USAARL-TECH-FR--2022-03

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Designing a Trust Manipulation for Unmanned Aerial Systems

Michael Wilson, Christopher Long, Elizabeth Niedbala, Juan Colón, Christopher Aura, William Irvin, Kevin Andres, Kevin Baugher, Cristina Delgado-Howard, & Kevin O’Brien

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 Statement

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) dated 8 November 2011.

IRB Determination and Number

The USAARL Exempt Determination Official has determined that this research is exempt [32 CFR 219.104(d) (3) (i) and (ii)]. This research involves behavioral interventions in conjunction with verbal and written responses, and audiovisual recordings. This data is collected in a manner in which the identity of the subject cannot readily be ascertained. The benign behavioral interventions are brief in duration, harmless, painless, not physically invasive, not likely to have a significant adverse lasting impact on the subjects, and the investigator has no reason to think the subjects will find the interventions offensive or embarrassing.

This data set was collected under an Exempt Determination by the U.S. Army Medical Research and Materiel Command, now the U.S. Army Medical Research and Development Command, Institutional Review Board and U.S. Army Aeromedical Research Laboratory (USAARL); USAARL study #2020-002. Exempt Determination memorandum dated 25 Feb 2020.

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

1. REPORT DATE *(DD-MM-YYYY)*
13-10-2021

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*
OCT 2018 - JUL 2021

4. TITLE AND SUBTITLE
Designing a Trust Manipulation for Unmanned Aerial Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
6.2

5d. PROJECT NUMBER
2020-002

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Wilson, M.[1], Long, C.[1], Niedbala, E.[1,3], Colón, J.[1,2], Aura, C.[1,2], Irvin, W.[1,3], Andres, K.[1,2], Baugher, K.[1], Delgado-Howard, C.[1,3,] & O'Brien, K.[1]

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-FR--2022-03

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

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

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

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

13. SUPPLEMENTARY NOTES
[1]U.S. Army Aeromedical Research Laboratory, [2]Goldbelt Frontier, LLC; [3]Oak Ridge Institute for Science and Education

14. ABSTRACT
The purpose of this study is to establish a validated trust in automation manipulation to use in future research. Specifically, this experiment compared a high reliability system to a low reliability system with the intention of creating significant and meaningful differences in system trust and system use. Twenty-two unmanned aircraft system operators were randomly assigned to one of two (high and low) system reliability conditions of a hypothetical target identification system. Subjective trust in the system and behavioral use of the system were measured. The goal of this experiment is to validate a novel method of manipulating operator trust using an operationally relevant device, the Universal Mission Simulator. This trust manipulation was evaluated for use in future research testing exogenous factors that influence unmanned aircraft system operator trust in automation. Without modification, the trust manipulation was not effective in producing differences in individual trust between the reliability groups. Further research is necessary to determine how the trust manipulation should be changed in order to yield the desired results.

15. SUBJECT TERMS
unmanned aerial systems, UAS, trust in automation, universal mission simulator, UMS

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	19	Loraine St. Onge, PhD
					19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

This page is intentionally blank.

Summary

The purpose of this study was to establish a validated trust in automation manipulation to use in future research. Specifically, this experiment compared a high reliability system to a low reliability system with the intention of creating significant and meaningful differences in system trust and system use. Twenty-two unmanned aircraft system operators were assigned to one of two (high and low) system reliability conditions of a hypothetical Target Identification System (TIS). Subjective trust in the system and behavioral use of the system were measured. Although the hypothesized effect size of the trust manipulation was not achieved, results obtained suggest that subjects' level of fatigue moderated the effects of trust in the automated system. Across fatigue levels, subjects in the high system accuracy condition trusted the automated system more than subjects in the low system accuracy condition. Moreover, the most fatigued group trusted the high accuracy system, and distrusted the low accuracy system, more than the low and medium fatigued groups.

This page is intentionally blank.

Acknowledgements

This research was supported in part by an appointment to Postgraduate Research 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

List of Figures

List of Tables

This page is intentionally blank.

Introduction

Historically, technology assessments such as the U.S. Department of Defense's (DoD) Unmanned Systems Roadmap have asserted that greater autonomy produces more effective military systems. However, this path has not been as straightforward as hoped. There are serious questions about traditional views on how autonomous systems function alongside humans. For example, an analysis of fratricide incidents involving the Patriot Missile system concluded, counterintuitively, that complex technologies increase the need for operator expertise (Hoffman et al., 2014). Reversing years of precedent in federally funded autonomy research, the 2012 Defense Science Board report, recommended military technology procurement programs abandon the focus on supervisory control and "levels of autonomy" and develop a reference framework around human-computer collaboration (Systems et al., 2012). Better understanding the operator's trust in automation is essential to developing such a framework.

