



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**EXPLORING THE DEVELOPMENT OF TRUST
BETWEEN HUMAN AND AUTONOMOUS TEAMMATES**

by

Peter J. Coombe

September 2020

Thesis Advisor:
Second Reader:

Mollie R. McGuire
Steven J. Iatrou

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 September 2020	3. REPORT TYPE AND DATES COVERED Master's thesis		
4. TITLE AND SUBTITLE EXPLORING THE DEVELOPMENT OF TRUST BETWEEN HUMAN AND AUTONOMOUS TEAMMATES			5. FUNDING NUMBERS	
6. AUTHOR(S) Peter J. Coombe				
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) This research aimed to identify and combine objective performance-based metrics of trust with subjective measures in order to better understand the development of trust between humans and autonomous systems during complex tasks that include risk. A virtual reality game developed by the Johns Hopkins University Applied Physics Laboratory (JHU-APL) was the platform used to measure trust scores between both homogeneous human teams and human-autonomous teams (utilizing a bot teammate designed by JHU-APL). Of interest in this study was whether objective and subjective measures of trust differed depending on whether the teammate was (a) actually a human or a bot and (b) perceived to be a human or a bot. Additionally, this study examined how objective performance metrics compare to subjective trust scores among varying teammate conditions. The objective performance metrics identified in this thesis were not indicative of overall trust alone but did shed some light on the development of trust and would be beneficial if used in conjunction with subjectively derived metrics to provide a more complete set of measures that can contribute to a greater understanding of the development of trust between man and machine.				
14. SUBJECT TERMS trust, trust measures, teaming, autonomy			15. NUMBER OF PAGES 147	
			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.

**EXPLORING THE DEVELOPMENT OF TRUST BETWEEN
HUMAN AND AUTONOMOUS TEAMMATES**

Peter J. Coombe
Major, United States Marine Corps
BA, Eastern Washington University, 2003
MA, Marshall University, 2016

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

**NAVAL POSTGRADUATE SCHOOL
September 2020**

Approved by: Mollie R. McGuire
Advisor

Steven J. Iatrou
Second Reader

Thomas J. Housel
Chair, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This research aimed to identify and combine objective performance-based metrics of trust with subjective measures in order to better understand the development of trust between humans and autonomous systems during complex tasks that include risk. A virtual reality game developed by the Johns Hopkins University Applied Physics Laboratory (JHU-APL) was the platform used to measure trust scores between both homogeneous human teams and human-autonomous teams (utilizing a bot teammate designed by JHU-APL). Of interest in this study was whether objective and subjective measures of trust differed depending on whether the teammate was (a) actually a human or a bot and (b) perceived to be a human or a bot. Additionally, this study examined how objective performance metrics compare to subjective trust scores among varying teammate conditions. The objective performance metrics identified in this thesis were not indicative of overall trust alone but did shed some light on the development of trust and would be beneficial if used in conjunction with subjectively derived metrics to provide a more complete set of measures that can contribute to a greater understanding of the development of trust between man and machine.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	PROBLEM STATEMENT	1
B.	PURPOSE STATEMENT.....	2
C.	HYPOTHESES AND RESEARCH QUESTIONS	2
1.	Hypotheses.....	2
2.	Research Questions.....	3
D.	RESEARCH METHOD	3
E.	PROPOSED DATA, OBSERVATION AND ANALYSIS	
	METHODS	4
F.	THESIS ORGANIZATION.....	4
 II.	 LITERATURE REVIEW	 7
A.	BACKGROUND	7
B.	SYSTEMS AND TEAMING.....	9
1.	Systems.....	9
2.	Teaming	11
C.	FOUNDATIONS OF TRUST	14
D.	FACTORS THAT INFLUENCE TRUST	18
1.	Interpersonal Trust.....	18
2.	Trust in Technology.....	18
E.	RELIANCE ON AUTOMATION.....	22
F.	SIMILARITIES AND DIFFERENCES BETWEEN HUMAN-	
	HUMAN AND HUMAN AUTOMATION TRUST	24
G.	EMPIRICAL RESEARCH	28
H.	SUMMARY	33
 III.	 METHODOLOGY	 35
A.	PARTICIPANTS AND DESIGN	35
1.	Participants.....	35
2.	Experiment Design.....	38
B.	MATERIALS	43
1.	Oculus Rift-S VR-HMD	43
2.	Oculus Touch Controllers	44
3.	ESCAPE.....	46
4.	Questionnaires.....	51
C.	PILOT TESTING	52
D.	PROCEDURE	54

