USAARL-TECH-TR--2026-02

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Designing Adaptive Automation for Aviation: Transparency and Handoff Effects on Cognitive Workload, Situational Awareness, and Trust

Jon Vogl, J. Andrew Atchley, Jared Basso, Sharon Bommer, & Charles D. McCurry

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 Subjects 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).

IRB Determination and Number

This study, USAARL 2025-16, was approved by the Medical Research and Development Command Institutional Review Board on 4 June 2025.

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

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
26-11-2025	Technical Report	10/01/2023 - 9/30/2025

4. TITLE AND SUBTITLE

Designing Adaptive Automation for Aviation: Transparency and Handoff Effects on Cognitive Workload, Situational Awareness, and Trust

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
6.2

6. AUTHOR(S)

Vogl, J.[1], Atchley, A.[1,2], Basso, J.[1], Bommer, S.[1,3], McCurry, C.[1,3]

5d. PROJECT NUMBER
MO240066

5e. TASK NUMBER

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--2026-02

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]US Army Aeromedical Research Laboratory, [2]Chickasaw Nation Industries, [3]Oak Ridge Institute for Science and Education

14. ABSTRACT

Military aviation increasingly depends on automation to manage complex missions, yet static or pilot-initiated systems risk misuse, disuse, or over-reliance. Adaptive automation offers a potential solution by dynamically adjusting autonomy in response to operator state. This study examined two foundational design features, system transparency (transparent vs. opaque displays of automation state and rationale) and handoff method (voluntary vs. forced activation), to examine their effects on pilot cognitive workload (CWL), situational awareness (SA), and trust in automation.

15. SUBJECT TERMS

adaptive automation, operator state, cognitive workload, situational awareness

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLAS	UNCLAS	UNCLAS	SAR	49	Loraine St. Onge, PhD
					19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

This page is intentionally blank.

Summary

Military aviation increasingly depends on automation to manage complex missions, yet static or pilot-initiated systems risk misuse, disuse, or over-reliance. Adaptive automation offers a potential solution by dynamically adjusting autonomy in response to operator state. This study examined two foundational design features, system transparency (transparent vs. opaque displays of automation state/rationale) and handoff method (voluntary vs. forced activation), to assess their effects on pilot cognitive workload (CWL), situational awareness (SA), and trust in automation.

Twenty-four rated Army aviators performed four sessions of the U.S. Army Aeromedical Research Laboratory (USAARL) Multi-Attribute Task Battery (MATB) augmented with an adaptive automation controller driven by workload modeling. A within-subjects 2 × 2 design crossed transparency and handoff manipulations. Multimodal outcomes were collected, including subjective workload ratings, behavioral performance, physiological measures (electrocardiogram, pupillometry, eye tracking), and validated SA (Situation Awareness Global Assessment Tool, Situational Awareness Rating Technique, visual entropy metrics) and trust (Trust of Automation Systems Test, reliance, dwell metrics) assessments.

Results revealed a differentiated pattern across constructs. Transparency was the dominant driver of workload, but in the opposite direction of the initial hypothesis; transparent displays increased CWL ($p = .002$, $\eta^2 = .35$), degrading multitasking efficiency and system monitoring. A significant transparency × handoff interaction also emerged ($p = .006$, $\eta^2 = .29$). For SA, both factors mattered; transparent conditions shaped scanning strategies but opaque displays produced higher comprehension and projection ($p = .005$, $\eta^2 = .29$), while voluntary handoffs strongly preserved comprehension across transitions ($p < .001$, $\eta^2 = .58$). Their interaction was significant ($p = .012$, $\eta^2 = .24$), with opaque-voluntary yielding the most favorable SA profile. Trust was governed primarily by handoff method ($p = .002$, $\eta^2 = .34$); voluntary handoffs improved reliance calibration and reduced over-monitoring while a task was automated whereas transparency alone did not increase trust ($p = .098$). Discriminant analyses corroborated these patterns (e.g., ~97% transparency classification for CWL; ~80–83% for SA transparency/handoff; ~74% for trust handoff).

Together, the results highlight a central design principle: Transparency influences cognitive workload, voluntariness governs trust calibration, and SA is optimized when display mode and handoff method are aligned with optimal configurations. At the condition level, opaque-voluntary produced the best overall operator state (lower workload, stronger SA, higher trust), whereas transparent-forced produced the poorest outcomes. These findings provide concrete guidance for adaptive automation in high-demand Army aviation: Default to opaque displays with voluntary, autonomy-by-consent handoffs, and deploy transparency selectively/phase-specifically or on demand to avoid workload penalties while preserving awareness and trust.

This page is intentionally blank.

Acknowledgements

This research was supported in part by an appointment to the Research Participation Program at 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)

Page

List of Figures

List of Tables

Introduction

Few domains have embraced automation as rapidly and pervasively as aviation. From the earliest decades of flight, system designers and operators recognized that the cognitive demands of flying could quickly exceed human limits. This recognition fueled a continuous push to embed automated processes wherever they could provide relief, precision, and consistency. Over time, automation permeated virtually every corner of aviation; manufacturing lines leveraged robotic assembly (Peck, 1959), recordkeeping systems shifted to digital automation to streamline logistics (Zimmerman et al., 1964), and even air traffic control adopted automated conflict-detection aids (Hink, 1974; Couluris et al., 1978). Each of these advances reflected a shared logic; by removing some portion of human burden, automation could reduce errors, extend capability, and enhance safety.

This orientation toward automation has only accelerated in the military sector. The 2022 milestone flight of a fully autonomous UH-60 Black Hawk helicopter was more than just a technological demonstration, it represented a glimpse of the near-future operator role (Lockheed Martin, 2022). Rather than being stick-and-rudder pilots, pilots are increasingly positioned as supervisory controllers of complex, semi-autonomous systems. Such a shift brings both opportunity and tension; operators may be freed from micromanagement, but they also risk being distanced from the system's underlying state and deprived of engagement. As Black Hawk development shows, supervisory control is not speculative; it is already here. The question is no longer *whether* automation should be integrated but *how* the human and machine should share control.

The rationale for such developments is longstanding. Warren (1956) observed nearly seventy years ago that the demands of piloting could routinely overwhelm capacity, particularly under high workload phases such as takeoff, landing, and combat maneuvering. Since then, those demands have only grown; modern military aviators must integrate weapons management, cooperative teaming with unmanned systems, communications, and navigation, all while maintaining continuous vigilance over a dynamic environment. The result is a workload environment in which human limits are routinely tested, if not exceeded.

Yet the paradox of automation is equally as old. As Thackray (1980) and de Waard (1996) noted, when automation relieves too much demand, operators can drift into underload, boredom, and lapses in vigilance. Excessive automation reliance may also degrade pilot skills through lack of engaged practice. The accident record confirms this. A widely cited review by Kayes and Yoon (2022) shows that cognitive offloading to automation has contributed to both systemic safety issues and specific aviation mishaps. Perhaps most striking is the counterfactual: the Apollo 11 landing in 1969, where astronaut Neil Armstrong's manual override of the automated descent system averted what could have been a catastrophic crash (David, 2019). Here automation provided invaluable support, but human intervention was still decisive. Such anecdotes underscore the paradox — automation can save lives, but only when designed in a way that balances relief with engagement.

This double-edged legacy frames the central problem, automation can extend human capacity, but it can also reconfigure the cognitive ecology of the cockpit in ways that introduce new vulnerabilities. Parasuraman and Riley (1997) famously characterized this problem as one of "use, misuse, and disuse." Sheridan and Parasuraman (2005) elaborated that supervisory control environments are particularly fraught, as operators must allocate attention between system monitoring and their own direct tasks, often without clear guidance about how authority should be shared. This tension drives the present focus, if automation is inevitable, then the challenge is not whether to automate but *how to design the transitions of control* so that optimal workload levels, awareness, and trust are preserved.

The Adaptive Automation Problem

The rise of adaptive automation has been one of the most promising responses to this challenge. Instead of setting automation levels statically, adaptive systems can adjust task allocation dynamically based on the operator's predicted cognitive state. The conceptual framework is straightforward:

1. **Monitoring.** Physiological signals (e.g., heart rate variability, pupil diameter, electroencephalogram frequency bands), behavioral cues (e.g., gaze scan patterns, affective speech analyses), and contextual cues (e.g., task demand, environmental phase of flight, operation within performance envelopes) are continuously observed.

2. **Estimation.** These signals are transformed into interpretable features known to change with cognitive states and fed into models that estimate operator states such as workload, fatigue, or vigilance.

3. **Adaptation.** When estimates exceed or approach thresholds (e.g., overload, underload), automation is engaged or disengaged to bring the operator back into a target zone that denotes an optimal arousal-performance balance.

This loop promises a way to keep operators "in the sweet spot," engaged enough to sustain vigilance but supported enough to avoid overload. In theory, such a system could prevent both catastrophic errors from task saturation and subtle degradations from underload.

Decades of work reveal that building such systems is far easier in principle than in practice. Real-time physiological monitoring is noisy, individualized, and context-dependent. Models that classify cognitive states often struggle to generalize across operators or environments. Even when reliable signals exist, translating them into automation actions introduces another layer of uncertainty: *When should it be automated? Which subtask should be automated? How should it be automated? How long should it be automated?* The literature is rife with cautionary tales of adaptive automation that relieved one burden while inadvertently creating another, either by disengaging operators from critical monitoring or by re-engaging them abruptly without context (Ruff et al., 2002).

For the current study, we developed a pseudo-adaptive automation system that allows us to bypass the current technically infeasible portions of the adaptive automation problem (e.g., translating from raw and noisy physiological signals to actionable output) and get to the heart of the when, which, how, and how long questions plaguing the field of automation. We have tackled this challenge by modifying a standard low-fidelity aviation software platform, the U.S. Army Aeromedical Research Laboratory (USAARL) Multi-Attribute Task Battery (MATB; see Vogl et al. [2024b] for further details on the software), to include a simulated adaptive automation system based on computational task analysis methods. To do this, we integrated the USAARL MATB with the Improved Performance Research Integration Tool (IMPRINT; see Bommer et al. [2025] for more details). IMPRINT, originally developed by the U.S. Army Research Laboratory and Micro Analysis & Design, models cognitive workload by quantifying interference across simultaneous task demands (Buck-Gengler et al., 2012). Within the USAARL MATB, all possible combinations of subtasks were modeled (192 in total) to produce workload scores that can be assigned in real time during simulation. This allows the system to display workload over the preceding 10 seconds and project workload 5 seconds into the future, based on known task states in the scenario file. In effect, the MATB uses IMPRINT as a stand-in for physiological monitoring, creating a simulated adaptive automation environment.

The resulting logic is both technical and oddly animate. By setting thresholds for overload and underload, the system can dynamically automate or revoke subtasks to bring workload back into a “comfort zone.” Crucially, these choices are not random; the algorithm selects the subtask predicted to yield the largest workload benefit in that context. This process gives the system a compelling quasi-intentional quality, as though the automation were actively “choosing” its actions. In practice, and with apologies to the author Mary Shelley, it evokes something akin to Frankenstein’s monster, stitched together from models and thresholds, not truly alive but moving and adapting in ways that suggest autonomy. While this remains a simulation rather than a true physiology-driven system, it allows researchers the ability to start answering questions related to automation handoffs that may otherwise not be presently approachable with the current state-of-the-art of physiologically driven adaptive automation pipelines. In this study, we employ the use of our MATB’s simulated adaptive automation system to consider design principles that may help shape the advancement of human-automation interactions, especially regarding control transitions.

Why Transitions Matter

Even if workload estimation were solved, one problem remains particularly intractable, control transitions. Early frameworks such as Parasuraman et al.’s (2000) taxonomy of levels of automation and Endsley’s (1995) three-level model of situational awareness warned that abrupt shifts of authority can undermine performance, particularly if operators are deprived of context. Empirical studies since then have reinforced the point. Molloy and Parasuraman (1992) showed that forced automation takeovers degraded awareness and slowed responses. Ruff et al. (2002) demonstrated that abrupt supervisory control transfers disrupted communication and coordination. Chen and Barnes (2014) found that in unmanned aerial vehicle operations, forced handoffs increased workload and eroded trust. The recurring message is simple; transitions are not neutral events. They reconfigure attention, awareness, and relational dynamics, often at the very moment when performance is most fragile.

Two factors dominate these discussions, **transparency** and **handoff method.** Transparency refers to how much the automation reveals about its logic, state, and intentions. Handoff method refers to whether control transfers are voluntary (initiated or confirmed by the operator) or forced (initiated by the system), or scaled somewhere in between. Together, these factors determine how operators experience transitions: Do they understand why the system is acting, and do they retain agency over whether the action occurs?

The literature leans heavily toward the view that *more* transparency and *more* voluntariness are better. Sheridan et al. (1978) argued that automation must not be a "black box." Billings (1996) framed transparency as essential to human-centered design. Endsley (1995) showed that transparency supports all three levels of SA, perception, comprehension, and projection. Parasuraman and Riley (1997) warned that forced handoffs risk both misuse and disuse by eroding trust. More recently, reviews such as Vogl et al. (2024a) have reaffirmed these themes, highlighting transparency and handoff as the two most consistent recommendations across three decades of adaptive automation research.

And yet, the picture may not be so one-sided. As supervisory control becomes the norm, several scholars have warned that less cluttered displays and more decisive automation actions might actually be beneficial, particularly under high workload. While transparency can improve comprehension, it also introduces another panel to scan, another channel to process, and additional mental workload, raising the risk that the cost of added visibility may outweigh its benefits (Van de Merwe et al., 2024; Gegoff et al., 2024). Similarly, although voluntary handoffs preserve operator agency, they also impose a decision burden, requiring the operator to evaluate prompts and take manual action under competing demands. In time-critical contexts, forced handoffs may therefore reduce cognitive load by removing that decision point and freeing the operator to focus on primary tasks (Akash et al., 2020; Villani et al., 2018).

Together, these findings suggest that the design of transitions cannot be reduced to a binary judgment of "more is better" for either transparency or voluntariness. Instead, the effectiveness of these features appears to be deeply context-dependent, shaped by the cognitive demands of the task, the timing of transitions, and the operator's mental model of system behavior. This recognition sets the stage for a closer examination of transparency and handoff method not as isolated design principles, but as dynamic levers that interact with workload, situational awareness, and trust in complex ways. The following sections consider each in turn, beginning with transparency, tracing its theoretical foundations, empirical evidence, and implications for adaptive automation.

Transparency in Adaptive Automation

Transparency has been described as the lifeblood of human-automation teaming. From Sheridan et al. (1978) early insistence that operators must "see inside" system logic, to Billings' (1996) framing of human-centered automation, the principle has remained remarkably stable, operators need visibility into automation's reasoning and intentions if they are to work effectively alongside it.

The operational justification is straightforward. Aviation tasks are inherently multichannel and time sensitive. If operators are left uncertain about why an automated system acted, or about what it will do next, they must divert scarce attentional resources to reconstruct system state. Transparency relieves this demand by making the automation's logic visible, narrowing the range of possibilities the operator must entertain, and allowing attention to be directed where it is most needed.

In terms of cognitive workload (CWL), transparency can be thought of as a regulator. By revealing what the system is monitoring and what thresholds it is applying, transparency reduces the uncertainty that drives inefficient scanning and redundant monitoring. Rather than continuously checking whether automation is functioning, the operator can offload some monitoring to the system itself, while retaining confidence in its criteria. Under Multiple Resource Theory (Wickens, 2002), this reduction in cognitive resource interference directly lightens workload. The Region Model (de Waard, 1996) reinforces this idea; by lowering the background cost of uncertainty, transparency helps keep workload within the optimal performance band rather than drifting into overload or underload (see Figure 1).

