



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**EFFECTS OF STRESS AND MOTIVATION ON
HUMAN RELIANCE ON AUTOMATED
DECISION-SUPPORT SYSTEMS**

by

Miroslav Bernkopf

June 2021

Thesis Advisor:
Second Reader:

Mollie R. McGuire
Glenn R. Cook

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE June 2021	3. REPORT TYPE AND DATES COVERED Master's thesis		
4. TITLE AND SUBTITLE EFFECTS OF STRESS AND MOTIVATION ON HUMAN RELIANCE ON AUTOMATED DECISION-SUPPORT SYSTEMS			5. FUNDING NUMBERS	
6. AUTHOR(S) Miroslav Bernkopf				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			12b. DISTRIBUTION CODE A	
13. ABSTRACT (maximum 200 words) Advancements in technology have accelerated the utilization of unmanned systems, which has led to an increased adoption rate of military human-machine teams. This study explored human reliance on automation in the context of stress, motivation, and automation reliability. A laboratory experiment was conducted with participants who were exposed to either a high- or low-stress condition and motivation or non-motivation condition. The participants executed a series of tasks alongside two automated decision aids that differed in reliability, and their decision on whether or not to rely on the decision aids was the measure of interest. The results indicate that the aid's reliability was the primary driver for reliance decisions such that the participants relied on the high-reliability aid more than on the low-reliability aid. The aid's reliability also affected the participants' reaction times, such that they made decisions with the high-reliability aid faster than with the low-reliability aid. Finally, motivation had a significant effect on reliance decisions of the male participants such that in the stress condition, and when they were motivated, males relied on the decision aids more than when they were not motivated. More research is needed to fully understand the effects investigated in this experiment, and future studies should include a larger female population to allow for exploration of gender differences.				
14. SUBJECT TERMS automation, reliance, stress, decision making, decision support, artificial intelligence, motivation, cortisol, saliva			15. NUMBER OF PAGES 105	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**EFFECTS OF STRESS AND MOTIVATION ON HUMAN RELIANCE ON
AUTOMATED DECISION-SUPPORT SYSTEMS**

Miroslav Bernkopf
Captain, United States Marine Corps
B, Brno University of Technology, 2006
M, Brno University of Technology, 2008

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

**NAVAL POSTGRADUATE SCHOOL
June 2021**

Approved by: Mollie R. McGuire
Advisor

Glenn R. Cook
Second Reader

Alex Bordetsky
Chair, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Advancements in technology have accelerated the utilization of unmanned systems, which has led to an increased adoption rate of military human-machine teams. This study explored human reliance on automation in the context of stress, motivation, and automation reliability. A laboratory experiment was conducted with participants who were exposed to either a high- or low-stress condition and motivation or non-motivation condition. The participants executed a series of tasks alongside two automated decision aids that differed in reliability, and their decision on whether or not to rely on the decision aids was the measure of interest. The results indicate that the aid's reliability was the primary driver for reliance decisions such that the participants relied on the high-reliability aid more than on the low-reliability aid. The aid's reliability also affected the participants' reaction times, such that they made decisions with the high-reliability aid faster than with the low-reliability aid. Finally, motivation had a significant effect on reliance decisions of the male participants such that in the stress condition, and when they were motivated, males relied on the decision aids more than when they were not motivated. More research is needed to fully understand the effects investigated in this experiment, and future studies should include a larger female population to allow for exploration of gender differences.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	OVERVIEW OF THE STUDY	1
B.	PROBLEM STATEMENT	3
C.	PURPOSE STATEMENT	4
D.	HYPOTHESES	4
E.	RESEARCH METHOD AND ANALYSIS	6
F.	POTENTIAL BENEFITS OF THE STUDY	6
G.	THESIS ORGANIZATION.....	7
II.	LITERATURE REVIEW	9
A.	AUTOMATION, AUTONOMY, AND ARTIFICIAL INTELLIGENCE.....	9
B.	HUMAN-MACHINE TEAMING	11
1.	Revolution in Military Affairs	11
2.	History of Human-Machine Teaming	12
3.	Employment and Benefits of Human-Machine Teams.....	14
C.	EFFECT OF TRUST AND RELIABILITY ON RELIANCE	16
1.	Trust.....	19
2.	Automation Reliability	20
D.	DECISION MAKING	23
1.	Memory and Information Processing	24
2.	Learning and Feedback.....	25
3.	Impact of Motivation	26
4.	Types of Decision Making	27
E.	EFFECT OF STRESS ON DECISION MAKING	28
1.	Stress—Definition and Response.....	28
2.	Impact of Stress on Cognition.....	30
3.	Impact of Stress on Decision Making.....	30
F.	SUMMARY	31
III.	METHODS	33
A.	PARTICIPANTS AND DESIGN	33
1.	Participants.....	33
2.	Experimental Design.....	35
B.	STRESS MANIPULATION	36
1.	Trier Social Stress Test.....	36
2.	Cortisol as a Quantifier of Stress.....	38

C.	MOTIVATION MANIPULATION	39
D.	MATERIALS	40
1.	Empatica E4 Wristband	40
2.	Tobii Pro X3-120 Eye Tracker.....	40
3.	Presentation Software.....	42
4.	Laboratory Configuration.....	43
5.	Pattern Learning Task	45
6.	Exit Questionnaire	47
E.	PROCEDURE	47
1.	Experiment	48
2.	Data and Collection Method	50
IV.	EXPERIMENT RESULTS	51
A.	STRESS MANIPULATION CHECK.....	51
1.	Cortisol Manipulation	51
2.	STAI Manipulation	53
B.	RELIANCE ANALYSIS	53
C.	REACTION TIMES ANALYSIS.....	54
D.	EXPLORATORY ANALYSES	56
V.	DISCUSSION	59
A.	FINDINGS DISCUSSION	59
1.	Reliance.....	59
2.	Reaction Times	60
3.	Stress	61
B.	LIMITATIONS	63
C.	RECOMMENDATIONS.....	63
D.	CONCLUSION	64
	APPENDIX A. TSST PROTOCOL INSTRUCTIONS.....	67
	APPENDIX B. STATE-TRAIT ANXIETY INVENTORY	69
	APPENDIX C. COGNITIVE REFLECTION TEST INSTRUCTIONS	71
	APPENDIX D. EXIT QUESTIONNAIRE	73
	APPENDIX E. OPERATION SPAN TASK.....	75
	LIST OF REFERENCES	77
	INITIAL DISTRIBUTION LIST	87

LIST OF FIGURES

Figure 1.	Visual representation of hypotheses H1, H2, and H3	5
Figure 2.	Visual representation of hypothesis H4	5
Figure 3.	A simplified model of a human-machine system. Source: Salvendy (2012).	14
Figure 4.	Interactions between factors influencing automation use. Source: Parasuraman and Riley (1997).	17
Figure 5.	Framework of reliance decision making. Source: van Dongen and van Maanen (2013).	18
Figure 6.	The relationship between subjective trust and automation reliability. Source: Wickens et al. (2013).	22
Figure 7.	Distribution of participants by age and gender	33
Figure 8.	Breakdown of participants by service	34
Figure 9.	Highest achieved education	35
Figure 10.	Experimental design conditions	35
Figure 11.	The TSST timeline. Adapted from Allen et al. (2017).	36
Figure 12.	A cryovial and a collection box used for saliva samples	39
Figure 13.	Empatica E4 heart rate monitor	40
Figure 14.	Tobii Pro X3-120 eye tracker with the external processing unit	41
Figure 15.	Presentation Software scenario editor	42
Figure 16.	Room A, Area A—rest area	43
Figure 17.	Room A, Area B—the participants’ workstation	44
Figure 18.	Room B	45
Figure 19.	An example of one pattern learning task iteration	47
Figure 20.	Cortisol level per sample between the stress conditions	52
Figure 21.	Effect of motivation on male reliance under stress	57
Figure 22.	Yerkes-Dodson law. Source: Arnsten (2009).	62

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	Level of autonomy assessment scale. Source: Proud et al. (2003).	11
Table 2.	Descriptive statistics for cortisol samples between stress conditions	52
Table 3.	Reliance descriptive statistics for all participants	54
Table 4.	AI1 and AI2 reaction times	54
Table 5.	Agreement and disagreement reaction times	55
Table 6.	Reaction times per AI partner between reliance conditions	55
Table 7.	Descriptive statistics for male participants only	57

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACTH	adrenocorticotropin
AI	artificial intelligence
ALEX	Agile Learning for Enhanced Experimentation
ANOVA	analysis of variance
CRT	cognitive reflection test
DOD	United States Department of Defense
F	f-statistic
HMT(s)	human-machine team(s)
HPA	hypothalamic-pituitary-adrenocortical
IBM	International Business Machines
M	mean
η_p^2	partial eta squared
NASA	National Aeronautics and Space Administration
NPS	Naval Postgraduate School
OSPAN	Operation Span Task
p	p-value
PFC	prefrontal cortex
pTSST	placebo version of the Trier Social Stress Test
RMA	revolution in military affairs
SAML	Systematic Algorithm for Machine Learning
SD	standard deviation
SPSS	Statistical Package for Social Sciences
STAI	State-Trait Anxiety Inventory
t	t-statistic
TSST	Trier Social Stress Test
UAV(s)	unmanned aerial vehicle(s)
U.S.	United States of America

USA	United State Army
USAF	United States Air Force
USMC	United States Marine Corps
USN	United States Navy

ACKNOWLEDGMENTS

It is an emotional moment for me to be typing these words. My journey started in Czechoslovakia, during the darkest time periods Europe has ever experienced—communism. From there, my personal goals, dreams, and desires have led me here—to the greatest country on Earth. Through hard work, a little bit of luck, and help from my loved ones, I reached my goal of becoming a United States Marine. My career gives me a purpose in life and an immeasurable amount of pride every day. The Marine Corps has provided me with this incredible opportunity to study at the Naval Postgraduate School and receive some of the world's best education. The Marines also allowed me to extend my time in Monterey and earn a second degree from the National University of Singapore. Therefore, my most immense gratitude goes to all members of this remarkable organization that continues to defend our way of life and those who cannot defend themselves.

I would also like to thank my thesis advisor, Dr. Mollie McGuire. Her impressive drive, academic skills, and everyday engagement were crucial to the experiment's success and allowed me to complete this thesis on time. Her patience with me during the conduct of the experiment was remarkable. Without her guidance, mentorship, and contributions, I would not have produced such a high-quality thesis.

Lastly, but most importantly, I want to thank my wife, Iva, for her patience, support, and perseverance during my three years in Monterey. Iva has been a wonderful mother to our three children and has provided them with all the care and love they needed, which allowed me to focus entirely on my career. From the bottom of my heart, I thank you.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW OF THE STUDY

Robots and unmanned systems that can provide autonomous actions or aid in decision making have been accompanying military forces for more than a century. The first remote-controlled system to deliver an explosive payload appeared at the end of World War I (McFadden, 2018). The proliferation of such systems increased significantly after World War II. This trend was further accelerated by advancements in science, physics, and modern technology. The rise of asymmetric warfare and modern-day technologies' incredible affordability led to warfare becoming less about the human warrior and more about the machine. Long gone are the days of armies full of uniformed soldiers facing one another on a battlefield. While conflicts in the twentieth century were mostly defined by struggles among large regional players and superpowers, conflicts in the twenty-first century will most likely be driven by rogue nations, revisionist powers, and failed states (Singer, 2009). These actors will rely on the concept of asymmetric, hybrid warfare and keep conflicts below the threshold of open hostilities. Due to their limited manpower and economic resources, these bad actors will utilize cheap unmanned technologies to disrupt, deny, degrade, and destroy their enemies' forces and infrastructure. Therefore, the United States of America (U.S.) needs to equip its forces with the same technology to counter its adversaries.

The U.S. Department of Defense (DOD) has been deploying humans and machines together in so-called human-machine teams (HMTs) for decades (McFadden, 2018). Humans and machines complement each other in many ways. However, certain things that humans execute with ease (e.g., making a joke) are almost impossible for machines, and what is easy for machines (e.g., complex computational tasks) continues to be very challenging for humans (Wilson & Daugherty, 2018). These complementing features, unique to HMTs, led to a steady increase in the number of HMTs deployed within the U.S. military. Some of the benefits of utilizing HMTs within the DOD are higher effectiveness, better mission design, reduced operational costs, increased adaptability, and accelerated adoption rate (Air Force Research Laboratory, 2015). Therefore, it is imperative for the

DOD to thoroughly understand how humans work with, interact, and respond to autonomous systems designed to aid them and what factors drive specific human behavior toward autonomous systems.

The existing literature on human-machine teaming, reliability, reliance, and decision making is extensive and provides a solid foundation for the thesis. Specifically, human interaction and cooperation with a machine have been studied broadly. However, the intangible aspects and the driving factors behind human-machine interaction, such as trust or reliance, are not entirely understood, yet they are critical to the human-machine dynamic. Although similar, reliance and trust carry slightly different meanings. According to Lee and See (2004), trust is “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” (p. 51). While trust is an attitude, reliance is a behavior that is affected by trust (Lee & See, 2004). Reliance describes the level of dependence of one agent on another (Rice & Keller, 2009).

On the topic of reliance, Van Dongen and van Maanen (2013) proposed a framework to explain reliance on decision aids and postulated that “reliance is determined by the difference between a decision maker’s trust in a decision aid and the confidence he has in his own performance” (p. 411). In the same study, however, the two authors also argued that trust by itself is not the only determining factor. Instead, more variables exist that affect human reliance on a decision aid. As Rice and Keller (2009) demonstrated, under time pressure, trust has little to no impact on human reliance on automation. The authors proposed that stress was an additional factor affecting human reliance on automation.

The effect of stress on human performance has also been researched extensively. Sandi (2013) concluded that high-stress conditions negatively impact an operator’s performance because they limit the operator’s cognitive abilities. This impairment of cognitive abilities may decrease the perceived reliability of own performance and increase reliance on an automated decision aid. If stress does affect human reliance on automation, it is essential to understand how, because reliance is one of the critical components of successful human-machine teaming and thus the focus of this study. Therefore, this study explored if stress-induced cognitive limitations could lead to an overall increase in reliance on automated decision aids.

When studying human reliance decisions, it is critical to consider automation reliability as well. van Dongen and van Maanen (2006) argued that human reliance on an automated decision aid depends on the difference between one's perception of the aid's reliability and their own performance. Using a laboratory experiment, the authors found that humans were less forgiving toward machines than themselves. In other words, human trust in themselves recovered faster than their trust in a machine making errors. Because of these findings, the experiment conducted as a part of this thesis implemented two levels of reliability into the two automated decision aids with which the study participants interacted.

Another important factor to consider is motivation, since motivation affects performance (Sansone & Harackiewicz, 2000). Therefore, in addition to stress and reliability, this study implemented the element of motivation to see if motivation might have any mitigating effects on the effect of stress. As defined by Ryan et al. (1983), performance-contingent rewards are tied to a specific task and designed to stimulate one's motivation in the performance of the task. Adding incentive-based performance as a condition to this thesis' experiment allowed for the exploration into the effect motivation might have on human reliance on automation.

In summary, the current study explored human reliance on automated decision aids in the context of stress and motivation, along with varying levels of the aid's reliability. Specifically, the study participants worked with two decision aids (one with high and one with low reliability) after having undergone a high- or low-stress manipulation. Additionally, half of the participants were given the motivation condition through explicit instructions and accounting of performance-contingent rewards, while the other half were not aware that they were earning rewards for their performance.

B. PROBLEM STATEMENT

Advancements in technology and science have accelerated the utilization of unmanned systems, which has led to the large-scale deployment of HMTs by military forces around the world. According to Sandi (2013), high-stress conditions have a deleterious effect on human cognitive abilities. This finding makes stress an important concept to understand when considering critical factors affecting the performance of

military HMTs. Specifically, stress may impact the human reliance on the machine in an HMT. This is a problem because, as Rice and Keller (2009) stated, humans tend to rely on automated decision aids more under pressure, regardless of the automated system's reliability. This tendency may lead to catastrophic results when the automated system's reliability level is low.

C. PURPOSE STATEMENT

The purpose of this study was to explore (a) the effect of stress on human reliance on an automated decision aid; (b) the effect of the aid's reliability on the level of human reliance on the given aid in both high- and low-stress conditions; and (c) the effect of motivation on human reliance on automation. Findings from this study will contribute to the overall body of knowledge on human-machine teaming. Specifically, these findings will help the DOD expand its understanding of the dynamics that exist within an HMT. The study may also expose potential risks associated with how HMTs make decisions in stressful situations in which the human cognition abilities are impaired. This knowledge may then be used to further improve the design, efficiency, and effectiveness of the U.S. military's autonomous systems. Additionally, this study's findings may help advance the training of service members on the interaction with and operation of autonomous systems under stress.

D. HYPOTHESES

This study tested the following hypotheses:

- H1: Human reliance on automated decision aids is higher in the high-stress than low-stress condition, regardless of the system's reliability level.
- H2: In the low-stress condition, it is predicted that human reliance on automated decision aids will be higher in the high-reliability than low-reliability condition.

- H3: In the high-stress condition, no difference in human reliance on automated decision aids is predicted between high and low-reliability conditions.
- H4: The effect of motivation is predicted to lessen the impact stress has on human reliance on automation.

Figure 1 and Figure 2 help in visualizing the four hypotheses using a two-by-two matrix. Stress, motivation, and automated decision aid's reliability was the independent variables, while human reliance on said aid was the dependent variable.