Present unmanned aerial vehicle (UAV) operations are automated to such a degree that the operator largely monitors the flight with few manual inputs (Cummings et al., 2014). While highly automated systems offer several advantages, such as reduced crewing requirements and reduced workload, improper levels of trust in automation can lead to catastrophic consequences. Improper trust calibration is typically categorized as "misuse" or "disuse" of automation, both of which are behaviors that undermine performance and safety (Parasuraman & Riley, 1997). Misuse is characterized by an inappropriate reliance on automation, such as when a pilot over-trusts the autopilot function and fails to intervene when necessary. Disuse of automation is the inappropriate rejection of automation, for instance, when Army UAV operators reject using Shadow's highly reliable auto-landing feature in favor of a manual landing, which requires a complex multi-step checklist, increasing the risk of human error. Successful use of automation requires understanding the factors that encourage misuse and disuse of automated agents.

Trust researchers have identified numerous endogenous and exogenous factors that contribute to an operator's trust in automation (for reviews, see [Hoff & Bashir, 2015; Lee & See, 2004]). However, one of the strongest factors that predicts trust in automation is the automation's performance (Hancock et al., 2011; McLeod et al., 2005; Schaefer et al., 2016). Generally, automation that is more reliable or dependable is more likely to be trusted. Hancock et al.'s (2011) meta-analysis on the factors influencing trust in robots found that robot performance (e.g., dependability, reliability, predictability) had a stronger effect on trust compared to the user's ability (e.g., expertise, prior experience), characteristics (e.g., personality, propensity to trust, demographics), the robot's attributes (e.g., robot type, adaptability, anthropomorphism), and environmental factors (e.g., team communication, task type). Similarly, a meta-analysis conducted by Schaefer et al. (2016) on the factors influencing trust in autonomous systems found that automation capability and behavior had a greater effect on trust than human cognitive factors and demographics. Human emotive factors (e.g., the user's confidence in the automation, attitudes toward the automation, and/or commitment to the automation) were the only human dimension that showed a comparable effect size to automation capability and behavior.

Other research has found significant effects on trust when the automated system's reliability level is manipulated (Bagheri & Jamieson, 2004; Bailey & Scerbo, 2007; Bliss & Acton, 2003; Cahour & Forzy, 2009; Cummings et al., 2010; Donmez et al., 2006; Ho et al., 2005; Kazi et al., 2007; Moray et al., 2000; Stedmon et al., 2007). Schaefer et al. provided a succinct summary, “systems with highly reliable behavior and communication engender trust across differing task contexts” (2016, p. 17). Although errors occur in human-centered human automation, the reduction of all types of error increases the reliability of a system and thus positively influences trust development (de Vries et al., 2003). Conversely, miss rate and false alarm rate degrades trust and adversely affects performance (Wickens et al., 2009; Yamada & Kuchar, 2006). Further research suggests that the degree of task difficulty influences trust level much more than the type of error that appears. Trust in the automation system increases when the targets detected are perceived by the operator to be a difficult target to identify. Madhavan and Wiegmann (2007) suggested that trust degrades when the automation system misses or provides a false alarm while detecting what the operator perceives to be an easily identifiable target. “Master and colleagues (2005) found a significant change in trust between conservative and risky systems but no difference in trust based on error rates” (Schaefer et al., 2016, p. 16). Critically, the level of trust that an operator feels toward an automated agent is independent from how appropriate their level of trust is for the agent. Increasing an operator’s trust level should not be the ultimate goal if it leads to misuse by over-trusting the agent when it should not be used.

One factor that may critically influence trust in automation is fatigue. Fatigue is associated with increased reaction times, reduced reaction accuracy, and reduced attention levels, all of which increase the risk for a mishap. It is possible that fatigue also reduces the decision-making capability required to aptly rely on an automated agent, leading to disastrous outcomes. However, before exogenous factors like fatigue can be tested as elements influencing unmanned aircraft system (UAS) operator trust in automation, we must first establish a valid method of inducing and measuring operator trust. The purpose of this experiment is to validate a new method of manipulating trust in automation in a UAS for use in future research.

Prior research has typically manipulated trust in automation by varying the automated system’s reliability level. Generally, the more reliable the system is at performing its job, the more trust a human feels toward the system, and the more the human uses the system. The relationship between system reliance and operator trust is well documented in experiments using an automatic space shuttle cabin air management system simulation (Chavaillaz et al., 2016), an auto-correcting multi-task flight simulator (Bagheri & Jamieson, 2004), and automated target identification aids in simulated combat scenarios (Ross, 2008; Wang et al., 2009). While prior research shows a consistent positive relationship between agent reliability and human trust, there is currently no established method of systematically varying an automated agent’s reliability in the Universal Mission Simulator (UMS) to induce different levels of UAS operator trust. The UMS is the Army’s flight simulator for the Shadow and Gray Eagle UAS platforms.