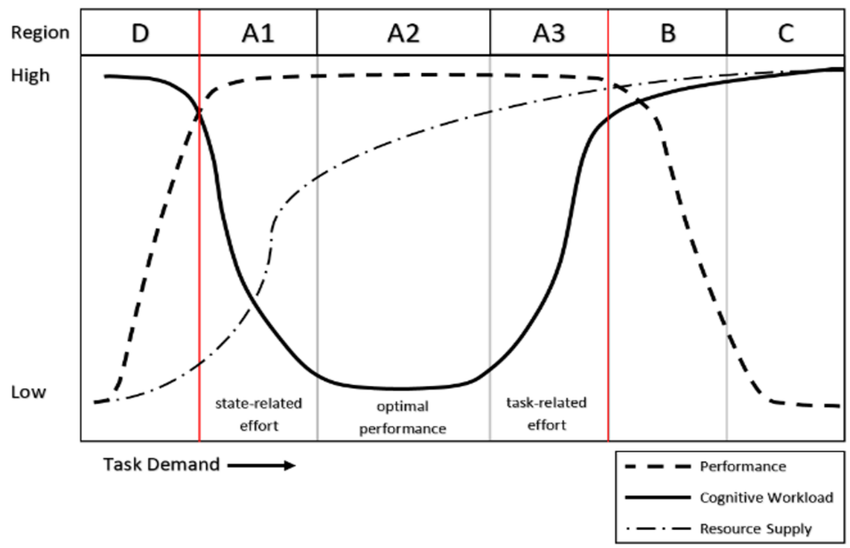


Figure 1. The Region Model (left; adapted from de Waard, 1996; Young et al., 2015).

For SA, transparency provides scaffolding. Endsley's (1995) three-level model emphasizes perception, comprehension, and projection. Transparency contributes at all three levels, it exposes perceptual cues about current state, clarifies the rationale for why those cues matter, and makes system intentions explicit to support projection. In supervisory environments, where operators may not be hands-on with controls, projection is particularly vulnerable. Transparency helps maintain forward-looking awareness even when the operator's primary role is monitoring rather than manual control.

The link between transparency and trust is more contested. Transparency has often been proposed as a calibration mechanism, allowing operators to align their reliance with system capability (Lyons, 2013). In principle, visibility into system logic should prevent overtrust (delegating when the system is weak) and undertrust (withholding reliance when it is strong). But the literature has shown mixed results. Too much transparency can overwhelm, confuse, or frustrate operators. Wiener (1989) warned that information overload can be just as damaging as opacity. Transparency is valuable not as an unbounded stream but as tailored, comprehensible, and actionable cues.

Handoff Methods in Adaptive Automation

If transparency determines what the operator *knows* about automation, handoff determines what the operator *controls*. Handoff refers to the manner in which authority is transferred between human and system. The contrast is sharp. Voluntary handoffs occur when the system recommends an action but leaves the decision to the operator. Forced handoffs occur when the system acts unilaterally.

The implications for cognitive workload are intuitive but not always straightforward. Voluntary handoffs preserve operator agency but impose an additional decision cost. Each prompt requires attention, evaluation, and response. In a busy multitasking environment, these micro-decisions can accumulate. Forced handoffs, by contrast, relieve this decision burden. The operator can remain focused on current tasks while the system reallocates control in the background. Yet this very relief can also create disruption. When control shifts without preparation, workload can spike as the operator scrambles to catch up. Whether voluntary handoffs genuinely reduce workload, or whether forced handoffs can be protective of workload by removing decision load, remains an empirical question.

For SA, the trade-offs are sharper. Endsley (1995) emphasized that awareness is most vulnerable at points of transition, when mental models must be updated. Voluntary handoffs allow operators to anticipate transitions and align their mental models in advance. Forced handoffs risk leaving operators disoriented, struggling to reacquire comprehension of system state. Ruff et al. (2002) showed that abrupt supervisory control transfers degraded situational awareness in communication-heavy tasks. In this sense, voluntariness preserves continuity. Yet here too a nuance exists, in some cases, forced handoffs may act as a safeguard, ensuring that automation engages when the operator is inattentive or overloaded. Whether the preservation of comprehension outweighs the potential safety benefits is precisely the kind of tension that must be examined systematically.

For trust, handoff is arguably the decisive factor. Trust is fundamentally relational; it reflects whether the operator perceives the automation as a dependable teammate (Lee & See, 2004). Forced handoffs can be interpreted as automation overreach, undermining perceptions of controllability and damaging trust. Voluntary handoffs, by contrast, affirm the operator's role as decision-maker, reinforcing perceptions of agency and aligning trust with reliability. Chen and Barnes (2014) found that voluntary handoffs enhanced trust calibration in unmanned aerial vehicle control, while forced handoffs eroded reliance even when system performance was high.

Still, even here there are counterpoints. Casner and Schooler (2014) argued that forced handoffs can be vital safety nets, particularly in cases of incapacitation or delayed operator response. From this perspective, voluntariness is not universally superior; its advantages must be weighed against the system's obligation to act when human performance falters.

Gaps and Rationale for the Present Study

Despite decades of research, guidance on transition design remains underdeveloped. Most studies have focused on *when* automation should act, for example, triggering based on workload thresholds or system failures. Far fewer studies have examined *how* automation should act with multiple facets of design principles, specifically, how much context it should provide (transparency) and whether control should be confirmed or imposed (handoff). Vogl et al. (2024a) emphasized this gap in their review, noting that while transparency and handoff are the most frequently recommended design principles, systematic tests of these factors within the same study and together are sparse.

The present study was designed to fill this gap by systematically manipulating automation transparency (transparent versus [vs.] opaque) and handoff method (voluntary vs. forced) within a low-fidelity simulation environment. Using the USAARL MATB augmented with IMPRINT workload modeling, we were able to precisely control task demands and automation behavior while engaging rated aviators in complex supervisory control tasks that approximate operational aviation contexts. Outcomes were assessed across three core constructs, CWL, SA, and trust in automation, using a multimodal measurement strategy that combined subjective scales, physiological indices, and behavioral performance. This comprehensive approach ensured that each condition could be evaluated not only in isolation but also in terms of how transparency and handoff jointly shaped the dynamics of human-automation teaming. Critically, this multidimensional framework reflects the interdependence of CWL, SA, and trust in automation; reducing workload does not automatically preserve awareness, and maintaining awareness does not necessarily yield calibrated trust. By integrating these constructs, the present study provides a more holistic evaluation of how transition design features influence operator state and system performance.

From this framework, six hypotheses were developed. The first three hypotheses concern the effect of transparency, in line with the longstanding claims that transparent displays reduce uncertainty, narrow attentional demands, and support trust calibration (Wickens, 2002; Lyons, 2013).

- Hypothesis 1: Automation displays that provide more context of the automated state (i.e., transparent) will yield lower levels of CWL relative to less verbose (i.e., opaque) automated systems.

- Hypothesis 2: Automation displays that provide more context of the automated state (i.e., transparent) will yield higher levels of SA relative to less verbose (i.e., opaque) automated systems.

- Hypothesis 3: Automation displays that provide more context of the automation state decisions (i.e., transparent) will yield higher levels of trust in the automated system relative to silent (i.e., opaque) automated systems.

The second set of hypotheses state that voluntary transitions preserve operator agency, enhance comprehension, and improve trust calibration by avoiding perceptions of automation dominance (Ruff et al., 2002; Chen & Barnes, 2014).

- Hypothesis 4: Automation systems that suggest manual activation of automation (i.e., voluntary handoff) will yield lower levels of CWL relative to automation systems that automatically take control of a subtask (i.e., forced handoff).

- Hypothesis 5: Automation systems that suggest manual activation of automation (i.e., voluntary handoff) will yield higher levels of SA relative to automation systems that automatically take control of a subtask (i.e., forced handoff).

- Hypothesis 6: Automation systems that suggest manual activation of automation (i.e., voluntary handoff) will yield higher levels of trust in automation relative to automation systems that automatically take control of a subtask (i.e., forced handoff).

In summary, the present research investigates how transparency and handoff method shape the three pillars of human-automation teaming assessment. By systematically testing these factors in a controlled aviation simulation with U.S. Army rated aviators, the study provides empirical evidence to inform the design of next-generation adaptive automation systems. The hypotheses are structured to test both the main effects of transparency and handoff and explore their combined consequences (as little work had been done to explore their interaction), producing findings that can be directly mapped to practical guidelines for system designers.

Methods

The study was reviewed and approved by the U.S. Army Medical Research and Development Command Institutional Review Board prior to execution. This study employed a within-subjects design across two factors each with two levels. A 2 (handoff method: voluntary vs. forced) by 2 (transparency level: transparent vs. opaque) study design was utilized to address the research objectives.

Participants

A total of 24 rated aviators (male = 23, female = 1; m_{age} = 39.46 [standard deviation [*SD*] = 7.11]) participated in this study. All participants self-reported being in good health, free from medications that could induce drowsiness, and abstinent from alcohol or sedatives for 24 hours, caffeine for 16 hours, and nicotine for 2 hours prior to data collection. Recruitment occurred in the Fort Rucker area through word-of-mouth, flyers, social media, and e-mail communications. Participants provided written informed consent prior to study enrollment. Upon completion of the study, participants who were in a “leave” status received $200 in gift cards as compensation.

Participants had an average career flight time of 2589.75 (*SD* = 1929.23) hours. Participants reported an average of 7.58 (*SD* = 0.84) hours of sleep the night before and the average Karolinska Sleepiness Scale score across all participants was 2.58 (*SD* = 1.10). All participants reported to be alert at the start of the study. All participants passed the required training thresholds and finished the study with complete data sets.

Materials

USAARL MATB.

The USAARL MATB was used as the experimental platform. The USAARL MATB is an aviation-like simulation environment consisting of four concurrently performed subtasks designed to mimic activities commonly performed in a cockpit. Figure 2 displays the graphical user interface (GUI) of the simulation. A full description of the software and paradigm is provided in Vogl et al. (2024b), with brief descriptions below.

System monitoring (SYS) subtask.

Participants completed a discrete visual vigilance task involving four lights, each mapped to a unique joystick button. When a light illuminated, the participant pressed the corresponding button to extinguish it before it timed out (5 seconds). Accuracy and reaction time were recorded, combined, and normalized to yield a SYS score for the subtask.

Communications (COM) subtask.

Participants monitored auditory channels for their callsign (e.g., “NASA 504”) and responded by adjusting radio, channel, and frequency settings to match the instructions they heard. Distractor messages directed at other callsigns were to be ignored. Accuracy and response time were recorded, combined, and normalized to yield a COM score for the subtask.

Tracking (TRK) subtask.

Participants performed a continuous compensatory tracking task, using a joystick to control a randomly moving circle and keep it aligned with a central target square. Deviations were recorded and normalized (against a deviation window of four units) to yield a TRK score for the subtask.

Resource management (RM) subtask.

Participants used a mouse to manage fuel levels across two primary tanks, maintaining them near a target fuel level of 2000 fuel units while they continuously drained throughout the simulation. Pumps with varying flow rates transferred fuel, while occasional pump shutoffs and failures (i.e., disabled the pump for 10 seconds) added dynamic decision demands. Fuel levels were continuously recorded, combined, and normalized (against a deviation window of 500 fuel units) to yield a RM score for the subtask.

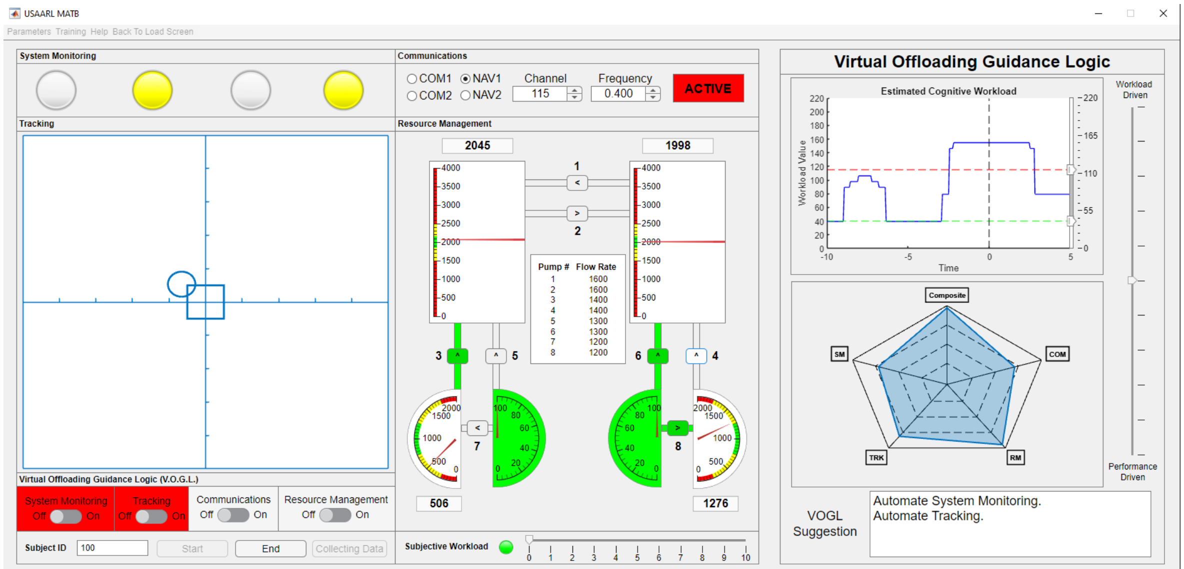


Figure 2. The USAARL MATB graphical user interface (GUI). The four subtasks, the subjective workload prompt, and the transparent Virtual Offloading Guidance Logic (VOGL) system are displayed.

Virtual Offloading Guidance Logic (VOGL) - USAARL MATB automation system.

The VOGL system enables experimental manipulations of automation reliability, transparency, and handoff method. VOGL was designed to provide automation assistance across all four of the primary USAARL MATB subtasks, allowing for dynamic control and varied operator-system interactions to address research questions in the field of human-automation teaming. The VOGL system was enhanced from the standard system presented in Vogl et al. (2024), to address the needs of the current study. Here, a brief overview of the relevant updates will be provided.

Reliability.

Automation systems come with varying levels of reliability in terms of how well they can perform the task they were designed to automate. An automation system can likely achieve minimum thresholds of performance under ideal conditions, but reliability may deteriorate under real-world conditions. Within the VOGL automation system, we defined reliability as the subtask score that is achieved by the system when it is automated. Formulas and algorithms were integrated into the VOGL system to derive these scores in real-time and to give the illusion of variable reliability levels. For example, while under automation, the tracking task would drift less erratically as if under the control of the VOGL system, with some random but system-deliberate drifts to maintain the appearance of active task performance while also controlling for the score that the automation system could achieve. The discrete tasks (SYS and COM) were more difficult to enable 'real-time' performance, so instead, negative feedback was incorporated to highlight incorrect responses made by the system (e.g., static would be played if the automation system 'incorrectly' put the wrong radio channel in). These updates made the automation feel more realistic, allowing the USAARL MATB program to bridge the gap from a simple desktop program to a more involved aviation simulation.

For the current study, reliability itself was not changed between conditions presented in the USAARL MATB program. Instead, static reliability was targeted at 80% for all subtasks across all conditions, with the exception of TRK which was targeted at 85%. While these reliability levels may appear low relative to our general understanding of what makes a good score (e.g., grade school percentages), they were specifically chosen to mimic the response profile similar to, if not slightly better than, a novice performer of the USAARL MATB task. This ensured that the automation system was perceived at known reliability levels for each participant consistently throughout the study and that it would not always be the best choice to automate everything all at once. Instead, these automation levels ensured a balance of human and automation performance that could be leveraged to achieve a high score in the program.

Transparency.

The VOGL system was designed to be able to change its transparency level, i.e., how much information it provides the operator regarding its decision-making process. Two transparency levels were tested. In the transparent condition, the operator would be provided an additional panel with information regarding the current experienced workload, the thresholds the system uses to determine when automation should be turned on or off, and the ability to alter how the automation system makes its decisions (to an extent). A depiction of the VOGL workload panel with the transparent automation window open is available in Figure 2. In the opaque conditions, the VOGL panel on the righthand side of the GUI is not visible, resulting in the only interface with the automation system being the switches in the lower left corner of the GUI.

Handoff method.