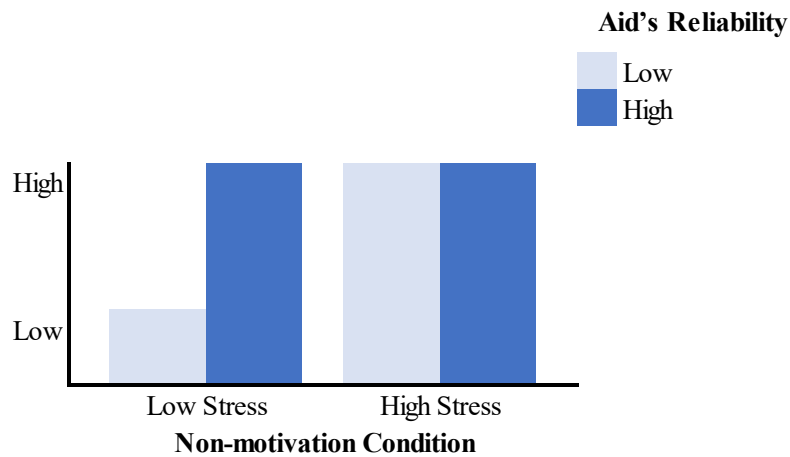


Figure 1. Visual representation of hypotheses H1, H2, and H3

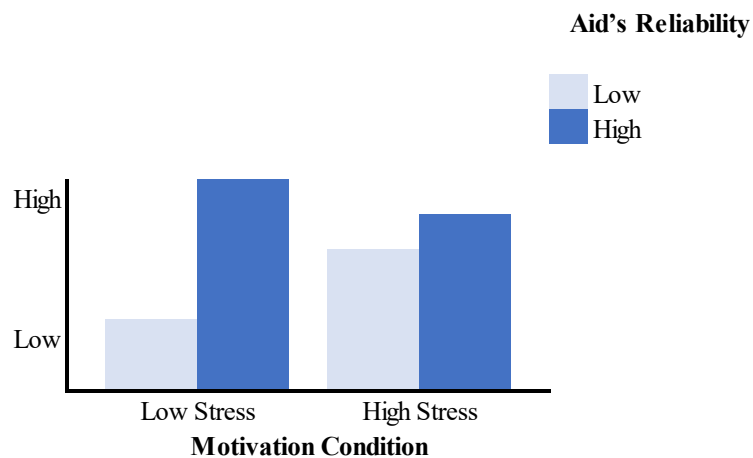


Figure 2. Visual representation of hypothesis H4

E. RESEARCH METHOD AND ANALYSIS

The study was an experiment to investigate how stress, motivation, and automation reliability affect human decisions to rely on automation and utilized data collected through an experiment from volunteer participants. The experiment was conducted in a laboratory setting with volunteer participants recruited from the Naval Postgraduate School. The study's primary analysis was reliance: how often each participant relied on the decision aid between the stress, reliability, and motivation conditions. During the experiment, stress was assessed through physiological indicators such as heart rate variability, electro-dermal activity, salivary samples collected to measure cortisol levels, and subjective questionnaires to confirm participants responded as intended to the stress and motivation manipulation. Additional data such as demographics or perceived reliability were collected via a questionnaire at the end of the experiment. During the data analysis, a two-by-two analysis of variance (ANOVA) was conducted to determine any difference in reliance between the stress, reliability, and motivation conditions. Finally, *t*-tests were conducted to determine the nature of those differences. Other analyses included cortisol levels, reaction times, stress manipulation checks, and qualitative data collected through questionnaires. Secondary analyses investigated if there were any differences based on age or gender.

F. POTENTIAL BENEFITS OF THE STUDY

The main benefit of this study is that it will further the DOD's understanding of human-machine teaming. Specifically, it will provide a better insight into how stress and motivation affect human decision making and cognition and how human reliance on a friendly automated decision aid is altered when stress and (or) motivation are present and (or) the aid's reliability changes. This research yielded data that will help the DOD understand the interactions between a human and a machine in an HMT. The results can be used to inform future decisions and drive policies within the DOD. Additionally, the improved understanding of the HMT dynamics may result in a better design of unmanned systems, training scenarios, and missions. The study may also spark further research into these areas, which may increase the lethality and effectiveness of future HMTs.

G. THESIS ORGANIZATION

The thesis is divided into five chapters. Chapter I provides an insight into the factors and motivations behind this study. This chapter also summarizes the main problem, the study's purpose, and introduces the hypotheses. It also briefly explains the research and the analysis methods and presents the potential benefits of the research for the military community.

Chapter II provides a detailed look at the current literature on the main topic and related issues. This chapter starts with a summary of the most common terms related to autonomous systems. Next, an overview of the history of human-machine teaming and the benefits of HMTs for the U.S. military is provided. The chapter then continues with an overview of contemporary research on reliance, trust, and reliability. It ends with a comprehensive review of the current literature on stress, cognition, and decision making.

Chapter III explains the design of the experiment and the method behind selecting the study's participants. This chapter also covers all hardware and software used to conduct the experiment and how these tools enabled the study. A summary of the overall procedure and processes of the experiment concludes the chapter.

Chapter IV covers the study results and provides an in-depth analysis of the focus of the study—human reliance on automation. Also described in this chapter are the manipulation effectiveness results, such as cortisol levels and other indicators. Reaction times and exploratory analyses focused on demographics and individual participant differences conclude the chapter.

Finally, Chapter V discusses the findings of the experiment and conclusions. This chapter also covers the study's limitations, offers recommendations for the DOD, and provides suggestions for future research into this topic.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

For nearly 200 years, the tools and tactics of how we fight have evolved with military technologies. Now, fundamental changes are affecting the very character of war. Who can make war is changing as a result of weapons proliferation and the fact that the tools of war increasingly are marketplace commodities.

—Cebrowski and Garstka (1998, p. 28)

A. AUTOMATION, AUTONOMY, AND ARTIFICIAL INTELLIGENCE

For better understanding of the terminology used in this thesis, it is essential to first define a few terms. The word *machine* used in this thesis refers to an electronic system powered by computer software. The term *automation* can be defined as “the execution by a machine agent (usually a computer) of a function that was previously carried out by a human” (Parasuraman & Riley, 1997, p. 231). Automation does not describe simple machine operations (e.g., an automatic elevator) but rather actions by devices such as the flight management system in aircraft, which performs a function that humans can manually perform in an emergency (Parasuraman & Riley, 1997). In this thesis, automation is an all-encompassing term used to describe machine operations at various levels of autonomy.

The term *autonomy* can be defined as “the ability of the system to attain a set of goals under a set of uncertainties” (Antsaklis & Rahnema, 2018, p. 23). A system can exhibit different degrees of autonomous behavior (Antsaklis & Rahnema, 2018). The least autonomous systems are called *automated* systems, and their modes of operation are restricted to a pre-defined set (Matteson, 2019). The most autonomous systems, sometimes called synthetic agents (McNeese et al., 2018), are typically complex, sophisticated, and cooperative systems. These machines are usually powered by *artificial intelligence* (AI), which can be defined as “systems that combine sophisticated hardware and software with elaborate databases and knowledge-based processing models to demonstrate characteristics of effective human decision making” (Shubhendu & Vijay, 2013, p. 28). The final important term is *machine learning*, which is a subset of AI that is concerned with designing algorithms to simulate human learning; that is to take information from the

external environment and use it to improve one's performance (Wang et al., 2009). AI-powered systems with a high degree of autonomy typically leverage machine learning to learn from their environment and to adapt to changing conditions, subsequently becoming more effective at their operation (Matteson, 2019).

The complexity of the AI algorithm built into the synthetic agent dictates its level of autonomy. For example, as Matteson (2019) stated, a simple algorithm that matches data from internal sensors to a pre-determined set of actions loaded into the system will exhibit a low autonomy level. On the other hand, an algorithm that can create new courses of action using machine learning applied to environment data fused from both internal and external sensors will exhibit a high degree of autonomy (Matteson, 2019). A system's autonomy level is not binary but rather a spectrum, ranging from basic automation to fully autonomous, self-aware systems (Proud et al., 2003). Many evaluation methods have been developed to measure a system's level of autonomy. Some methods are more suitable for self-driving cars, while others, such as the National Aeronautics and Space Administration's (NASA) autonomy scale, are more appropriate for unmanned aerial vehicles (UAVs).

Table 1 depicts NASA's approach to autonomy levels for aircraft systems. As can be seen, the most rudimentary autonomy level (level 1) requires 100% human involvement. This level can be described as an *automated* system. This term is used to describe systems that are limited in executing their tasks by a set of pre-defined parameters and courses of action (Matteson, 2019). On the other end of the spectrum is an *autonomous* system, depicted by levels 7 and 8. Such systems leverage machine learning to adapt to dynamic environments around them and decide on an appropriate course of action without human intervention (Matteson, 2019). Clark et al. (2015) defined an autonomous system as a system that "has a set of intelligence-based capabilities that allows it to respond to situations that were not pre-programmed or anticipated (i.e., decision-based responses) prior to system deployment" (p. 13). The authors also stated that such systems are self-governed and do not require much human assistance. Loitering munition is an example of a system at this end of the spectrum. A loitering munition system, also known as a *suicide*

drone, is a UAV that operates autonomously over the target area, scans its environment for the target, and then strikes the target once positive target identification has been acquired.

Table 1. Level of autonomy assessment scale. Source: Proud et al. (2003).

Level	Observe	Orient	Decide	Act
8	The computer gathers, filters, and prioritizes data without displaying any information to the human.	The computer predicts, interprets, and integrates data into a result which is not displayed to the human.	The computer performs ranking tasks. The computer performs final ranking, but does not display results to the human.	Computer executes automatically and does not allow any human interaction.
7	The computer gathers, filters, and prioritizes data without displaying any information to the human. Though, a "program functioning" flag is displayed.	The computer analyzes, predicts, interprets, and integrates data into a result which is only displayed to the human if result fits programmed context (context dependant summaries).	The computer performs ranking tasks. The computer performs final ranking and displays a reduced set of ranked options without displaying "why" decisions were made to the human.	Computer executes automatically and only informs the human if required by context. It allows for override ability after execution. Human is shadow for contingencies.
6	The computer gathers, filters, and prioritizes information displayed to the human.	The computer overlays predictions with analysis and interprets the data. The human is shown all results.	The computer performs ranking tasks and displays a reduced set of ranked options while displaying "why" decisions were made to the human.	Computer executes automatically, informs the human, and allows for override ability after execution. Human is shadow for contingencies.
5	The computer is responsible for gathering the information for the human, but it only displays non-prioritized, filtered information.	The computer overlays predictions with analysis and interprets the data. The human shadows the interpretation for contingencies.	The computer performs ranking tasks. All results, including "why" decisions were made, are displayed to the human.	Computer allows the human a context-dependant restricted time to veto before execution. Human shadows for contingencies.
4	The computer is responsible for gathering the information for the human and for displaying all information, but it highlights the non-prioritized, relevant information for the user.	The computer analyzes the data and makes predictions, though the human is responsible for interpretation of the data.	Both human and computer perform ranking tasks, the results from the computer are considered prime.	Computer allows the human a pre-programmed restricted time to veto before execution. Human shadows for contingencies.
3	The computer is responsible for gathering and displaying unfiltered, unprioritized information for the human. The human still is the prime monitor for all information.	Computer is the prime source of analysis and predictions, with human shadow for contingencies. The human is responsible for interpretation of the data.	Both human and computer perform ranking tasks, the results from the human are considered prime.	Computer executes decision after human approval. Human shadows for contingencies.
2	Human is the prime source for gathering and monitoring all data, with computer shadow for emergencies.	Human is the prime source of analysis and predictions, with computer shadow for contingencies. The human is responsible for interpretation of the data.	The human performs all ranking tasks, but the computer can be used as a tool for assistance.	Human is the prime source of execution, with computer shadow for contingencies.
1	Human is the only source for gathering and monitoring (defined as filtering, prioritizing and understanding) all data.	Human is responsible for analyzing all data, making predictions, and interpretation of the data.	The computer does not assist in or perform ranking tasks. Human must do it all.	Human alone can execute decision.

B. HUMAN-MACHINE TEAMING

1. Revolution in Military Affairs

In warfare, failure to adapt to the changing nature of war often results in a defeat. History provides many examples. For example, in the book *On War*, von Clausewitz (2012)

highlighted one such example by explaining why Napoleon defeated the Prussians in the 1800s. According to the author, the Prussians became complacent, underestimated Napoleon, and ignored the changing character of war, which ultimately led to their demise. Napoleon, one of the most prominent military commanders of all time, reformed many aspects of warfare by introducing organizational and doctrine changes. Napoleon's creative employment of new and existing technologies, such as artillery, and his introduction of staff officers into his organization forever changed the character of contemporary warfare (Creveld, 1985). Emerging technologies, or ingenious employment of existing ones, are usually the main driving force behind any significant change in the character of war. The inevitable paradigm shift that follows any introduction of disruptive military technology is called the revolution in military affairs (RMA) (Singer, 2009).

The most recent example of an RMA that has fundamentally transformed the character of war and redefined the way nations think about warfare is the introduction of human-machine teaming (M. Ryan, 2018). Davis (1997) also argued that this RMA is a natural product of the tendencies that have driven the U.S. defense policy since the Korean War. He backed his opinion by the fact that the world's society has shifted from being industrial-based to information-based and further supported his argument by referring to Korean War military leaders' desire to replace manpower with firepower. Davis concluded that the aversion to human casualties combined with the arrival of the information age gave rise to military HMTs.

2. History of Human-Machine Teaming

In the book *Wired for War*, Singer (2009) extensively explored human attraction to mechanical inventions. According to the author, the desire to create human-like machines dates to antiquity. Singer explained that the reason behind this fascination is the idea of replacing a human worker with a non-human device that would do the work for them. He continued by stating that machines have evolved from simple, automated systems serving as tools designed to execute easy tasks to autonomous, self-learning, independent systems acting as partners to humans.

The first mechanical creations can be traced back to the Greek mathematician Archytas of Tarentum, who built a self-propelled bird called *the Pigeon* (Singer, 2009). This pigeon was the first example of an *automated* system—a system that follows a pre-defined set of actions without any ability to respond to external events. At the turn of the twentieth century, technology advanced enough that machines could be controlled remotely via wireless communication (Singer, 2009). One could argue that the first team consisting of a human and a machine came into existence during this time. However, as Thompson (2008) posited, a true team consists of members who pursue the same objective by sharing information, resources, knowledge, and skills. Therefore, upon closer examination of Thompson’s definition of a team, it becomes evident that remote-controlled machines from the beginning of the twentieth century did not qualify as HMTs.

True HMTs did not appear until the end of the twentieth century. Specifically, from 1983 to 1993, the DOD spent \$1 billion on research and development efforts focused on AI through its Strategic Computing Initiative program (Roland & Shiman, 2002). This program was launched with the goal of achieving machine intelligence. It was the catalyst that gave rise to fully- and semi-autonomous systems at the beginning of the twenty-first century. These technologies then ushered in the era of true HMTs due to their ability to share information, resources, and knowledge with their human counterparts, which qualified them as teammates. One of the more recent interpretations of an HMT was provided by Salvendy (2012), who created a system model to represent modern human-machine systems’ complexities. The model, shown in Figure 3, considers human factors as well as social, organizational, and environmental factors. It also depicts the two-way cognitive interaction between humans and machines, representing modern HMTs (Salvendy, 2012).

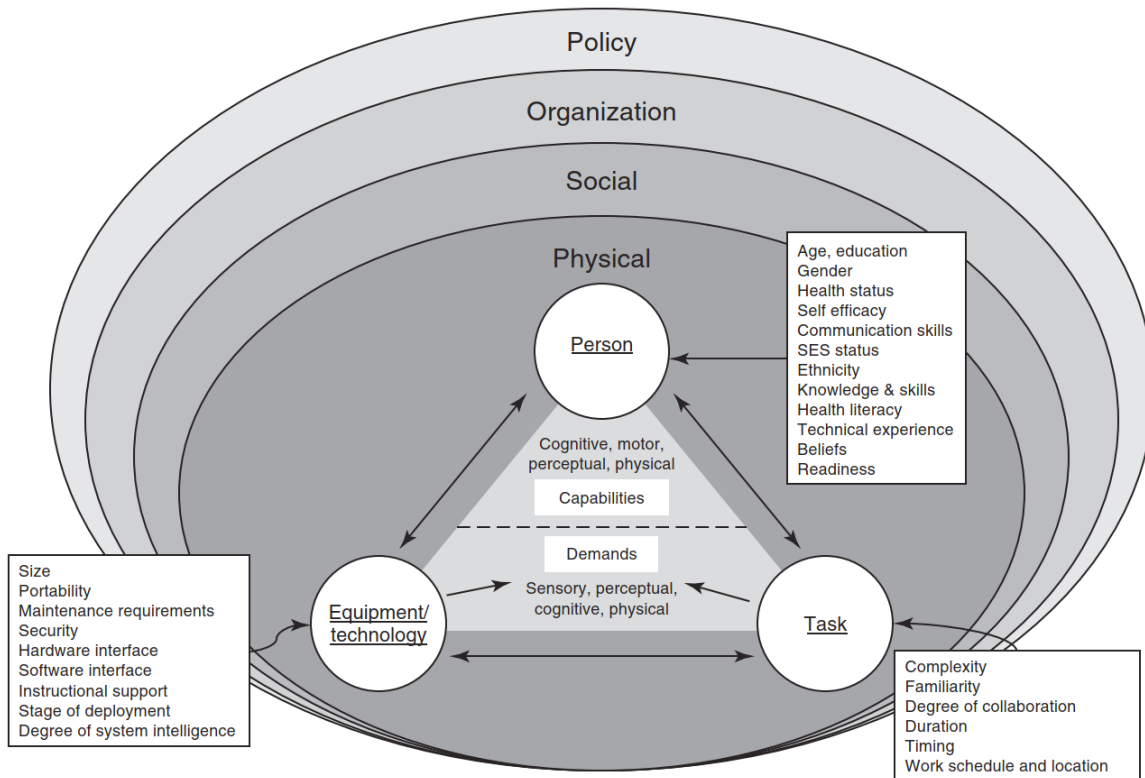


Figure 3. A simplified model of a human-machine system.
Source: Salvendy (2012).

3. Employment and Benefits of Human-Machine Teams

Tossell et al. (2020) argued that national defense would not be possible without smart machines integrated into the military in the twenty-first century. The 38th Commandant of the Marine Corps, General Berger, has made autonomous systems an integral part of his vision for the force in 2030 (United States Marine Corps, 2020). Today's most sophisticated machines (also referred to as *robots*) are powered by advanced AI algorithms and already fulfill essential military roles. Wilson and Daugherty (2018) proposed that military operators can expand their abilities through the employment of smart, AI-powered machines. The authors explored ways in which machines can assist humans and synthesized their research into the following three categories:

1. AI can *amplify* human cognitive abilities. Specifically, by supplying the right information at the right time, AI can enhance human decision-making and analytical skills.

2. Collaborative machines can *interact* with the environment on behalf of the human operator. By taking over specific rudimentary tasks, the human operator (e.g., a pilot) is then free to focus on higher-level responsibilities.
3. Machines can *embody* human skills to extend operators' physical capabilities. For example, a robot-assisted assembly of heavy industrial components for ships can leverage the physical strength of a machine while using human dexterity and judgment for more sophisticated complementary tasks at the same time.

The benefits of deploying such smart systems within the U.S. military are apparent. Contemporary autonomous military systems can perform complex functions such as independently collecting, fusing, and analyzing data, navigating the operational environment, proposing alternative courses of action, or providing advanced analytics capabilities to their human partners. Recent growth in the quantity of generated data from humans and sensors connected to the Internet has resulted in an enormous amount of information available to the DOD (van den Bosch & Bronkhorst, 2018). Military decision-makers can easily get overwhelmed and experience what is called *information overload* (Hemp, 2009) and subsequent *analysis paralysis* (Bensoussan, 2008) when making decisions. Therefore, having a system that can quickly fuse, analyze, and interpret data from sensors, media, and intelligence is paramount. The military benefits are even more evident upon realizing that conflicts in the twenty-first century are rapidly shifting into the cyber domain (van den Bosch & Bronkhorst, 2018).

