USAARL-TECH-TR--2026-01

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

The Relationship Between Trust, Control, and Automation Reliance in High-Demand Aviation Tasks

Bethany Ranes, Jon Vogl, & J. Andrew Atchley

Notice

Qualified Requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Fort Belvoir, Virginia 22060. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of Address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

Human Subject Use

In the conduct of research involving human subjects, the investigator(s) adhered to the policies regarding the protection of human subjects as prescribed by Department of Defense Instruction 3216.02 (Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research) and 32 CFR 219 (Protection of Human Subjects).

This study, USAARL 2023-019, was approved by the Medical Research and Development Command Institutional Review Board on 23 April 2024.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
24-11-2025	Technical Report	APR 2024 - JUL 2025

4. TITLE AND SUBTITLE
The Relationship Between Trust, Control, and Automation Reliance in High-Demand Aviation Tasks

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
6.3

6. AUTHOR(S)
Ranes, B.[1,2], Vogl, J.[1], & Atchley, J. A.[1,2]

5d. PROJECT NUMBER
MO230061

5e. TASK NUMBER
3.4.2

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL-TECH-TR--2026-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Military Operational Medicine Research Program
504 Scott Street
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
USAMRDC MOMRP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited.

13. SUPPLEMENTARY NOTES
[1]U.S. Army Aeromedical Research Laboratory; [2]Chickasaw Nation Industries

14. ABSTRACT
Trust in automation (TIA) is a critical factor in aviation safety and performance, yet the extent to which self-reported trust aligns with actual automation use remains unclear. This study investigated relationships among trait, state, and behavioral measures of TIA in a simulated multitasking aviation environment. Seventeen active-duty military aviators completed four trials of the USAARL Multi-Attribute Task Battery-II (MATB-II), during which automation was available across four subtasks. Automation reliability (70-percent vs. 90-percent) and task load order were varied within subjects. Participants completed standardized self-report measures of trait TIA (Adapted Propensity to Trust in Technology Questionnaire), state TIA (Checklist for Trust between People and Automation), and workload, alongside behavioral indicators of automation engagement: Reliance (time automation was engaged), delegation (user-initiated use), and compliance (prompt-initiated use). Data were analyzed using linear and generalized linear mixed-effects models. Results showed that reliance on automation significantly improved task performance, with a 10-percent increase in automation use corresponding to a 1.06-point improvement in task score ($p < .001$). In contrast, reliance did not reduce perceived workload ($p = .957$). Trait TIA significantly predicted state TIA ($\beta = 1.38$, $p = .047$), but neither predicted behavioral use of automation (all $p > .48$).

15. SUBJECT TERMS
trust in automation, cognitive workload, beliefs, attitudes, behavior, aviators

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Loraine St. Onge, PhD
UNCLAS	UNCLAS	UNCLAS	SAR	18	19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

Reset

REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheet)

14. Abstract (continued)

Instead, contextual factors drove automation engagement: Delegation was more likely in dynamic subtasks such as resource management and tracking (both $p < .001$), whereas higher cognitive workload increased compliance ($p < .001$). Delegation exerted a stronger effect on sustained reliance than compliance ($\beta = 0.309$ vs. $\beta = 0.193$, both $p < .001$), underscoring the importance of operator agency when considering automation use behaviors.

These findings highlight a belief-behavior gap in TIA, suggesting that TIA self-report measures alone are insufficient to predict aviators' automation use. System designs that support operator autonomy while providing workload-sensitive prompting may better calibrate reliance and optimize human-machine teaming in aviation contexts.

Summary

Trust in automation (TIA) is a critical factor in aviation safety and performance, yet the extent to which self-reported trust aligns with actual automation use remains unclear. This study investigated relationships among trait, state, and behavioral measures of TIA in a simulated multitasking aviation environment. Seventeen active-duty military aviators completed four trials of the U.S. Army Aeromedical Research Laboratory (USAARL) Multi-Attribute Task Battery-II (MATB-II), during which automation was available across four subtasks. Automation reliability (70-percent vs. 90-percent) and task load order were varied within subjects. Participants completed standardized self-report measures of trait TIA (Adapted Propensity to Trust in Technology Questionnaire), state TIA (Checklist for Trust between People and Automation), and workload, alongside behavioral indicators of automation engagement: reliance (time automation was engaged), delegation (user-initiated use), and compliance (prompt-initiated use). Data were analyzed using linear and generalized linear mixed-effects models.

Results showed that reliance on automation significantly improved task performance, with a 10-percent increase in automation use corresponding to a 1.06-point improvement in task score ($p < .001$). In contrast, reliance did not reduce perceived workload ($p = .957$). Trait TIA significantly predicted state TIA ($\beta = 1.38$, $p = .047$), but neither predicted behavioral use of automation (all $p > .48$). Instead, contextual factors drove automation engagement: delegation was more likely in dynamic subtasks such as resource management and tracking (both $p < .001$), whereas higher cognitive workload increased compliance ($p < .001$). Delegation exerted a stronger effect on sustained reliance than compliance ($\beta = 0.309$ vs. $\beta = 0.193$, both $p < .001$), underscoring the importance of operator agency when considering automation use behaviors.

These findings highlight a belief-behavior gap in TIA, suggesting that TIA self-report measures alone are insufficient to predict aviators' automation use. System designs that support operator autonomy while providing workload-sensitive prompting may better calibrate reliance and optimize human-machine teaming in aviation contexts.

This page is intentionally blank.

Acknowledgements

This research was supported in part by an appointment to the U.S. Army Aeromedical Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Army Medical Research and Development Command.

This page is intentionally blank.