For this study, we used prior research as a guide for developing our own trust manipulation in the United States Army Aeromedical Research Laboratory (USAARL) UAS simulator, the UMS at USAARL. These experiments generally have the operator complete a complex task with the help of an automated aid that performs at contrasting reliability levels. For instance, Bagheri and Jamieson (2004) used the Multi-Attribute Task Battery (MAT-B), where subjects performed a tracking, fuel management, and system-monitoring task with the help of an automated aid that assisted with the system-monitoring requirement. The reliability of the aid was manipulated to successfully alert a certain percentage of system failures (87.5% vs. 56.25%). In Chavaillaz et al.’s (2016) experiment, subjects interacted with a cabin air management system simulation modified from previous research. An automated aid assisted the operator with system faults during their four-part complex process control task. Reliability level of the aid was defined as the percentage of correct diagnoses made by the system (100% vs. 80% vs. 60%). Wiegmann et al. (2001) asked subjects to diagnose the validity of pump failures in a waste processing facility simulation with the help of a diagnostic aid that was 100%, 80%, or 60% reliable. While there would be benefits to using one of these previously validated trust manipulations in the current study, there are two main reasons for developing our own manipulation in the UMS. First, because of logistical limitations (e.g., the restrictions guarding USAARL's UMS software), we are unable to acquire and apply the exact trust manipulations validated in prior research to our simulator. Therefore, we created a new target identification aid using the built-in functions within the UMS software. Second, to increase realism and to most directly benefit the population of interest (Army UAS operators), we believe it is essential to test subjects in the most realistic UAS setting possible as opposed to testing them in a non-UAS simulator or in a less realistic UAS simulator taken from previous research. Validating a trust manipulation in USAARL's UMS will lay the foundation for future studies investigating exogenous factors that cause mishaps resulting directly from an operator's misuse or disuse of automation.

As unmanned systems have proven their worth on the battlefield, DoD has allocated an increasing percentage of its budget to developing and acquiring these systems (see: *Unmanned Systems Integrated Roadmap FY2017-2042*). Unmanned systems will continue to be critical to U.S. operations because of the capability and performance advantages these systems provide the Soldier. Additionally, these systems permit operations in riskier conditions than manned systems. Further, Future Vertical Lift (FVL) efforts are working toward the capability of single operator control of multiple UAVs simultaneously, which will require automated systems in the UAS to offset operator task load. Because of this inevitable future, we must deeply understand the nature of trust between humans and automated systems. As the DoD continues to acquire new systems, it is imperative that the science continues to test human-system interactions to achieve the highest possible level of operator performance.

Methods

Research Design

In this study, a between-subjects experimental design evaluated operator trust in an automated target identification system (TIS) developed for use in USAARL research projects using the UMS. The TIS employed two levels of reliability for accurately detecting targets, high reliability (90% accuracy, 10% false positives) and low reliability (70% accuracy, 30% false positives). The TIS provided an alert on the mission manager that a target was detected and the operator had to verify the existence of the target. False positives were when there was an alert and no target.

Subjects

Twenty-two healthy (see COVID-19 Impacts below) rated UAS operators with Military Occupational Code (MOS) 15C (MQ-1, Gray Eagle) or 15W (RQ-7, Shadow) were recruited from resident non-commissioned officer professional military education classes conducted at

Fort Rucker, AL. Volunteers were eligible for participation if they currently held either MOS 15C or 15W. Demographics information is provided in Table 1. All flights were conducted after duty hours and subjects were compensated with a $100 gift card for participation.

All methods (recruitment and experimental) were carried out in accordance with current regulations and guidelines. Prior to the execution, this study was reviewed and approved in accordance with the Medical Research Development Command policies. This project was determined to be exempt according to 32 CFR 219.104(d) (3) (i) and (ii).

COVID-19 Impacts

Planning and active recruitment was temporarily halted when the population of potential volunteers at the Fort Rucker Non-Commissioned Officer Academy (NCOA) became unavailable due to COVID-19 precautions mandated by changes in the Health Protection Posture of the Fort Rucker Garrison. The NCOA initially cancelled, and then subsequently pivoted their courses to an online instruction format. When in-person classes resumed, we coordinated COVID-19 protection measures with the NCOA to ensure subject safety and restarted recruitment actions. Subjects were pre-screened telephonically with a COVID-19 questionnaire 24 hours prior to their scheduled appointment, re-screened with a COVID-19 questionnaire upon arrival at the lab, scanned for a fever, required to wear a mask, wash their hands prior to the experiment, and answered a post experiment follow-up questionnaire 72 hours following the experimental session to check for symptoms.

Equipment

Universal Mission Simulator.

USAARL employs the Army's UMS, a computer based device that replicates the operational controls within a UAS operator shelter for the Shadow or Gray Eagle weapon systems. This device was modified with three video cameras, which simultaneously recorded operator actions and instrument indications during flight. The UMS was also modified to capture all operator keyboard input and generate a repeatable threat scenario across all subjects. Thus, the only change in the scenario between conditions was TIS accuracy.

Figure 1. Army's Universal Mission Simulator with the operator's console in the foreground.

Questionnaires

Propensity to Trust.