In addition to the data processing benefits, which might lead to better decisions and lower operational costs, there are numerous other benefits. For example, force protection and risk mitigation are two of the top priorities of the U.S. military. Collaborative military robots can save lives by aiding with explosive ordnance disposal, carrying supplies through rugged terrain or hostile environment, patrolling dangerous areas, or detecting anomalies in the operational environment (Lyons et al., 2019). Exo-skeletons—wearable robotic devices designed to enhance human strength and stamina—can turn U.S. troops into superhumans (Wilson & Daugherty, 2018). Machines, such as UAVs, extend humans' reach in terms of time-on-station or intelligence collection capabilities. Automated

airborne systems can prevent loss of life by taking over control of the aircraft from the pilot when a collision course is calculated (Tossell et al., 2020). AI-powered teammates can also help military decision-makers in non-routine situations or during times of increased uncertainty (van den Bosch & Bronkhorst, 2018). In summary, the number of use scenarios and benefits to the U.S. military is simply enormous, but without a complete understanding of factors affecting human reliance decisions, the operational effectiveness and safety of HMTs will likely be suboptimal.

C. EFFECT OF TRUST AND RELIABILITY ON RELIANCE

The U.S. military's reliance on automation has increased considerably over the last two decades (Tossell et al., 2020). Reliance is a behavior (Lee & See, 2004) that is closely related to trust, which is one of the factors that help explain reliance decisions (Dzindolet et al., 2003). Reliance behavior, also linked with automation reliability (Riley, 1996), describes one agent's level of dependence on another (Rice & Keller, 2009). However, despite the available research on reliance the precise understanding of the relationship between trust, automation reliability, and other factors remains vague (Lee & See, 2004).

The body of research on reliance is extensive. For example, Riley (1996) stated that reliance is a factor that “represents the probability that an operator will use automation” (p. 21). In the same work, Riley hypothesized directions of influence between trust, reliance, reliability, and other factors affecting the human-machine relationship. He outlined a relationship diagram showing the proposed directions of influence between factors that play a role in HMT dynamics. Parasuraman and Riley (1997) later updated this diagram based on empirical data, as shown in Figure 4. Solid arrows indicate relationships supported by experimental data. Dashed arrows represent hypothesized relationships initially outlined by Riley (1996). As shown in Figure 4, reliance is influenced by trust, which is affected by machine accuracy—reliability. Others have also subscribed to the theory that reliance is correlated with trust, both of which affect the system's overall performance (Lee & Moray, 1992; Muir, 1987; Wickens et al., 2013).

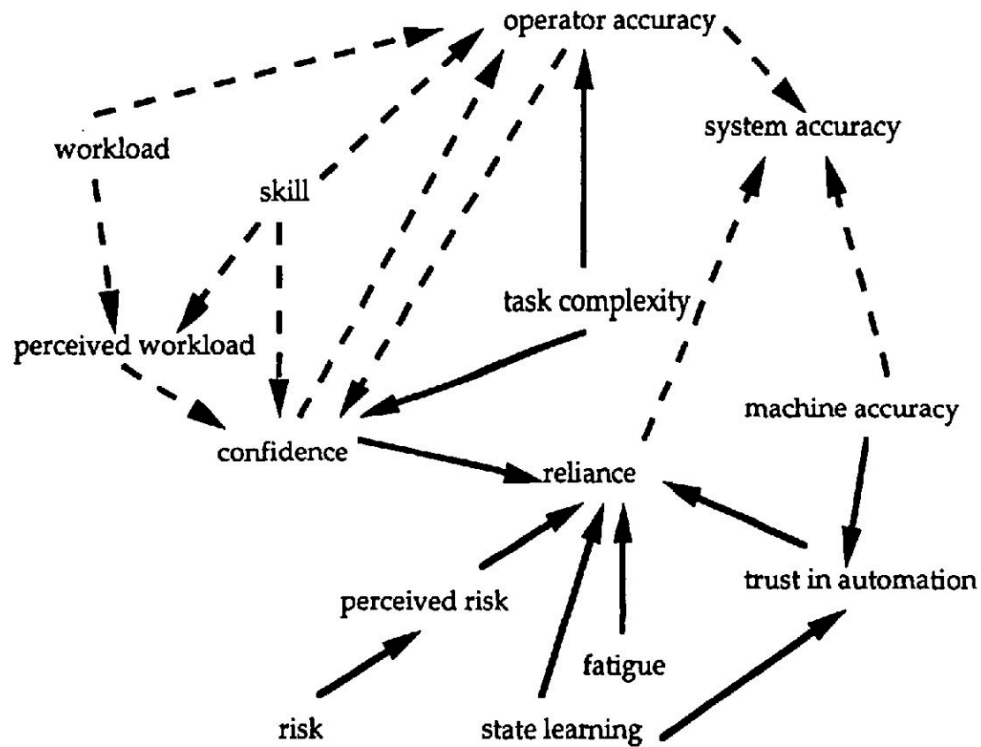


Figure 4. Interactions between factors influencing automation use.
Source: Parasuraman and Riley (1997).

Besides factors shown in Figure 4, reliance decisions can be driven by various other factors. On this topic, van Dongen and van Maanen (2013) submitted a framework for understanding these factors, which is shown in Figure 5. The framework shows that reliance decisions are affected by several specific task and psychological factors. For example, the type of decision making, cognitive availability, reasoning ability, information processing speed, performance feedback, or perceived reliability and trust.

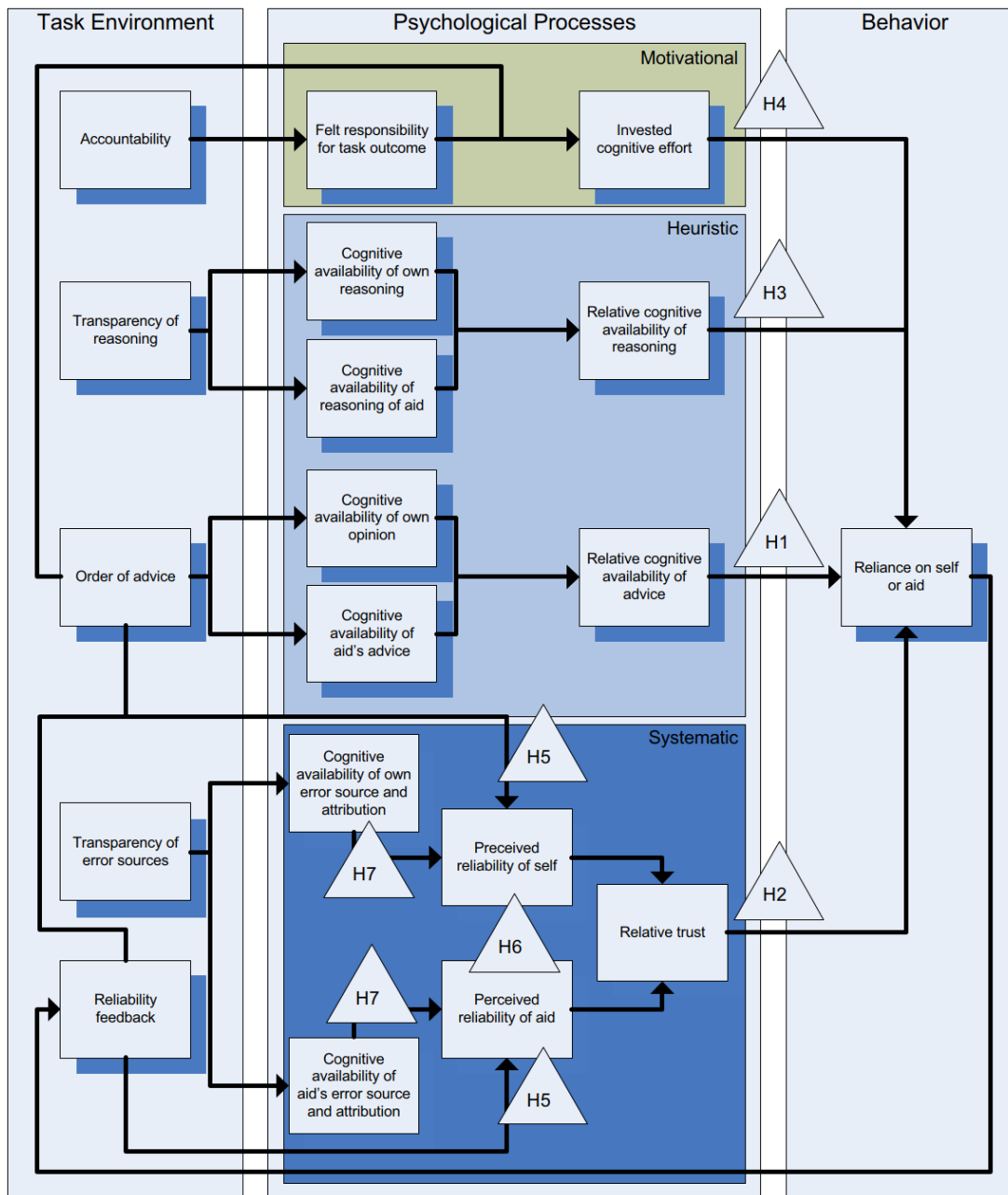


Figure 5. Framework of reliance decision making. Source: van Dongen and van Maanen (2013).

Appropriate reliance is essential to the human use of automation (Parasuraman & Riley, 1997). Properly calibrated reliance, the authors argued, leads to appropriate *use* of automation, while an improper reliance calibration results in *misuse*, *disuse*, or *abuse* of automation. Too much reliance on automation leads to automation *misuse*, which may lead

to operator complacency and monitoring failures. Too little reliance leads to automation *disuse* or “underutilization of automation” (Parasuraman & Riley, 1997, p. 230). Finally, automation *abuse* is “inappropriate application of automation” (Parasuraman & Riley, 1997, p. 233). It is evident that these terms are relevant and applicable to the use of modern autonomous weapons systems being deployed by the U.S. military. Also evident is that effects of trust and automation reliability need to be explored more closely.

1. Trust

No other variable is as essential to the interactions between humans and machines as trust (Wickens et al., 2013). Trust “guides reliance when complexity and unanticipated situations make a complete understanding of the automation impractical” (Lee & See, 2004, p. 50). Although similar, trust and reliance carry different meanings. Trust is a mental state (Rice & Keller, 2009) that can be defined as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” (Lee & See, 2004, p. 51). Trust is also closely tied to reliance as it “is the psychological precursor to engaging in reliance behavior” (Lyons & Stokes, 2012, p. 113). Mayer et al. (1995), who focused on interpersonal trust, provided their general definition of trust as follows:

the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party. (Mayer et al., 1995, p. 712)

de Visser et al. (2016) submitted that the increased proliferation of intelligent autonomous systems embedded into daily human life, such as voice assistants Alexa or Cortana, led humans to establish relationships with these systems and that dynamic factors, such as interpersonal trust, can be observed in the interaction between the two. The authors conducted a study in which they paired participants with agent advisors with varying human appearance in order to explore how trust is formed, violated, and repaired during decision-making tasks. The authors provided evidence suggesting that when machines that are acting as teammates are given human-like (anthropomorphic) characteristics, trust resilience (resistance to breakdowns in trust) improves. Dzindolet et al. (2003) also

observed that when human operators were informed about why the automation failed, their trust in automation increased, and so did their reliance on it. This observation is closely related to the concept of *explainable AI*, which describes how autonomous machines should be designed. According to Gunning and Aha (2019), machines that can explain their decisions and actions will make humans trust them more appropriately. Groom and Nass (2007) benchmarked several HMT configurations and challenged the notion that machines could replace humans as teammates. The authors mentioned the importance of trust and claimed that humans might find machine teammates untrustworthy because they lack human-like features. Because trust drives reliance, explainable AI with human-like features might result in a more appropriate level of reliance, as discussed earlier.

Additionally, Mayer et al. (1995) proposed that the level of trust one has toward a teammate can be perceived as the amount of willingness to take a risk when working together with the teammate. This proposition is an important one as it helps clarify the relationship between trust and cooperation, which are two factors relevant to HMTs. The research by Mayer et al. can be extended to HMTs because humans might be more likely to cooperate with a machine they do not trust under low-risk conditions than high-risk conditions. Trust violations (caused by low machine reliability) are also a critical to reliance decisions. Low machine reliability leads to the loss of trust (de Visser et al., 2018), which affects reliance. The magnitude of trust violation consequences depends on the type of a machine. For example, a trust violation of a health care machine is perceived differently and requires longer repair than a trust violation by a digital personal assistant (de Visser et al., 2018). Finally, Muir (1987), who argued that trust between humans and machines depends on the same factors that influence interpersonal trust, claimed that automation reliability, like human reliability, is critical to the relationship.

2. Automation Reliability

Automation reliability is one of the most critical factors impacting an HMT's overall performance (Gao & Lee, 2006; Lee & See, 2004) because unreliable machines can lead to unintended outcomes, property damage, or a loss of life (Dhillon, 1986). Wickens et al. (2013) stated that “of all the variables to affect trust/dependence, probably the most

critical is automation reliability” (p. 341). Dzindolet et al. (2003) added that human trust in a machine is largely affected by the machine’s reliability. In general, reliability can be defined as “the probability of accomplishing a specified task successfully by humans at any required stage in system operation within a defined minimum time limit (if the time requirement is specified)” (Dhillon, 2009, p. 4).

Automation reliability affects the operator’s attitudes toward it. Research shows that, when teamed with a machine, humans tend to perceive their automated partners as more accurate (or reliable) than humans, which affects their trust in and reliance on them (Lyons & Stokes, 2012). However, unreliable machines lead to behavioral changes. For example, Sorkin (1988) examined why operators turn off automated alarms or block out flight assistant warning indicators, which was a behavior observed by investigators following several air and train crashes. Sorkin argued that this behavior results from a high false alarm rate, which is an indicator of low system reliability. Additionally, Parasuraman and Riley (1997) mentioned kitchen smoke detectors as an example of a device that often gets disconnected because of its typically high rate of false alarms. Using this example, the authors argued that automation reliability shapes the operator’s attitude toward it, which then drives its use (or disuse).

A recent trend has been to improve the overall HMT by perfecting automation reliability (Dzindolet et al., 2003). However, Dzindolet et al. (2003) argued that increasing automation reliability alone might not improve the HMT’s performance. On the topic, Greengard (2009) posited that as the automation reliability increases in the future, so will human reliance on it, to the point that humans may blindly rely on it. Greengard argued that this trend might lead to the *automation paradox*, which may, in fact, decrease the overall system performance. This phenomenon might lead to catastrophic results when the machine’s reliability, on which human military operators rely, suddenly decreases during combat operations. Dzindolet et al. (2003) added that when humans see a reliable machine make an error, their trust level decreases. Therefore, the authors argued that knowing the reason why the automated decision aid may fail increases both trust in and reliance on automation, which affects the overall system performance. In other words, even though the machine teammate might be reliable, if it cannot provide explanations for its actions or

potential failures, human trust and reliance may remain suboptimal. This observation is in line with the benefits of explainable AI described by Gunning and Aha (2019) earlier in this chapter.

Further on the topic of automation reliability, Lee and Moray (1992) noted that the frequency of fault occurrence in a system affects human trust in it, driving human reliance behavior. Along those lines, Wickens et al. (2013) presented the calibration curve, shown in Figure 6. The curve is “a critical concept in the relation between automation trust/dependence and reliability” (Wickens et al., 2013, p. 341). The diagonal line in Figure 6 represents *perfect calibration* between over-trust (upper left-hand corner) and under-trust (lower right-hand corner). Wickens et al. used this figure to illustrate what happens when a human operator works with a modern, highly reliable system. Initially, the system works without errors for some time, building the operator’s trust in it. If the time period is long enough, the operator’s trust in the system reaches the point of complacency. Then, the system makes an error, and the operator loses most of his trust in it. According to the authors, this initial reaction is typically dramatic and is often associated with automation-based accidents. Following the initial reaction, the operator’s trust crosses the calibration curve into the under-trust region. Over time, the operator’s trust in the system slowly recovers toward the calibration curve, which represents long-term reliability.

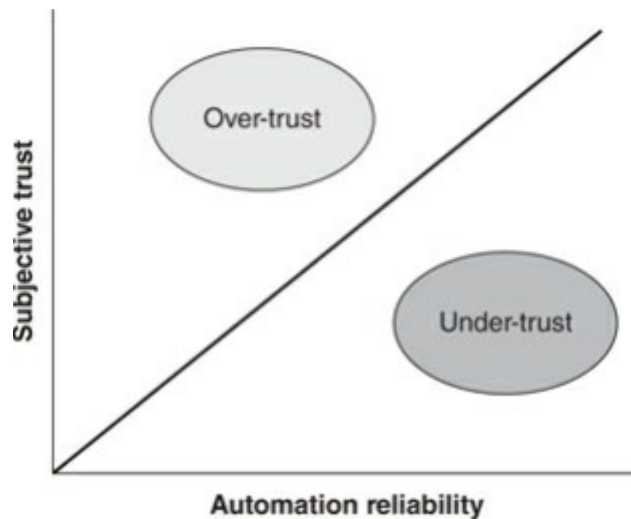


Figure 6. The relationship between subjective trust and automation reliability. Source: Wickens et al. (2013).

Finally, Wiegmann et al. (2001) examined how different levels of automation reliability affect human trust in it. Results of their study indicated that participants were sensitive to various levels of automation reliability, which also affected their trust in it. Dixon and Wickens (2006) took a similar approach in studying the effects of variable UAV reliability on military operators' performance in high-workload environments. Their results indicated that the desired automation reliability should be above 70% because below this threshold task performance is worse than having no automated aid at all. Humans also have to exert more attentional effort when working with a machine that exhibits reliability below this threshold (Onnasch, 2015).

In summary, the decision to rely on automation is driven by multiple psychological and situational factors. This study is primarily focused on the concept of trust and explores the effect of automation reliability on reliance decisions. However, to fully understand reliance decisions, it is important to explore how decisions are made in general.

D. DECISION MAKING

Because decision making is an integral part of everyday life, the literature on this topic is extensive. For example, Balleine (2007) conducted a study on the neural basis of decision making and provided the following explanation of the decision-making process:

Decision making refers to the ability of humans and other animals to choose between competing courses of action based on the relative value of their consequences. This capacity is, therefore, fundamentally integrative, melding the complex cognitive processes through which causal relations between actions and consequences are encoded, retrieved, and maintained in working memory with the motivational processes that determine the value, or utility, of actions or sequences of actions. (p. 8159)

From Balleine's description, it is evident that decision making is a complex process that requires significant cognitive resources, such as attention. Decision making is a continuous activity that is constantly adjusted based on new inputs from the environment (van den Bosch & Bronkhorst, 2018). Before exploring decision making, and variables that affect it in greater detail, it is essential first to understand the concept of memory, how people learn, and how they process information.