Two handoff methods were also utilized in the current study. In the voluntary handoff condition, the operator would be presented a suggestion to turn on automation for a specific subtask. This suggestion would occur when the tracked workload value (derived by IMPRINT scores) crossed the red-line threshold (e.g., see the Estimated Cognitive Workload plot in the upper right of Figure 2). The suggestion would be indicated by turning the corresponding automation switch to a red color (i.e., red to remove the task from their control responsibilities) in the lower left VOGL switch panel. In the case of a revocation suggestion (i.e., the workload value is below the green line indicating underload), the corresponding automation switch color would turn green (i.e., green to add the task to their control responsibility). These suggestions were then to be considered by the operator but the decision to automate or revoke was ultimately left to them. Conversely, in the forced condition, automation or revocation actions would be automatically applied, without any input from the operator. Additionally, the operator would have no control over automating any subtask or revoking control unless the system made the action itself. This would yield a dynamic passing back and forth of the subtask controls between the VOGL system and the operator without the operator having to make the decision.

Physiological devices.

Two physiological devices were used to estimate cognitive workload and situational awareness changes during task performance. Standard operating procedures were followed for the use of each device in accordance with the device user manuals. Each of the recorded physiological measures were synchronized utilizing the open-source Lab Streaming Layer (LSL) protocol.

Cardiac activity.

The Polar H10 electrocardiogram (ECG) system was used to monitor cardiac activity during the USAARL MATB simulations. The Polar H10 consists of a single strap that is placed around the mid-point of the participant's chest and a Bluetooth sensor module that samples ECG data at a rate of 130 Hertz (Hz). For the current study, the Polar H10 was remotely linked to the data collection computer. The ECG data were processed to derive metrics known to correlate with changes in cognitive workload. These metrics included heart rate (in beats per minute) and heart rate variability (i.e., low frequency - high frequency ratio). These metrics were derived as average values across each testing condition.

Eye-tracking.

The participants' eye movement activity was recorded using the Gazepoint 3 High Definition (GP3 HD) eye-tracking system. The GP3 HD system is a remote video-based eye-tracking system that was mounted securely under the monitor on which the USAARL MATB GUI was displayed. The GP3 HD eye tracker collected eye movement data to determine which subtask in the USAARL MATB the participant was looking at, as well as pupil diameter data. Eye-tracking data were collected at a sample rate of 150 Hz, with 0.5-1.0 degree of visual angle accuracy. The GP3 HD data were collected using both an LSL-synchronized approach and using the accompanying Gazepoint Analysis software for video playback.

Eye-tracking data were processed to generate metrics aligned with CWL, SA, and trust in automation. Cognitive workload was assessed using pupil diameter, averaged across the left and right eye. Because increases in pupil diameter are positively associated with greater cognitive effort under luminance-controlled conditions, values were expressed as a percent change from baseline to enhance comparability across participants. Situational awareness was evaluated through the relative visual entropy across the monitor panel and within the regions of interest (ROIs) of the USAARL MATB interface, providing an index of attentional distribution. Finally, trust in automation was inferred from dwell time on subtasks under automated control. These values were reverse-coded, such that greater monitoring of automated subtasks reflected lower trust, indicating that operators felt the need to visually verify system behavior rather than delegate confidently.

Subjective scales.

A standard demographics survey and the Karolinska Sleepiness Scale (KSS) were administered prior to task engagement. The KSS is a widely validated, single-item measure in which participants rate their current level of sleepiness (Kaida et al., 2006). Scores on the KSS reflect momentary daytime sleepiness, with higher values indicating greater subjective sleepiness. Several standardized questionnaires were also administered to assess perceived cognitive workload, situational awareness, and trust in the automation system used in the study.

CWL scales.

Instantaneous Self-Assessment (ISA) Scale.

During the USAARL MATB simulation, operators were prompted every 30 seconds to provide a subjective workload rating on a 1–10 scale. Each prompt was signaled both visually, via a light on the GUI, and auditorily, via a 1000 Hz tone. Operators entered their responses using the on-screen sliding scale, selecting the desired value with the mouse. The software automatically recorded both the chosen rating and the associated response time. If no response was provided within 10 seconds, the prompt timed out and the omission was logged. Prior to the experiment, operators were introduced to the ISA scale during training to ensure familiarity with the 1–10 anchors. This preparation minimized the need for external reference during task performance, thereby reducing intrusiveness of the measure.

NASA-Task Load Index (NASA-TLX).

The NASA-TLX is a widely used multidimensional questionnaire designed to assess perceived workload across six domains: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). Each dimension is rated on a 100-point scale, and the scores are averaged to yield a composite workload index in addition to the six individual subscale scores. In the present study, the NASA-TLX was administered following each block of the USAARL MATB simulation to capture participants' retrospective assessment of task demand. Importantly, the NASA-TLX is not suitable for administration during task performance, as the multi-item scale is time-intensive and would substantially disrupt ongoing performance if embedded within the task. Thus, post-task administration allowed for comprehensive workload assessment without interfering with operators' real-time task execution.

SA scales.

Situation Awareness Global Assessment Tool.

The Situation Awareness Global Assessment Technique (SAGAT) was used to provide an objective assessment of operator situational awareness during task performance (Endsley, 1988). Unlike many situational awareness measures that are administered post-task, where responses are vulnerable to recall errors and bias, the SAGAT was embedded directly within the USAARL MATB simulation. During each trial, the simulation was randomly frozen at pre-specified intervals, and operators were presented with queries targeting their current awareness of the task environment. These queries spanned all three levels of Endsley's model of situational awareness, perception, comprehension, and projection.

For example, perception-level queries asked operators to identify immediate system states such as, "What is your current system monitoring light configuration?" Comprehension-level questions probed understanding of the broader context, such as, "Which tasks were automated at any point over the previous round?" Projection-level items required operators to anticipate future events, such as, "How many seconds until the next subjective workload prompt?" Operators responded using a mouse to select or enter answers via the simulation interface.

Responses were automatically recorded and scored against the actual system state at the time of the query, yielding objective indices of situational awareness. Scoring was normalized between 0 and 1, either as a percent correct (for categorical queries) or percent deviation from the true value relative to task performance window thresholds (for continuous metrics). By pausing the simulation and querying awareness in real-time, the SAGAT provided a direct, unbiased measure of perception, comprehension, and projection, ensuring that all three levels of situational awareness were captured during task performance.

Situational Awareness Rating Technique.

The Situational Awareness Rating Technique (SART) is a self-assessment technique designed to measure an operator's level of SA (Taylor, 2017). To capture key features of situational awareness, the items were generated by presenting scenarios to aviators and having them identify features of situational awareness that were pertinent to two of the three scenarios. The three primary factors are: 1) demands on attentional resources, 2) the supply of attentional resources, and 3) the operator's understanding of the situation. These factors correspond to perception, comprehension, and projection, respectively. In this study, the 10-item SART using a 7-point Likert scale was administered upon completion of each testing condition.

Trust in Automation Scales.

Adapted Propensity to Trust in Technology Questionnaire.

The Adapted Propensity to Trust in Technology Questionnaire (APTQ) was used to measure each participant's baseline tendency to trust automation (Jessup et al., 2019). The APTQ is a six-item instrument developed specifically to assess propensity to trust automated systems, building on earlier, more general trust-in-technology measures. It has demonstrated strong reliability and predictive validity for both perceived trustworthiness of automation and behavioral reliance on it (Jessup et al., 2019). Participants rated their agreement with six statements about automation (e.g., "Generally, I trust automated systems") on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The APTQ was administered prior to any interaction with the USAARL MATB tasks or automation features to capture dispositional biases. These scores were then used as a covariate by regressing trust outcome measures onto APTQ scores, thereby residualizing dispositional trust and isolating state-based trust effects attributable to the experimental manipulations.

Trust of Automated Systems Test.

The Trust of Automated Systems Test (TOAST) was used as a 9-item scale to assess trust in automation (Wojton et al., 2020). Each item was rated on a 7-point Likert scale. The TOAST was originally developed to capture three theorized foundations of trust identified in earlier literature—purpose, performance, and process (Lee & Moray, 1994; Lee & See, 2004). However, confirmatory factor analysis supported a two-factor solution, resulting in subscales for system performance and system understanding. The scale has demonstrated strong concurrent validity with existing trust measures and has been validated across both military and civilian contexts (Wojton et al., 2020).

Procedure

Participants completed informed consent, eligibility screening, and baseline surveys (demographics, APTQ, Karolinska Sleepiness Scale). Then, the participant was instrumented with the Polar H10 device and the GP3 eye-tracking system, following standard operating procedures. Physiological data were recorded throughout the remainder of the task (i.e., during baseline, training, and testing condition blocks). A 5-minute resting baseline was administered, split between opaque (first 2.5 minutes) and transparent (last 2.5 minutes) MATB conditions to control for luminance differences.

Training included instructing the participant on the USAARL MATB subtasks, subjective workload prompts, and automation procedures. Participants had to achieve ≥ 60% on all subtasks before proceeding, with up to three training failures permitted before being removed from the study. Next, participants were trained on VOGL automation displays and handoff mechanisms, followed by four short practice runs (one per condition).

The experimental phase consisted of four counterbalanced 10-minute MATB blocks representing the 2 by 2 design (transparent vs. opaque by voluntary vs. forced, as seen in Table 1). During task performance, the USAARL MATB would freeze, color the screen white, and present the SAGAT questions. Once the SAGAT questions were answered, the USAARL MATB GUI would appear, offer a count down, and allow the operator to proceed with the simulation.

After each block, participants completed the NASA-TLX, TOAST, and SART. At study completion, participants were debriefed, compensated, and escorted from the USAARL facility.

Table 1. Experimental Conditions by Transparency and Handoff Method Factors

	Voluntary Handoff	**Forced Handoff**
Transparent Display	VOGL panel visible + automation suggestions	VOGL panel visible + forced automation activation
Opaque Display	VOGL panel not visible + automation suggestions	VOGL panel not visible + forced automation activation

Note. Conditions represent a 2 by 2 design crossing transparency (transparent vs. opaque) with handoff method (voluntary vs. forced).

Data Quality and Statistical Analysis

Prior to hypothesis testing, all physiological and performance data streams were inspected for quality and completeness. Raw ECG, pupillometry, and eye-tracking data were visually screened and algorithmically checked for artifacts (e.g., blink-related pupil spikes, ectopic heartbeats, or tracking dropouts). Segments with signal loss or artifact contamination were excluded using standard preprocessing routines, and remaining data were baseline-corrected relative to each participant's pre-task resting period. Task performance metrics derived from the USAARL MATB were verified for consistency against system logs to ensure accuracy of automation state tagging and handoff events.

To reduce dimensionality and limit Type I error inflation, metrics were grouped into three construct-specific sets aligned with the study hypotheses: CWL, situational awareness (SA), and trust in automation. Within each set, principal components analysis (PCA) was applied to identify latent dimensions and minimize redundancy prior to conducting omnibus tests.

Multivariate analysis of variance (MANOVA) was used as the primary inferential framework, with transparency, handoff method, and their interaction treated as within-subjects factors. Pillai's Trace was selected as the omnibus statistic due to its robustness against departures from multivariate normality and variance-covariance heterogeneity. Assumption checks included Mardia's test of multivariate normality, Mahalanobis distance for multivariate outliers, and correlation matrices to screen for multicollinearity.

For significant omnibus effects, linear discriminant analysis (LDA) was employed to derive discriminant functions and visualize separation among factor levels. Leave-one-subject-out cross-validation (LOSO-CV) and receiver operating characteristic (ROC) area under the curve (AUC) were used to evaluate classification accuracy and generalizability. Univariate repeated-measures ANOVAs, corrected with the Benjamini-Hochberg false discovery rate (FDR) procedure, were then conducted as follow-ups to clarify the contribution of individual measures.

Finally, trust-related outcome variables were residualized against baseline propensity to trust automation scores to isolate variance attributable to the experimental manipulations rather than trait-level predispositions. Together, these procedures ensured that subsequent results reflect reliable, well-calibrated estimates of how transparency and handoff manipulations influenced CWL, SA, and trust in automation.

Results

Measures

The collected metrics were separated into three sets relative to the hypotheses put forth. The three groups included clusters of metrics known to change as a function of cognitive workload, situational awareness, and trust in automation to ensure accurate model development for each construct. Table 2 details the descriptive data for each group of metrics across the four conditions. Table 9 compiles the definitions of the multivariate functions derived from these metric sets for quick reference.

The CWL metric set included a combination of subjective, physiological, and performance-based measures to capture workload from multiple perspectives. Subjective workload was assessed using the NASA-TLX composite score (averaged across six subscales) and the within task-prompted ISA scale. Physiological indices included pupil diameter and heart rate variability in the low-to-high frequency band ratio. Each physiological metric was baselined relative to the at-rest baseline recorded prior to the testing session. Task-based performance metrics included the MATB subtask scores (SYS, TRK, COM, RM), the IMPRINT model-derived workload estimate (to provide an overall score of the average workload experienced throughout the simulation), and a multitasking efficiency coefficient (all of which are standard output of the USAARL MATB platform). To maintain directional consistency across the set, metrics where higher raw values reflected improved performance were reverse scored, such that larger values always indicated greater workload. These included the multitasking efficiency coefficient and each MATB subtask score. This scoring approach ensured that all measures could be meaningfully integrated for the multivariate analyses, with higher scores uniformly representing higher cognitive workload.

The SA metric set was designed to capture perceptual, cognitive, and attentional aspects of awareness using both direct probes and behavioral indices. Subjective measures included SAGAT-based perception, comprehension, and projection scores, along with the SART dimensions of demand, supply, and understanding. In addition, entropy-based eye-tracking metrics were incorporated, including relative entropy both in general across the visual space and within ROI, which reflect the consistency and distribution of attentional scanning. Behavioral efficiency was further indexed by mean revisit rates to the tracking and resource management subtasks. Because these tasks are high-bandwidth and event-driven, greater revisit frequencies were interpreted as reflecting better SA, indicating that operators were actively monitoring dynamic channels where rapid changes occur. As with other SA measures, higher values consistently represented improved situational awareness; thus, no reverse scoring was applied within this set.

This space is intentionally blank.

Table 2. Descriptive Data from Each Condition Across Metric Sets: CWL, SA, and Trust

Metric	**Opaque x Forced**		**Opaque x Voluntary**		**Transparent x Forced**		**Transparent x Voluntary**	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
CWL Metric Set								
NASA-TLX	61.09	9.69	61.76	9.26	62.20	9.92	59.88	9.17
ISA Score	5.42	1.30	4.99	1.28	5.51	1.28	5.18	1.45
Pupil Diameter (% Δ)	10.03	8.79	8.49	7.77	5.42	8.02	5.87	9.63
ECG LF/HF (% Δ)	-23.85	43.77	-13.58	52.96	-13.10	50.55	-15.13	49.95
Multitasking Coefficient*	-0.62	0.12	-0.64	0.06	-0.76	0.11	-0.83	0.04
SYS Score*	-63.78	5.93	-66.56	8.88	-67.62	6.28	-67.3	7.77
TRK Score*	-79.02	4.31	-79.66	4.13	-79.30	4.08	-79.88	3.74
COM Score*	-47.74	18.12	-61.69	18.31	-40.79	26.92	-76.68	22.17
RM Score*	-77.48	11.84	-82.69	6.81	-77.67	11.54	-79.12	8.93
IMPRINT	22.90	1.17	24.66	8.32	22.48	1.64	26.02	8.48
SA Metric Set								
SAGAT: Perception	0.62	0.20	0.51	0.21	0.66	0.13	0.6	0.20
SAGAT: Comprehension	0.70	0.10	0.85	0.08	0.57	0.16	0.82	0.14
SAGAT: Projection	0.21	0.41	0.36	0.21	0.72	0.14	0.54	0.24
SART: Demand	5.18	0.79	5.11	0.81	5.35	0.78	5.19	0.83
SART: Supply	5.83	0.58	5.65	0.83	5.54	0.78	5.58	0.97
SART: Understanding	4.46	0.83	4.22	0.92	4.33	0.89	4.43	0.89
Relative Entropy	0.86	0.01	0.86	0.02	0.86	0.02	0.87	0.02
ROI Relative Entropy	0.64	0.10	0.65	0.11	0.61	0.10	0.67	0.10
TRK Revisit	4.08	1.67	3.19	1.45	3.05	0.78	2.82	1.05
RM Revisit	2.59	1.21	2.24	1.02	2.13	0.83	2.14	0.82
Trust Metric Set								
TOAST: Understanding	5.16	1.39	5.10	0.88	4.96	1.12	5.05	0.86
TOAST: Performance	3.87	1.65	4.14	1.55	4.03	1.52	4.24	1.43
Automation Reliance	0.06	0.06	0.15	0.28	0.14	0.10	0.09	0.20
SYS Auto Dwell %*	-5.33	6.62	-1.38	2.36	-7.16	4.00	-4.27	13.72
TRK Auto Dwell %*	-9.00	8.84	-10.25	8.90	-14.12	7.50	-9.85	9.88
COM Auto Dwell %*	-19.84	10.27	-12.78	17.74	-14.43	11.15	-11.01	13.11
RM Auto Dwell %*	-18.04	20.83	-12.42	12.31	-33.25	24.71	-13.59	10.88

Note. Metrics marked * were reverse scored so that higher values consistently reflect a higher level of the construct (higher CWL, higher SA, or higher trust).