Table of Contents

Table of Contents (continued)

List of Figures and Tables

Introduction

Complex technology advancements are rapidly increasing and being integrated into a wide variety of operational environments; this is especially true in military environments. As the complexity of these technologies increases and as their role becomes more centralized to critical tasks, humans who are interacting with advanced technology on a regular basis may find it more difficult to oversee successful operation without assistance. Automated systems have the potential to significantly improve the interactions between advanced technology and human operators, optimizing the strengths of both human and non-human components to maximize the frequency of successful outcomes. However, for an automation to be effective, it must be used appropriately by the human operator, and appropriate use is largely dictated by the human operator's level of trust in the automation (TIA). For this reason, TIA has emerged as one of the most significant considerations for engineering the next generation of complex technological innovations.

Defining Trust in Automation

Trust is generally defined as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004); an "agent" was traditionally considered to be another person that actively interacted with the environment on behalf of an operator, but this definition has been found through decades of research to be suitable for describing TIA as well (Kohn et al., 2021). Trust is a critical component in ensuring proper human engagement and performance in any type of collaboration, and this is no different with collaborations with automation, which includes any technology that "actively selects data, transforms information, makes decisions, or controls processes" (Lee & See, 2004, p. 50). Just like trust among humans, TIA is highly complex and is dynamically influenced by many intertwining factors that can have immediate and significant impacts on a user's subsequent behavior during an automation-enhanced task (Lee & See, 2004; Hoff & Bashir, 2015; Mayer et al., 1995).

In general, these factors have been identified by researchers as human-based, automation-based, and environment-based (Lee & See, 2004; Hoff & Bashir, 2015). Human-based trust factors are those that are related to the operator themselves, and include personality traits, pre-existing knowledge, ethnicity, age, and gender (Lu & Sarter, 2019). Automation-based trust factors are those that are inherent to the system being used, and include system reliability, ability, robustness, and predictability. Environment-based factors are perhaps the most complicated and difficult to measure element of trust in an interaction, and include things like prior experience, societal impact, culture, team collaboration, and task type. TIA is also a highly dynamic construct based upon one's experience across time, and experiences can influence trust-related beliefs and behaviors in both the short-term (e.g., during an automated task) and long-term (e.g., a long career with experiences across multiple types of automations).

TIA is also complicated by its bi-directional nature, as optimal TIA lies in the middle of a spectrum between under-trusting and over-trusting a system. The overall level of TIA impacts the level of vigilance and sustainable attention an operator will give toward automation (Krausman et al., 2022). When operators place too much trust in a system (also called complacency), it can lead to an increased risk of mistakes, incidents, and accidents related to the user being “out of the loop” (Krausman et al., 2022; Lu & Sarter, 2019). Low levels of trust cause disuse of an automated system and can lead to unnecessarily high levels of user workload driven by the need to constantly (and unnecessarily) monitor automated systems to ensure safety and accuracy (Lu & Sarter, 2019). Increased workload can significantly increase the risk of mistakes and oversights that would otherwise be identified and accounted for by the automation (Lee & See, 2004). Several researchers have hypothesized that calibrating appropriate levels of user trust through measurement and modification of human-based, automation-based, and environment-based TIA factors will result in ideal user reliance for optimal performance outcomes (Lee & See, 2004; Mayer et al., 1995; Sanders et al., 2011).

Measuring TIA across Domains

Valid and reliable measurement of TIA across human-based, automation-based, and environment-based domains is a difficult task. Trust is an emotional construct. As with any other emotional variable, it is perceptual, which means the experience of trust is context-dependent, based on a complex interaction of environmental and psychophysiological components, and entirely unique to the individual experiencing it (Barrett, 2017). In the case of TIA, measures are often used to inform the automation itself and therefore require data that are not only valid and reliable but are also continuously collected and able to be analyzed and modeled at an interval or ratio level of measurement (Wei et al., 2020). Researchers have attempted to overcome the inherent obstacles of properly measuring TIA by developing a wide array of instruments and metrics, ranging from simple single-item self-report measures to algorithmically constructed values representing the integration of several different elements related to trust. While most researchers advocate the use of a multi-modal approach to capturing TIA (a combination of different types of TIA metrics collected simultaneously and interpreted holistically), single-type TIA metrics generally fall into three broad categories: user-reported, physiological, and behavioral.

A systematic review of measures across each of these three categories, and their suitability for use in an aeromedical environment, was conducted by Ranes, Wilkins, Kenser, & Caid-Loos (2023); behavioral measures of TIA were deemed to be of the highest quality and the most suitable for aeromedical applications (see Table 1 below), although a number of user-reported measures were also deemed valuable and feasible for capturing trait-based measures of TIA prior to or immediately following a flight mission. The two highest-rated user-reported TIA measures for an aeromedical environment were the Adapted Propensity to Trust Questionnaire (Jessup, 2019) and the Checklist for Trust between People and Automation (Jian, 2000). Both of these measures were included in the present study alongside passively-collected behavioral measures. While several physiological measures of TIA were also found to be appropriate for aeromedical applications (Ranes, Wilkins, Kenser, & Caid-Loos, 2023), these variables were all related to a user’s level of physiological arousal, which is more appropriately aligned with levels of stress rather than as a direct measure of TIA (Mayer et al., 1995; Lee & See, 2004; Hoff & Bashir, 2015); for that reason, physiological measures of TIA were not included in the present study.

Table 1. Behavioral Indicators of TIA With High Recommendation for Aeromedical Environments (Ranes et al., 2023)

Behavioral Measures	**Description**
Compliance with automation recommendations	The operator uses recommendations given by the automation
Delegation	Allowing automation to handle a task when the operator could do it manually
Intervention	Overriding automation/taking manual control (even when the automation is accurate)
Reliance	The operator makes no attempt to override the automation (even when mistakes are made)
Response Time	How long it takes the operator to act after a prompt or alert from the automation

Automation Reliability

A significant factor in an operator's TIA levels is how reliable the automation is in achieving its intended aims; unreliable automations decrease TIA, and vice-versa (Hoff & Bashir, 2015). Operators can also experience low trait TIA before ever engaging with an automated system if they have established a belief that the automation is unreliable prior to an interaction (from peer reports, previous negative experiences with other automations, etc.). Conversely, operators are more likely to go into an interaction with high trait TIA if they have established a prior belief that the automation is reliable (Hoff & Bashir, 2015). Once an interaction begins, operators are constantly fine-tuning their level of TIA based on feedback from the system – if an automation helps them to achieve their aims and/or reduce the effort necessary to achieve the aims, then trust in the system is naturally expected to increase. If the automation fails to achieve the operator's aims, then trust diminishes. Automation reliability also provides context for establishing the appropriateness of an operator's level of TIA – high TIA in a low-reliability automation condition and/or low TIA in a high-reliability automation condition are inappropriate levels of TIA that both present risks to subsequent task performance.