1. Memory and Information Processing

One's ability to remember and process information is tied to their mental capacity (Baddeley, 2013). The term *memory* was introduced in the 1880s (Chai et al., 2018) to describe a part of the brain where information is stored. Memory enables information processing, which consists of three phases—learning, consolidation, and retrieval (Sandi, 2013). Atkinson and Shiffrin (1968) proposed that memory could be broadly divided into short-term memory, long-term memory, and sensory memory. Short-term memory describes “the temporary storage of material necessary for performing a range of complex tasks such as comprehension, reasoning, and long-term learning” (Baddeley, 2013, p. 17) while long-term represents “a more durable encoding and storage system” (Baddeley, 2013, p. 17). Sensory memory is involved in the processing of sensory information such as audiovisual information (Baddeley, 2013).

From the idea of short-term memory, Baddeley and Hitch (1974) proposed the idea of working memory, which is critical to consciousness because it allows one to think about the pieces of information stored in it and give them attention (Alloway & Alloway, 2012). Working memory is essential to learning and processing new information, but it has a limited storage and processing capacity (Baddeley, 2013). On average, Baddeley found that most people can reliably memorize and recall the order of digits in a string between six and seven digits long and that task is made more complicated when paired with another task, especially reasoning or comprehension.

Working memory can be represented by a four-component model consisting of the central executive, the phonological loop, the visuospatial sketch pad, and the episodic buffer (Baddeley, 2012). According to Baddeley, the phonological loop is responsible for processing and storing auditory information, while the visuospatial sketch pad aids with the processing and manipulation of images. The episodic buffer was added to the working memory model later and represents a temporary store and a link between working memory and long-term memory (Baddeley, 2000). The components in the model are controlled by the central executive (Baddeley, 2013), whose functions are performed by the prefrontal cortex (PFC), which manipulates incoming sensory information for further execution (Chai et al., 2018).

The PFC is considered the most evolved part of the human brain and provides cognitive functioning at the highest level (Arnsten, 2009). Dickerson and Kemeny (2004) added that the PFC, together with the thalamus, integrates, assesses the significance of, and assigns meaning to sensory information. The PFC connects multiple regions in the brain and helps individuals make intelligent decisions by regulating thoughts, actions, and emotions (Arnsten, 2009). The PFC also monitors for errors; therefore, people can better understand which decisions may lead to undesirable consequences and adjust their strategy accordingly (Arnsten, 2009). This ability allows one to learn and improve using feedback from the external environment. Therefore, it is essential to explore next how humans learn and how feedback affects the learning process.

2. Learning and Feedback

Feedback is essential to learning, which is a complex and intricate process (Starcke & Brand, 2016). The effect of feedback on learning is especially pronounced during uncertain situations when outcomes of a decision are unknown (Starcke & Brand, 2016). To help explain the role of feedback in human decision making, Banis and Lorist (2012) submitted that humans set their internal goals through decision making. These goals then drive the human behavior toward them, while inputs from the environment—feedback—adjust the behavior to reach these goals. Feedback is primarily defined by valence, which is an indicator of a positive or a negative outcome, and magnitude, which describes the degree of positivity or negativity (Banis & Lorist, 2012).

Feedback is critical to human-machine teaming for several reasons. Feedback affects human trust in automation, which, in turn, drives reliance decisions (Gao & Lee, 2006; Lee & See, 2004). van Dongen and van Maanen (2013) stated that trust in the machine teammate “depends upon one’s own previous performance and that of the decision aid” (p. 411). This vital feedback about past performance affects human trust. Additionally, de Visser et al. (2016) conducted an experiment in which they manipulated the feedback delay given to humans working with synthetic agent advisors—a computer, an avatar, and a human. Their findings suggest that immediate feedback from an advisor improves the overall performance of the HMT. Further on the topic of trust, Gunning and Aha (2019)

argued for the concept of explainable AI and described the benefits of an HMT in which the machine teammate can explain its actions and intentions to the human. Such a feedback loop allows the human to learn faster and build trust in the machine sooner than if no feedback was available. Finally, it is also essential to consider the effect of motivation on decision making.

3. Impact of Motivation

Decision making is influenced by motivation such that the value or utility of a particular action affects one's decisions (Balleine, 2007). Additionally, motivation is critical to learning and information processing (Baddeley, 2013). Pochon et al. (2002) argued that motivation could be viewed as an invisible force “that pushes one to act” (p. 5669), while Verharen et al. (2020) defined motivation as “the willingness to invest resources (such as time or effort) in order to receive a reward or to avoid a punisher” (p. 93). Verharen et al. also noted the distinction made in earlier research that defined motivation as *wanting* a reward instead of *liking* a reward, which equates to enjoyment.

When performing a task, one can be motivated in two ways, as described by Sansone and Harackiewicz (2000). The first type of motivation—*intrinsic*—refers to the natural inclination of a person to engage in an activity that interests them, while the second type—*extrinsic*—refers to motivations that are external to the individual, such as rewards or praises. According to the authors, both types of motivation could be at work at the same time. For example, a child might be reading a book not only because they like the activity (intrinsic motivation) but also because they want to gain a parent's or a teacher's approval (extrinsic motivation). It is important to consider these findings for the purpose of this study because during the experiment some participants were under the influence of extrinsic motivation created by performance-based rewards. However, these participants might have also been influenced by their own interest in the study (intrinsic motivation).

Pochon et al. (2002) demonstrated that the human brain responds to external motivational stimuli. Later, Locke and Braver (2008) came to a similar conclusion when they found evidence of increased activity in the parts of the brain associated with reward. However, the authors argued that individuals exhibit variations in their motivation that can

be attributed to (a) the individual's motivational state, (b) their sensitivity to rewards, (c) their ability to utilize the increased motivation to improve their cognitive performance, or (d) the combination of any of the first three factors.

Research also shows that motivation can improve cognitive performance (Ferdinand & Czernochowski, 2018). In the study done by Massar et al. (2016), the authors explored the effect of motivation on sustained attention performance using the Effort Allocation theory. This theory, proposed in earlier research, states that one's cognitive resources are allocated based on a simple cost-benefit analysis (Massar et al., 2016). In other words, an individual would allocate more resources to a more rewarding task, leading to better performance and fewer resources to a more taxing task, leading to lower performance. Massar et al. also argued that "the subjective value of reward is discounted by the effort required to receive the reward" (Massar et al., 2016, p. 27).

In summary, learning is a complex process that involves several parts of the brain (Baddeley, 2013) and is influenced by feedback and motivation. Having explored the fundamental parts of information processing, learning, and the impact of feedback and motivation, it is now necessary to examine the process of decision making itself.

4. Types of Decision Making

According to Balleine (2007), decision making is "the ability of humans and other animals to choose between competing courses of action based on the relative value of their consequences" (p. 8159). Essential to this ability is working memory, which encodes, retrieves, and maintains information about actions and their consequences (Balleine, 2007). The PFC is also a crucial part of the brain involved in decision making as it can shift one's attention from one task to another based on the change in reward (Arnsten, 2009). Combined together, these parts of the brain play a central role in decision making.

A variety of factors affect how people approach problems, reason, and make decisions. In his book, Kahneman (2011) proposed a dual-system theory of decision making, arguing that two fundamental (hypothetical) systems drive human reasoning, thinking, and decision making. Kahneman labeled the two systems as System 1 and System 2. System 1 represents an intuitive (heuristic) approach to problem solving in this dual-

system theory, while System 2 refers to a more deliberate (effortful) thinking. Kahneman explained that System 1 is typically used to make decisions guided by emotions, intuition, or past experiences. It acts quickly (almost automatically) and requires little to no effort. Systems 2, on the other hand, is employed when a quick, intuitive solution fails and the answer to the problem requires attention, complex computations, or concentration. System 2 is typically associated with terms such as agency, choice, and concentration. While System 1 is quicker than System 2, it is also prone to cognitive biases that can lead to undesirable consequences (Kahneman, 2011).

Most military forces intentionally approach decision making analytically (deliberately), utilizing System 2 to create structure, order, and uniformity in problem solving across all levels of warfare (van den Bosch & Bronkhorst, 2018). However, making decisions that are tied to consequences often induces psychological stress, impacting the decision-making process itself (Gathmann et al., 2014). Many military operators also make tactical-level decisions in situations that involve risk and uncertainty. Therefore, it is critical to understand how stress affects decision making.

E. EFFECT OF STRESS ON DECISION MAKING

Research shows that stress can negatively affect human cognition as well as decision making (Arnsten, 2009; Sandi, 2013), both of which are essential to human-machine teaming. This potential negative impact stress has on human cognitive abilities may have a debilitating effect on the operational effectiveness and efficiency of HMTs. Hence, it is critical to understand what stress is, how the human body responds to it, and what is the exact effect of stress on human cognition and decision making.

1. Stress—Definition and Response

The term stress originally comes from engineering, where it is used to describe the effect of external physical forces being applied to an object (Lupien et al., 2015). Similarly, Lupien et al. submitted, stress in psychology describes the effect of stressors (forces), often other than physical, on the human body. According to Lupien et al., most people associate stress with *time pressure*, but time pressure is not equivalent to stress because if it were, every individual would exhibit a stress response in the presence of a time limit. While some

individuals feel stressed in the presence of a time limit, some report feeling stressed in the absence of a time limit. Therefore, the authors argued, a stress response is usually triggered by specific psychological determinants.

The stress determinants were outlined by Mason (1968), who wanted to define a universal set of psychological characteristics that would elicit a stress response. He submitted that for a situation to induce a stress response in an individual, it must be perceived as stressful by them. In other words, the level of stressfulness is a result of the individual's cognitive analysis of the given situation and its circumstances (Lupien et al., 2015). Mason outlined three determining factors that make a situation stressful. Firstly, the situation has to be interpreted by the person as *novel* (i.e., something that the individual has not encountered before). For example, when a fundamentally different operating system is installed on a UAV control panel. Secondly, the situation has to be interpreted by the person as *unpredictable*, which is a situation that the individual did not anticipate—for example, a sudden change in the enemy defensive posture. Finally, the situation has to be interpreted as being *out of control* (e.g., receiving an electric shock for an incorrect answer). More recently, Dickerson and Kemeny (2004) added the presence of a *social-evaluative threat* to the three stress determinant characteristics originally postulated by Mason (1968). A social-evaluative threat is a form of psychosocial stress that occurs when an individual's self-identity, competence, or performance could be negatively judged by others.

In general, researchers agree that stress is a real (or perceived) challenge (or threat) to one's normal state (i.e., homeostasis) (de Kloet et al., 2005; Sandi, 2013; Smith & Vale, 2006). Homeostasis is a natural balance that all living organisms try to achieve (de Kloet et al., 2005). When stressors (physical or psychological) are applied to an individual, and the individual interprets the situation as stressful, their body's natural (stress) response will be to trigger behavioral and physiological changes aimed at restoring homeostasis (de Kloet et al., 2005). According to Lupien et al. (2015), one of the systems that a stressful situation activates in the body is the hypothalamic-pituitary-adrenocortical (HPA) axis. The hypothalamus (body's master gland) releases the corticotropin-releasing hormone, which triggers the secretion of adrenocorticotropin (ACTH) from the pituitary. Lupien et al. further explained that ACTH then travels through the bloodstream to the adrenal glands,

triggering the release of glucocorticoids (i.e., cortisol in humans) and catecholamines (e.g., epinephrine and norepinephrine). However, this process does not start if the situation is not perceived as stressful by the individual.

2. Impact of Stress on Cognition

Cognition refers to the variety of processes that enable functions such as information perception, retention, and learning (Sandi, 2013) and stress has been shown to affect cognition (Lupien et al., 2015). Specifically, stress-induced cortisol affects areas containing cortisol receptors, which can be found in the frontal lobes of the brain (Lupien et al., 2015). The frontal brain lobes are involved in learning and memory and include the PFC, which is considered the center for higher cognitive functions and the region most sensitive to stress (Sandi, 2013). Although, in general, stress has been shown to impair PFC functioning, mild stress might improve the performance of simple, habitual tasks that do not depend on the PFC (Arnsten, 2009).

Memory, another component critical to human cognition, is also prone to the effects of stress. Margittai et al. (2016) argued that the detrimental effect of acute stress impacts working memory and attention, making it more difficult for an individual to separate relevant information from the rest of the sensory information coming into the brain. Stress has also been found to affect one's ability to retrieve information stored in their memory (Lupien et al., 2007). Additionally, Sandi (2013) submitted that acute stress of high intensity improves implicit memory (skills and habits) performance while impairing explicit memory (facts and events). However, Sandi also stated that the effects of stress on memory vary due to factors such as the stress intensity, cognitive workload, or the time between the onset of stress and the start of the task. Finally, it is vital to understand how stress affects human decision making.

3. Impact of Stress on Decision Making

According to Starcke and Brand (2012), "stress and decision making are intricately connected" (p. 1228). Margittai et al. (2016) explored how elevated cortisol levels affect the dual-system theory about information processing and decision making postulated by Kahneman (2011). In their study, Margittai et al. elevated cortisol in the participants by

orally administering hydrocortisone, thus excluding other factors typically associated with different psychological stress induction methods. The authors found evidence suggesting that cortisol affects one's decision making, inducing switch from deliberate thinking (System 2) to intuitive thinking (System 1).

Additionally, other studies have found evidence supporting that stress alters decision making (Starcke & Brand, 2012). For example, individuals under stress tend to prefer a decision that brings an instant reward while neglecting the future negative consequences of such a decision (Starcke & Brand, 2012). Picking up habits such as smoking or excessive drinking when stressed is an example of such a tendency. Although the precise effect of stress differs based on individual traits such as self-esteem, stress has been found to affect various daily decisions people make, such as risk strategy decisions, moral dilemma choices, or reward/punishment decisions (Starcke & Brand, 2012). Other authors also described the effects of stress on decision making. For example, Arnsten and Goldman-Rakic (1998) argued that stress impairs decision making to the point of taking the PFC "off-line." Otto et al. (2013) claimed that complex decisions requiring working memory were affected by stress while simple decisions were not and that the negative effect of stress was also more pronounced in individuals with smaller working memory capacity.

F. SUMMARY

Machines can compensate for human limitations by performing complex tasks such as advanced computations. They can also act as decision-support agents by providing pattern recognition or big data visualization to guide military decision makers. However, many factors impact the overall performance of HMTs and, as society shifts from treating machines as tools to treating them as collaborative teammates (Lee & See, 2004), a complete understanding of potentially negative factors is required to maximize the benefits and minimize the adverse outcomes of these technologies.

Despite a large body of research on the topic, many human-machine teaming areas remain unexplored or poorly understood. Specifically, gaps exist in our current understanding of the dynamics and factors that affect the interaction between a human and a machine acting as an autonomous collaborative teammate. This chapter provided a brief

overview of the history of HMTs and explored some of these factors, such as human trust, reliance, or machine reliability. Additionally, this chapter explored the effect of stress on human cognition and decision making. Topics researched and presented in this chapter provide a solid foundation for the experiment conducted for this thesis and show gaps in the research that need to be addressed.

III. METHODS

A. PARTICIPANTS AND DESIGN

1. Participants

A total of 80 participants (61 male, 17 female, two unspecified), recruited from the Naval Postgraduate School (NPS), were run through the experiment between October 15, 2019, and March 13, 2020. The age of the participants ranged from 22 to 44 years ($M_{\text{age}} = 32.47$, $SD = 4.87$, one participant did not specify). Figure 7 depicts the age and gender distribution of the participants.

Exclusionary criteria included psychiatric illness, cardiovascular disease, neuroendocrine disorder, or pregnancy. All participants were 18 years or older and were recruited through campus-wide emails, flyers, and word of mouth. Interested participants signed up for a two-hour time slot through a free signup service at <https://www.signupgenius.com>. Experimentation was approved by the NPS Institutional Review Board.

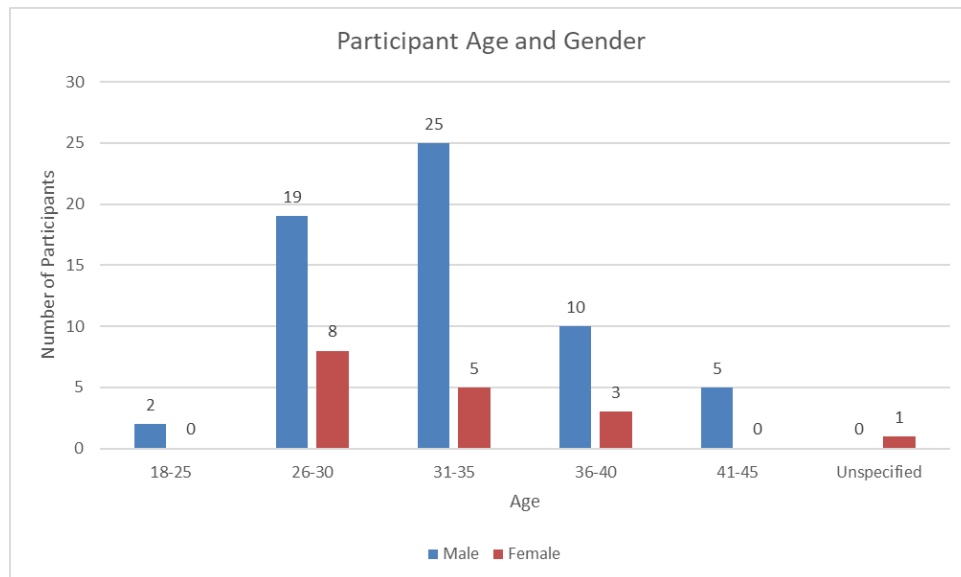


Figure 7. Distribution of participants by age and gender

Figure 8 depicts the distribution of the participants by the U.S. military service. Most participants were members of the United States Navy (USN) at 43%. The United States Marine Corps (USMC) represented 23% percent, while the United States Army (USA) and the United States Air Force (USAF) represented 18% and 11%, respectively. No other U.S. uniformed service was represented. Also, five participants were civilians not associated with a uniformed service. All participants associated with a uniformed service were on active duty at the time.

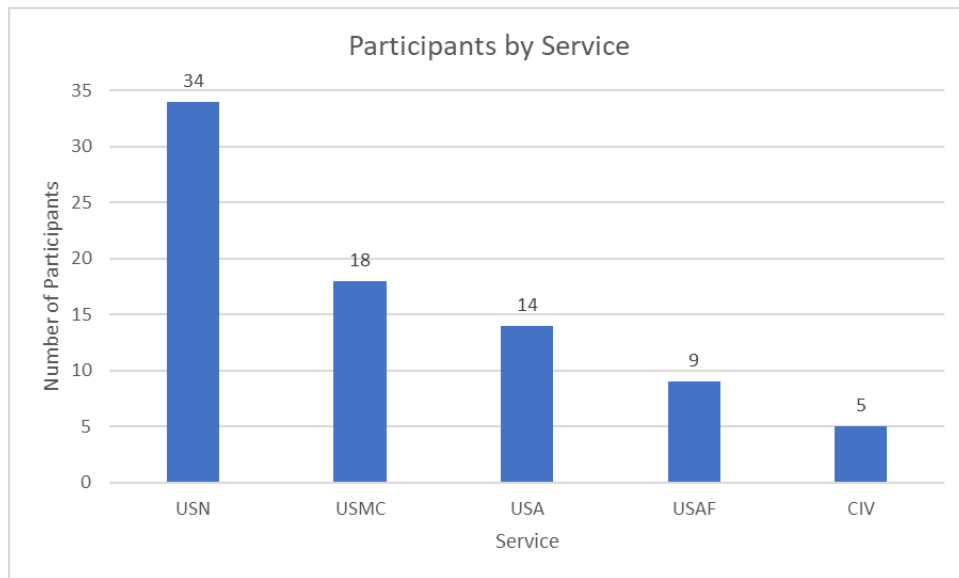


Figure 8. Breakdown of participants by service

Forty-four percent of participants reported having no prior stress training, while 56% of participants reported having undergone some stress training in the past. Finally, Figure 9 shows the distribution of the highest achieved education among the participants.