The trust metric set captured both global trust attitudes and subsystem-specific reliance behaviors during automation. Subjective trust was measured using the TOAST scales for understanding and performance. Behavioral reliance was calculated as the percentage of time to which operators deferred to automation versus manual control. In addition, trust calibration was assessed via dwell time percentages for each subtask, which represented the proportion of dwell time operators spent monitoring a task while it was automated. Lower dwell times within the subtask were assumed to correlate with more trust in the system to do its job without being actively monitored. Because greater reliance and lower dwell time are indicative of higher trust, all dwell time metrics were reverse scored so that lower dwell values reflected higher trust. This ensured directional consistency across the set, with higher scores uniformly representing greater trust in automation.

Multivariate Analysis - Cognitive Workload

Assumption checks.

Prior to analysis, all dependent variables were examined for multivariate assumptions. Robust Mardia's test of multivariate normality indicated no significant skew (skew = 1781.17, $p = 1.00$), though significant kurtosis was detected (kurtosis = -21.17, $p < .001$), suggesting some departure from multivariate normality. However, no multivariate outliers were identified at the 97.5% Mahalanobis distance cutoff, and no dependent variable pairwise correlations exceeded |0.90|, indicating the absence of problematic multicollinearity. Together, these results supported the suitability of the dataset for multivariate analysis of variance (MANOVA), with Pillai's Trace selected because it is the most robust multivariate test statistic under violations of normality and heterogeneity of variance-covariance assumptions (Olson, 1974; Tabachnick & Fidell, 2019).

Omnibus MANOVA.

To reduce redundancy and limit error inflation, PCA was applied to the 10 selected workload measures. The PCA retained seven components, explaining 93.6% of the total variance. A repeated-measures MANOVA was then conducted with transparency, handoff, and their interaction as within-subjects factors. Results revealed a significant multivariate effect of transparency on the combined dependent variables, Pillai's Trace = .35, $F(1, 23) = 12.52$, $p = .002$, $\eta^2 = .35$. No significant main effect of handoff was found, Pillai's Trace = .07, $F(1, 23) = 1.70$, $p = .205$, $\eta^2 = .07$. However, the transparency × handoff interaction was significant, Pillai's Trace = .29, $F(1, 23) = 9.34$, $p = .006$, $\eta^2 = .29$. Figure 3 depicts the estimated marginal means of the interaction effect.

Table 3. Omnibus MANOVA Results for Cognitive Workload PCA Set (7 Principal Components)

Effect	Statistic	Value	F	df1	df2	p	η^2
Transparency	Pillai's Trace	0.35	12.52	1	23	.002	.35
Handoff	Pillai's Trace	0.07	1.70	1	23	.205	.07
Transparency x Handoff	Pillai's Trace	0.29	9.34	1	23	.006	.29

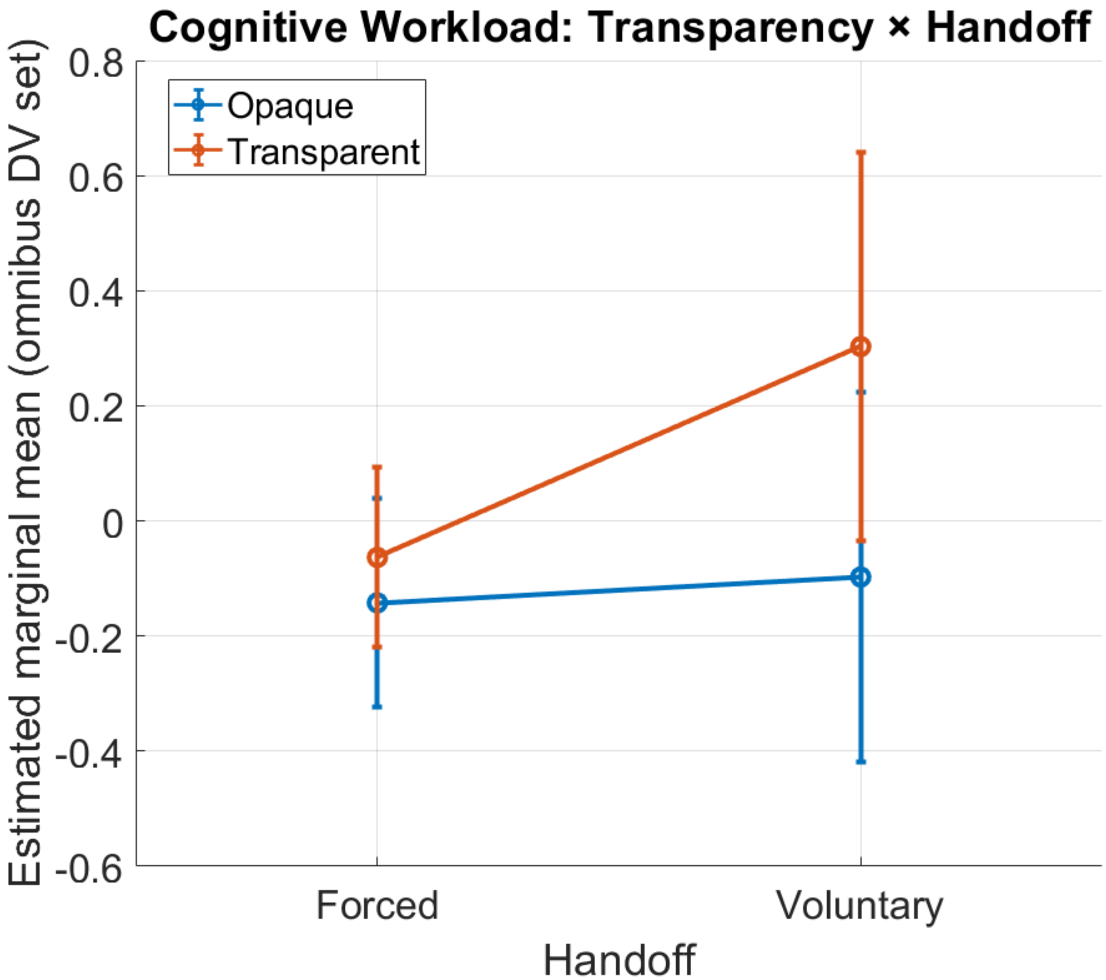


Figure 3. Transparency and handoff differences in multivariate cognitive workload results (higher values indicate higher workload).

Linear Discriminant Analysis (LDA).

LDA was performed to identify which metrics had the largest multivariate effect and determine the classification accuracy (as a percentage of correct classifications and AUC for ROCs). For the significant transparency and interaction effects, the leave-one-subject-out cross-validation demonstrated high classification accuracy for transparency (96.9%, AUC = 0.99) and good accuracy for the combined condition-level, interaction classification (69.8%, AUC = 0.92; relative to random chance of 25%).

Transparency functions.

One discriminant function was extracted for transparency. Back-projected PCA-LDA coefficients confirmed that multitasking efficiency (weight = 5.50), SYS score (4.06), and pupil diameter (1.06) loaded positively, whereas RM score (-2.65) and ISA scores (-1.58) loaded negatively. This pattern defines what we termed the efficiency-monitoring function, which captures operators' ability to manage concurrent tasks while sustaining vigilance on vision-based monitoring tasks, with lower reliance on subjective self-reports of strain. Higher scores on this function indicate stronger multitasking efficiency and system monitoring coupled with reduced physiological load. As depicted in Figure 4, classification results revealed that opaque automation conditions aligned with superior profiles on this function, as operators in the opaque condition demonstrated higher multitasking efficiency and vigilance with lower self-reported strain, whereas transparent conditions clustered in the opposite direction. Note that this discriminant axis is oriented such that higher function scores reflect a more efficient/low-strain profile, even though individual CWL measures were aligned as higher = higher workload for the MANOVA. The near-perfect separation of opaque and transparent conditions (AUC = 0.99) highlights the robustness of this function in distinguishing automation transparency effects on workload regulation.

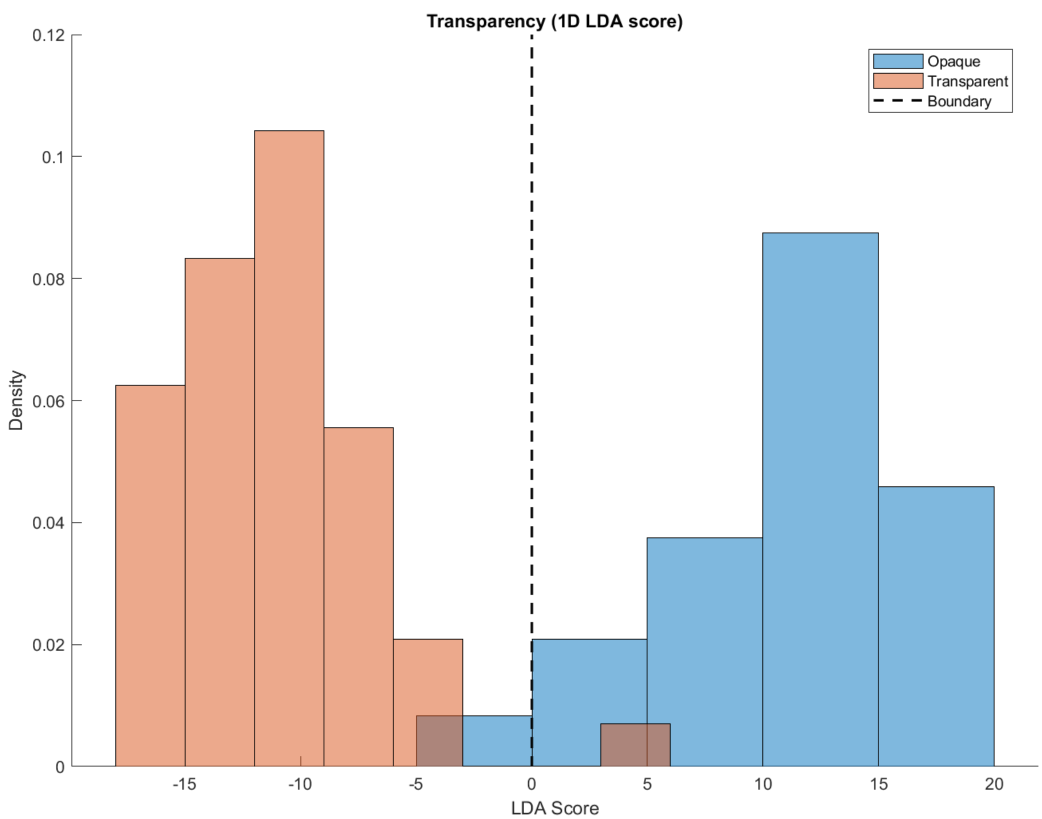


Figure 4. LDA function scores for transparency efficiency-monitoring function.

Condition-level functions.

For the four-level condition classification (transparency × handoff), three primary discriminant axes were extracted, but only the first two are examined and plotted in Figure 5 for ease of visualization. The first axis, efficiency-coordination, weighted positively on multitasking efficiency (-.157), SYS score (.111), and COM score (.049), but negatively on RM score (-.063). This axis distinguished operators who successfully integrated multitasking, vigilance, and communication performance from those who faltered under resource management demands.

Inspection of group centroids (see 'X' markers in Figure 5) indicated that the opaque-forced condition scored highest on this axis, reflecting strong integrated performance, whereas the transparent-voluntary condition scored lowest, showing breakdowns in coordination and resource management score balance due to the extra display and decision-making management requirements of the condition.

The second axis, resource-workload tradeoff, contrasted negative loadings on RM score (-.056) and COM score (-.104) with a positive loading on IMPRINT workload estimates (.075). High scores on this axis indicated stable resource management scores despite elevated modeled workload, while low scores reflected resource management task strain paired with deceptively low workload predictions. Centroid separation showed that voluntary conditions leaned higher on this axis, suggesting that voluntary takeovers preserved resource management stability even at the cost of higher modeled workload. Conversely, forced handoffs trended lower, reflecting fragile resource management task control despite lighter workload estimates.

These functions (AUC macro = 0.920; Figure 5) provided robust discrimination of the four condition profiles, with opaque-voluntary emerging as the most favorable overall for cognitive workload reduction (highest combined scores on LD1 and LD2), transparent-forced as the least favorable, and the other two conditions diverging depending on whether operator control (opaque-forced) or resource stability (transparent-voluntary) was prioritized.

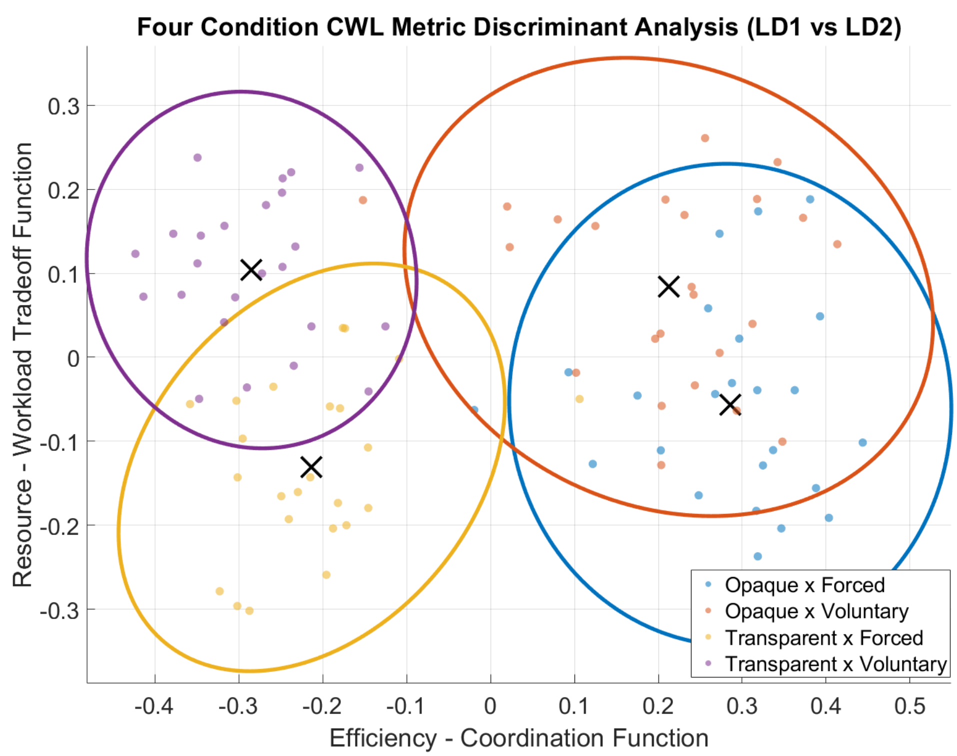


Figure 5. Four condition discriminant analysis plot for the CWL metric set. X markers indicate group centroids.

Univariate follow-ups.