Before the mission, subjects completed the Adapted Propensity to Trust Questionnaire (Jessup et al., 2019), which is a 6-item measure of the propensity to trust automated agents. Example items include, “Generally, I trust automated agents,” “Automated agents are reliable.”

State Fatigue.

Before and after the mission, subjects answered a five-item measure of fatigue (the fatigue subscale) from the Profile of Mood States (POMS) questionnaire (Biehl & Landauer, 1975; McNair et al., 1971). This questionnaire asks subjects to rate how exhausted, weary, sluggish, fatigued, and worn-out they feel “right now.”

Subjective Workload.

After the mission, subjects completed the NASA-Task Load Index (Hart & Staveland, 1988) V 1.0.3 as a measure of subjective workload via an iPad App.

Self-Reported Trust Measures.

After the mission, subjects were asked to complete the Checklist of Trust between People and Automation (CTPA) questionnaire (Jian et al., 2000), and the Human Computer Trust (HCT) questionnaire (Madsen & Gregor, 2000). Data from these instruments provided a measure of subjective trust in the TIS.

Post-Mission Questionnaire.

After the trust questionnaires, subjects answered several questions about their general experience with TIS, such as “What did you like about the TIS?”/“What did you dislike about the TIS?,” etc. These data allowed the researchers to perform a text analysis on trust-related language.

Demographic Questionnaire.

After the trust questionnaires, subjects were asked to complete a demographic questionnaire with age, rank, gender, months as a UAS operator, and total flight time.

Procedure

All interested persons who responded to the recruitment efforts were instructed to report to the laboratory at a time convenient to the volunteer and the research staff. All research periods were held after the volunteer’s normal duty day. Prior to participating in the study, volunteers were briefed by a member of the research team on the study procedures in a quiet conference room. They were then given an unlimited amount of time to read and review the informed consent document; volunteers were provided a copy of the informed consent document for their records. After providing informed consent, subjects completed the pre-measure of the Propensity to Trust Scale and the State Fatigue surveys and were assigned to either the high or the low system reliability condition. Condition assignment was counter-balanced.

Next subjects were provided a mission briefing (Appendix B) following the Army's mission, enemy, troops, terrain, time, and civilian concerns (METT-TC) format. The primary mission was searching for a high-value target in a civilian high foot-traffic area. The secondary mission was to test the new TIS. Subjects were briefed that the TIS would identify threat armored personnel carriers, providing them a grid location. They were asked to verify the presence of threat vehicles and provide a spot report either confirming or denying the presence of the threat armored personnel carriers, prior to the TIS alert disappearing from their mission manager. The subject's use of the TIS was recorded via video and screen captures, and was quantified by the investigators (e.g., number of times the subject clicked on the TIS's suggested target). The subjects were briefed that other assets were managing the aircraft and they were only tasked with the control of the payload and the mission as briefed.

Once seated, the eye-tracking cameras were calibrated, and then the mission started when the operator assumed control of an MQ-1 Gray Eagle loitering at 10,000 feet above ground level (AGL) of the mission area of interest. Mission duration was 60 minutes. After completing the mission, the Subjective Workload, Self-Reported Trust Measures, Post-Mission Questionnaire, and Demographic Questionnaire were administered.

Results

Demographics

Twenty-two healthy, rated UAS operators (18 male) with MOS 15C or 15W participated in the study. This resulted in a normally distributed sample of 11 subjects in each condition. The demographic summary is provided in Table 1. There were no statistically significant differences between the groups in age, $t(20) = .17$, $p = .87$; rank, $t(20) = .49$, $p = .63$; gender, $t(20) = 1.1$, $p = .29$; months experience, $t(20) = .16$, $p = .88$; and total hours, $t(20) = .15$, $p = .88$.

Table 1. Study Group Demographic Summary

Demographic Element	**Combined Groups**	**90% Accuracy Group**	**70% Accuracy Group**
Age in years *(Mean/SD)*	35.6/5.5	28.5/3.4	28.1/6.2
Rank *(SGT/SSG)*	17/5	9/2	8/3
Gender *(M/F)*	18/4	8/3	10/1
Months Experience *(Mean/SD)*	73.7/26.0	72.8/34.0	77.4/14.1
Total Hours	**878/577**	**897/742**	**860/386**
Note: No statistical difference found between High Accuracy and Low Accuracy groups, p≤.05 (2-tailed)			

Propensity to Trust

There were no statistical differences in pre-existing attitudes to system trust between the groups based on the analysis of the a priori administered propensity to trust scale, $t(20) = .17$, $p = .86$.

Checklist of Trust between People and Automation (CPTA)

Analysis of the CPTA mean values showed no difference in the system trust between the groups, $t(20) = 3.0$, $p = .77$.

Human Computer Trust Questionnaire (HCT)

Analysis of the HCT mean values showed no difference in the system trust between the groups, $t(20) = .24$, $p = .82$.