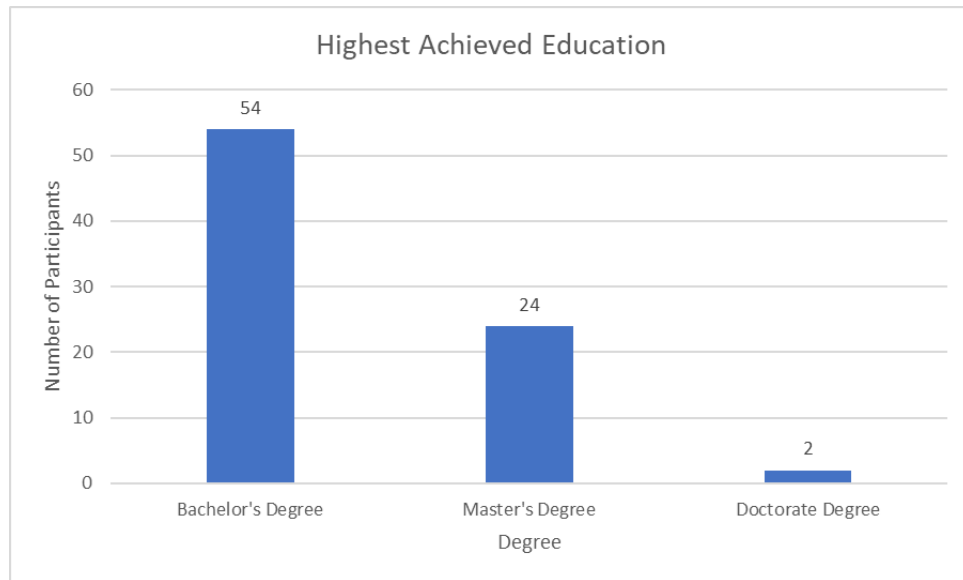


Figure 9. Highest achieved education

2. Experimental Design

The experiment was a 2(Stress: high, low) x 2(Motivation: present, absent) x 2(Reliability: High, low) repeated measures design with stress and motivation manipulated between the subjects and automation reliability manipulated within the subjects. First, the participants were screened for exclusionary criteria. Those who passed were randomly assigned to either a high-stress or a low-stress condition and either motivation or non-motivation condition. Figure 10 depicts the experimental design conditions for stress and motivation. It also shows the two AI partners (AI1 and AI2) the participants worked with in each condition during the experiment.

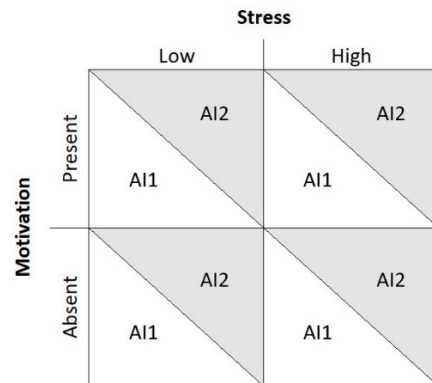


Figure 10. Experimental design conditions

B. STRESS MANIPULATION

1. Trier Social Stress Test

Many laboratory stress induction techniques, such as noise exposure, pain induction, or cognitive tasks, do not always elicit a reliable response (Allen et al., 2017; Narvaez Linares et al., 2020). The most effective way to produce a stress response in a laboratory setting is to combine uncontrollability and a social-evaluative threat (Dickerson & Kemeny, 2004). The Trier Social Stress Test (TSST) is a protocol developed by Kirschbaum et al. (1993) that maximizes the body's endocrine stress response in a reliable, repeatable, and measurable way by utilizing both uncontrollability and a social-evaluative threat. The TSST, which reliably increases the level of cortisol in saliva by two to four times over the baseline (Kirschbaum et al., 1993), has become the most widely used method to study human response to acute stress in laboratory conditions (Dickerson & Kemeny, 2004; Kirschbaum et al., 1993). Therefore, the TSST was the stress induction method used in the current experiment.

Figure 11 depicts the flow of the TSST protocol used in the current study. The procedure started with the anticipation phase. The participants received instructions on their next task, which was to deliver a speech (without notes) on why they should be selected for a position of their choosing in front of a committee they had never met. The participants were told that the committee comprised two experts in psychology, specialized in non-verbal communication and that they would be recorded. The participants were allowed five minutes to prepare. They were allowed to prepare by taking notes but not allowed to bring the notes with them during their speech.

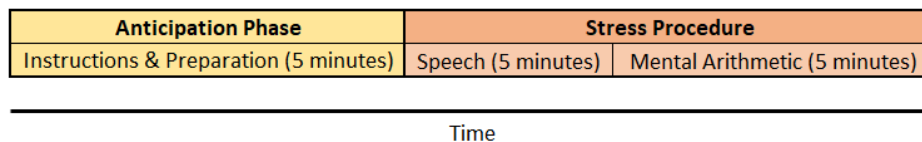


Figure 11. The TSST timeline. Adapted from Allen et al. (2017).

Following the five-minute preparation period, the participants were asked to leave their notes at the workstation and were escorted to another room by the researcher. In this room, the participants were directed to stand on a designated mark and face the committee. The committee then instructed the participants to begin their speech and started a five-minute timer. The committee members refrained from encouraging the participants, showing emotions, or interacting with them in any way outside the script. The committee members asked the participant to continue speaking if they stopped before the five-minute timer expired. If they stopped again, the committee members waited 30 seconds before speaking and were permitted to ask them a few interview questions (e.g., the biggest personal weakness in the workplace). After the five-minute speech, the committee gave the participants an unexpected five-minute mental arithmetic task, which introduced the element of uncontrollability. The task was to continuously subtract number 13 from number 1022 until the participant reached zero as fast and as accurately as possible. When the participant performed an incorrect subtraction, they were asked to start from the beginning. After five minutes of this mental arithmetic task, the participants were asked to leave the room and return to the workstation.

The participants in the control group (low-stress condition) have undergone a placebo version of the TSST (pTSST) (Het et al., 2009). The pTSST is similar to the TSST but is not intended to elicit a stress response because the two main stressors—uncontrollability and socio-evaluative threat—are removed. The participants were instructed to prepare to talk about any subject they wished. They were informed that they would be talking in a room for five minutes about the topic but that nobody would be able to hear them—they would be talking to themselves for five minutes. After the five-minute preparation period, the participants were escorted to another room by the researcher and left alone to talk for five minutes. After the speech, the researcher came back in and told the participant to continuously add by 15 aloud for five minutes. After the mental arithmetic task, the researcher retrieved the participant from the room and asked what number they reached to assess whether they followed the researcher's instructions. See Appendix A for the instructions given to the participants and the script for the TSST committee.

2. Cortisol as a Quantifier of Stress

When stressors are applied to an individual, specific behavioral and physiological changes are triggered through the release of stress hormones (Dickerson & Kemeny, 2004). One of the stress hormones, cortisol, can be used as a viable physiological measure of stress (Hermans et al., 2014). Cortisol was measured through saliva samples collected from the participants during the experiment to ensure the stress induction was successful.

In total, five saliva samples were collected. Sample #1 was collected as a baseline immediately after a ten-minute rest period. Sample #2 was collected at the beginning of the stress procedure (before the anticipation phase), sample #3 was collected at the end of the stress procedure, and samples #4 and #5 were collected at 10-minute intervals throughout the rest of the experiment.

Each experiment session was conducted in the afternoon because cortisol levels typically peak in the morning (Dickerson & Kemeny, 2004). Additionally, the participants were instructed to refrain from eating, drinking, smoking, chewing tobacco, or exercising one hour before the experiment as well as throughout the experiment as these activities could affect their salivary cortisol levels. The participants were also asked at the end of the experiment whether they adhered to these requests and encouraged to be honest as their answers would not negatively affect them in any way.

Saliva samples were collected periodically throughout the experiment via the passive drool method using cryovials and SalivaBio collection aids. This technique, developed by Salimetrics (n.d.), has become the preferred method of saliva collection and allowed the participants to provide up to 1.8 mL of saliva on their own. Each cryovial with a sample was immediately sealed and put in a freezer. After samples for nine participants were collected, the box with the samples was packaged according to Salimetrics shipping instructions and sent to a Salimetrics laboratory, where the samples were examined and the cortisol level in each sample captured. After the cortisol level analysis, the Salimetrics laboratory sent the data back to the researcher. Figure 12 depicts the cryovial used for saliva samples and the collection/storage box used for shipping. Each cryovial was identified by a participant identification number and a sample number.



Figure 12. A cryovial and a collection box used for saliva samples

Finally, the participants were asked to complete the state portion of the State-Trait Anxiety Inventory (STAI) twice throughout the experiment—once immediately following a 10-minute rest period and once after the stress induction. The goal of the STAI was to measure the participants’ level of stress at the particular moment by asking them to self-assess statements such as “I feel calm” or “I feel confused.” This was done to verify the stress induction was successful between the conditions. See Appendix B for the STAI.

C. MOTIVATION MANIPULATION

Motivation was manipulated between the subjects during the main pattern recognition task and the cognitive reflection test (CRT). Specifically, the participants in the motivation condition were informed before they began the CRT that they would earn \$0.50 for every correct answer. They were then told before the primary pattern recognition task that they would earn \$0.40 for a correct answer. They were also informed that the two decision aids would be earning points for every correct answer they suggest. The control group (non-motivation) participants were rewarded using the same rules and conditions, but they did not receive this information until completing the experiment. See Appendix C for the CRT instructions.

D. MATERIALS

1. Empatica E4 Wristband

The participants wore an Empatica E4 wristband during the experiment (Garbarino et al., 2014). The wristband is a wireless, battery-operated, wrist-monitoring device that allows for real-time streaming and visualization of several different types of physiological data (Empatica, n.d.). In this experiment, data collected via the wristband included heart rate variability, heart rate, and electrodermal activity. Heart rate and heart rate variability were collected by the built-in photoplethysmography sensors, while the electrodermal activity was captured by two electrodes attached to the bottom of the participant's palm connected to the built-in skin conductance sensor. Figure 13 depicts how the participants wore the wristband during the experiment.

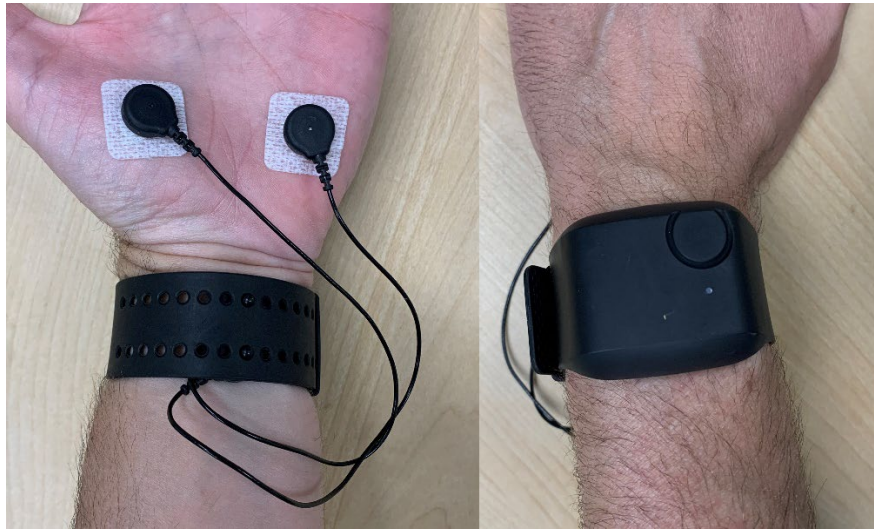


Figure 13. Empatica E4 heart rate monitor

2. Tobii Pro X3-120 Eye Tracker

Tobii Pro X3-120 eye tracker was used during the experiment to track the position and gaze of participants' eyes. This eye tracker was mounted on the workstation monitor and captured where the participants were looking and how long. The eye tracker uses infrared illuminators to capture information about participant's eyes and can also track eye

movement through prescription glasses (Tobii Pro, 2016). The sampling rate of the eye tracker is 120 Hz, and its wide capture area allowed the subjects to move around during the experiment and maintain tracking accuracy instead of sitting in a fixed position.

Eye-tracking involves complex mathematical calculations, which can consume significant processing resources of the data capturing computer. Therefore, the eye tracker was connected to an external processing unit via a USB 3 cable (Tobii Pro, 2015). This external processing unit performed all eye-tracking calculations and sent its output to the data capturing computer via an Ethernet cable. This configuration decreased the processing demands required from the data capturing computer. Figure 14 shows the eye tracker mounted at the bottom of the participants' screen and connected to the external processing unit via a USB 3 cable.

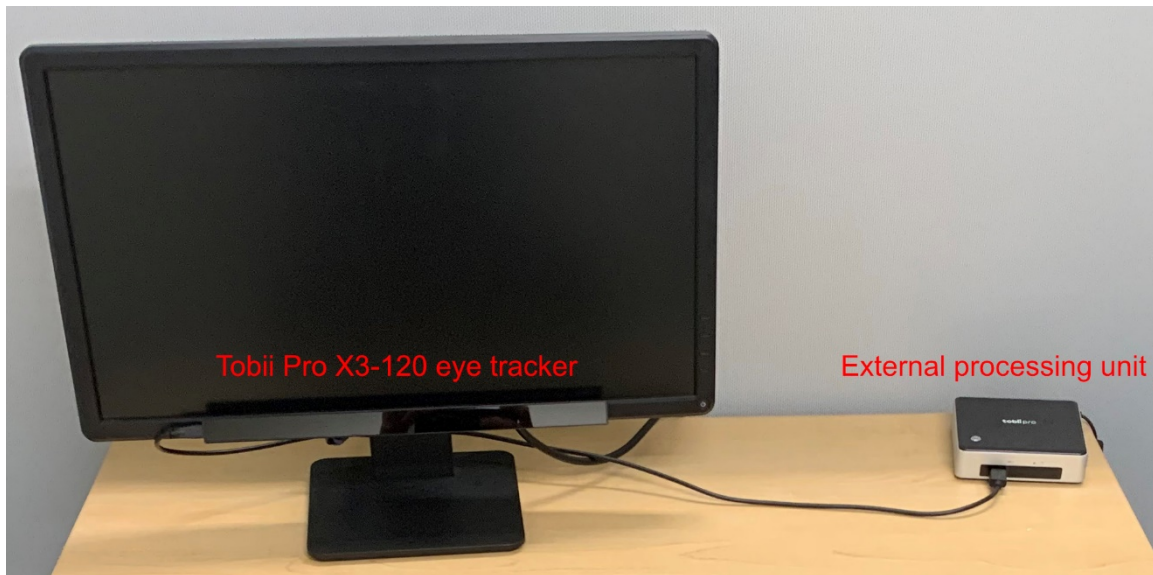


Figure 14. Tobii Pro X3-120 eye tracker with the external processing unit

Eye-tracking data, wristband sensor information, and the participant's screen were captured, combined, and synchronized using iMotions® Software (Version 8, iMotions Inc., Boston, MA, www.imotions.com). This software is an analysis platform that allows for the integration of multiple biometric sensors. It also includes an application

programming interface to import and export data from external sources. iMotions ran on the data capturing computer located at the researcher's desk.

3. Presentation Software

All computer-based tasks in the experiment were delivered using Presentation® Software (Version 21.0, Neurobehavioral Systems Inc., Berkley, CA, www.neurobs.com). This Windows application was developed for conducting psychological and neurobehavioral experiments. It allows researchers to create their own auditory and visual stimuli, deliver them to test subjects, and record their responses. The instructions, exercises, and questionnaires that the participants worked with during the experiment were written using Presentation Software scenarios. At the end of each experiment session, Presentation Software exported participants' keyboard responses into log files. These log files were in a plain text format and were used in the data analysis at the end of the experimentation period. As an example, Figure 15 depicts the scenario editor in Presentation Software.

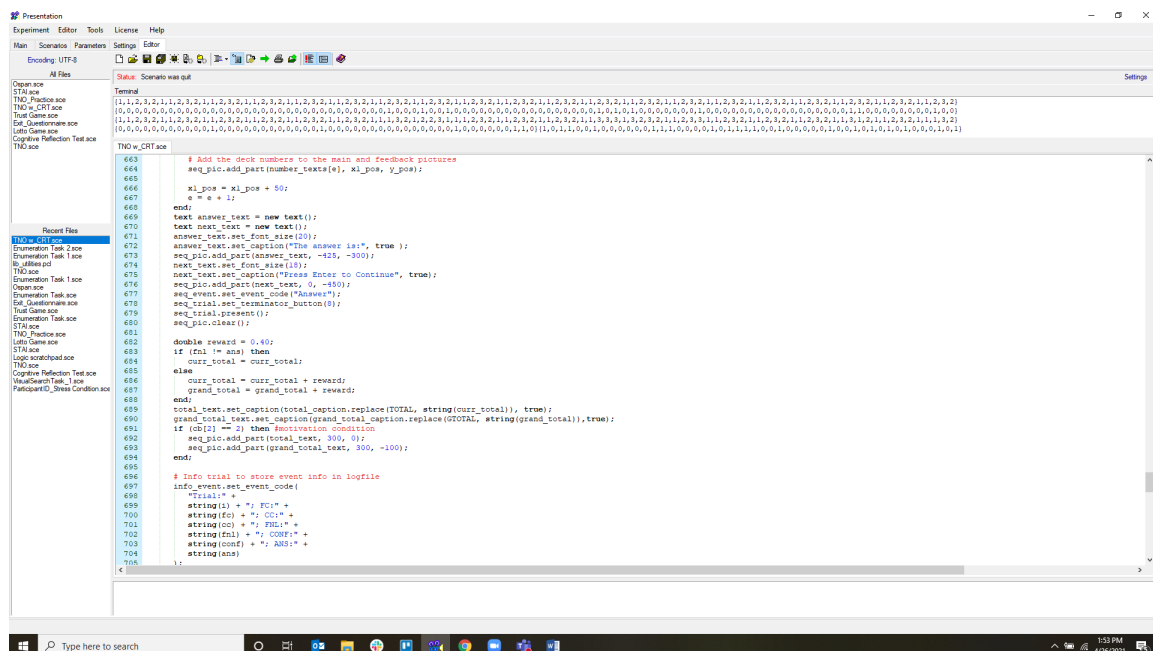


Figure 15. Presentation Software scenario editor

4. Laboratory Configuration

The laboratory comprises two rooms separated by a door. Room B was used for the speech and mental arithmetic tasks only. Room A was used for everything else and consisted of three individual areas separated by cubicle walls. Area A was a rest area where participants spent their 10-minute test period. This area included a sofa and a book library and was used for debriefs as well. Area A is depicted in Figure 16.