Follow-up repeated-measures ANOVAs were conducted for each dependent measure to clarify the source of significant multivariate effects. For transparency, significant effects were observed for multitasking coefficient, $F(1, 23) = 229.34$, $p < .001$, $\eta_p^2 = .92$, and SYS score, $F(1, 23) = 17.97$, $p = .001$, $\eta p^2 = .46$. No other performance, physiological, or subjective measures reached significance after FDR correction. For handoff, the strongest effect was found for COM score, $F(1, 23) = 43.38$, $p < .001$, $\eta p^2 = .65$, indicating that communication task reliance differed reliably across voluntary versus forced handoffs. For the transparency × handoff interaction, no effects reached significance after FDR correction, indicating that transparency and handoff primarily exerted independent rather than interactive influences on workload at the univariate level. No other univariate effects were significant.

Table 4. Repeated-Measures ANOVA Results for CWL Metrics with Benjamini-Hochberg FDR Correction for Multiple Comparisons

	Transparency		**Handoff**		**Interaction**	
Measure	***p* (FDR)**	***ηp²***	***p* (FDR)**	***ηp²***	***p* (FDR)**	***ηp²***
Performance Metrics						
IMPRINT	.563	.02	.244	.10	.247	.09
COM Score	.563	.03	**< .001**	.65	.069	.24
RM Score	.417	.08	.108	.20	.247	.10
SYS Score	**.001**	.46	.479	.04	.069	.25
TRK Score	.563	.02	.479	.04	.938	0
Multitasking Coefficient	**< .001**	.92	.244	.18	.247	.12
Physiological Metrics						
ECG LF:HF (% Δ)	.102	.06	.479	.02	.852	.10
Pupil Diameter (% Δ)	.455	.19	.490	.03	.247	0
Subjective Metrics						
NASA-TLX	.765	0	.490	.02	.317	.07
ISA Score	.563	.03	.206	.13	.852	.01

This space is intentionally blank.

Multivariate Analysis – SA

Assumption checks.

Multivariate assumptions were first evaluated. Robust Mardia's test indicated significant departures from multivariate normality in terms of kurtosis (skew = 1781.31, $p = 1.00$; kurtosis = -21.17, $p < .001$). However, no multivariate outliers were detected at the 97.5% Mahalanobis distance cutoff, and no dependent variable correlations exceeded |.90|, suggesting the absence of multicollinearity. The dataset was therefore deemed appropriate for multivariate analysis using Pillai's Trace as a metric robust to the multivariate normality violation.

Omnibus MANOVA.

To reduce redundancy among the 10 SA measures, PCA was conducted. Seven components were retained, explaining 91.9% of the variance. A repeated-measures MANOVA indicated a significant multivariate effect of transparency, Pillai's Trace = .29, $F(1, 23) = 9.60$, $p = .005$, $\eta^2 = .29$, and a significant multivariate effect of handoff, Pillai's Trace = .58, $F(1, 23) = 31.75$, $p < .001$, $\eta^2 = .58$. The transparency × handoff interaction was also significant, Pillai's Trace = .24, $F(1, 23) = 7.36$, $p = .012$, $\eta^2 = .24$. These results are presented in Table 5 and depicted in Figure 6.

Table 5. Omnibus MANOVA Results for Situational Awareness PCA Set (*n* = 7 PCs)

Effect	Statistic	Value	F	df1	df2	*p*	η^2
Transparency	Pillai's Trace	0.29	9.60	1	23	.005	.29
Handoff	Pillai's Trace	0.58	31.75	1	23	< .001	.58
Transparency x Handoff	Pillai's Trace	0.24	7.36	1	23	.012	.24

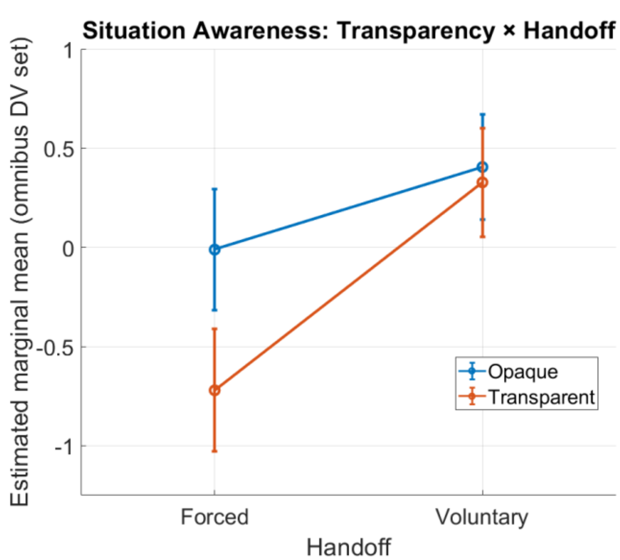


Figure 6. Transparency and handoff differences in multivariate SA results (higher values indicate higher SA).

LDA.

Cross-validated linear discriminant analysis with leave-one-subject-out train/test splits showed good classification accuracy for transparency (80.2%, AUC = 0.90), handoff (83.3%, AUC = 0.93), and the four-level condition interaction classification (74.0%, macro/micro AUC = 0.92; relative to chance at 25%).

Transparency functions.

One discriminant function was retained for transparency. Back-projected PCA-LDA coefficients indicated that relative entropy (-2.18), SAGAT projection (-2.03), SAGAT comprehension (+1.28), and TRK revisit mean (+0.99) were the primary contributors, with smaller effects from SART dimensions and relative entropy ROI. This pattern defines a projection-comprehension balance function; it distinguishes operators who sustain forward-looking projection and systematic scanning, paired with comprehension of task states, from those whose projection deteriorates despite attempts to maintain scanning breadth. The distribution of scores (as seen in Figure 7) confirmed that opaque automation displays consistently yielded higher discriminant scores than transparent automation displays, indicating that opacity enhanced projection and comprehension while supporting balanced attentional allocation. Classification performance was strong, with an AUC of .90, demonstrating reliable separation of transparent from opaque conditions.

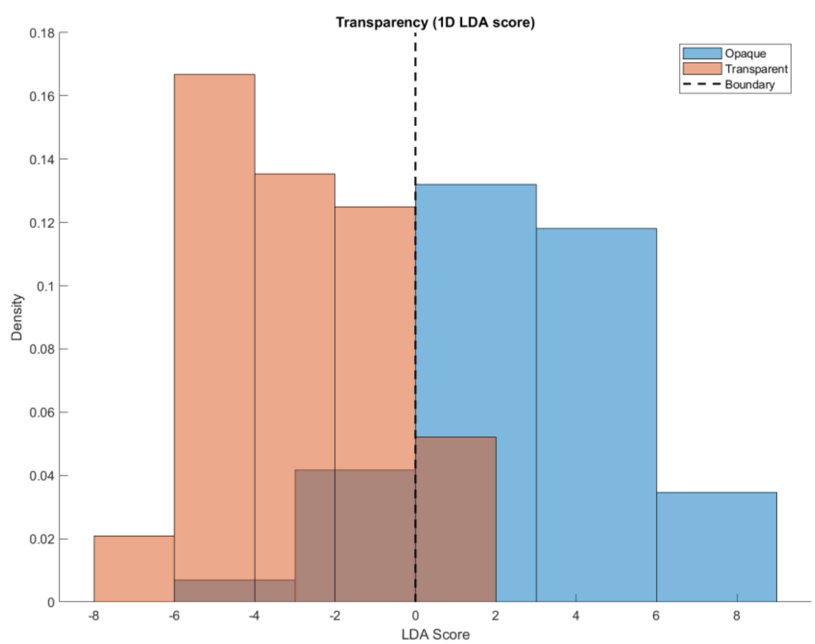


Figure 7. LDA function scores for transparency projection-comprehension balance function.

Handoff functions.

A single discriminant function was extracted for handoff classification in the SA domain. Back-projected coefficients revealed that SAGAT comprehension (-3.65), TRK revisit mean (+1.78), and SAGAT perception (+1.62) were the dominant loadings, with smaller contributions from SART demand, supply, and entropy-based scanning indices. This function, which we label comprehension-reacquisition, reflects the extent to which comprehension and perceptual continuity are preserved versus disrupted during transitions. Voluntary handoffs yielded more negative function scores, consistent with stronger comprehension and more frequent revisits to dynamic subtasks, while forced handoffs clustered positively, indicating that comprehension was degraded and had to be reacquired after the system took control. The ROC analysis confirmed robust discrimination between voluntary and forced conditions (AUC = 0.93), emphasizing that SA during transitions hinges on whether operators are permitted to anticipate and manage the handoff or are forced to rebuild understanding afterward.

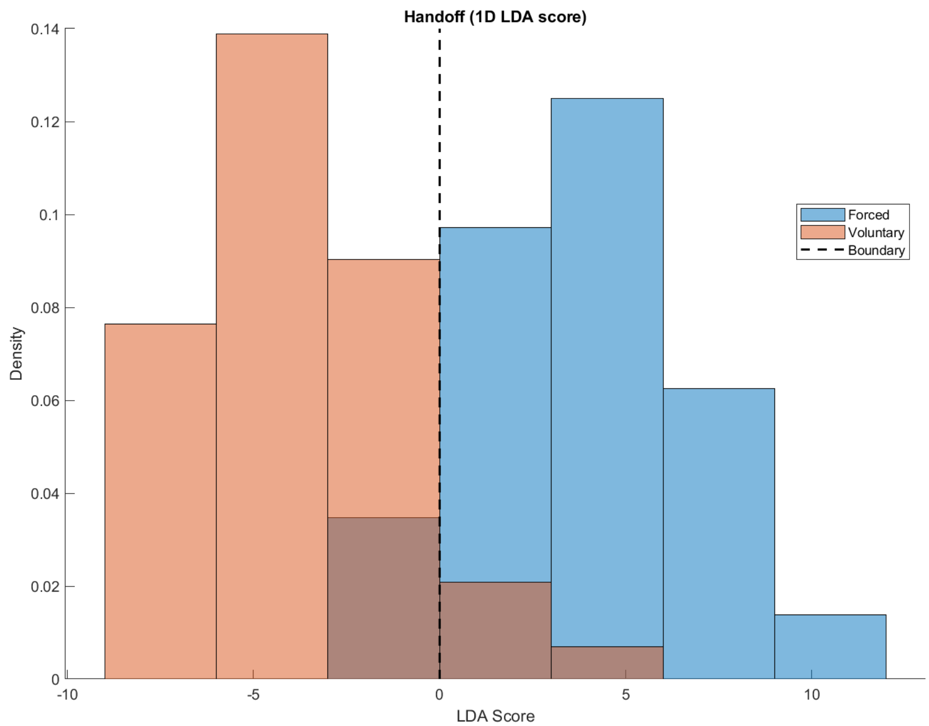


Figure 8. LDA function scores for handoff comprehension-reacquisition function.

Condition-level functions.

For the four-level condition classification (transparency × handoff), two primary discriminant functions emerged that accounted for the majority of discrimination among conditions (overall classification accuracy = 73.9%, macro AUC = .921). The first axis (*x*-axis; LD1) reflected a comprehension-perception tradeoff, distinguishing conditions where deeper comprehension of system state was preserved versus those that relied more on surface-level perceptual cues. Positive loadings for SAGAT comprehension (+0.148) contrasted with negative weights for SAGAT perception (-0.063) and TRK revisit mean (-0.046), indicating that higher LD1 scores corresponded to conditions supporting richer comprehension and more stable monitoring strategies. Lower scores, in contrast, suggested fragmented perceptual monitoring with reduced revisiting of dynamic subtasks.

The second axis (y-axis; LD2) captured a projection-attention balance, emphasizing forward-looking situational awareness and systematic attentional distribution. Strong positive weightings for SAGAT projection (+0.107) and relative entropy (+0.064), combined with a negative contribution from TRK revisit mean (-0.070), indicated that higher LD2 values reflected conditions fostering future-oriented awareness and balanced scanning. Lower values suggested more reactive strategies, with attention distributed unevenly across subtasks.

These two functions highlight that opaque displays and forced handoffs were associated with richer comprehension and projection-oriented awareness, whereas transparent or voluntary handoff conditions shifted operators toward more fragmented and reactive monitoring patterns. Note that the condition ordering here reflects multivariate discriminant space (LD1-LD2); it may differ from omnibus MANOVA or univariate ANOVA SA effects that emphasized voluntary handoffs for preserving comprehension.

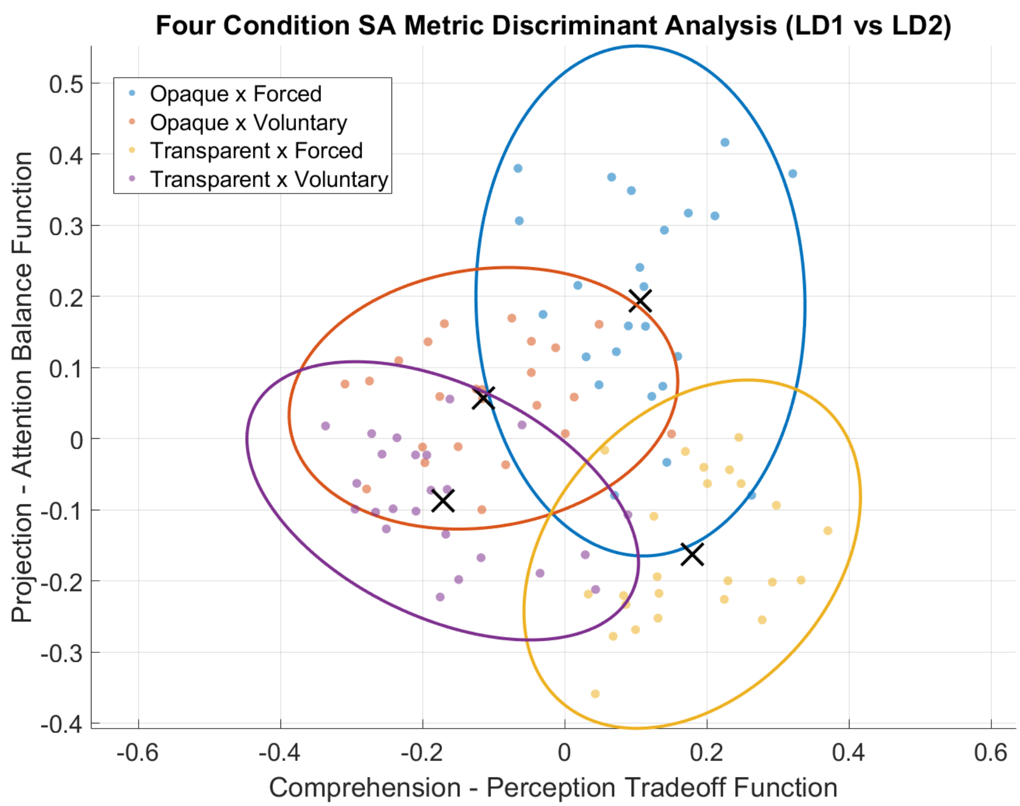


Figure 9. Discriminant space plot for SA metric discriminant analysis across conditions. X markers indicate group centroids.

Univariate follow-ups.

Follow-up repeated-measures ANOVAs, corrected for multiple comparisons using the Benjamini-Hochberg FDR procedure, clarified the sources of these multivariate effects (Table 6).

For transparency, significant effects emerged for relative entropy, $F(1, 23) = 29.91$, $p < .001$, $\eta p^2 = .55$; SAGAT comprehension, $F(1, 23) = 11.95$, $p = .006$, $\eta p^2 = .35$; SAGAT projection, $F(1, 23) = 34.98$, $p < .001$, $\eta p^2 = .62$; RM revisit mean, $F(1, 23) = 7.42$, $p = .025$, $\eta p^2 = .24$; and TRK revisit mean, $F(1, 23) = 10.48$, $p = .009$, $\eta p^2 = .32$.