Trust Manipulation Findings

To assess how system accuracy affects trust in automation, we first performed a manipulation check using independent samples t-test (one-tailed). Specifically, we asked subjects to report the extent to which: "The system analyzes problems consistently." Subjects in the 90% accuracy condition reported that their system analyzed problems more consistently than subjects in the 70% accuracy condition (mean difference = .91, $p < .05$).

To evaluate the extent to which the subjects trusted the automated system, we asked them to assess the extent to which: "When I am uncertain about a decision, I believe the system rather than myself." Independent sample t-tests indicated that, when uncertain, subjects in the 90% accuracy condition believed the system more than subjects in the 70% accuracy condition (mean difference = 1.36, $p < .05$).

Influence of Fatigue on Trust

This study was designed as a precursor to a follow-on study that will examine whether the impact of system accuracy on trust-in-automation differed according to experienced levels of fatigue. Therefore, fatigue questions were included and their effects examined. To accomplish this, we first developed a three-item measure of fatigue from subjects' answers to three post-study questions. Specifically, subjects were asked on a 0-4 scale to report on the extent to which they felt "Fatigued," "Sluggish," and "Worn-out." The Cronbach's alpha of this three-item measure was .86. To develop meaningful indicators of fatigue, we collapsed raw fatigue scores into three categories ("Low" = Aggregate fatigue score = 0-1; "Medium" = Aggregate fatigue score = 1-2; "High" = Aggregate fatigue score = 3-4).

We then employed a univariate general linear model (GLM) analysis where we entered condition and fatigue levels as fixed effects (Table 2 & Figure 2). We also added the number of months of simulator experience that subjects reported as a covariate. In our model, we created an interaction term by combining the study condition to which each subject was assigned with their reported level of fatigue.

The results we obtained suggest that the level of fatigue subjects experienced in the study moderated the effects of system accuracy on trust in the automated system ($F = 3.8$; $p < .05$). An inspection of the plot of these values reveals that, across fatigue levels, subjects in the 90% accuracy condition trusted the automated system more than subjects in the 70% accuracy condition (mean difference = 2.27, $p < .01$). The most marked difference was among subjects exhibiting 'high" levels of fatigue (mean difference = 5.02), although subjects experiencing "medium" (mean difference = 1.38) and "low" (mean difference = .41) levels of fatigue also exhibited differences that were consistent with this pattern.

Table 2. Assessing Trust -in-Automation Across Study Conditions and Fatigue Levels

Managerial Control Condition	Trust in Automated System Model 1	Model 2
Intercept	6.5*	3.7+
Simulator Experience	0.5	2.2
Condition	4.1+	12**
Fatigue Level	0.9	0.8
Condition X Fatigue Level	N/A	3.8*
F values reported; +p<.10; *p<.05; **p<.01 (two tailed)		

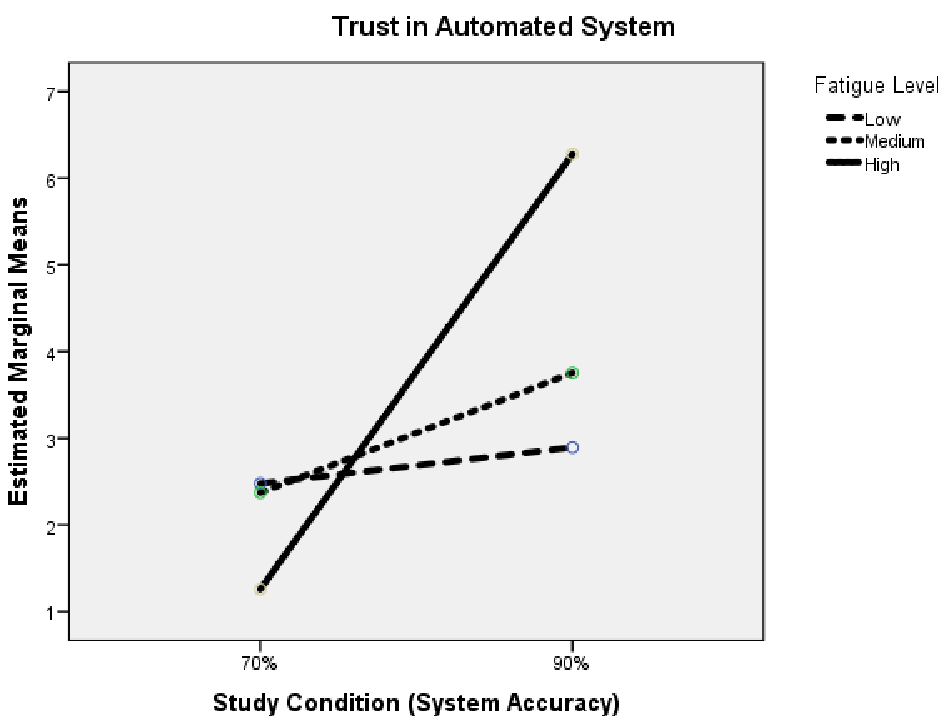

Figure 2. Estimated marginal means of trust in automated system by study condition and fatigue level.