Figure 16. Room A, Area A—rest area

Area B contained the workstation that the participants used during the experiment. The workstation comprises a computer screen with an eye tracker, a keyboard, and an external processing unit for the eye tracker. The computer screen was connected to the keyboard/video/mouse switch at the researcher's desk, and the external processing unit was connected to the researcher's data capturing computer. Area B is depicted in Figure 17.



Figure 17. Room A, Area B—the participants' workstation

Area C was the researcher's space. This area included the data capturing computer with an external monitor. A keyboard/video/mouse switch was also connected to the computer, which allowed for switching of content on the participant's screen. At the beginning of the experiment, the participants were shown an eye-tracking calibration pattern. Once calibrated, the participants were shown the output of the computer-based tasks created in Presentation Software.

Room B was designated for the speech and mental arithmetic tasks. As shown in Figure 18, the room included one desk with two chairs where the two committee members sat for the TSST participants. Also visible in Figure 18 is a blue mark on the floor designating the place where the participants were directed to stand for the duration of the TSST procedure. Additionally, there were two cameras placed on the desk and aimed at the participants. The cameras were not recording any audio or video but were plugged into a power strip to give the impression they were functioning. For the pTSST participants, the cameras were not plugged in, the committee was not present, and the participants could sit anywhere in the room or walk around as they wished.



Figure 18. Room B

5. Pattern Learning Task

The main experimental task was a pattern learning and prediction task, originally created by van Dongen and van Maanen (2013). The participants were asked to find a pattern with the help of two decision aids—AI partners. The first AI partner was called Agile Learning for Enhanced Experimentation (ALEX), and the second partner was called Systematic Algorithm for Machine Learning (SAML). The participants were informed that the AI partners were being evaluated for their performance, and one would be recommended for integration into more complex systems. They were also informed that both decision aids were expected to have 90% reliability. However, SAML's actual reliability level was programmed to be only 60%. This deceptive instruction was implemented to observe whether the participants can distinguish between the two AI partners' actual versus expected reliability. The AI reliability levels were chosen on

purpose to be above and below 70%, which is a threshold generally considered to be the border between a reliable aid and an unreliable aid (Dixon & Wickens, 2006).

The task consisted of 100 trials (iterations). In each iteration, the participants were asked to predict the following number in the underlying pattern (i.e., 1, 2, or 3). This prediction required the participants to remember the results of the previous iterations. The pattern used in the task was a sequence of five digits (i.e., 1, 1, 2, 3, 2). This pattern was repeated over 100 iterations. However, due to the simplicity of the pattern, a 10% error was programmed into the sequence to ensure the participants could not detect the pattern with absolute certainty. This detail was not included in the instructions given to the participants.

First, the participants performed 20 practice iterations—10 iterations with each AI partner—to ensure their familiarity and understanding of the procedure. They started learning the pattern from the answers on each iteration. In the beginning, the participants were guessing until they started to see a pattern. The participants were also informed that the AI partners did not know the pattern and were learning with them. The participants typed one of three numbers (1, 2, or 3) in each iteration, depending on which one they thought came next. After that, the AI partner recommended which number it thought was next in the pattern. After seeing the AI partner's input, the participants made their final decision by pressing 1, 2, or 3 on the keyboard. Finally, the participants rated their confidence in their decision using a seven-point Likert scale, and the correct answer was displayed to them. Following the practice iterations, the participants performed the main task consisting of 100 iterations. As an example, Figure 19 depicts one iteration of the pattern learning task with the selected answers highlighted by red boxes.

Please make your prediction:

SAML's Recommendation:

Please make your final prediction:

AI Total:\$0.8

Your Total:\$0.8

Please rate your confidence in your final answer: 1 2 3 4 5 6 7

The answer is:

Press Enter to Continue

Progress: 3/100

Figure 19. An example of one pattern learning task iteration

6. Exit Questionnaire

At the end of the experiment, the participants were asked to fill out an exit questionnaire. The purpose of this voluntary questionnaire was to reveal factors that might have potentially altered the participant's cortisol levels during the experiment and to collect demographic data. Through the questionnaire, the participants self-reported information such as their engagement during the experiment, the current level of stress, or how much sleep they had the night before. Additional demographic questions included data such as age, gender, prior enlistment, or the current branch of service. See Appendix D for the exit questionnaire.

E. PROCEDURE

One experiment session lasted between 80 and 120 minutes, during which the participants worked alone in a separated area devoid of audio/visual distractions. Stress induction was performed in a room adjacent to the laboratory. Stress was assessed through physiological indicators such as heart rate variability, electrodermal activity, salivary

samples collected to measure cortisol levels, and subjective questionnaires to confirm the participants responded as intended to the stress and motivation manipulation. Additional data, such as demographics or perceived reliability, were collected via a questionnaire at the end of the experiment.

The main risk in the experiment was the potential for the participants to feel discomfort after stress induction due to the effects of mild stress. Physiological and psychological monitoring was implemented in the experiment to mitigate risk. The participants wore a wristband to monitor their heart rate and filled out self-reporting questionnaires. Additionally, all participants were informed before starting the experiment that they could initiate the termination of the session at any point. Finally, signs of confusion, panic, or other severe stress response symptoms were assessed through a subjective observation by the researcher.

1. Experiment

The experiment consisted of four phases. In the first phase, the participants self-scheduled for a session through a free web-based signup service. In the second phase, the participants received an introductory brief and signed all required forms. The third phase was dedicated to the actual experiment, and the final phase concluded the experiment with a verbal debrief and the re-signing of all required documents.

Each participant received an introductory email with instructions 24 hours before their scheduled time slot. In the same email, the participants also received the NPS informed consent form and the California bill of rights form to review. After the participants arrived, they were given another opportunity to read the two forms and ask questions. After signing the forms, the participants were given a verbal in-brief on the sequence of events and instructions on using their workstation. At this time, the heart rate monitor with two electrodes to sense electrodermal activity was fitted on the participant's wrist and hand, and a random participant identification number was assigned to them by the researcher. Next, the participants were asked to perform an eye-tracking calibration procedure on their workstation to ensure the eye-tracking sensor was calibrated to their height and distance from it.

Following the eye tracker calibration, the participants performed the Operation Span Task (OSPAN) to account for individual differences due to working memory capacity (Baddeley, 2013). See Appendix E for the OSPAN instructions. After the OSPAN, the participants were escorted to the rest area and told that they were to rest for 10 minutes because they may have felt some anxiety from the OSPAN task, and the researchers did not want that to affect their performance on the rest of the tasks. At this time, each participant was reminded to refrain from consuming any liquids or food for the duration of the experiment. After 10 minutes of resting, the participants were asked to provide the first saliva sample, which served as the baseline for their cortisol levels. Next, the participants were instructed to return to the workstation.

At the workstation, the participants completed the STAI to assess their current level of perceived stress. After the STAI, the participants completed a practice run of the pattern recognition task to familiarize themselves with the controls and the design of the task, and to start learning the pattern. Following the practice, the participants provided another saliva sample. At this point, the high-stress condition participants were given instructions for the TSST procedure, while the participants in the control group were given instructions for the pTSST. Once the instructions were administered and all questions answered, participants started the first phase of the TSST/pTSST protocol.

Once the TSST/pTSST protocol was complete, the participants returned to the workstation, where they provided another saliva sample, and a 10-minute timer for the next sample began. After providing the sample, the participants completed another STAI, followed by the CRT and the main pattern recognition task. After the pattern recognition task, the participants completed a trust game, which was designed to measure the participants' level of trust in the two AI partners. It also asked the participants to evaluate each AI partner's reliability level on a scale from 0% to 100%.

After the trust game, the participants completed an exit questionnaire. Once complete, the experiment concluded, data collection was stopped, and the participants were debriefed. During the debrief, the participants were informed about the deceptive elements and the true nature of the experiment and asked to sign the informed consent form again if

they wished to do so. Finally, the participants received an Amazon gift card valued at the amount they earned during the main pattern recognition task, the CRT, and the trust game.

2. Data and Collection Method

Before each session, a random identification number was assigned to the participant and noted in the experiment log. The log also contained additional information such as the motivation condition, the participants' earnings, and any notes on significant unplanned events during the experiment that could be useful during the final data analysis.

During each session, data from the wristband monitor, the eye tracker, and the participant's screen were collected using iMotions software. The main output of the software was a screen recording of the participant's screen with their eye-tracking data plotted over it. iMotions also output the wristband data and eye-tracking data into individual text files. Additionally, Presentation Software collected the participants' keyboard inputs and saved them in log files. Each log file (one per task per participant) was in a tab-delimited plain-text format and included other information such as participants' keyboard response times.

After each session, the log files were converted into the Microsoft Excel format using a Python script written by the principal investigator. The saliva samples were put in a cryogenic freezer and, once samples for nine participants were collected, shipped to a Salimetrics laboratory for analysis. After the experimentation period, all experiment data were combined in the International Business Machines (IBM) Statistical Package for Social Sciences (SPSS) Statistics Premium Grad Pack version 26 for the final processing.

IV. EXPERIMENT RESULTS

This chapter is divided into four sections. The first section covers the checks for stress manipulation built into the experiment to ensure the participants responded to the stress manipulation as intended. The second section presents the analysis of reliance on automation, which was the dependent variable in this experiment, and the main focus of this thesis. The third section examines the participants' reaction times during the experiment, and the fourth section provides an insight into the exploratory analyses conducted during the research.

A. STRESS MANIPULATION CHECK

1. Cortisol Manipulation

It is essential to assess the cortisol response to ensure the participants responded to stress as intended. Results indicate that there was a significant interaction between stress and cortisol levels, $F(4, 75) = 8.17, p < .001, \eta_p^2 = .30$, which suggests that the TSST elicited the desired stress response. Specifically, a significantly higher cortisol level was found in the high- than low-stress condition in both cortisol sample #4 and sample #5; $t(78) = 3.42, p = .001, d = 0.760$ and $t(78) = 4.32, p < .001, d = 0.957$, respectively. Table 2 provides descriptive statistics for each cortisol sample collected during the experiment separated by the stress conditions. Figure 20 depicts these values graphically.

Table 2. Descriptive statistics for cortisol samples between stress conditions

Cortisol Sample	Stress Condition	Mean	Std. Deviation	N	95% Confidence Interval	
					Lower Bound	Upper Bound
Sample 1	Low Stress	0.16	0.08	41	0.13	0.19
	High Stress	0.17	0.09	39	0.14	0.20
Sample 2	Low Stress	0.16	0.09	41	0.13	0.19
	High Stress	0.14	0.10	39	0.12	0.17
Sample 3	Low Stress	0.16	0.11	41	0.13	0.19
	High Stress	0.18	0.11	39	0.14	0.21
Sample 4	Low Stress	0.16	0.10	41	0.12	0.20
	High Stress	0.26	0.16	39	0.22	0.30
Sample 5	Low Stress	0.15	0.07	41	0.11	0.19
	High Stress	0.27	0.16	39	0.23	0.31

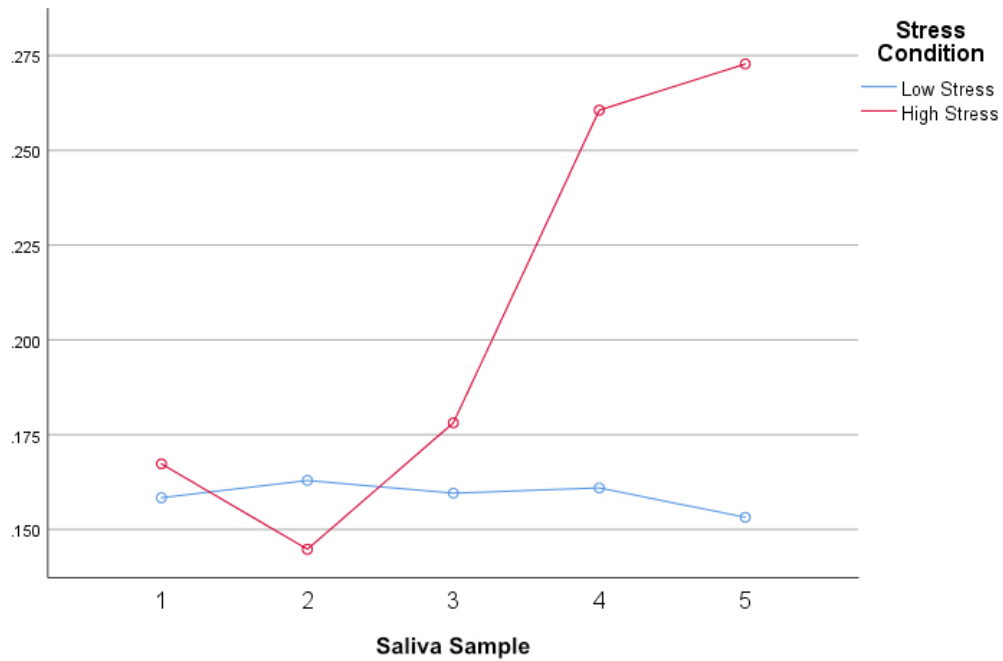


Figure 20. Cortisol level per sample between the stress conditions

2. STAI Manipulation

In addition to verifying the cortisol response, it is also important to assess the score difference between the subjective stress scores pre- and post-stressor. The difference in scores between the STAI administered before the stressor, and the STAI administered immediately after the stressor was analyzed between the stress conditions. An independent samples *t*-test revealed that the mean difference for the TSST participants was significantly higher ($M = 9.72$, $SD = 9.47$) than the mean of the STAI score difference for the pTSST participants ($M = .83$, $SD = 4.13$); $t(78) = 6.51$, $p < .001$. This finding indicates that the participants reported a higher increase in stress in the high- than low-stress conditions. Therefore, both subjective and objective measures of stress indicate that the TSST worked in raising stress levels, and the pTSST worked as a complimentary protocol that did not raise stress levels.

B. RELIANCE ANALYSIS

Reliance on automation was assessed from the pattern recognition task. Reliance was defined as how often the participant chose the AI partner's choice when there was a disagreement between the participant's initial answer and the AI partner's answer. The analysis revealed that there was a main effect of AI such that the participants relied on AI1 (the AI partner with higher reliability) more than AI2 (the AI partner with lower reliability), $F(1, 76) = 104.09$, $p < .001$, $\eta_p^2 = .58$. Table 3 provides mean, standard deviations, and confidence intervals. No other significant findings were found in this analysis.

Table 3. Reliance descriptive statistics for all participants

						95% Conf. Interval	
Reliance	Stress Condition	Motivation Condition	Mean	Std. Deviation	N	Lower Bound	Upper Bound
Reliance with AI1	Low Stress	No Motivation	0.75	0.20	20	0.65	0.84
		Motivation	0.74	0.23	21	0.65	0.83
	High Stress	No Motivation	0.62	0.22	19	0.52	0.71
		Motivation	0.78	0.18	20	0.69	0.88
		Total	0.72	0.21	80	0.67	0.77
Reliance with AI2	Low Stress	No Motivation	0.58	0.19	20	0.50	0.67
		Motivation	0.56	0.19	21	0.48	0.64
	High Stress	No Motivation	0.48	0.18	19	0.39	0.56
		Motivation	0.55	0.19	20	0.47	0.64
		Total	0.54	0.19	80	0.50	0.59

C. REACTION TIMES ANALYSIS

The time participants took to make their final decision in each iteration of the pattern recognition task can provide some valuable insight into their decision making. The participants' reaction times were measured as the duration between the point when a participant was asked to make their final prediction and the point when they selected their final answer. The analysis revealed that there was a main effect of AI such that there were longer reaction times with AI2 than AI1, $F(1, 76) = 29.98, p < .001, \eta_p^2 = .28$. Table 4 provides the descriptive statistics for this analysis.

Table 4. AI1 and AI2 reaction times

				95% Conf. Interval	
Reaction Time (RT)	Mean	Std. Deviation	N	Lower Bound	Upper Bound
RT with AI1	1100.62	315.55	80	1028.66	1168.74
RT with AI2	1231.16	419.98	80	1135.00	1320.68

Further analysis also revealed a main effect of agreement, $F(1, 76) = 175.04$, $p < .001$, $\eta_p^2 = .70$. That is, when the AI partner proposed the same answer as the participant (agreement), the participant's mean reaction time, reported in milliseconds, was faster ($M = 900.95$, $SD = 249.40$) than when the AI partner proposed a different answer than the participant (disagreement) ($M = 1385.44$, $SD = 451.22$). These results represent the aggregate mean reaction times with both AI partners, regardless of their reliability, and show that the participants spent more time deliberating when there was a disagreement. Table 5 provides the descriptive statistics for this analysis.

Table 5. Agreement and disagreement reaction times

				95% Conf. Interval	
Reaction Time (RT)	Mean	Std. Deviation	N	Lower Bound	Upper Bound
RT with Agreement	900.95	294.40	80	844.40	953.95
RT with Disagreement	1385.44	451.22	80	1281.87	1483.39

The results also indicate that there was a main effect of reliance, $F(1, 69) = 16.71$, $p < .001$, $\eta_p^2 = .195$ such that reaction times were faster when the participants decided to rely on the AI partners than when they decided to not rely on them. Table 6 provides the breakdown of reaction times per AI partner between the reliance conditions.

Table 6. Reaction times per AI partner between reliance conditions

				95% Conf. Interval	
Reliance Condition	Mean	Std. Deviation	N	Lower Bound	Upper Bound
Reliance on AI1	1343.77	423.17	73	1244.33	1442.35
Reliance on AI2	1470.40	490.60	73	1355.17	1583.49
Non-reliance on AI1	1616.10	690.75	73	1466.97	1777.99
Non-reliance on AI2	1535.58	617.88	73	1396.23	1680.78

In addition to the main effects, there were several significant interactions: (a) reliance and motivation, $F(1, 69) = 8.58$, $p = .005$, $\eta_p^2 = .11$; (b) reliance, stress, and motivation, $F(1, 69) = 4.04$, $p = .048$, $\eta_p^2 = .06$; (c) reliance and AI, $F(1, 69) = 8.07$, $p =$

.006, $\eta_p^2 = .11$; and (d) reliance, AI, and motivation, $F(1, 69) = 4.18, p = .045, \eta_p^2 = .06$. In follow-up t -tests it was revealed that motivation had an effect on reliance decision times, such that in the motivation condition reaction times were slower when participants chose not to rely on the AI partner's choice ($M = 1630.25, SD = 498.73$) than when they chose to rely on the AI partner's choice ($M = 1360.12, SD = 441.39$); $t(40) = 3.88, p < .001$.