For handoff, the strongest effect was observed for SAGAT comprehension, $F(1, 23) = 38.09$, $p < .001$, $\eta p^2 = .63$. Additional significant effects included relative entropy ROI, $F(1, 23) = 6.47$, $p = .020$, $\eta p^2 = .28$; TRK revisit mean, $F(1, 23) = 6.40$, $p = .020$, $\eta p^2 = .27$; and SAGAT perception, $F(1, 23) = 6.12$, $p = .020$, $\eta p^2 = .27$.

The transparency × handoff interaction reached significance for relative entropy, $F(1, 23) = 6.67$, $p = .042$, $\eta p^2 = .21$; relative entropy ROI, $F(1, 23) = 7.02$, $p = .042$, $\eta p^2 = .23$; SAGAT comprehension, $F(1, 23) = 5.22$, $p = .042$, $\eta p^2 = .21$; SAGAT projection, $F(1, 23) = 12.72$, $p = .012$, $\eta p^2 = .33$; and RM revisit mean, $F(1, 23) = 7.09$, $p = .042$, $\eta p^2 = .25$.

Overall, these results indicate that opaque displays most consistently enhanced projection, comprehension, and monitoring, whereas handoff effects were concentrated in comprehension and revisit behavior, with voluntary handoffs yielding superior performance. Importantly, several interaction effects highlight that the benefits of transparency depended on the handoff method, particularly for projection and resource management processes.

This space is intentionally blank.

Table 6. Repeated-Measures ANOVA Results for Situational Awareness Metrics with Benjamini-Hochberg FDR Correction for Multiple Comparisons

	Transparency		**Handoff**		**Interaction**	
Measure	***p* (FDR)**	***ηp²***	***p* (FDR)**	***ηp²***	***p* (FDR)**	***ηp²***
Entropy Metrics						
Relative Entropy	**< .001**	.55	.265	.08	**.042**	.21
Relative Entropy ROI	.706	.01	**.020**	.28	**.042**	.23
Revisit Metrics						
RM Revisit Mean	**.025**	.24	.079	.17	**.042**	.25
TRK Revisit Mean	**.009**	.32	**.020**	.27	.096	.15
SAGAT Metrics						
SAGAT: Comprehension	**.006**	.35	**< .001**	.63	**.042**	.21
SAGAT: Perception	.198	.10	**.020**	.27	.605	.02
SAGAT: Projection	**< .001**	.62	.760	0	**.012**	.33
SART Metrics						
SART: Demand	.386	.05	.405	.04	.773	0
SART: Supply	.268	.08	.617	.02	.372	.05
SART: Understanding	.742	.01	.662	.01	.222	.10

Multivariate Analysis – Trust in Automation

Assumption checks.

Multivariate assumptions were assessed prior to analysis. Robust Mardia's test indicated significant departures from multivariate normality in terms of kurtosis (skew = 1778.50, p = 1.00; kurtosis = -21.20, $p < .001$). However, no multivariate outliers were identified using a 97.5% Mahalanobis cutoff, and no dependent variable correlations exceeded |.90|, indicating acceptable levels of multicollinearity. The dataset was therefore considered suitable for MANOVA, with Pillai's Trace selected as the omnibus test statistic given its robustness to violations of multivariate normality.

Residualization of trust metrics.

Because individuals differ in their baseline propensity to trust automation, raw trust-related outcome measures were residualized prior to analysis. Specifically, each metric was regressed on scores from the Adapted Propensity to Trust in Technology Questionnaire (Jessup et al., 2019), which was administered before any task interaction to assess the participant's biases toward automated technology. By residualizing trust outcomes on this measure, subsequent MANOVA and discriminant analyses isolated the variance attributable to experimental manipulations of transparency and handoff, rather than preexisting individual differences in trust in automation propensity. This ensured that observed trust effects reflected task-driven influences rather than dispositional biases.

Omnibus MANOVA.

Residualized trust measures (n = 10) were entered into the MANOVA, including TOAST residuals, automation reliance, and eye gaze behavior metrics. Results showed no significant multivariate effect of transparency, Pillai's Trace = .11, $F(1, 23) = 2.98$, $p = .098$, $\eta^2 = .11$. A significant effect of handoff was observed, Pillai's Trace = .34, $F(1, 23) = 12.10$, $p = .002$, $\eta^2 = .34$. The transparency × handoff interaction did not reach significance, Pillai's Trace = .10, $F(1, 23) = 2.46$, $p = .131$, $\eta^2 = .10$. MANOVA results are presented in Table 7 and depicted in Figure 10.

Table 7. Omnibus MANOVA Results For Trust In Automation Metrics Regressed Onto APTQ Scores

Effect	Statistic	Value	F	df1	df2	p	η^2
Transparency	Pillai's Trace	0.11	2.98	1	23	.098	.11
Handoff	Pillai's Trace	0.34	12.10	1	23	.002	.34
Transparency x Handoff	Pillai's Trace	0.10	2.46	1	23	.131	.10

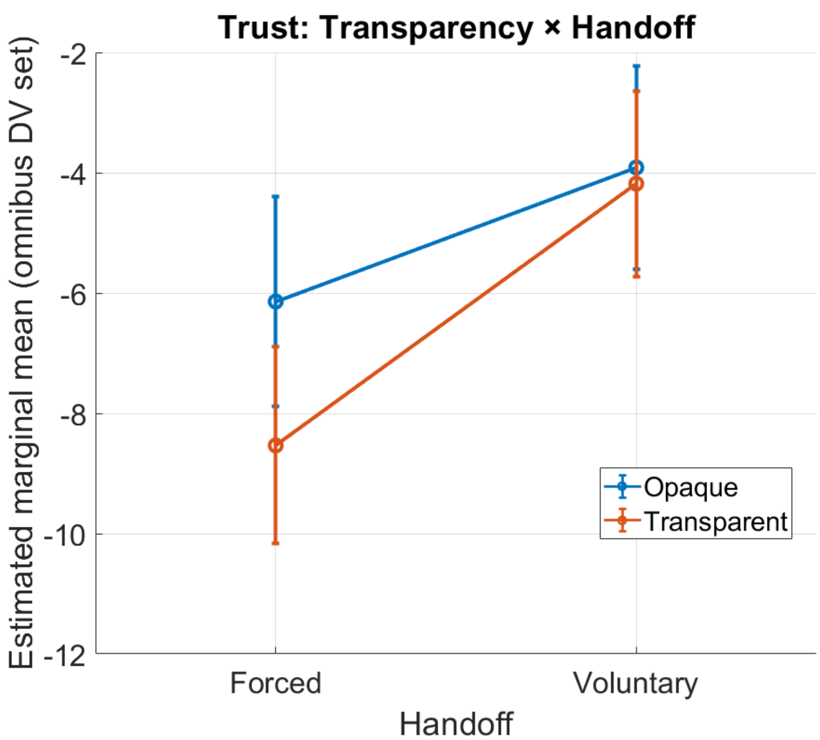


Figure 10. Transparency and handoff differences in multivariate trust results (higher values indicate higher trust).

Linear discriminant analysis.

Cross-validated LDA achieved a classification accuracy of 74.0% for the significant main effect of handoff, with an area under the ROC curve of AUC = 0.845. One discriminant function was retained, dominated by automation dwell-time measures. The strongest contributors were resource management dwell time (d = .67), system monitoring dwell time (d = .43), and communications dwell time (d = .39). Smaller but still meaningful influences came from automation reliance (d = -.27), tracking dwell time (d = .17), and TOAST performance (d = .16). The resulting discriminant function score distributions are depicted in Figure 11.

This discriminant function captured the degree to which trust behaviors aligned with automation use by contrasting genuine reliance with inflated, forced reliance. High scores reflected conditions in which operators continued to dwell extensively on automated subtasks despite high automation usage, a pattern indicative of mistrust. Low scores, in contrast, reflected reduced monitoring of automated subtasks when automation was engaged, consistent with increased trust.

In practice, voluntary handoffs were associated with lower scores on this axis, as operators reduced dwell time on automated subtasks and demonstrated trust that the automation could function without constant oversight. Forced handoffs, however, clustered at higher values, where automation reliance was inflated by design, but operators still devoted significant monitoring effort, undermining the calibration of trust. Thus, this reliance-monitoring Calibration function shows that voluntary handoff produced more authentic trust behaviors, whereas forced handoff yielded a mismatch between automation use and operator confidence.

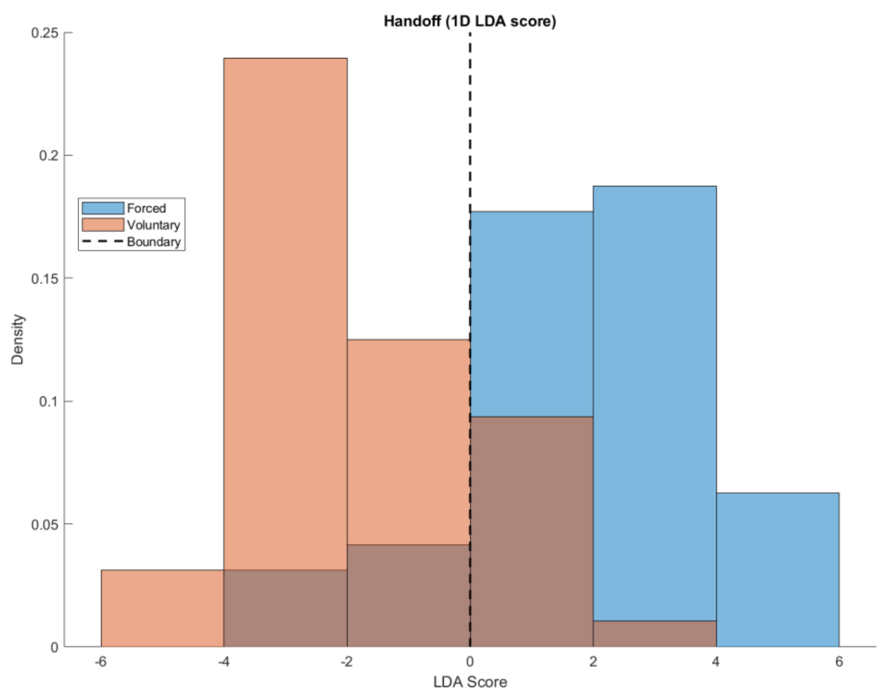


Figure 11. LDA function scores for handoff reliance-monitoring calibration function.

Univariate follow-ups.

Follow-up repeated-measures ANOVAs were conducted to explore individual variables. For handoff, only significant effects emerged for RM auto dwell %, $F(1, 23) = 9.76$, $p = .049$, $\eta p^2 = .28$, after correction. No transparency main effects were significant. The transparency × handoff interaction did not yield significant univariate effects.

Table 8. Repeated-Measures ANOVA Results for Trust in Automation Metrics with Benjamini-Hochberg FDR Correction for Multiple Comparisons

	Transparency		**Handoff**		**Interaction**	
Measure	***p* (FDR)**	***ηp²***	***p* (FDR)**	***ηp²***	***p* (FDR)**	***ηp²***
Reliance Metric						
Automation Reliance	.747	0	.475	.04	.121	.22
Gaze Metrics						
COM Auto Dwell %	.262	.11	.101	.19	.727	.02
RM Auto Dwell %	.211	.17	**.049**	.28	.121	.18
SYS Auto Dwell %	.266	.09	.153	.14	.843	.01
TRK Auto Dwell %	.211	.14	.638	.02	.257	.09
Subjective Metrics						
TOAST: Performance	.321	.06	.404	.06	.853	0
TOAST: Understanding	.321	.05	.886	0	.727	.02

Table 9. Discriminant Function Names, Definitions, and Loadings by Metric Set

Name	**Definition**	**Primary Loadings**
Cognitive Workload		
Transparency: Efficiency-Monitoring	Captures operators' ability to manage concurrent tasks while sustaining vigilance on vision-based monitoring. High values reflect strong multitasking efficiency and system monitoring, with lower reliance on self-reported strain.	MTC (+5.4), SYS (+4.0), RM (-2.6), ISA (-1.6)
Condition LD1: Efficiency-Coordination	Distinguishes operators who integrate multitasking, system monitoring, and communication performance against those faltering in resource management. High values indicate integrated performance; low values reflect breakdowns in coordination.	MTC (+.157), SYS (+.111), COM (+.49), RM (-.63)
Condition LD2: Resource-Workload Tradeoff	Captures the tension between resource management demands and modeled workload. High values reflect stable resource management with moderate workload; low values indicate resource strain despite relatively low modeled workload estimates.	RM (-.56), COM (-.104), IMPRINT (+.075)
Situational Awareness		
Transparency: Projection-Comprehension Balance	Differentiates conditions by emphasizing forward-looking projection and systematic scanning versus degraded comprehension. Transparency supports both projection of future states and balanced attentional allocation, allowing operators to sustain deeper situational models.	Rel. Entropy (-2.183), SAGAT Pro (-2.028), SAGAT Comp (+1.128), TRK Revisit (+0.993)
Handoff: Comprehension-Reacquisition	Captures the degree to which comprehension and perception of system state are preserved or disrupted during control transitions. Voluntary handoffs preserve comprehension and revisit frequency, whereas forced handoffs undermine integration, requiring operators to reacquire understanding.	SAGAT Comp (-3.648), TRK Revisit (+1.784), SAGAT Per (+1.618)
Condition LD1: Comprehension-Perception Tradeoff	Distinguishes conditions where comprehension is dominant from those where reliance shifts toward surface perception with reduced revisiting of dynamic subtasks. This reflects how awareness toggles between meaningful understanding and fragmented perceptual monitoring across automation conditions.	SAGAT Comp (+0.148), SAGAT Per (-0.063), TRK Revisit (-0.046)
Condition LD2: Projection-Attention Balance	Reflects differences in future-oriented awareness (projection) and systematic attentional distribution (entropy, revisits). Transparent and voluntary conditions promote projection with balanced scanning, whereas opaque or forced conditions yield fragmented, reactive monitoring strategies.	SAGAT Pro (+0.107), TRK Revisit (-0.070), Rel. Entropy (+0.064)
Trust in Automation		
Handoff: Reliance-Monitoring Calibration	Differentiates voluntary vs. forced handoff by capturing the contrast between genuine behavioral trust (lower dwell time on subtasks under voluntary handoff) and inflated automation reliance (forced handoff). Voluntary handoff was associated with reduced monitoring of automated subtasks (greater calibrated trust), whereas forced handoff inflated reliance values artificially while eroding genuine trust behaviors.	RM A. Dwell % (.67), SYS A. Dwell % (.43), COM A. Dwell % (.39), Auto Reliance (-.27)

This page is intentionally blank.

Discussion

The purpose of this study was to examine how transparency and handoff method, two foundational features of adaptive automation, jointly influence operator CWL, SA, and trust in automation. These three constructs represent core determinants of effective human-automation teaming, yet the ways in which they interact with system design choices are complex and often yield divergent outcomes. By employing a within-subjects 2 × 2 design, supported by multivariate analyses and discriminant function modeling, the present work sought to clarify how transparency and handoff individually shape operator states in a demanding aviation environment while also exploring their joint interaction.

Across the three assessment domains, the hypotheses received mixed but informative levels of support, revealing both convergence and divergence in how transparency and handoff shape operator states. For cognitive workload, transparency emerged as the dominant factor, but in the opposite direction of the original prediction, transparent conditions significantly increased workload, whereas opaque conditions reduced strain and supported efficiency-monitoring profiles. Voluntary handoff did not yield strong workload benefits overall, suggesting that transparency, not handoff, was the primary workload driver.

For situational awareness, both factors mattered. Opaque displays enhanced comprehension, projection, and attentional distribution, while voluntary handoff preserved comprehension and mitigated the disruptive effects of transitions. Together, these findings show that SA is best maintained when opacity is paired with user control (i.e., voluntary handoffs).

For trust, handoff was decisive. Voluntary transitions significantly enhanced reliance and trust behavior (automation reliance and reduced automated subtask monitoring) across subsystems, whereas transparency did not produce reliable trust gains.