Discussion & Recommendations

For this study, we evaluated a trust in automation manipulation for use in future research on the effects of fatigue on system trust. This experiment was designed to validate a novel TIS method for manipulating operator trust using an operationally relevant device, the UMS. We used this system to compare how high and low reliability systems produce significant and meaningful differences in system trust and its use by the subjects. Additionally, the experimental results suggest that the level of fatigue subjects reported moderated the effects of system accuracy on trust in the automation. Moreover, the trends showed that the fatigued subjects trusted the high accuracy system, and distrusted the low accuracy system, more than the other groups.

By testing these relationships in a controlled experiment, this study supports the general proposition that system accuracy is a key factor that individuals weigh when determining the extent to which they trust an automated system. A key finding of this study is that operators' fatigue levels moderate this effect in important ways. Specifically, the more fatigued that operators were, the more that they relied on system accuracy when determining the extent to which they trusted the automated system.

In contrast to some prior research that reports how individuals' fatigue levels encourage them to increasingly depend on automated systems, the findings from this study align with more targeted work that details how increased levels of fatigue actually heighten operator sensitivity to system failures. For example, Reichenbach et al. (2011)'s examination of how individuals employ decision-aids found that operators who were sleep-deprived were more sensitive to failures in their decisions to employ automation. In addition, Wohleber et al. (2019) found that as subjects in their multi-UAS simulation study fatigued over time, they tended to become more sensitive to system reliability. They specifically observed that in lower reliability conditions, subjects became less dependent on automation as their experiment progressed and fatigue levels rose.

Work by Banks et al. (2014) and Wohleber et al. (2019) suggest that this phenomenon may occur when actors view monitoring an automated system as an additional sub-task that they must manage while performing their primary tasks. As fatigue levels increase, individuals in these situations become more motivated to economize their efforts. Consistent with effort-regulation dynamics, this can lead operators to devote fewer cognitive resources to monitoring the automated system in order to devote more cognitive energies to their own decision-making (i.e., without the aid of the automated system) (Hockey, 2017; Sauer et al., 2003). This can create problems for operators, especially in situations where relying on an automated system could increase their decision accuracy and overall effectiveness.

These perspectives suggest that it is important for designers to construct automated systems that motivate individuals across fatigue levels to display an appropriate reliance on and trust in a given technology. To accomplish this, operators must perceive that a given technology removes some of their workload burden while placing few (if any) additional cognitive demands on them. Future research is needed to determine exactly what types of tasks and workload levels might increase individuals' motivations to trust a technology enough that they appropriately off-load task functions to a given automated system.

Study Fatigue

As noted above, all subjects participated in this study in the evenings after their scheduled duty day. To place this in context, the duty day for these individuals began around 0530 for physical training. Followed by a full day of in-person training at the Non-Commissioned Officer Academy. Most subjects started the informed consent process at USAARL at about 1800, after a 12+ hour duty day. To evaluate the impact of being well rested, we attempted to recruit subjects to participate during their off-duty days; however, these students only had one day of no scheduled activities per week and no one was willing to participate in this study on those days.

Study Limitations

Changes to the health protection posture within USAARL, NCOA, and at Ft. Rucker prevented the total recruitment effort of N=34 for this project. Historical data used to complete the power analysis indicated that similar manipulations resulted in a large effect size of Cohen's $d = 1.0$, however the current study yielded an effect size much lower, less than 0.2. Further research is necessary to determine how to modify the trust manipulation in order to yield a larger effect size.

References

Bagheri, N., & Jamieson, G. A. (2004). A sampling model to ascertain automation-induced complacency in multi-task environments. In C. W. Johnson & P. Palanque (Eds.), *Human Error, Safety and Systems Development* (pp. 131-145). Springer.

Bailey, N., & Scerbo, M. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science, 8,* 321-348. https://doi.org/10.1080/14639220500535301

Banks, V., Stanton, N., & Harvey, C. (2014). Sub-systems on the road to vehicle automation: Hands and feet free but not 'mind' free driving. *Safety Science, 62*, 505-514. https://doi.org/10.1016/j.ssci.2013.10.014

Biehl, B., & Landauer, A. (1975). Das profile of mood states (POMS). *Mannheim (Unveröffentlichtes Manuskript).*

Bliss, J., & Acton, S. (2003). Alarm mistrust in automobiles: How collision alarm reliability affects driving. *Applied Ergonomics, 34*, 499-509. https://doi.org/10.1016/j.apergo.2003.07.003

Cahour, B., & Forzy, J.-F. (2009). Does projection into use improve trust and exploration? An example with a cruise control system. *Safety Science*, *47*(9), 1260-1270. https://hal.archives-ouvertes.fr/hal-00471270

Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. *Applied Ergonomics, 52,* 333-342. https://doi.org/10.1016/j.apergo.2015.07.012

