1080/14639220110037452

Woods, D. D., Licu, T., Leonhardt, J., Rayo, M., Balkin, E. A., & Cioponea, R. (2021). *Patterns in how people think and work: Importance of patterns discovery for understanding complex adaptive systems.* European Organisation for the Safety of Air Navigation (EUROCONTROL).

Wusk, G. C., Abercromby, A. F., & Gabler, H. C. (2019). Psychophysiological monitoring of aerospace crew state. *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 404–407. https://doi.org/10.1145/3341162.3349309

Xu, Y., Graber, H., & Barbour, R. (2017). *NirsLAB user manual*. NIRX Medical Technologies. https://www.nitrc.org/frs/?group_id=651

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. https://doi.org/10.1080/00140139.2014.956151

Zak, Y., Parmet, Y., & Oron-Gilad, T. (2020). Subjective workload assessment technique (SWAT) in real time: Affordable methodology to continuously assess human operators' workload. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2687–2694. https://doi.org/10.1109/SMC42975.2020.9283168

Zhang, J., Cao, X., Wang, X., Pang, L., Liang, J., & Zhang, L. (2021). Physiological responses to elevated carbon dioxide concentration and mental workload during performing MATB tasks. *Building and Environment*, *195*, 107752. https://doi.org/10.1016/j.buildenv.2021.107752

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
WWW.NPS.EDU
WHERE SCIENCE MEETS THE ART OF WARFARE