Finally, the effect of AI beyond reliance and motivation was explored. A significant interaction between reliance and non-reliance for AI1 in the motivation condition was found. In the motivation condition with AI1, the time it took the participants to make their final choice was faster when they decided to rely on AI1 ($M = 1344.47, SD = 381.27$) than when they decided *not* to rely on AI1 ($M = 1817.07, SD = 755.46$); $t(35) = 4.31, p < .001$. Additionally, a significant interaction between AI1 and A2 for reliance decisions in the motivation condition was found. In the motivation condition, the time it took the participants to make their final choice was faster when they decided to rely on AI1 ($M = 1285.50, SD = 401.80$) than when they decided to rely on AI2 ($M = 1482.49, SD = 568.97$); $t(40) = 3.51, p = .001$.

D. EXPLORATORY ANALYSES

Gender differences have been explored in the decision-making literature (Starcke & Brand, 2016). One of the goals in this experiment was to attract enough female participants to explore whether there were gender differences. While this goal was not met, it is interesting to look at the analysis with only the male participants. The analysis of only male participants revealed significant results. Besides the effect of AI mentioned in the main reliance analysis, the results indicate that there was a significant interaction between AI reliability and motivation, $F(1, 57) = 4.10, p = .048, \eta_p^2 = .07$, for the male participants. Additionally, the interaction between AI reliability, motivation, and stress was marginally significant for the male participants, $F(1, 57) = 3.91, p = .053, \eta_p^2 = .06$. Table 7 provides the descriptive statistics for this analysis.

Table 7. Descriptive statistics for male participants only

						95% Conf. Interval	
AI Reliability	Stress Condition	Motivation Condition	Mean	Std. Deviation	N	Lower Bound	Upper Bound
Reliance with high reliability AI	Low Stress	No Motivation	0.75	0.20	16	0.66	0.85
		Motivation	0.73	0.26	17	0.64	0.83
	High Stress	No Motivation	0.60	0.19	14	0.50	0.71
		Motivation	0.85	0.09	14	0.74	0.95
		Total	0.73	0.21	61	0.68	0.79
Reliance with low reliability AI	Low Stress	No Motivation	0.59	0.20	16	0.50	0.68
		Motivation	0.57	0.20	17	0.48	0.65
	High Stress	No Motivation	0.51	0.12	14	0.41	0.60
		Motivation	0.59	0.15	14	0.50	0.69
		Total	0.56	0.17	61	0.52	0.61

Follow-up *t*-tests revealed that in the high stress group, reliance was higher in the motivation than no motivation condition; $t(26) = 3.44$, $p = .002$, $d = 1.30$. Further breakdown showed that the effect of motivation was significant specifically in the high-stress condition for AI1; $t(26) = 4.34$, $p < .001$, $d = 0.584$. Figure 21 depicts the effect of motivation on reliance for the male participants under stress.

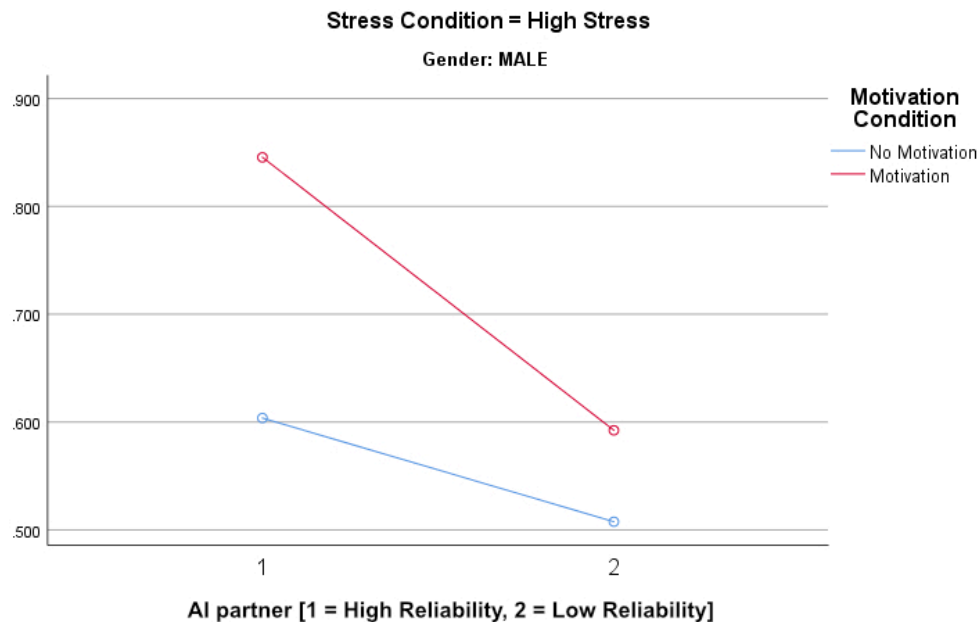


Figure 21. Effect of motivation on male reliance under stress

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

This study investigated how humans make reliance decisions when they are working with an AI partner. Specifically, the laboratory experiment explored the effect of stress, motivation, and automation reliability on the participants' reliance decisions. The participants were exposed to either a high- or low-stress condition. Additionally, some participants were also exposed to a motivation condition and were informed that they would receive a monetary reward for each correct answer. The main pattern recognition task required the participants to work with two AI partners that differed in reliability. In each iteration of the main task, the participants had to choose between their answer or the AI partner's answer. The findings are discussed in the sections below.

A. FINDINGS DISCUSSION

1. Reliance

The results indicate that no significant evidence was found to support this study's four hypotheses in the measure of reliance. Although some interactions were approaching the significance level, neither stress nor motivation were found to have a significant impact on how the participants relied on the AI partners. However, the AI partners' reliability was found to be the main driver of the participants' reliance decisions such that there was a clear distinction between AI1 and AI2. The total reliance on AI1 was 25% higher than on AI2. This finding indicates that humans rely on a high-reliability teammate more than on a low-reliability one.

The lack of evidence to support this study's hypotheses might be attributed to the fact that the final number of volunteers that participated in the experiment was lower than initially planned. This fact was mainly due to the Coronavirus restrictions put in place in the second half of the experimentation period requiring all experimentation to be suspended. It is possible that with a larger sample of participants, some interactions that are currently approaching the significance level could cross the threshold and become significant.

Additionally, the lack of female participants made it impossible to determine the presence of any gender differences or the impact of the female population on the overall reliance decisions data. However, some of these marginally significant interactions became significant when data for female participants were removed from the sample. Although this approach made the analysis scope narrower than initially intended, it led to a more homogenous sample, which provided interesting results. Mainly, motivation was the primary driver of the male participants' reliance decision. Overall, motivation caused the male participants' mean reliance on AI1 in both stress conditions to fluctuate from 0.60 (high stress, no motivation) to 0.85 (high stress, motivation). For reliance on AI2, motivation had a minimal effect as reliance only ranged from 0.51 (high stress, no motivation) to 0.59 (high stress, motivation).

Further analysis revealed that when motivated, the male participants relied on the high-stress condition on AI1 31% more than on AI2. This result was a substantial difference from when the male participants were not motivated in the high-stress condition, in which case they relied on AI1 only 15% more than on AI2. Additionally, as can be seen in Table 7, the reliance values in the high-stress condition with motivation present approximated the AI partners' true reliability. In this case, the mean of reliance on AI1 for the male participants was 0.85, which was close to AI1's actual reliability of 90%, while the mean of reliance on AI2 was 0.59, which was only one percentage point away from its actual reliability of 60%.

2. Reaction Times

While reliance reflects the decision that was made, reaction times provide insight into the decision-making process. Overall reaction times for the participants, regardless of gender, were 11% slower with AI2 than with AI1. This result indicates that a higher-reliability machine teammate can accelerate one's decision-making cycle, while a low-reliability machine teammate might introduce an additional delay. Furthermore, when there was a disagreement between the participant's proposed answer and the AI partner's answer, the reaction time was 35% slower than when both answers matched. This finding indicates

that the participants were more deliberate in their decision making when faced with a disagreement from the AI partner, regardless of its reliability.

Next, the analysis revealed that reaction times were faster when the participants decided to rely on the AI partners than when they decided to not rely on them. However, differences were observed between AI1 and AI2. Specifically, when the participants decided to rely on AI1, reaction times were 17% faster than when they decided not to rely on AI1. However, the decision to rely on AI2 was made only 4% faster than the decision not to rely on AI2. The results also highlighted the effect of motivation on reliance decision reaction times. Notably, the reaction time when making the decision whether to rely or not to rely on the AI partner was 74% slower when the participants were motivated than when the participants were not motivated. This finding suggests that, when motivated, the participants were engaged in more deliberate thinking when there was a disagreement between their proposed answer and the AI partner's proposed answer. Furthermore, when motivated, the reaction time for the final choice was 26% faster when the participants decided to rely on AI1 than when they decided not to rely on AI1. Finally, the time the participants took to rely on the AI partner was 13% faster for AI1 than for AI2. These findings could indicate that the participants questioned the reliability of AI2 in general. In other words, the participants were more hesitant to accept AI2's suggestions, which could imply that the amount of subjective trust they had with AI2 was smaller than with AI1.

3. Stress

As Figure 20 shows, the stress manipulation using the TSST protocol was successful. The cortisol level began to increase after stress induction and continued to rise for the high-stress participants. On average, the cortisol level in the last sample for the high-stress participants was 44% higher than for the low-stress participants. The effectiveness of stress induction was also confirmed by the difference in scores between the STAI administered before stress induction, and the STAI administered immediately after stress induction. On average, the STAI score difference for the high-stress participants was 92% higher than the score difference for the low-stress participants. This finding suggests that the high-stress participants felt substantially more stressed after stress

induction than the low-stress participants, confirming that the TSST protocol elicited the intended stress response.

As seen in Table 7, the reliance of motivated male participants in the high-stress condition approximates the reliability of the AI partners. This finding could indicate that the cognitive abilities of the high-stress male participants were not as impaired by elevated cortisol as expected. In other words, the TSST protocol might have increased the male participants' stress level to an optimal level that improved their cognitive performance. This theory would be in line with a study conducted by Mendl (1999), in which the author stated that "an inverted U-shaped relationship exists between an individual's state of stress or arousal and its ability to perform a cognitive task effectively, the so-called Yerkes–Dodson law, is commonly encountered" (p. 221). Figure 22 helps in visualizing this relationship. The fact that 56% of all participants indicated that they had undergone some type of stress training in their careers could be another contributing factor to this theory.

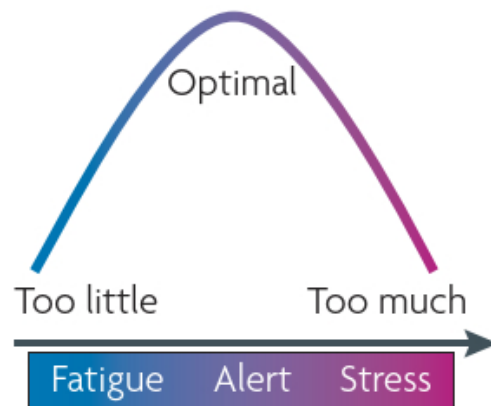


Figure 22. Yerkes-Dodson law. Source: Arnsten (2009).

In summary, AI reliability was the primary driver of the participants' reliance decisions in this experiment. The more reliable the AI partner was, the more the participants relied on it. AI reliability also affected the speed of the participants' decision making such that the participants took longer to arrive at their final decision with the low-reliability AI partner. This finding is also vital to the design of machine teammates since a machine with high reliability can help warfighter decide and act faster than the enemy. For

male participants, motivation significantly contributed to reliance decisions with AI1. Specifically, motivation not only substantially increased male reliance on AI1 in the high-stress condition, but it also led to appropriate reliance, which approximated the AI partners' reliability. This finding indicates the necessity of appropriate motivation of human operators because insufficient motivation might lead to automation disuse.

B. LIMITATIONS

The primary limitation of this study was the lack of participants, which resulted in fewer data collected than initially intended. This fact was mainly due to the unforeseen event of the Coronavirus pandemic, which ended the laboratory experiment prematurely. Unanticipated changes to the participants' work schedules or emergencies also contributed to the lower-than-intended participation.

Another limitation was that this study recruited volunteers from NPS only, which means the vast majority of the volunteers participating in the experiment were members of the U.S. military. This sample was not representative of an average individual, and the analysis results cannot be extrapolated to the general population for several reasons. Firstly, the vast majority of participants were male. Secondly, the mean age of a participant was only about 32.5 years old. Finally, more than half of all participants indicated that they had undergone some type of stress training before in their lives. Additionally, most participants were mid-grade officers with only several participants identifying as enlisted members. This composition resulted in a skewed sample that is not representative of the overall military population, in which most members are enlisted.

C. RECOMMENDATIONS

This study intended to explore several phenomena and analyze various factors related to human-machine teaming, which required over 100 participants to ensure statistical significance across the board. However, due to the reasons mentioned in the Limitations section, the final number of participants was only 80. Therefore, the primary recommendation for future research is to conduct this experiment with more participants. Extra effort should be made to recruit specifically female participants in order to have a more balanced representation of the two genders.

The analysis of reaction times in this study provided valuable insight into the participants' decision-making process. Likewise, a secondary analysis of gaze times and pupil dilation data collected during this experiment could prove helpful in analyzing participants' decision making. Such analysis might reveal deficiencies in interface design and provide suggestions for future improvements.

Follow-on research could also explore demographics differences. For example, research shows differences in decision making between males and females under stress (van den Bos et al., 2009). With more female participants, such gender differences in reliance decisions under stress could be explored in the context of human-machine teaming. As more females enter the military service, such research becomes valuable. Furthermore, future research should focus on the effects of stress on reliance decisions in the general population instead of limiting the sample to members of the U.S. military.

Additionally, more than one-half of the participants indicated having received some stress training. An exploratory analysis of differences between the participants with prior stress training and the participants without any stress training could also yield valuable insight into reliance decisions. As discussed in Chapter V, prior stress training might have affected the stress induction in this experiment such that the participants were in the optimal performance part of the U-shaped curve shown in Figure 22. Therefore, conducting this experiment outside a laboratory setting, such as during a simulated or even live-fire exercise, could provide a more realistic picture of the effect of stress on reliance decisions. Such results could then inform decision makers on the importance of stress training for humans working with machines.

D. CONCLUSION

The importance of smart machines for national defense contributed to the increased proliferation of HMTs into the U.S. military and its combat operations (Singer, 2009). Because stress is an integral part of combat and can negatively affect human performance (Sandi, 2013), the primary aim of this thesis was to explore the effect of stress on human reliance on automation. Additionally, research shows that motivation affects human cognitive performance (Locke & Braver, 2008). Therefore, this thesis also aimed to

investigate the effect of motivation on reliance decisions in both low- and high-stress conditions. Finally, automation reliability has been shown to affect human reliance on automation (Parasuraman & Riley, 1997). Therefore, both high- and low-reliability machine teammates were incorporated into the study. The results suggest that high machine reliability caused higher reliance and shortened decision reaction times. Further, the effect of motivation on reliance decisions was found to be significant for male participants such that motivation promoted appropriate reliance according to reliability. This thesis contributed to the body of knowledge on human factors and human-machine teaming. The results of this study expand the DOD's overall understanding of the HMT dynamics and can help inform the DOD decision makers. Future policies based on the findings from studies such as this one can improve the design of unmanned systems, training scenarios, and missions.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. TSST PROTOCOL INSTRUCTIONS

TSST Participant Instructions

“For the next task you will not be using a computer. Your task is to imagine yourself as a job applicant for a managerial position. You will present a speech to a committee, with the goal of convincing them you are the best candidate for the position. The speech should be five minutes long. There will be analyses run post-interview on vocal frequency and non-verbal behavior. You will have five minutes to prepare your speech. You are allowed to write down your ideas in those five minutes, but not allowed to take any notes with you. The committee is comprised of two doctoral students, one with expertise in organizational leadership, the other in behavioral sciences, specializing in non-verbal communication. There will also be two cameras pointed at you to record your performance. After your speech, you will be given a second task by the committee.”

pTSST Participant Instructions

“For the next task you will not be using a computer. You are going to talk about any topic you wish (e.g., recent/favorite movie, hobbies, recent vacation) for 5 minutes. Right now, you are going to spend 5 minutes thinking about that topic. When five minutes are up, you will go into the room and talk about it. No one will hear you, and you will not be recorded. Also, after your five minutes of talking are up, you will be given a second five-minute task.”

After five minutes:

“Your next task is to continuously add 15, starting at zero (i.e., 0, 15, 30, etc.).”

After the next five minutes:

“What number have you reached when counting by 15?”

TSST Committee Instructions

When a participant enters, direct them to the X on the floor. When participant is ready:

“You may begin your speech.”

If the participant drifts from the X on the floor, direct them back.

If they stop before five minutes:

“You still have some time left. Please continue.”

If the participant stops again, wait 20s, then ask questions:

“What would you say is your biggest weakness?”

“How do you handle conflict?”

“Where do you see yourself in 5 years?”

“What is something that is unique about you?”

At five minutes:

“That is enough, thank you. We would now like to give you a second task. Starting at 1022, continuously subtract by 13 until you get to 0. Please do this as quickly and accurately as possible.”

If the participant makes an error:

“That is incorrect, please start again from 1022.”

After five minutes:

“Thank you, your time is up. Please return to the other room.”

APPENDIX B. STATE-TRAIT ANXIETY INVENTORY

A number of statements which people have used to describe themselves are given on the following screen. Read each statement and type in the appropriate numerical response (i.e., 1 – Not at all, 2 – Somewhat, 3 – Moderately so, 4 – Very much so) to indicate how you feel right now, that is, at this moment. There is no right or wrong answer. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

Statement 1: I feel calm.

Statement 2: I feel secure.

Statement 3: I am tense.

Statement 4: I feel strained.

Statement 5: I feel at ease.

Statement 6: I feel upset.

Statement 7: I am presently worrying over possible misfortunes.

Statement 8: I feel satisfied.

Statement 9: I feel frightened.

Statement 10: I feel comfortable.

Statement 11: I feel self-confident.

Statement 12: I feel nervous.

Statement 13: I am jittery.

Statement 14: I feel indecisive.

Statement 15: I am relaxed.

Statement 16: I feel content.

Statement 17: I am worried.

Statement 18: I feel confused.

Statement 19: I feel steady.

Statement 20: I feel pleasant.

APPENDIX C. COGNITIVE REFLECTION TEST INSTRUCTIONS

Non-motivation Condition Instructions

In this task, you will be presented with 3 questions. Please type the answer and press enter to continue. Do not spend too much time on each question.

Motivation Condition Instructions

In this task, you will be presented with 3 questions. Please type the answer and press enter to continue. Do not spend too much time on each question. You will receive \$0.50 for each correct answer.