Cummings, M., Mastracchio, C., Thornburg, K. M., & Mkrtchyan, A. (2014). Boredom and distraction in multiple unmanned vehicle supervisory control. *Interacting with Computers, 25*. https://doi.org/10.1093/iwc/iws011

Cummings, M. L., Clare, A., & Hart, C. (2010). The Role of human-automation consensus in multiple unmanned vehicle scheduling. *Human Factors, 52*(1), 17-27. https://doi.org/10.1177/0018720810368674

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies*, *58*(6), 719-735. https://doi.org/https://doi.org/10.1016/S1071-5819(03)00039-9

Donmez, B., Boyle, L. N., & Lee, J. D. (2006). The impact of eistraction mitigation strategies on driving performance. *Human Factors, 48*(4), 785-804. https://doi.org/10.1518/001872006779166415

Hancock, P., Billings, D., Schaefer, K., Chen, J., de Visser, E., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors, 53,* 517-527. https://doi.org/10.1177/0018720811417254

Hart, S., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. North Holland Press.

Ho, G., Wheatley, D., & Scialfa, C. T. (2005). Age differences in trust and reliance of a medication management system. *Interacting with Computers, 17*(6), 690-710. https://doi.org/10.1016/j.intcom.2005.09.007

Hockey, B. (2017). In G. Matthews, P. A. Hancock, & P. A. Desmond (Eds.), *The Handbook of Operator Fatigue* (pp. 45-60). Ashagte.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407-434. https://doi.org/10.1177/0018720814547570

Hoffman, R., Hawley, J., & Bradshaw, J. (2014). Myths of automation, Part 2: Some very human consequences. *Intelligent Systems, IEEE, 29*, 82-85. https://doi.org/10.1109/MIS.2014.25

Jessup, S., Schneider, T., Alarcon, G., Ryan, T., & Capiola, A. (2019). The Measurement of the Propensity to Trust Automation. In J. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality. Applications and Case Studies. HCII 2019. Lecture Notes in Computer Science, vol 11575.* (pp. 476-489). Springer. https://doi.org/10.1007/978-3-030-21565-1_32

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

Kazi, T. A., Stanton, N. A., Walker, G. H., & Young, M. S. (2007). Designer driving: drivers' conceptual models and level of trust in adaptive cruise control. *International Journal of Vehicle Design, 45*(3), 339-360. https://doi.org/10.1504/ijvd.2007.014909

Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors, 46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Madhavan, P., & Wiegmann, D. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science, 8*, 277-301. https://doi.org/10.1080/14639220500337708

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. Proceedings of the 11 th Australasian Conference on Information Systems.

Master, R., Jiang, X., Khasawneh, M. T., Bowling, S. R., Grimes, L., Gramopadhye, A. K., & Melloy, B. J. (2005). Measurement of trust over time in hybrid inspection systems. *Human Factors and Ergonomics in Manufacturing & Service Industries, 15*(2), 177-196. https://doi.org/https://doi.org/10.1002/hfm.20021

McLeod, R. W., Walker, G. H., & Moray, N. (2005). Analysing and modelling train driver performance. *Applied Ergonomics, 36*(6), 671-680. https://doi.org/https://doi.org/10.1016/j.apergo.2005.05.006

McNair, D., Lorr, M., & Droppleman, L. (1971). *Manual for the Profile of Mood States.* Educational and Industrial Service.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology Applied, 6*(1), 44-58. https://doi.org/10.1037//1076-898x.6.1.44

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

Reichenbach, J., Onnasch, L., & Manzey, D. (2011). Human performance consequences of automated decision aids in states of sleep loss. *Human Factors, 53*(6), 717-728. https://doi.org/10.1177/0018720811418222

Ross, J. M. (2008). *Moderators Of Trust And Reliance Across Multiple Decision Aids* [Unpublished Doctoral Dissertation]. University of Central Florida.

Sauer, J., Wastell, D. G., Hockey, G. R., & Earle, F. (2003). Performance in a complex multiple-task environment during a laboratory-based simulation of occasional night work. *Human Factors, 45*(4), 657-669. https://doi.org/10.1518/hfes.45.4.657.27090

Schaefer, K., Chen, J., Szalma, J., & Hancock, P. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 58*. https://doi.org/10.1177/0018720816634228

Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 58*(3), 377-400. https://doi.org/10.1177/0018720816634228

Stedmon, A. W., Sharples, S., Littlewood, R., Cox, G., Patel, H., & Wilson, J. R. (2007). Datalink in air traffic management: Human factors issues in communications. *Applied Ergonomics, 38*(4), 473-480. https://doi.org/10.1016/j.apergo.2007.01.013

United States Defense Science Board. (2012). *Task force report: The role of autonomy in DoD systems*. Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. https://books.google.com/books?id=2fFLAQAACAAJ

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors, 51*(3), 281-291. https://doi.org/10.1177/0018720809338842