Mediators Between TIA and Performance

In order to properly measure and respond to dynamic changes in TIA during an automation-assisted task, it is critical to consider and quantify key mediators in the relationship between TIA and subsequent performance. Once patterns of mediation are properly understood, continuous monitoring of key mediators can be used to predict TIA and subsequent performance in real time, allowing for valuable dynamic adjustments to the automated system. While there are countless potential mediating variables that come into play for individual trust and performance scenarios, there are three critical TIA mediators that have been repeatedly identified across studies to influence the population at large: difficulty of the automated task, cognitive workload, and propensity to trust automation.

Difficulty of the automated task.

Difficulty of the task has also been identified as a critical environmental factor in shaping trust-related behaviors during human-automation interaction (Lee & See, 2004; Hoff & Bashir, 2015). While it may be assumed that operators are more likely to rely on automation when task demands are high, the relationship between task difficulty and automation use is complex and context dependent. Some studies suggest that users may engage automation more frequently in difficult or high-load tasks to reduce cognitive burden, while others indicate that high task demands can cause users to disengage from automation due to increased perceived vulnerability or reduced perceived control (Parasuraman et al., 2008; Drnec et al., 2016). These contradictory patterns suggest that task difficulty may interact with other factors, such as individual differences or the mode of automation engagement, to determine actual operator behavior. In the present study, we explored the extent to which task type influenced the use of automation, particularly in relation to user-initiated versus system-prompted behaviors.

Cognitive workload.

Increases in an operator's cognitive workload have been shown in multiple instances to significantly increase TIA (regardless of operators' self-reported trait-level TIA), and more specifically increase the likelihood of complacency even when an automation is unreliable (Parasuraman et al., 2008). Other studies have suggested a bi-directional relationship where workload is also influenced by TIA; as trust in an unreliable automation goes down, subjects report increased cognitive workload (Park et al., 2019). The present study observed changes to TIA (user-reported and behavioral) and cognitive workload in two types of conditions – unreliable automation and reliable automation – to evaluate this relationship. The U.S. Army Aeromedical Research Laboratory (USAARL) MATB-II has a built-in measure of perceived workload that users are prompted to respond to multiple times throughout each trial. This measure served as a subjective source of cognitive workload data for the present study.

Propensity to trust automated systems.

While frequently included as a measure of TIA, an individual's propensity to trust automated systems is less dynamic than the learned TIA that evolves during an automated task. Hoff and Bashir's (2015) theoretical model of TIA classifies an individual's propensity to trust as an initial learned aspect of trust that influences subsequent interactions with automated systems. It can be thought of as a baseline level of TIA that an operator brings with them prior to each interaction with an automated system. The Adapted Propensity to Trust Automation Questionnaire (APTQ; Jessup et al., 2019) was used in the present study to capture this initial TIA level prior to a subject's interaction with the tasks or automated systems.

While many publications on TIA have accounted for these mediators in a theoretical model, there are fewer studies that have quantified the effects of how TIA and its mediators impact task performance, and even fewer that take into account variations in the level of an automation's reliability. Automation reliability is an important component that offers a sense of whether an operator's level of TIA is appropriate to the conditions at hand; high TIA for a low-reliability automation (and vice versa) both present risks for degraded performance. Previous studies that have incorporated both mediators and variable automation quality are largely specific to tasks that have little to no relevance for aeromedical operations. In order to gain a sense of how TIA is affected by recordable mediators, and how it subsequently impacts downstream performance, we need to gain a sense of how these variables interact to predict when operators may experience a shift in TIA that potentially jeopardizes their ability to interact safely with automated technology.

Purpose of the Present Study

The overall goal of this study was to evaluate the relationship between operator TIA and actual use of automation on a series of simulated aeromedical tasks using USAARL Multi-Attribute Task Battery (MATB-II). In addition to the two primary variables of interest (TIA and automation use), the study included outcome measures of subjects' task performance scores and cognitive workload ratings to gain a sense of whether patterns in automation use reflected operator gains in performance or cognitive load. Analyses were designed to also examine the relationships among the type of task being automated, levels of operator control (user-initiated automation versus system-prompted automation), operators' cognitive workload, and operators' propensity to trust technology (or trait-level TIA) to determine how varying levels of TIA and different types of automation engagement strategies impact automation use patterns, including downstream task performance and workload.

Methods

This study employed a within-subjects design in which participants completed four trials of a simulated aviation multitasking platform. Each trial required participants to manage four concurrent tasks. Automation could be voluntarily initiated by the participant for any of the tasks at any time (referred to as delegation), or it could be engaged in response to a prompt that came on screen when task performance dipped below a certain threshold (referred to as compliance). Automation reliability (70-percent vs. 90-percent accurate) and task difficulty load order were systematically varied and randomized across trials. Each trial lasted approximately 5 minutes.