CRT Questions

Question 1:

A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

Question 2:

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

Question 3:

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. EXIT QUESTIONNAIRE

There are many factors that may affect cortisol levels, and therefore the ability to accurately interpret cortisol analysis. The questions in this section are included to assess for those factors. As a reminder, your responses are completely anonymous and voluntary, and will not affect any earnings you received as a part of your participation.

Question 1:

In the past week, how much stress have you experienced due to classwork/exams?

Please rate on a 1–9 scale where 1 = no stress at all, and 9 = extreme stress

Question 2:

Today, how much stress do you feel as a result of classwork/exams?

Please rate on a 1–9 scale where 1 = no stress at all, and 9 = extreme stress

Question 3:

How many hours of sleep did you get last night?

Question 4:

Is this more/less/same as usual?

Question 5:

What time did you wake up today?

Question 6:

Did you eat, drink, or smoke less than an hour before the experiment? If so, please list all that apply.

Question 7:

Do you smoke? If so, how many cigarettes a day?

Question 8:

How much alcohol do you consume in a typical week?

Question 9:

Please list any and all medications or drugs that you take, and how often.

Question 10:

Do you have now, or a history of high cholesterol?

Question 11:

Do you have diabetes?

Question 12:

Do you have now, or a history of high blood pressure?

Question 13:

Do you take prescribed steroids (e.g., asthma inhaler)?

Question 14:

Do you have any heart conditions? If so, what?

Question 15:

Do you have any illnesses? If so, what?

Question 16:

Overall, do you consider yourself a healthy person?

Question 17:

How often do you exercise?

Question 18:

For female participants: Are you on oral contraceptives?

Question 19:

For female participants: When was the first day of your last period?

APPENDIX E. OPERATION SPAN TASK

In this task, we are interested in how well you can remember a sequence of stimuli. Overall, you will see a math problem and then a letter, and at the end of each trial you need to recall the letters in the order they were presented. The training will be broken down into 3 sections:

1. Practice with letter recall
2. Practice with equation solving
3. Practice with both together

Practice #1 – Letter Recall

You will be presented with a sequence of letters; each letter will be shown for 800 milliseconds before moving on to the next letter. Afterwards, you will be prompted to enter the letters in the order in which they were presented.

Practice #2 - Equations

This practice is meant to familiarize you with the math portion on the task. You will be shown 10 equations.

1. You will see the equation. Once you solve the equation, press SPACEBAR.
2. You will next see an answer.

If the answer is correct, press 1 for true.

If the answer is incorrect, press 2 for false.

Practice #3 - Equations and Letters

This practice is meant to familiarize you with the entire sequence of the task. The entire sequence is:

1. You will see an equation, when you solve the equation, press SPACEBAR.
2. You will then decide if the given answer is True or False.
3. You will be presented with a letter to remember.
4. Repeat.

After the set of equations, you will be prompted to recall the letters in the order they were presented. You will then receive feedback on your math and recall accuracy. It is important to answer the equations quickly but accurately. During the math practice portion your average solve time was calculated.

Your maximum time allowed to solve the equation is your average solve time plus 2.5 standard deviations. If you take too long to solve the equation or fall below 85% in equation accuracy, the set will start over.

Main Task

You have now completed the practice trials and will move on to the actual task. Each set is 3–7 trials long. You will complete 15 sets in total. You are not allowed to write anything down throughout this task. If you have any questions, please ask at this time. Press ENTER when you are ready to begin.

LIST OF REFERENCES

- Air Force Research Laboratory. (2015, November 25). *Human machine teaming*. Defense Innovation Marketplace. https://defenseinnovationmarketplace.dtic.mil/wp-content/uploads/2018/02/88ABW-2014-1504_140312v2_-_Autonomy_Strategy_Part_2.pdf
- Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier Social Stress Test: Principles and practice. *Neurobiology of Stress*, 6, 113–126. <https://doi.org/10.1016/j.ynstr.2016.11.001>
- Alloway, T. P., & Alloway, R. (2012). *Working memory: The connected intelligence* (1st ed.). Psychology Press. <https://ebookcentral-proquest-com.libproxy.nps.edu/lib/ebook-nps/detail.action?pq-origsite=primo&docID=1517587>
- Antsaklis, P. J., & Rahnama, A. (2018). Control and machine intelligence for system autonomy. *Journal of Intelligent & Robotic Systems*, 91(1), 23–34. <https://doi.org/10.1007/s10846-018-0832-6>
- Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews. Neuroscience; London*, 10(6), 410–422. <http://dx.doi.org/10.1038/nrn2648>
- Arnsten, A. F. T., & Goldman-Rakic, P. S. (1998). Noise stress impairs prefrontal cortical cognitive function in monkeys: Evidence for a hyperdopaminergic mechanism. *Archives of General Psychiatry*, 55(4). <https://doi.org/10.1001/archpsyc.55.4.362>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63(1), 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, A. (2013). *Essentials of human memory* (1st edition). Psychology Press.
- Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)

- Balleine, B. W. (2007). The neural basis of choice and decision making. *Journal of Neuroscience*, 27(31), 8159–8160. <https://doi.org/10.1523/JNEUROSCI.1939-07.2007>
- Banis, S., & Lorist, M. M. (2012). Acute noise stress impairs feedback processing. *Biological Psychology*, 91(2), 163–171. <https://doi.org/10.1016/j.biopsycho.2012.06.009>
- Bensoussan, B. E. (2008). *Analysis without paralysis: 10 tools to make better strategic decisions*. FT Press.
- Cebrowski, A. K., & Garstka, J. H. (1998). Network-centric warfare: Its origin and future. *U.S. Naval Institute Proceedings*, 124(1), 28–35.
- Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00401>
- Clark, M., Alley, J., Deal, P., DePriest, J., Hansen, E., Heitmeyer, C. ... Corey, M. (2015). *Autonomy community of interest (COI) test and evaluation, verification and validation (TEVV) working group: Technology investment strategy 2015–2018* (No. AD1010194; p. 19). Office of the Assistant Secretary of Defense (research and engineering) Washington. <https://apps.dtic.mil/sti/citations/AD1010194>
- Crevelde, M. V. (1985). *Command in war*. Harvard University Press.
- Davis, N. C. (1997). An information-based revolution in military affairs. In D. Ronfeldt and J. Arquilla (Eds.), *In Athena's camp: Preparing for conflict in the information age* (1st ed., pp. 79–98). RAND Corporation. <https://doi.org/10.7249/mr880osd-rc.9>
- de Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience*, 6(6), 463–475. <http://dx.doi.org.libproxy.nps.edu/10.1038/nrn1683>
- de Visser, E. J. de, Pak, R., & Shaw, T. H. (2018). From ‘automation’ to ‘autonomy’: The importance of trust repair in human–machine interaction. *Ergonomics*, 61(10), 1409–1427. <https://doi.org/10.1080/00140139.2018.1457725>
- de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F. ... Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331–349. <https://doi.org/10.1037/xap0000092>
- Dhillon, B. S. (1986). *Human reliability: With human factors*. Pergamon Press.

- Dhillon, B. S. (2009). *Human reliability, error, and human factors in engineering maintenance: With reference to aviation and power generation*. CRC Press. <https://doi.org/10.1201/9781439803844>
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3), 355–391. <https://doi.org/10.1037/0033-2909.130.3.355>
- Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors*, 48(3), 474–486. <https://doi.org/10.1518/001872006778606822>
- 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. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)
- Empatica. (n.d.). *E4 wristband*. Retrieved April 26, 2021, from <https://www.empatica.com/research/e4/>
- Ferdinand, N. K., & Czernochowski, D. (2018). Motivational influences on performance monitoring and cognitive control across the adult lifespan. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01018>
- Gao, J., & Lee, J. D. (2006). Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 36(5), 943–959. <https://doi.org/10.1109/TSMCA.2005.855783>
- Garbarino, M., Lai, M., Tognetti, S., Picard, R., & Bender, D. (2014). Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming Healthcare through Innovations in Mobile and Wireless Technologies."* 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming healthcare through innovations in mobile and wireless technologies," Athens, Greece. <https://doi.org/10.4108/icst.mobihealth.2014.257418>
- Gathmann, B., Schulte, F. P., Maderwald, S., Pawlikowski, M., Starcke, K., Schäfer, L. ... Brand, M. (2014). Stress and decision making: Neural correlates of the interaction between stress, executive functions, and decision making under risk. *Experimental Brain Research*, 232(3), 957–973. <https://doi.org/10.1007/s00221-013-3808-6>
- Greengard, S. (2009). Making automation work. *Communications of the ACM*, 52(12), 18–19. <https://doi.org/10.1145/1610252.1610261>

- Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, 8(3), 483–500. <https://doi.org/10.1075/is.8.3.10gro>
- Gunning, D., & Aha, D. W. (2019). DARPA’s explainable artificial intelligence program. *AI Magazine*, 40(2), 44–58.
- Hemp, P. (2009). Death by information overload. *Harvard Business Review*, 87(9), 82–89.
- Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences*, 37(6), 304–314. <https://doi.org/10.1016/j.tins.2014.03.006>
- Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., & Wolf, O. T. (2009). Neuroendocrine and psychometric evaluation of a placebo version of the ‘Trier Social Stress Test.’ *Psychoneuroendocrinology*, 34(7), 1075–1086. <https://doi.org/10.1016/j.psyneuen.2009.02.008>
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan. https://books.google.com/books?id=SHvzzuCnuv8C&lr=&source=gbs_navlinks_s
- Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The ‘Trier Social Stress Test’—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1–2), 76–81. <https://doi.org/10.1159/000119004>
- Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270. <https://doi.org/10.1080/00140139208967392>
- 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
- Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: Behavior, brain activation, and individual differences. *Cognitive, Affective, & Behavioral Neuroscience*, 8(1), 99–112. <https://doi.org/10.3758/CABN.8.1.99>
- Lupien, S. J., Maheu, F. S., Tu, M. T., Fiocco, A. J., & Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and Cognition*, 65(3), 209–237. <https://doi.org/10.1016/j.bandc.2007.02.007>
- Lupien, S. J., Ouellet-Morin, I., Hupbach, A., Tu, M. T., Buss, C., Walker, D. ... McEwen, B. S. (2015). Beyond the stress concept: Allostatic load—a developmental biological and cognitive perspective. In *Developmental Psychopathology* (Vol. 2, pp. 578–628). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470939390.ch14>

- Lyons, J. B., & Stokes, C. K. (2012). Human–human reliance in the context of automation. *Human Factors*, 54(1), 112–121. <https://doi.org/10.1177/0018720811427034>
- Lyons, J. B., Wynne, K. T., Mahoney, S., & Roebke, M. A. (2019). Trust and human-machine teaming: A qualitative study. In W. Lawless, R. Mittu, D. Sofge, I. S. Moskowitz, and S. Russell (Eds.), *Artificial intelligence for the Internet of Everything* (1st edition, pp. 101–116). Academic Press. <https://doi.org/10.1016/B978-0-12-817636-8.00006-5>
- Margittai, Z., Nave, G., Strombach, T., van Wingerden, M., Schwabe, L., & Kalenscher, T. (2016). Exogenous cortisol causes a shift from deliberative to intuitive thinking. *Psychoneuroendocrinology*, 64, 131–135. <https://doi.org/10.1016/j.psyneuen.2015.11.018>
- Mason, J. W. (1968). A review of psychoendocrine research on the sympathetic-adrenal medullary system. *Psychosomatic Medicine*, 30(5), 631–653.
- Massar, S. A. A., Lim, J., Sasmita, K., & Chee, M. W. L. (2016). Rewards boost sustained attention through higher effort: A value-based decision making approach. *Biological Psychology*, 120, 21–27. <https://doi.org/10.1016/j.biopsycho.2016.07.019>
- Matteson, S. (2019, June 7). *Autonomous versus automated: What each means and why it matters* [Technology News]. TechRepublic. <https://www.techrepublic.com/article/autonomous-versus-automated-what-each-means-and-why-it-matters/>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.2307/258792>
- McFadden, C. (2018, November 6). *A brief history of military robots including autonomous systems*. <https://interestingengineering.com/a-brief-history-of-military-robots-including-autonomous-systems>
- McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 60(2), 262–273. <https://doi.org/10.1177/0018720817743223>
- Mendl, M. (1999). Performing under pressure: Stress and cognitive function. *Applied Animal Behaviour Science*, 65(3), 221–244. [https://doi.org/10.1016/S0168-1591\(99\)00088-X](https://doi.org/10.1016/S0168-1591(99)00088-X)
- Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5), 527–539. [https://doi.org/10.1016/S0020-7373\(87\)80013-5](https://doi.org/10.1016/S0020-7373(87)80013-5)

- Narvaez Linares, N. F., Charron, V., Ouimet, A. J., Labelle, P. R., & Plamondon, H. (2020). A systematic review of the Trier Social Stress Test methodology: Issues in promoting study comparison and replicable research. *Neurobiology of Stress*, 13, 100235. <https://doi.org/10.1016/j.ynstr.2020.100235>
- Onnasch, L. (2015). Crossing the boundaries of automation—Function allocation and reliability. *International Journal of Human-Computer Studies*, 76, 12–21. <https://doi.org/10.1016/j.ijhcs.2014.12.004>
- Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, 110(52), 20941–20946. <https://doi.org/10.1073/pnas.1312011110>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Pochon, J. B., Levy, R., Fossati, P., Lehericy, S., Poline, J. B., Pillon, B. ... Dubois, B. (2002). The neural system that bridges reward and cognition in humans: An fMRI study. *Proceedings of the National Academy of Sciences*, 99(8), 5669–5674. <https://doi.org/10.1073/pnas.082111099>
- Proud, R. W., Hart, J. J., & Mrozinski, R. B. (2003). *Methods for determining the level of autonomy to design into a human spaceflight vehicle: A function specific approach* (No. 20100017272; p. 15). NASA Johnson Space Center. <https://ntrs.nasa.gov/citations/20100017272>
- Rice, S., & Keller, D. (2009). Automation reliance under time pressure. *Cognitive Technology Journal*, 14(1), 36–44.
- Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (1st ed., pp. 19–35). CRC Press. 10.1201/9781315137957-2
- Roland, A., & Shiman, P. (2002). *Strategic computing: DARPA and the quest for machine intelligence, 1983–1993*. MIT Press.
- Ryan, M. (2018). *Human-machine teaming for future ground forces*. Center for Strategic and Budgetary Assessments. <https://csbaonline.org/research/publications/human-machine-teaming-for-future-ground-forces>
- Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, 45(4), 736–750. <https://doi.org/10.1037/0022-3514.45.4.736>

- Salimetrics. (n.d.). *SalivaBio's passive drool saliva collection method*. Retrieved April 26, 2021, from <https://salimetrics.com/collection-method/passive-drool-saliva-collection-device/>
- Salvendy, G. (2012). *Handbook of human factors and ergonomics* (4th ed.). Wiley.
- Sandi, C. (2013). Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3), 245–261. <https://doi.org/10.1002/wcs.1222>
- Sansone, C., & Harackiewicz, J. M. (2000). *Intrinsic and extrinsic motivation the search for optimal motivation and performance*. Academic Press.
- Shubhendu, S., & Vijay, J. (2013). Applicability of artificial intelligence in different fields of life. *International Journal of Scientific Engineering and Research*, 1(1), 28–35.
- Singer, P. W. (2009). *Wired for war: The robotics revolution and conflict in the twenty-first century*. Penguin Press.
- Smith, S. M., & Vale, W. W. (2006). The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. *Dialogues in Clinical Neuroscience*, 8(4), 383–395.
- Sorkin, R. D. (1988). Why are people turning off our alarms? *The Journal of the Acoustical Society of America*, 84(3), 1107–1108. <https://doi.org/10.1121/1.397232>
- Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, 36(4), 1228–1248. <https://doi.org/10.1016/j.neubiorev.2012.02.003>
- Starcke, K., & Brand, M. (2016). Effects of stress on decisions under uncertainty: A meta-analysis. *Psychological Bulletin*, 142(9), 909–933. <https://doi.org/10.1037/bul0000060>
- Thompson, L. L. (2008). *Making the team: A guide for managers* (Third, Vol. 3). Pearson/Prentice Hall. <http://archive.org/details/makingteamguidef0003thom>
- Tobii Pro. (2015, December 18). *Tobii Pro external processing unit* [Information]. <https://www.tobiipro.com/product-listing/external-processing-unit/>
- Tobii Pro. (2016, January 28). *Tobii Pro X3-120* [Information]. <https://www.tobiipro.com/product-listing/tobii-pro-x3-120/>

- Tossell, C. C., Kim, B., Donadio, B., de Visser, E. J., Holec, R., & Phillips, E. (2020). Appropriately representing military tasks for human-machine teaming research. In C. Stephanidis, J. Y. C. Chen, and G. Fragomeni (Eds.), *HCI International 2020 – Late Breaking Papers: Virtual and Augmented Reality* (pp. 245–265). Springer International Publishing. https://doi-org.libproxy.nps.edu/10.1007/978-3-030-59990-4_19
- United States Marine Corps. (2020). *Force design 2030*. Department of the Navy, United States Marine Corps. <https://www.hqmc.marines.mil/Portals/142/Docs/CMC38%20Force%20Design%202030%20Report%20Phase%20I%20and%20II.pdf?ver=2020-03-26-121328-460>
- van den Bos, R., Harteveld, M., & Stoop, H. (2009). Stress and decision-making in humans: Performance is related to cortisol reactivity, albeit differently in men and women. *Psychoneuroendocrinology*, 34(10), 1449–1458. <https://doi.org/10.1016/j.psyneuen.2009.04.016>
- van den Bosch, K., & Bronkhorst, A. (2018). Human-AI cooperation to benefit military decision making. *Proceedings of the NATO Science and Technology Organization Specialists Meeting on Big Data & Artificial Intelligence for Military Decision Making*, 13. <https://doi.org/10.14339/STO-MP-IST-160-S3-1-PDF>
- van Dongen, K., & van Maanen, P.-P. (2006). Under-reliance on the decision aid: A difference in calibration and attribution between self and aid. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(3), 225–229. <https://doi.org/10.1177/154193120605000304>
- van Dongen, K., & van Maanen, P.-P. (2013). A framework for explaining reliance on decision aids. *International Journal of Human-Computer Studies*, 71(4), 410–424. <https://doi.org/10.1016/j.ijhcs.2012.10.018>
- Verharen, J. P. H., Adan, R. A. H., & Vanderschuren, L. J. M. J. (2020). How reward and aversion shape motivation and decision making: A computational account. *The Neuroscientist*, 26(1), 87–99. <https://doi.org/10.1177/1073858419834517>
- von Clausewitz, C. (2012). *On War*. CreateSpace Independent Publishing Platform.
- Wang, H., Ma, C., & Zhou, L. (2009). A brief review of machine learning and its application. *2009 International Conference on Information Engineering and Computer Science*, 1–4. <https://doi.org/10.1109/ICIECS.2009.5362936>
- Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (Fourth edition.). Pearson.

- 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>
- Wilson, H. J., & Daugherty, P. R. (2018, July 1). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, July–August 2018. <https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces>

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California