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a “cry wolf” effect? *Human Factors, 51*(4), 446-462. https://doi.org/10.1177/0018720809344720

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, *2*(4), 352-367. https://doi.org/10.1080/14639220110110306

Wohleber, R. W., Matthews, G., Lin, J., Szalma, J. L., Calhoun, G. L., Funke, G. J., Chiu, C.-Y. P., & Ruff, H. A. (2019). Vigilance and automation dependence in operation of multiple unmanned aerial systems (UAS): A simulation study. *Human Factors*, *61*(3), 488-505. https://doi.org/10.1177/0018720818799468

Yamada, K., & Kuchar, J. K. (2006). Preliminary study of behavioral and safety effects of driver dependence on a warning system in a driving simulator. *IEEE Transactions on Systems, Man, and Cybernetics Part A, 36*(3), 602-610. https://doi.org/10.1109/TSMCA.2006.871646

Appendix A. Acronyms and Abbreviations

AGL	above ground level
CTPA	Checklist of Trust between People and Automation
DoD	Department of Defense
FVL	Future Vertical Lift
HCT	Human Computer Trust
M	male
MAT-B	Multi-attribute Task Battery
METT-TC	mission, enemy, terrain, troops, time, civil considerations
MOS	Military Occupational Specialty
NCOA	Noncommissioned Officers Academy
POMS	Profile of Mood States
SD	standard deviation
TIS	target identification system
UAS	unmanned aircraft system
UAV	unmanned aerial vehicle
UMS	universal mission simulator
USAARL	U.S. Army Aeromedical Research Laboratory

Appendix B. USAARL Mission Brief

USAARL MISSION BRIEF

Universal Mission Simulator

Juan Colón, UAS Research Operator, juan.a.coloncruz.ctr@mail.mil

21-08-2019

Version #: 2
Date: 28 August 2020

UNCLASSIFIED

A/C, Crew, and Flight Condition

» Aircraft information – Gray Eagle 70118; Maintenance Status – Green; Fuel – 575 lbs; Ramp weight – 3200 lbs

» Aircraft Commander – You will be the only operator in the UGCS. An asset manager will assist with all mission parameters. The aircraft will be remotely monitored by an Aircraft Commander from a centralized location. This is to evaluate if operators can effectively monitor multiple aircrafts on different missions. Your priority is the completion of this mission.

» Flight Condition - Day

Juan Colón, UAS Research Operator UNCLASSIFIED 2

METT-TC

- Mission – Flight will be conducted out of Gray Army Airfield, Fort Hood, Texas. Area Reconnaissance will be performed on Area Alpha with the intent on identifying the High Value Target (HVT) and to record all enemy activity. Due to an upcoming mission in the target area the aircraft will be fitted with the Target Identification System (TIS) to help locate enemy targets. You will have to confirm the enemy target positions with the payload. Once the TIS has identified the enemy targets you will have 15 seconds before the target information is deleted from the CUCS map and Mission Manager. The TIS is designed this way for evaluation purposes.

- Enemy – All enemy personnel will be military aged males and their uniform ranges from multi-cam to all black as well as masks ranging from black to gray in color. The HVT will have black attire (no sleeves), with a gray mask. Reports indicate up to 50 enemy personnel in Area Alpha as well as enemy vehicles in the vicinity of target area.

Juan Colón, UAS Research Operator UNCLASSIFIED 3

METT-TC (cont.)

- Terrain and Weather – No issues/hazards with local terrain. Current METAR - 08004KT 10SM BKN200 32/22 A3006

- Time Available - +1:00

- Troops – None

- Civilian Considerations –There will be foot traffic in and out of multiple buildings and the majority of civilians will be centered around the market area.

Juan Colón, UAS Research Operator UNCLASSIFIED 4

High Value Target

Juan Colón, UAS Research Operator UNCLASSIFIED 5

Enemy Targets

Juan Colón, UAS Research Operator UNCLASSIFIED 6

CCIRs

- Stay in Area Alpha unless you are verifying TIS targets. Do not follow vehicles leaving the Area Alpha.
- You will have to visually identify the targets that have been located by the TIS.
- You will be the only asset on this mission. There are no friendlies in Area Alpha at this time.
- Reporting –
  - While in Area Alpha the focus will be on foot traffic. Capture as much movement as possible. The report given to the asset manager will be in the SALT format: “1 person (state enemy or civilian), walking toward market area, coming from the Mosque.”
  - Enemy Armored Personnel Carrier (APC) – “Enemy APC located, grid location 14R PV 14500 25000.”

Mandatory Brief Items

- Please leave the crew station displays alone. Do not rearrange any menus.
- Do not add any marked targets on the mission manager.

Risk Assessment

LOW

Juan Colón, UAS Research Operator UNCLASSIFIED 9

QUESTIONS?

?

Juan Colón, UAS Research Operator UNCLASSIFIED 10

U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama

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

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

Army Futures Command
U.S. Army Medical Research and Development Command