Subjects

The study protocol was reviewed and approved by the USAARL Scientific Review Committee (protocol number 2023-019) and USAMRDC Institutional Review Board (IRB) (IRB number M-11078). Study participants were 17 adult U.S. Army-rated aviators who possessed a valid DD-2292 ("up-slip" indicating clearance for flight duties). Participants were eligible if they were at least 18 years of age, in good health per Army aviation standards, and able to follow verbal and written instructions in English. Exclusion criteria included current use of sedating medications, alcohol consumption within the prior 24 hours, nicotine use within 2 hours, or caffeine consumption within 16 hours of data collection, in accordance with Army Regulation (AR) 40-8 guidance and prior recommendations for reducing variability in physiological measures (Department of the Army, 2022). Compliance with inclusion and exclusion criteria was assessed via self-report prior to participation. The final sample had a mean age of 39.4 years (SD = 5.1) and were predominantly male (n = 16). Mean total flight hours for the sample was 2187.6 (SD = 1221.0).

Procedures

Participants scheduled individual study sessions at the USAARL and reported to the MATB-II computer laboratory. Upon arrival, a study technician provided an overview of the study, explained the activities to be performed, and reviewed the informed consent form. After providing written consent, participants completed a brief set of computer-administered questionnaires, including a demographics and personal history survey and an adapted Propensity to Trust Questionnaire.

Participants then received training on the subtasks of the USAARL MATB-II platform, following the procedures outlined in the MATB-II manual (Vogl et al., 2023). Training covered task rules, self-reported workload procedures, and the use of automation prompts and controls. Participants were given practice opportunities until they demonstrated baseline competency on each subtask.

Following training, participants completed four experimental trials, each lasting approximately five minutes. During each trial, participants managed four concurrent subtasks while automation support was available for all tasks. Automation reliability was manipulated between trials, with two blocks presented at 70-percent reliability and two at 90-percent reliability, randomized across participants. After each trial, participants completed the Checklist for Trust between People and Automation. Participants were given a two-minute seated rest break between trials.

Upon completion of the fourth trial, participants were debriefed and released. Total participation time ranged from 90 to 120 minutes. Note that physiological measures such as electrocardiogram and eye tracking were collected during the study but are not included in the present analyses.

Data Analysis

Analyses were conducted to examine the effects of automation use, automation reliability, task load, and individual differences on performance, workload, and trust in automation. Data were preprocessed to exclude incomplete trials and to align behavioral and self-report measures by trial. All statistical analyses were performed in R (version 4.5.0). Internal consistency of multi-item questionnaires was evaluated using Cronbach's alpha (see Materials section for specific questionnaires used in the analyses). Linear mixed-effects models (LMMs) were used to examine relationships between automation use and overall task performance. Subjective workload ratings from MATB-II were analyzed using LMMs with percentage of overall task time with automation engaged (reliance) as a predictor. Random intercepts were specified for participants to account for repeated measures. Demographic and experiential variables (e.g., age, flight hours, video game experience) were entered as predictors in models of task performance.

Video game experience was treated categorically (low, moderate, high) to test for nonlinear associations. Trial-level predictors included automation reliability (70-percent vs. 90-percent) and task load order (3-9-6-9-3 vs. 9-3-9-6-9). These were modeled as fixed effects in analyses of reliance and other behavioral measures. Relationships between trait-level trust, state-level trust, and behavioral indicators of trust were examined. Automation engagement was separated into two categories for analysis: compliance and delegation. Compliance was defined as engaging automation following a system prompt, while delegation was defined as engaging automation without a prompt. Mixed-effects logistic regression models were specified for binary outcomes, while LMMs were used for continuous outcomes. Models included task type, workload, and trust measures as predictors. In cases where outcome variability was low, penalized logistic regression (Firth method) was used to verify model estimates. Reliance models were further stratified by type of automation engagement (delegation versus compliance) to assess their differential contributions to sustained automation use.

Materials

Data were collected on demographic characteristics, trust in automation, behavioral indicators of trust, task performance, cognitive workload, and propensity to trust automation. Instruments and systems used in the study are described below.

Demographics and Personal History

Participants completed a brief demographics and personal history questionnaire developed by the investigators. Items included age, sex, rank, race/ethnicity, education history, flight hours, prior video game experience, use of automated technology, and confirmation of eligibility criteria. These variables were collected to describe the sample and to serve as potential covariates or moderators in analyses.

Trust in Automation Data

Propensity to trust in automation (trait-level TIA).

The Adapted Propensity to Trust in Technology Questionnaire (APTQ; Jessup et al., 2019) was administered prior to the experimental trials. This 6-item instrument assesses individual differences in propensity to trust automation, with items rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The APTQ has demonstrated superior reliability and predictive validity relative to broader technology trust scales (Jessup et al., 2019).

Post-trial trust in automation checklist (state-level TIA).

Trust in automation during task performance was measured with the 12-item Checklist for Trust between People and Automation (Jian et al., 2000). This measure, widely cited in the literature (e.g., Kohn et al., 2021; Ranes et al., 2023), includes items reflecting both trust and distrust. Responses are provided on a 7-point scale, and items may be combined into a single score or scored separately for trust and distrust. The checklist was administered following each of the four experimental trials.

Cognitive Workload

Cognitive workload was assessed both subjectively and behaviorally. Within MATB-II, operators were periodically prompted to provide subjective workload ratings using a graphical user interface (GUI)-based sliding scale. These ratings, along with response latencies, were recorded automatically by the system. Physiological data were collected for exploratory purposes related to workload, but were not included in the present analyses. Pupillometry and gaze behavior were recorded using the Pupil Labs Core eye-tracking system (200 hertz [Hz] sampling rate), and cardiac activity was measured with the Shimmer3 electrocardiogram (ECG) system (512 Hz sampling rate). Data streams were synchronized through Lab Streaming Layer. These measures were intended to serve as indices of perceived risk and cognitive workload in future analyses.

Simulated Task Environment

Computerized aviation task battery.

Experimental trials were conducted on the USAARL MATB-II, a multitasking simulation platform originally derived from the NASA MATB and customized for aviation research (Vogl et al., 2023). The MATB-II engages participants in four concurrent subtasks typical of the aviation domain: system monitoring, communications, target tracking, and resource management. Participants interacted with the simulation using a joystick (nondominant hand) and computer mouse (dominant hand). Task demand was manipulated via pre-generated parameter files. Performance (task score) and workload measures were automatically captured and time-stamped within the MATB-II system.