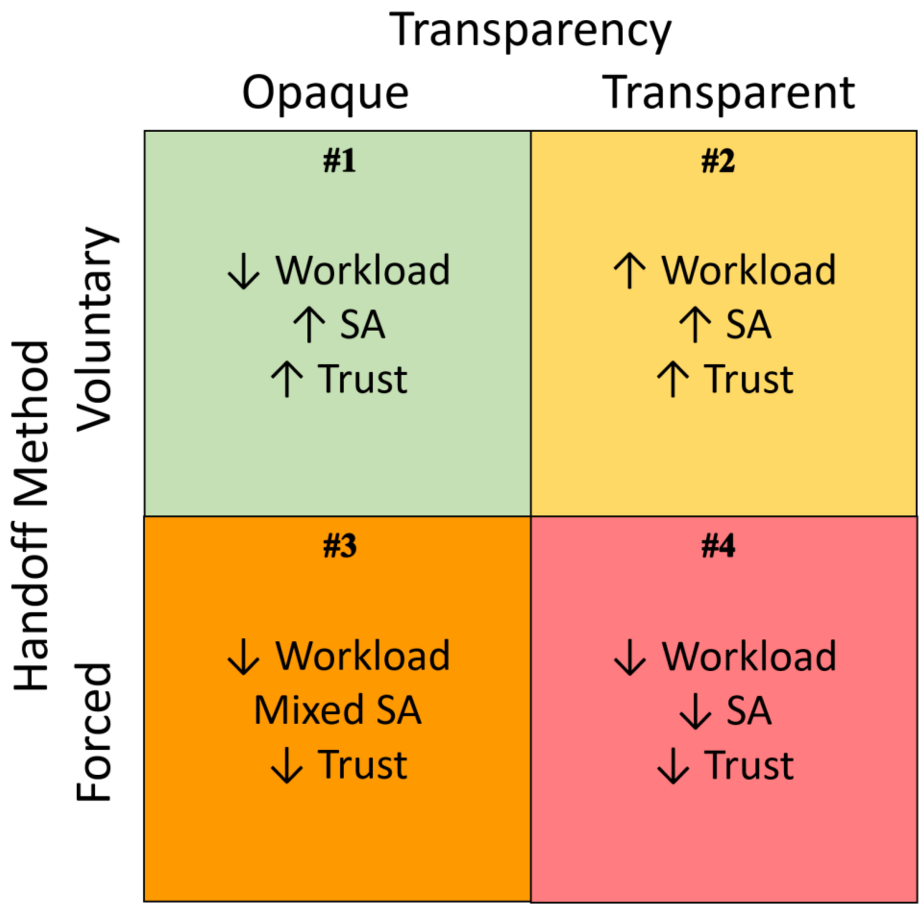
These patterns indicate that transparency primarily regulates cognitive load and attentional balance, handoff governs the relational dynamics of trust, and situational awareness emerges at their intersection. The strongest outcomes were not achieved under transparency, but rather when opaque displays were paired with voluntary handoffs, which consistently produced favorable operator states across workload, SA, and trust. Conversely, the transparent-forced combination undermined performance across multiple constructs, while the mixed conditions revealed tradeoffs depending on which factor dominated. This pattern of results is illustrated in Figure 12, with the main effects summarized in Table 10 alongside corresponding design recommendations, and the exploratory interaction findings detailed in Table 11. These findings highlight that adaptive automation design must avoid privileging one construct at the expense of others.

Table 10. Summary of Results Relative to Hypotheses (H) 1-6

Hypothesis	Factor Level	Observed Change	Design Recommendations
Transparency			
H1: CWL	Transparent	↑ $p = .002$, $\eta^2 = .35$	✗ Avoid continuous transparency in high-demand phases; use opaque displays operationally
H2: SA	Transparent	↓ $p = .005$, $\eta^2 = .29$	✗ Avoid persistent confidence readouts; provide transparency only on demand
H3: Trust	Transparent	× $p = .098$, $\eta^2 = .11$	⚠ Transparency alone does not build trust; combine with reliability cues and voluntary handoffs
Handoff Method			
H4: CWL	Voluntary	× $p = .205$, $\eta^2 = .07$	⚠ No strong workload benefit, but ✓ retain voluntary handoff for flexibility
H5: SA	Voluntary	↑ $p < .001$, $\eta^2 = .58$	✓ Always prefer voluntary handoff; preserves comprehension/projection
H6: Trust	Voluntary	↑ $p = .002$, $\eta^2 = .34$	✓ Adopt voluntary/autonomy-by-consent handoffs to foster calibrated reliance

Table 11. Summary of Exploratory Interaction Effects Between Transparency and Handoff Method

Transparency Level	Handoff Method	Level	Design Recommendations
CWL ($p = .006$, $\eta^2 = .29$)			
Transparent	Voluntary	↑	⚠ Use only in training/rehearsal; elevates workload in operations
Transparent	Forced	↓	✗ Short-term CWL benefit but harms SA, do not deploy operationally
Opaque	Voluntary	↓	✓ Best pairing for workload reduction and agency; set as default
Opaque	Forced	↓	⚠ Lowers CWL but ✗ harms trust, limit to emergency overrides
SA $p = .012$, $\eta^2 = .24$			
Transparent	Voluntary	↑	⚠ Useful in training/low tempo, selective use only
Transparent	Forced	↓	✗ Worst-case, avoid entirely
Opaque	Voluntary	↑	✓ Optimal, supports projection and comprehension
Opaque	Forced	~	⚠ Neutral but weaker than voluntary, secondary option only



1. **Opaque-Voluntary:** The optimal profile. CWL decreased, SA improved across all facets: higher comprehension, stronger projection, and balanced scanning (entropy and revisits). Trust also increased, with voluntariness preserving reliance and limiting unnecessary monitoring.
2. **Transparent-Voluntary:** CWL increased due to transparency's attentional costs, but SA improved in comprehension and projection. Attentional balance was maintained at moderate levels, with voluntariness buffering strain. Trust also rose, as voluntary handoffs sustained calibrated reliance despite workload elevation.
3. **Opaque-Forced:** Opacity lowered workload and partially supported SA by preserving projection and some scanning balance. Comprehension declined under forced handoffs, requiring reacquisition. Trust dropped, with forced transitions eroding reliance despite efficiency gains.
4. **Transparent-Forced:** The poorest profile. SA declined across comprehension, projection, and visual scanning, with fewer revisits to dynamic subtasks. Trust also fell, as operators over-monitored automation under reduced agency. Workload was also lower relative to the transparent-voluntary condition and in-line with the other conditions.

Figure 12. Summary of operator states across a 2 × 2 design crossing transparency (transparent vs. opaque) with handoff method (voluntary vs. forced).

Cognitive Workload

The CWL results demonstrate that transparency was the dominant determinant of workload regulation, but in the opposite direction of expectations. Rather than reducing workload, transparency significantly increased strain, while opaque conditions supported better multitasking efficiency, system monitoring, and overall vigilance. This finding directly contradicted H1 and suggests that providing continuous system visibility may impose additional attentional and coordination costs in high-demand supervisory control settings.

By contrast, handoff method alone did not significantly alter workload, offering only partial support for H4. Still, its influence was evident in certain subtasks. Voluntary handoffs preserved resource management stability, while forced handoffs disrupted coordination in communication-heavy domains. Thus, transparency drove global workload differences, while handoff shaped how those effects were distributed across tasks.

These results diverge from classic predictions that transparency should reduce uncertainty and thereby lower workload (Kaber & Endsley, 2004; Wickens, 2002). Instead, they align with more recent critiques of the "transparency paradox." Explanatory content can clarify automation logic but simultaneously increase cognitive load when delivered during complex, time-sensitive tasks. The exceptionally strong effect on multitasking efficiency ($\eta p^2 = .92$) further reinforces that excessive transparency undermines operators' ability to coordinate across subtasks, a critical competency in aviation and other multitasking domains.

From a design perspective, these findings suggest that opaque displays should be prioritized during high-demand operational phases (e.g., degraded visual environments, high communication load), where reducing strain and supporting vigilance is paramount. Transparency may still have value, but primarily in training, mission rehearsal, or on-demand contexts, where attentional costs are less consequential and explanatory cues can be leveraged for learning. Handoff design remains secondary but important. Voluntary transitions should be the default, as they help preserve coordination, while forced overrides should be reserved for emergencies.

In sum, the CWL results reveal that transparency increased workload rather than reduced it, while opacity enabled more efficient multitasking and monitoring. Voluntary handoffs offered localized benefits but could not offset the elevated workload induced by transparency. These findings refine workload theory by showing that context, task demand, and timing determine whether transparency alleviates or exacerbates operator strain.

Situational Awareness

The SA results show that both transparency and handoff method strongly influenced awareness, but with different roles. Transparency increased attentional strain, such that opaque displays yielded higher comprehension, projection, and balanced scanning, while voluntary handoffs preserved continuity of comprehension during transitions. The large effect sizes associated with handoff highlight that transition quality is the primary determinant of SA preservation, consistent with Endsley's (1995) model that awareness is most fragile at control transfer points.

The second hypothesis, that transparency would improve SA, was not supported; instead, opaque displays provided clearer scaffolding for comprehension and projection while transparent displays primarily shaped scanning (entropy/revisits) without delivering net SA gains. The fifth hypothesis, that voluntary handoff would improve SA, was supported, with voluntariness protecting comprehension and reducing disruption during transitions. Importantly, the interaction between transparency and handoff showed that the best outcomes occurred under opaque-voluntary conditions, while the worst outcomes occurred under transparent-forced conditions, where both attentional load and control disruptions undermined SA.

These findings refine SA theory in several ways. First, they confirm that projection and comprehension are the most vulnerable levels of SA (Endsley, 1995), and demonstrate that opacity, by reducing attentional overhead, better supports both. Second, they align with prior evidence that forced transitions degrade comprehension and projection (Chen & Barnes, 2014; Wright et al., 2018), while voluntary handoffs allow operators to sustain mental models across shifts in control. Finally, the eye-tracking results support work linking entropy balance and revisit frequency to higher awareness (Jones & Endsley, 2004; Pan et al., 2025), showing that opaque-voluntary conditions fostered systematic monitoring strategies.

From a design perspective, these findings suggest that opaque displays should be the default in operational phases, where minimizing workload and preserving comprehension are critical. Transparency can still serve a role in training or low-demand contexts, but operational systems should instead emphasize voluntary, predictable handoffs that preserve comprehension and continuity of SA. Eye-tracking indices could also serve as real-time triggers for adaptive display management, detecting when SA begins to degrade and delaying or adjusting handoffs accordingly.

In sum, the SA results indicate that opacity scaffolds projection and voluntary handoff preserves comprehension, with the two factors jointly determining awareness outcomes. The most favorable conditions combined both, while the transparent-forced pairing consistently undermined SA.

Trust in Automation

The trust results revealed a different pattern than CWL and SA. Handoff quality, not transparency, was the dominant determinant of trust calibration. Transparency showed only a weak, nonsignificant trend, whereas voluntary handoffs consistently preserved reliance and reduced over-monitoring, especially in resource management tasks. This underscores that trust is shaped less by informational context and more by operators' perceived autonomy and control during transitions.

The third hypothesis, that transparency would enhance trust, was not supported. By contrast, the sixth hypothesis, that voluntary handoff would foster higher trust, was supported, with voluntariness distinguishing conditions through reliance-based behaviors. This aligns with theories that describe trust as fundamentally relational, rooted in autonomy, predictability, and controllability (Lee & See, 2004; Hoff & Bashir, 2015). Operators appeared to interpret forced handoffs as violations of agency, which eroded trust even when automation performed competently.

Behavioral measures such as dwell time on automated subtasks proved more sensitive than global trust ratings, echoing Dzindolet et al.'s (2003) observation that operators may report trust but still monitor excessively when trust is fragile. The discriminant analysis highlighted this distinction through a reliance-monitoring calibration function, where voluntary handoffs were associated with genuine calibrated reliance (less monitoring), and forced handoffs reflected inflated reliance paired with persistent oversight.

Condition-level comparisons reinforced this conclusion. Opaque-voluntary and transparent-voluntary handoffs produced the highest trust, with little difference between them, while both forced handoff conditions yielded the lowest trust outcomes. This ordering shows that handoff quality outweighed transparency in shaping trust, distinguishing trust patterns from CWL (dominated by transparency) and SA (shaped by both factors).

From a design standpoint, these findings indicate that handoff processes should be the primary lever for trust management. Effective systems should prioritize voluntary, predictable, and reversible transitions, reserving forced overrides for emergencies. Transparency can supplement these processes, but it cannot compensate for poor handoff design. Monitoring reliance behaviors such as dwell time in real-time may allow adaptive automation to detect erosion of trust and dynamically adjust handoff timing or transparency cues to recalibrate confidence.

In sum, the trust findings confirm that handoff quality is central to building and preserving trust, while transparency plays only a supporting role. Trust is relational rather than informational, dependent on how automation engages operators, not just on what it reveals about its internal state.

Future Adaptive Automation Design Guidelines

The synthesis of CWL, SA, and trust findings provides clear design guidance for next-generation adaptive automation systems in Army aviation. Unlike earlier frameworks that emphasized one construct at the expense of others, the present results show that CWL, SA, and trust are shaped differently by transparency and handoff method. Effective design must therefore adopt a multidimensional perspective, ensuring that interventions in one domain do not unintentionally degrade another.

The first principle derived from this work is that transparency elevates workload under complex multitasking conditions. Contrary to prior assumptions (Kaber & Endsley, 2004; Wickens, 2002), continuous transparency increased strain and degraded multitasking efficiency, suggesting that added contextual cues can overload attention when presented during high-demand phases (simulated by the high multitasking demand of the USAARL MATB). In Army aviation, where pilots must manage degraded visual environments, dense communications, and rapid decision cycles, continuous transparency may act as a cognitive distractor rather than a resource regulator. Consistent with recent critiques of the "transparency paradox" (Wright & Barber, 2021; Wright et al., 2020), transparency should therefore be restricted to training, rehearsal, or on-demand contexts, rather than being the default for full-time operational displays. For line operations, opaque interfaces with selectively triggered transparency (e.g., query-driven explanations or phase-selective overlays) are more likely to safeguard workload efficiency.

The second principle is that handoff voluntariness is the primary determinant of trust. Forced handoffs consistently eroded calibrated reliance, even when transparency cues were present, while voluntary handoffs preserved trust regardless of display condition. This highlights that trust in automation is less about what the system reveals and more about whether the operator retains agency over transitions (Lee & See, 2004; Hoff & Bashir, 2015). For Army aviation, this means that automation should operate on an "autonomy-by-consent" basis, where aircrew initiate or confirm transitions. Emergency overrides using forced control takeover remain necessary in safety-critical cases (e.g., collision avoidance, loss of control), but systems should provide clear justifications for forced handoffs and rapid paths for re-engagement.

The third principle is that SA requires coordination between transparency and voluntariness. Opaque displays supported comprehension and projection by reducing attentional overhead, while transparent displays primarily shaped scanning strategies and can raise strain under high tempo tasks such as the USAARL MATB. Voluntariness preserved comprehension across transitions. Only when opacity and voluntary handoffs were combined did operators demonstrate optimal awareness, reflected in higher SAGAT scores, balanced gaze entropy, and revisit behaviors. These findings extend Endsley's (1995) model by showing that projection is best scaffolded by display opacity, while comprehension continuity depends on voluntary control of transitions. For Army aviation, this implies that explanatory cues should be delivered in synchrony with voluntary handoffs, ensuring that awareness is carried forward seamlessly across the human-automation boundary.

The fourth principle is that condition-level tradeoffs must be anticipated. As shown in Tables 10 and 11, intermediate configurations (e.g., transparent-forced, opaque-voluntary) yielded partial benefits but also significant deficits. Transparent-forced reduced workload in some subtasks but undermined awareness and trust, while opaque-voluntary reduced workload and supported trust but did not fully optimize awareness. For Army aviation, this underscores the need for multi-objective optimization, where transparency and handoff are dynamically tuned to balance CWL, SA, and trust simultaneously. This approach aligns with recent recommendations for integrated adaptive automation frameworks (Pharmer et al., 2025).