Automation system and behavioral TIA.

The Virtual Offloading Guidance Logic (VOGL) panel embedded in MATB-II provided automation support for each subtask. Participants could engage or disengage automation manually via mouse controls, or automation could be initiated by system prompts or forced handovers coded in the parameter files. Automation reliability was defined by target accuracy thresholds set at either 70-percent or 90-percent and implemented separately across subtasks. Reliability manipulations were operationalized by adjusting response latencies and performance bounds to approximate the target reliability level.

Behavioral trust in automation was captured through automated logs of operator interactions with VOGL. Measures included compliance (automation engaged following a prompt), delegation (automation engaged without a prompt), and reliance (total time automation engaged).

This space is intentionally blank.

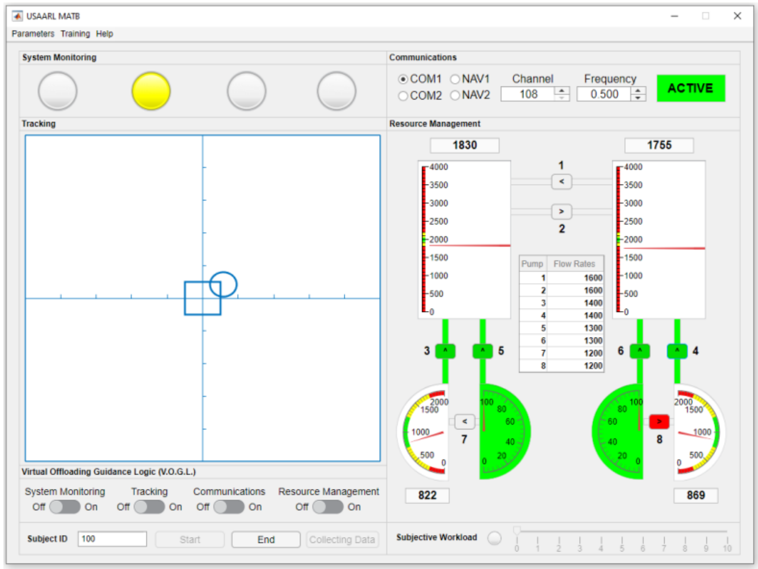

Figure 1. Screenshot from the MATB-II with VOGL automation dashboard (bottom left).

Results

Analyses examined the effects of automation use, automation reliability, task load, and individual differences on performance, workload, and trust in automation. Results are presented in five sections: (1) task performance and workload, (2) effects of individual differences, (3) automation reliability and task load conditions, (4) relationships among trust measures, and (5) behavioral trust outcomes (reliance, delegation, compliance).

Task Performance and Workload

Reliance on automation significantly improved overall task performance. A 10-percent increase in automation use was associated with an average increase of 1.06 points in task score ($p < .001$), indicating a strong positive effect of automation on multitasking accuracy. By contrast, automation use did not significantly reduce cognitive workload ratings. The estimated effect of reliance on subjective workload was near zero ($\beta = -0.059$), and the relationship was nonsignificant ($p = 0.957$), suggesting that participants did not perceive the use of automation during the task as reducing overall cognitive demands.

This space is intentionally blank.

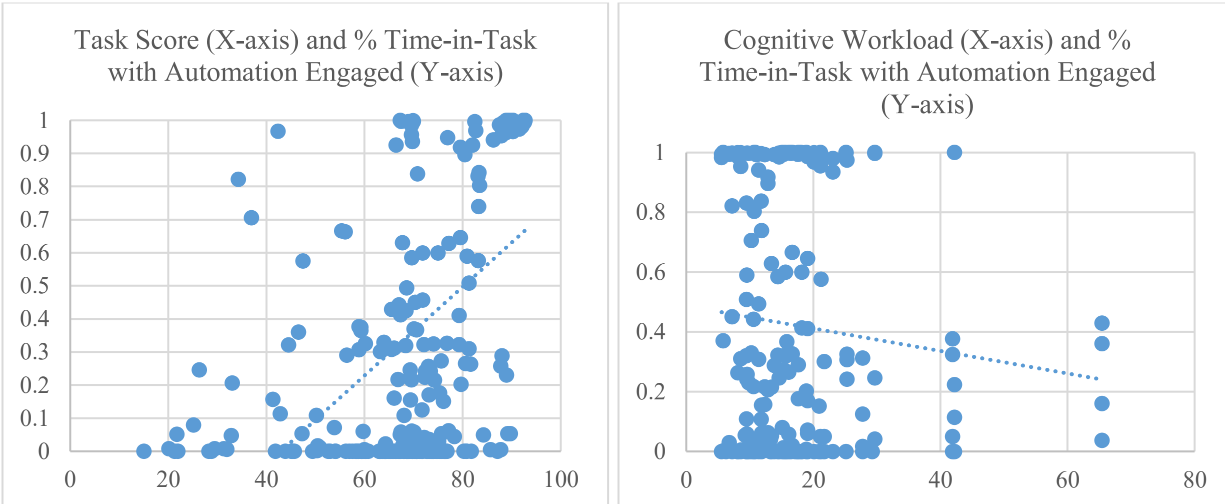

Figure 2. Performance (task score) and cognitive workload effects based on reliance (percent of task time with automation engaged).

Effects of Individual Differences

Neither age nor total flight hours significantly predicted task score. However, video game experience was a significant predictor in certain cases. Participants reporting moderate video game experience (5-15 hours/week) achieved task scores approximately 5.83 points higher than participants with little or no gaming experience (less than 5 hours per week; $p < .05$). This effect did not hold true for the heavy gaming group (more than 15 hours per week), who did not demonstrate a meaningful task score difference from the other groups. However, uneven group membership is likely contributing to this effect (see table 2), and findings should be interpreted with caution.

Table 2. Counts of Self-Reported Video Game Experience Categories

Video Game Experience	Count Of Subjects
Less than 5 hours per week	13
5-15 hours per week	2
More than 15 hours per week	2

Automation Reliability and Task Load Conditions

Automation reliability (70-percent vs. 90-percent) and task load condition (order of difficulty: 3-9-6-9-3 vs. 9-3-9-6-9) did not significantly predict automation use. Across participants, automation was engaged at similar rates regardless of programmed reliability levels or the trial sequence of task difficulty (all $p > .10$).

Relationships Among Trust Measures

A linear mixed-effects model was fitted to examine the relationship between trait trust (APTQ) and state trust (TIA Checklist total score), with random intercepts for subjects and tasks.

Results indicated a significant positive effect of trait trust on state trust (β = 1.380, *SE* = 0.637, $t(15) = 2.165$, $p = .047$), suggesting that higher APTQ scores were associated with greater state trust scores, even when considering situational factors like task scores and automation use (reliance). The model intercept was also significant (β = 34.740, *SE* = 14.024, $t(15) = 2.477$, $p = .026$), reflecting a moderate baseline level of state trust when trait trust equals zero. Random effects revealed substantial variability in trust scores across subjects ($\sigma^2 = 105.39$, *SD* = 10.27), suggesting that individual differences explained a considerable portion of the variance in trust scores. Self-reported automation use frequency did not significantly predict state trust scores nor behavioral use of automation, and differences between frequent, moderate, and infrequent users were not statistically significant. Finally, neither trait trust nor state trust significantly predicted behavioral Reliance (all $p > .48$), indicating no relationships between self-reported TIA beliefs and actual automation use.

Behavioral Trust Outcomes (Delegation, Compliance, Reliance)

Delegation.

The likelihood to delegate a task to automation (i.e., user-initiated automation use) was not significantly associated with trait trust ($p = 0.738$), automation use frequency ($p > 0.75$), or state trust ($p = 0.990$). However, task type did significantly predict delegation. Participants were more likely to delegate in resource management and tracking tasks compared to communications or system monitoring (both $p < .001$). These tasks involve continuous monitoring or dynamic resource allocation, which may encourage greater automation use patterns.

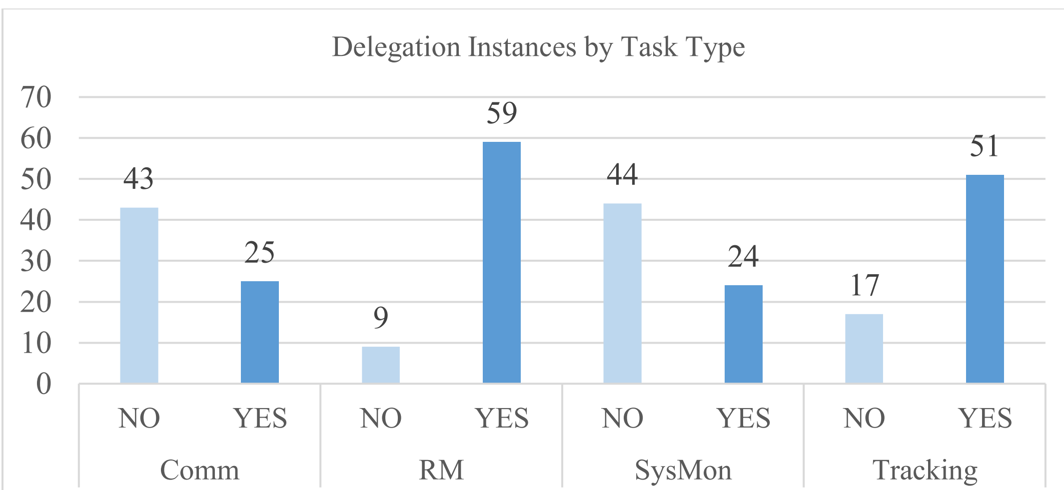

Figure 3. Count of delegation instances across different MATB-II task types.

Compliance.

Compliance with automation prompts was not significantly associated with trait trust ($p = 0.716$) or state trust ($p = 0.386$). Cognitive workload significantly predicted compliance: Higher workload was associated with greater likelihood of complying with prompts ($p < .001$). A penalized Firth logistic regression confirmed workload as a robust predictor ($p = 0.036$). Unlike delegation, compliance did not vary by task type ($p > .10$).

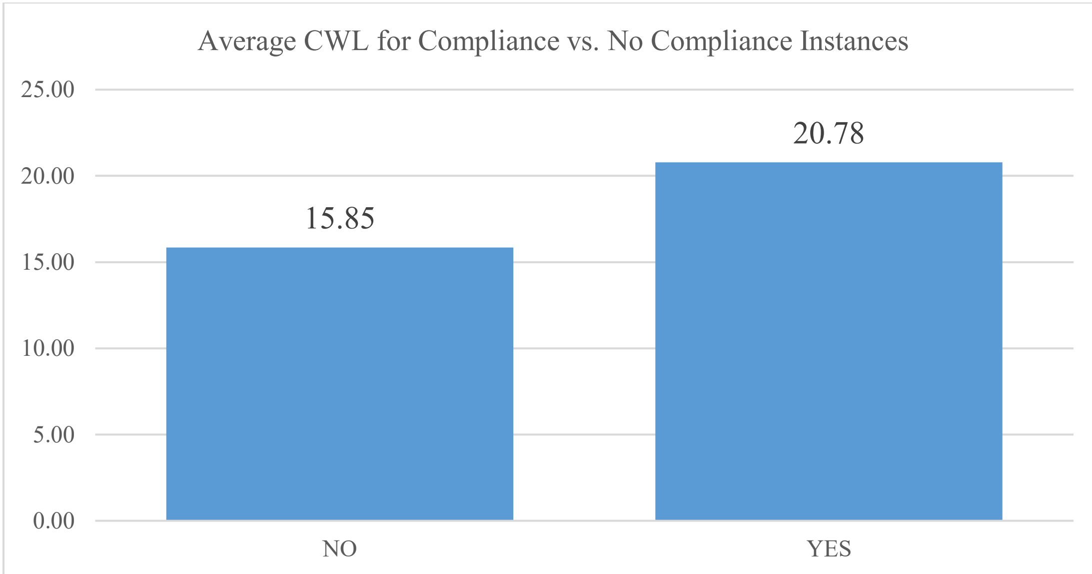

Figure 4. Average cognitive workload ratings between tasks where automation prompts were followed (compliance) versus tasks where prompts were not followed.