Finally, the results highlight the importance of dynamic, context-sensitive adaptation. Operator states fluctuate across mission phases, task demands, and environmental stressors. Fixed transparency settings or rigid handoff policies are insufficient. Instead, adaptive automation should employ real-time classifiers based on CWL efficiency, SA indicators (e.g., entropy, revisit patterns), and trust behaviors (e.g., dwell time) to guide adjustments. In Army aviation, this means that during high-tempo phases (e.g., nap-of-the-earth flight or degraded visibility), systems should default to opaque-voluntary, while in lower-tempo or training contexts, transparency-on-demand can enhance learning and projection without overloading.

In conclusion, the present findings suggest that the future of adaptive automation in Army aviation lies in integration rather than isolation of design principles. Transparency must be carefully phase-selected, voluntariness must be prioritized to preserve trust, and the coordinated use of both is necessary to sustain SA. Designing systems that dynamically balance these principles will enhance not only efficiency and resilience but also operator confidence and safety in high-demand operational environments.

Limitations

Although the present study offers novel insights into the effects of transparency and handoff on CWL, SA, and trust in automation, several limitations must be acknowledged. First, the relatively small sample size of 24 aviators constrains statistical power and limits generalizability. Larger and more diverse samples would strengthen confidence in the robustness of the observed effects and allow further examination of potential moderating variables such as flight experience, age, or propensity to trust automation.

Second, the study employed categorical manipulations of transparency (transparent vs. opaque) and handoff (voluntary vs. forced), rather than graded or continuous levels. In practice, transparency and autonomy management often exist on a spectrum, with varying degrees of system explanation, predictability, and operator control. The dichotomous manipulation therefore simplifies complex design variables and may not capture the full range of real-world system behavior. It is likely that the true optimal configuration will fluctuate between individuals and with various stressors. Future research should explore more granular manipulations that reflect the continuum of transparency and autonomy.

Third, trust in automation was measured within a relatively short experimental timeframe. Although residualization on the Adapted Propensity to Trust in Technology Questionnaire controlled for trait-level trust tendencies, state trust itself may evolve over longer periods of exposure, shaped by repeated cycles of reliability and performance feedback. The weak condition-level discriminability of trust observed in this study may therefore reflect its slower-moving nature. Longitudinal studies are needed to determine how transparency and handoff jointly influence trust trajectories over extended use as experience with the system is developed. Additionally, the sample collected for this study was represented by aviators with a skewed distribution of flight experience, potentially underrepresenting mid-career aviators, as seen in Figure 13.

This space is intentionally blank.

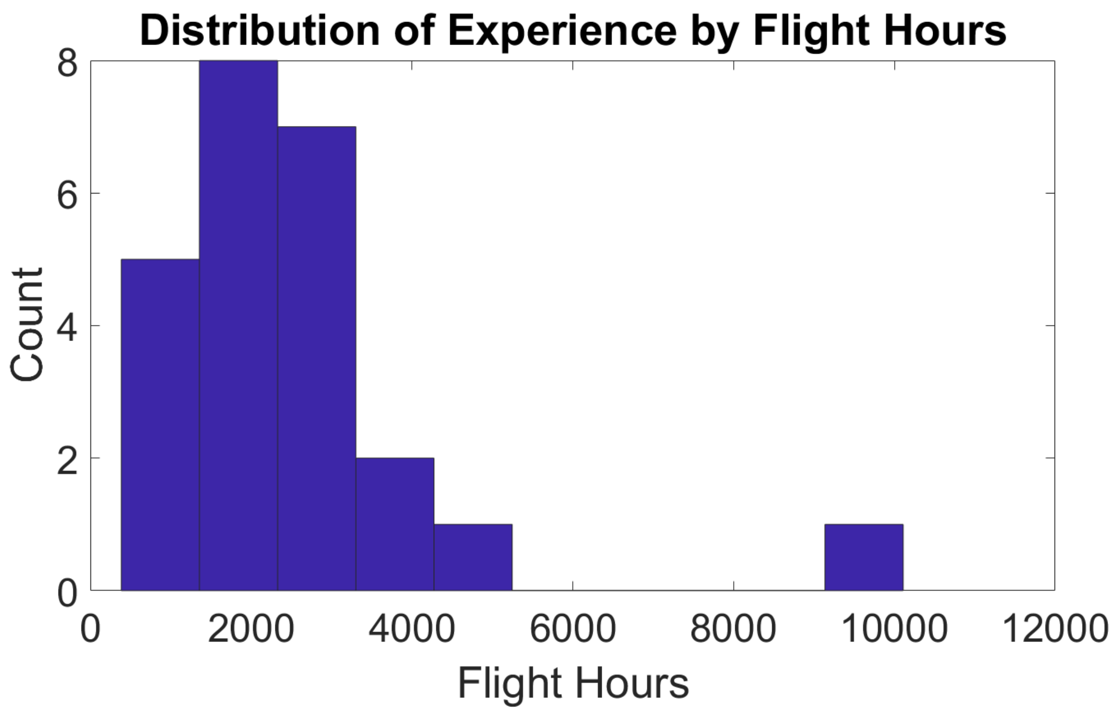


Figure 13. Distribution of aviator experience by flight hours.

Finally, although the study leveraged multiple converging measures within each construct, ecological validity remains limited. The USAARL MATB platform captures important features of active and supervisory control but does not fully replicate the complexity, stress, or stakes of operational aviation missions. Additionally, the laboratory environment in which the USAARL MATB was presented does not simulate the motion artifacts that are common in applied settings. Translating these findings into applied contexts will require testing in higher-fidelity simulators, in actual aircraft (in training centers or using our research Black Hawk), and eventually in field settings.

Conclusion

This study examined how transparency and handoff methods jointly shape CWL, SA, and trust in automation within a demanding aviation task environment. Multivariate analyses and discriminant modeling revealed that while these three constructs are tightly interrelated, they respond to automation design features in distinct yet overlapping ways. Transparency consistently increased CWL by taxing attentional resources, yet it also shaped aspects of SA by supporting projection and influencing scanning strategies. Opacity provided more stable benefits for comprehension and projection, while voluntariness of handoff emerged as the cornerstone of trust, preserving calibrated reliance across conditions. Together, these patterns highlight that CWL, SA, and trust are not governed by a single design lever, but by the dynamic interplay of transparency and handoff.

Across conditions, the rankings revealed that the opaque-voluntary combination was globally optimal, producing the lowest CWL, the highest SA, and the strongest trust. At the opposite extreme, the transparent-forced condition yielded the poorest outcomes, with degraded

SA, diminished trust, and elevated CWL demands. The intermediate combinations highlighted important tradeoffs; transparent-voluntary preserved trust and bolstered SA but carried higher workload costs, while opaque-forced reduced workload and partially preserved SA but undermined trust through abrupt transitions. These findings emphasize that transparency and handoff are not interchangeable levers; rather, they address different operator needs and must be carefully balanced.

From an applied perspective, the results point to several design imperatives for adaptive automation in Army aviation. Transparency should not default to full-time operation, as continuous displays increase workload under high-demand conditions; instead, transparency should be phase-selective or on-demand, used primarily in training, rehearsal, or lower-tempo phases. Handoff mechanisms should default to voluntariness, as operator agency is the hallmark of calibrated trust, with forced overrides reserved for emergencies and accompanied by clear justifications. Finally, SA requires coordination between display mode and handoff process; opacity supports comprehension and projection, while voluntariness preserves comprehension across transitions. Synchronizing these design features ensures that awareness is maintained even during dynamic handoffs.

The broader implication of this study is clear, effective adaptive automation is not about choosing transparency or voluntariness in isolation, but about integrating both within a cohesive, context-sensitive design philosophy. Transparency influences cognitive strain, voluntariness governs trust in the system, and awareness emerges when these dimensions are aligned. By recognizing this interdependence, future adaptive automation can move beyond piecemeal interventions toward holistic architectures that enhance operator performance and resilience.

The takeaway message is simple yet consequential; automation that explains itself but seizes control will not be trusted; automation that cedes control but burdens operators with transparency will elevate workload; and automation that conceals its logic while forcing transitions will degrade awareness. In short, the future of adaptive automation will not be won by transparency or autonomy alone, but by learning when to stay opaque, when to yield, and when to let the human lead.

References

Akash, K., McMahon, G., Reid, T., & Jain, N. (2020). Human trust-based feedback control: Dynamically varying automation transparency to optimize human–machine interactions. *IEEE Control Systems, 40*(6), 98–116.

Billings, C. E. (1996). *Human-centered aviation automation: Principles and guidelines* (No. NASA-TM-110381). NASA.

Bommer, S., Vogl, J., Atchley, J. A., & McCurry, C. D. (2025). *Advancing adaptive automation in aviation: A simulation approach integrating IMPRINT and MATB* [Manuscript in preparation].

Buck-Gengler, C. J., Raymond, W. D., Healy, A. F., & Bourne, L. E. (2012). Modeling cognitive tasks in IMPRINT. In *Training Cognition* (pp. 201-224). Psychology Press.

Casner, S. M., & Schooler, J. W. (2014). Thoughts in flight: Automation use and pilots' task-related and task-unrelated thought. *Human Factors, 56*(3), 433–442.

Chen, J. Y. C., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems, 44*(4), 671–681.

Couluris, G. J., Tashker, M. G., & Penick, M. C. (1978). *Policy impacts of ATC automation: Human factors considerations* (FAA-AVP-78-1). Federal Aviation Administration.

David, L. (2019, July 17). *Here's what Neil Armstrong saw as he landed the Eagle*. Space.com. Retrieved from https://www.space.com/neil-armstrong-apollo-11-landing-view-video.html

de Waard, D. (1996). *The measurement of drivers' mental workload* [Unpublished doctoral dissertation, University of Groningen].

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.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting* (Vol. 32, No. 2, pp. 97-101). Sage Publications.

Ensley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 85–104.

Gegoff, I., Tatasciore, M., Bowden, V., McCarley, J. S., & Loft, S. (2024). Transparent automated advice to mitigate the impact of variation in automation reliability. *Human Factors, 66*(8), 2008–2024.

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). North-Holland.

Hink, L. L. (1974). Automation in air traffic control. In *1974 IEEE Conference on Decision and Control including the 13th Symposium on Adaptive Processes* (pp. 545-547). IEEE.

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., & Capiola, A. (2019). The measurement of the propensity to trust automation. In *International conference on human-computer interaction* (pp. 476-489). Cham: Springer International Publishing.

Jones, D. G., & Endsley, M. R. (2004). Use of real-time probes for measuring situation awareness. *The international journal of aviation psychology*, *14*(4), 343–367.

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.

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, *117*(7), 1574–1581.

Kayes, D. C., & Yoon, J. (2022). Cognitive offloading strategies and decrements in learning: Lessons from aviation and aerospace crises. *Journal of Human Performance in Extreme Environments, 17*(1), 2.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International journal of human-computer studies*, *40*(1), 153–184.

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

Lockheed Martin. (2022, February 8). *Safe, reliable, and uninhabited: First autonomous BLACK HAWK® helicopter flight*. Lockheed Martin. Retrieved from https://www.lockheedmartin.com/en-us/news/features/2022/safe-reliable-and-uninhabited-first-autonomous-black-hawk-flight.html

Lyons, J. B. (2013). Being transparent about transparency: A model for human-automation trust. In *Proceedings of the AAAI Spring Symposium on Trust in Autonomous Systems* (pp. 48–53). AAAI.

Molloy, R., & Parasuraman, R. (1992). Monitoring automation failures: Effects of automation reliability and task complexity. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 36, No. 18, pp. 1518-1521). SAGE Publications.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, *69*(348), 894–908.

Pan, W., Liang, R., Wang, Y., Song, D., & Yin, Z. (2025). Situational awareness prediction for remote tower controllers based on eye-tracking and heart rate variability data. *Sensors*, *25*(7), 2052.

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. (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.

Peck, L. S. (1959). Machine tool automation: A description of the numerical tool control system developed by the Autonetics Division of North American Aviation, Inc. *Aircraft Engineering and Aerospace Technology, 31*(11), 345–346.

Pharmer, R. L., Wickens, C. D., & Clegg, B. A. (2025). Transparent systems, opaque results: A study on automation compliance and task performance. *Cognitive Research: Principles and Implications, 10*(1), 8.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators & Virtual Environments, 11*(4), 335–351.

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of Human Factors and Ergonomics, 1*(1), 89–129.

Sheridan, T. B., Verplank, W. L., & Brooks, T. L. (1978). Human/computer control of undersea teleoperators. In NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Taylor, R. M. (2017). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness* (pp. 111-128). Routledge.

Thackray, R. I. (1980). *Boredom and monotony as a consequence of automation: Consideration of the evidence relating boredom and monotony to stress* (FAA-AM-80-1). FAA Office of Aviation Medicine.

van de Merwe, K., Mallam, S., & Nazir, S. (2024). Agent transparency, situation awareness, mental workload, and operator performance: A systematic literature review. *Human Factors, 66*(1), 180–208.

Villani, V., Pini, F., Leali, F., & Secchi, C. (2018). Survey on human-robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics, 55*, 248–266.

Vogl, J., D'Alessandro, M., Wilkins, J., Ranes, B., Persson, I., McCurry, C. D., Bommer, S. (2024a). *Optimizing Adaptive automation in aviation: a literature review on dynamic automation system interaction* (2025-19). U.S. Army Aeromedical Research Laboratory.

Vogl, J., McCurry, C. D., Bommer, S., & Atchley, J.A. (2024b). The United States Army Aeromedical Research Laboratory Multi-Attribute Task Battery. *Frontiers in Neuroergonomics, 5*,1435588.

Warren, N. D. (1956). Automation, human engineering, and psychology. *American Psychologist, 11*(10), 531–535.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177.

Wiener, E. L. (1989). *Human factors of advanced technology (glass cockpit) transport aircraft* (No. NAS 1.26: 177528). NASA.

Wojton, H. M., Porter, D., T. Lane, S., Bieber, C., & Madhavan, P. (2020). Initial validation of the trust of automated systems test (TOAST). *The Journal of social psychology*, *160*(6), 735–750.

Wright, J. L., Chen, J. Y., & Barnes, M. J. (2018). Human-automation interaction for multiple robot control: The effect of varying automation assistance and individual differences on operator performance. *Ergonomics*, *61*(8), 1033–1045.

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

Zimmerman, B., Lefkovitz, D., & Prywes, N. S. (1964). The Naval Aviation Supply Office inventory retrieval system - a case study in file automation. *Management Science, 10*(3), 421–428.

Appendix A. Acronyms and Abbreviations

ANOVA	Analysis of Variance
APTQ	Adaptive Propensity to Trust in Technology Questionnaire
COM	Communications Task
CWL	Cognitive Workload
df	Degrees of Freedom
ECG	Electrocardiogram
EEG	Electroencephalography
FDR	False Discovery Rate
GUI	Graphical User Interface
HRV	Heart Rate Variability
IMPRINT	Improved Performance Research Integration Tool
ISA	Instantaneous Self-Assessment (of workload)
KSS	Karolinska Sleepiness Scale
LD	Linear Discriminant
LDA	Linear Discriminant Analysis
LF:HF	Low Frequency:High Frequency (band ratio)
LSL	Lab Streaming Layer
MANOVA	Multivariate Analysis of Variance
MATB	Multi-Attribute Task Battery
NASA	National Aeronautics and Space Administration
PCA	Principal Component Analysis
PVT	Psychomotor Vigilance Task
RM	Resource Management Task
RMSSD	Root Mean Square of Successive Differences
ROC	Receiver Operating Curve
ROI	Region of Interest
SA	Situational Awareness
SAGAT	Situation Awareness Global Assessment Tool
SART	Situational Awareness Rating Technique
SYS	System Monitoring Task
TLX	Task Load Index
TRK	Tracking Task
UAV	Unmanned Aerial Vehicle
UH-60	Utility Helicopter 60 (Black Hawk)
USAARL	U.S. Army Aeromedical Research Laboratory
VOGL	Virtual Offloading Guidance Logic
ηp^2	Partial Eta Squared

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

FORGE THE FUTURE

MRDC