Reliance.

Reliance was more strongly influenced by delegation (β = 0.309, *SE* = 0.041, *t*(267) = 7.545, $p < .001$) than compliance (β = 0.193, *SE* = 0.052, *t*(261) = 3.744, $p < .001$). These results suggest that user-initiated automation engagement (delegation) produces greater sustained automation use than system-prompted engagement (compliance).

Summary

Automation use enhanced performance but did not alleviate perceived workload. Individual differences such as age and flight hours were not predictive of outcomes, though participants reporting moderate video game experience performed better than those with little or no gaming experience (an effect that should be interpreted cautiously given uneven group sizes). Neither automation reliability nor task load condition influenced automation use, indicating that engagement with automation was stable across these manipulations.

Trait trust in automation (APTQ) was positively associated with state trust ratings (TIA Checklist), but neither trait nor state trust measures (nor self-reported prior automation use) predicted behavioral reliance. Instead, behavioral trust outcomes were shaped by task and workload demands. Delegation (user-initiated automation use) occurred more frequently in dynamic tasks such as resource management and tracking, while compliance (prompted automation use) increased with higher cognitive workload regardless of task type. Moreover, delegation exerted a stronger influence on sustained reliance than compliance, underscoring the importance of user agency in shaping automation use patterns.

Discussion

The present study examined how TIA relates to operator performance, workload, and automation use behaviors in a simulated aviation multitasking environment. Analyses focused on five domains: task performance and workload, individual differences, automation reliability and task load conditions, relationships among trust measures, and behavioral trust outcomes (reliance, delegation, and compliance). Overall, findings highlight a disconnect between self-reported measures of trust and actual operator behavior, underscoring the importance of situational and task-based factors in shaping automation use.

Reliance on automation significantly improved task performance, reinforcing the established finding that well-designed automation can enhance objective performance outcomes in complex environments (e.g., Parasuraman et al., 2008; Sato et al., 2020). However, automation use did not reduce subjective workload, suggesting that aviators continued to monitor automation even when it performed reliably. This discrepancy between objective task relief and subjective experience aligns with prior research on vigilance and "out-of-the-loop" phenomena, which indicate that operators may not perceive automation as reducing cognitive demands because oversight responsibilities remain (Krausman et al., 2022; Lu & Sarter, 2019). These findings suggest that while automation may free attentional resources, it does not necessarily make the task feel easier to the operator, which is a critical distinction for aviation contexts where workload management is essential.

Neither age nor flight hours predicted performance outcomes, indicating that aviation experience alone does not shape automation use patterns in this context. However, moderate video game experience (5-15 hours per week) was associated with significantly better multitasking performance relative to participants with little or no gaming experience. Although weakened by uneven group sizes within the sample, this result tentatively supports prior evidence that action video game play may enhance attentional control, visual-motor coordination, and task-switching ability (Green & Bavelier, 2003; Dye et al., 2009; Alzahabi & Becker, 2013). Interestingly, this effect did not extend to heavy gamers, who did not differ significantly from other groups. Given the very small subgroup sizes, these findings should be interpreted with caution, though they suggest that moderate levels of gaming may provide transferable cognitive benefits without introducing potential downsides of heavy gaming (e.g., fatigue or desensitization).

Automation reliability (70-percent vs. 90-percent) and task load condition (easy-to-hard vs. hard-to-easy orderings) did not significantly predict reliance. This suggests that within the relatively high reliability ranges tested, aviators engaged automation at similar rates regardless of small differences in accuracy. Prior studies have found that larger manipulations of reliability (particularly those involving low or variable reliability) are more likely to affect operator reliance and trust (Dzindolet et al., 2003; Wickens & Dixon, 2007). The lack of sensitivity to moderate reliability differences in this study may indicate that aviators treated both 70-percent and 90-percent as "good enough" for offloading in multitasking environments.

The study revealed a consistent link between trait trust (APTQ) and state trust (TIA Checklist), echoing prior work showing that dispositional trust tendencies influence situational trust ratings (Hoff & Bashir, 2015; Lee & See, 2004). However, neither trait nor state trust predicted actual automation use behaviors (reliance, delegation, or compliance). This gap between self-reported beliefs and observed behaviors reflects a broader attitude-behavior inconsistency well-documented in social psychology (Ajzen et al., 2018; Armitage & Conner, 2005). For TIA research, this finding is critical: Relying solely on questionnaires risks misrepresenting operator trust levels, especially in dynamic, high-load environments where situational risk and task demands may override individual predispositions.

Behavioral measures of TIA revealed that contextual factors, not self-reported trust, drove automation use. Task type significantly predicted delegation, with operators more likely to self-engage automation in dynamic tasks such as resource management and tracking. These tasks require continuous monitoring and dynamic resource allocation, making them natural candidates for offloading. This pattern is consistent with ecological models of trust, which emphasize context-dependent trust behaviors shaped by task demands (Drnec et al., 2016). In contrast, compliance was driven by workload rather than task type. Aviators were more likely to accept automation prompts under high cognitive load conditions, reflecting a pragmatic strategy to manage limited attentional resources. This aligns with theories of decision-making under load, where individuals adopt simplified strategies or accept external guidance to cope with high task demands (Parasuraman & Riley, 1997).

When it came to sustained automation use over the course of a task, delegation had a stronger impact on sustained reliance than compliance. When operators chose to engage automation proactively, they were more likely to leave it engaged for longer durations compared to when automation was activated in response to system prompts. This suggests that operator agency plays a critical role in shaping automation use. Prior research in human-machine teaming has emphasized the importance of perceived control in fostering trust and sustained engagement (de Visser et al., 2019). Designing systems that preserve operator autonomy while providing well-calibrated prompts may therefore optimize both performance and reliance outcomes.

The study has several limitations. The sample size was small ($N = 17$), with uneven subgroup distributions for individual difference variables such as video game experience. The reliability manipulation covered a narrow range (70-percent vs. 90-percent), which may not generalize to scenarios involving highly unreliable or adaptive automation. Although physiological data (ECG, pupillometry) were collected, they were not analyzed in this report, limiting triangulation of trust and workload measures. Finally, the MATB-II is a validated multitasking simulation but does not fully replicate the operational complexity of real-world aviation environments.

Future research should examine broader ranges of automation reliability and task difficulty to better understand thresholds at which operators modulate reliance. Improved workload measures (including physiological indicators) may provide richer insight into how trust and workload interact. Larger and more diverse samples of aviators are needed to validate individual difference effects, including potential transfer benefits from video gaming. Finally, future studies should explore system designs that support operator agency while leveraging prompts to encourage timely automation use, with an emphasis on developing trust measures that capture behavior rather than belief.

Conclusion

This study demonstrates that while automation reliably improves performance, its influence on workload and trust is more complex. Self-reported trust measures were poor predictors of actual automation use, highlighting a belief-behavior gap with direct implications for the design of human-machine teams. Instead, task demands, workload, and operator agency were the strongest predictors of automation engagement and use patterns. These findings emphasize the need to move beyond questionnaire-based trust assessments toward behavioral and context-sensitive measures that more accurately capture how operators interact with automated systems in dynamic environments.

References

Ajzen, I., Fishbein, M., Lohmann, S., & Albarracín, D. (2018). The influence of attitudes on behavior. In *The Handbook of Attitudes, Vol. 1: Basic Principles (2nd ed.)*. Routledge.

Alzahabi, R., & Becker, M. W. (2013). The association between media multitasking, task-switching, and dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1485–1495.

Armitage, C. J. & Conner, M. (2005). The effects of attitudinal ambivalence on attitude-intention-behavior relations. In Haddock, G. & Maio, G. R. *Contemporary Perspectives on the Psychology of Attitudes*. Psychology Press, Inc.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

de Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M.A. (2019). Towards a theory of longitudinal trust calibration in human-robot teams. *International Journal of Social Robotics, 12,* 459–478.

Department of the Army. (2022). Temporary flying restriction due to exogenous factors affecting aircrew efficiency (Army Regulation 40-8).

Drnec, K., Marathe, A. R., Lukos, J. R., & Metcalfe, J. S. (2016). From trust in automation to decision neuroscience: Applying cognitive neuroscience methods to understand and improve interaction decisions involved in human automation interaction. *Frontiers in Human Neuroscience, 10*, 290.

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science*, *18*(6), 321–326.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L.G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697–718.

Green, C. & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423,* 534–537.

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

Jessup, S. A., Schneider, T. R., Alarcon, G. M., Ryan, T. J., & Capiloa, A. (2019). The measurement of the propensity to trust automation. *International Conference on Human-Computer Interaction, 11575*, 476–489.

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

Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y. C., & Shaw, T. H. (2021). Measurement of trust in automation: A narrative review and reference guide. *Frontiers in Psychology, 12*, e604977.

Krausman, A., Neubauer, C., Forster, D., Lakhmani, S., Baker, A. L., Fitzhugh, S. M., & Schaefer, K. E. (2022). Trust measurement in human-autonomy teams: Development of a conceptual toolkit. *ACM Transactions on Human-Robot Interaction (THRI), 11*(3), 1–58

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

Lu, Y., & Sarter, N. (2019). Eye tracking: A process-oriented method for inferring trust in automation as a function of priming and system reliability. *IEEE Transactions on Human-Machine Systems, 49*(6), 560–568.

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

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

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.

Park, J., Jeong, H., Park, J., & Lee, B. C. (2019). Relationships between cognitive workload and physiological response under reliable and unreliable automation. In *Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2018 International Conference on Human Factors and Systems Interaction*, July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9 (pp. 3-8). Springer International Publishing.

Ranes, B., Wilkins, J., Kenser, E., & Caid-Loos, M. (2023). *Measuring trust in automation (TIA) in operational aeromedical settings: A systematic review of the literature* (USAARL-TECH-TR--2024-10). U.S. Army Aeromedical Research Laboratory.

Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y., & Hancock, P. A. (2011, September). A model of human-robot trust: Theoretical model development. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 1432-1436). SAGE Publications.

Sato, T., Yamani, Y., Liechty, M., & Chancey, E. T. (2020). Automation trust increases under high-workload multitasking scenarios involving risk. *Cognition, Technology & Work, 22*(2), 399–407.

Vogl, J., Bommer, S., McCurry, C. D. (2023). *The United States Army Aeromedical Research Laboratory (USAARL) Multi-Attribute Task Battery (MATB-II)* [Unpublished manuscript].U.S. Army Aeromedical Research Laboratory.

Wei, J., Bolton, M. L., & Humphrey, L. (2020). The level of measurement of trust in automation. *Theoretical Issues in Ergonomics Science, 22*(3), 274–295.

Wickens, C. D. & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201–212.

U.S. ARMY AEROMEDICAL RESEARCH LABORATORY

FORT RUCKER, ALABAMA

Optimizing

HUMAN PROTECTION AND PERFORMANCE

since 1962

All of USAARL's science and technical informational documents are available for download from the Defense Technical Information Center.

https://discover.dtic.mil/results/?q=USAARL

U.S. ARMY

FUTURES COMMAND

MRDC