1.	Experiment	54
2.	Data and Collection.....	57
IV.	EXPERIMENT RESULTS	59
A.	OBJECTIVE MEASURES OF TRUST	59
1.	Consistent and Inconsistent Conditions.....	59
2.	Perception Conditions	63
3.	Actual Conditions.....	65
B.	SUBJECTIVE MEASURES OF TRUST	67
1.	Consistent and Inconsistent Conditions.....	67
2.	Perception Conditions	69
3.	Actual Conditions.....	70
C.	PAIRED SAMPLES TESTING	71
D.	EXPLORATORY FINDINGS.....	72
V.	DISCUSSION AND THOUGHTS FOR FUTURE RESEARCH	79
A.	DISCUSSION	79
1.	Hypotheses.....	79
2.	Research Questions.....	83
3.	Limitations.....	85
B.	CONCLUSIONS AND THOUGHTS FOR FUTURE RESEARCH	86
1.	Conclusions.....	86
2.	Thoughts for Future Research.....	87
	APPENDIX A. ESCAPE OBJECTS AND OBSTACLES	91
	APPENDIX B. QUESTIONNAIRES	95
A.	PRE-ROUND 1 QUESTIONNAIRE.....	95
B.	END OF ROUND 1 QUESTIONNAIRE	95
C.	END OF ROUND 2 QUESTIONNAIRE	99
D.	END OF ROUND 3 QUESTIONNAIRE	102
E.	END OF ROUND 4 QUESTIONNAIRE	106
	APPENDIX C. HUMAN-BOT PROTOCOLS.....	113
	APPENDIX D. INFORMED CONSENT FORM	115
	APPENDIX E. EVENT BRIEF	117

LIST OF REFERENCES	121
INITIAL DISTRIBUTION LIST	127

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1.	Dyadic Model of Trust in Relationships. Source: Simpson (2007).	17
Figure 2.	Factors of Trust Development in Human-Robot Interaction. Source: Hancock et al. (2011).	20
Figure 3.	The Full Model of Factors That Influence Trust in Automation. Source: Hoff and Bashir (2015).	21
Figure 4.	Participant Service Breakdown.....	36
Figure 5.	Participant Gaming Experience	36
Figure 6.	Participant VR Gaming Experience.....	37
Figure 7.	Participant Game Genre Preference.....	37
Figure 8.	Room A.....	39
Figure 9.	Room B	40
Figure 10.	Scenario Sync Matrix.....	42
Figure 11.	Oculus Rift-S Headset. Source: Oculus (n.d.-a).	43
Figure 12.	Oculus Touch Controllers. Source: Oculus (n.d.-b).	45
Figure 13.	ESCAPE Round 1	47
Figure 14.	ESCAPE Round 2	48
Figure 15.	ESCAPE Round 3	49
Figure 16.	ESCAPE Round 4.....	50
Figure 17.	Example Oculus Rift-S Guardian Virtual Play Area. Source: Baker (2020).	55
Figure 18.	ESCAPE Training Session.....	56

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	The Characteristics of Effective Teams. Source: Baker et al. (2006).....	12
Table 2.	Trust Measures by Type. Source: Brzowski and Nathan-Roberts (2019).....	33
Table 3.	Pairwise Comparison of Time by Round.....	61
Table 4.	Pairwise Comparison of VR Deaths by Round.....	62
Table 5.	Pairwise Comparison of Pass/Fail by Round.....	63
Table 6.	Round Time Table of Means (Perception Condition).....	64
Table 7.	VR Deaths Table of Means (Perception Condition).....	64
Table 8.	Pass/Fail Table of Means (Perception Condition)	65
Table 9.	Round Time Table of Means (Actual Condition).....	65
Table 10.	VR Deaths Table of Means (Actual Condition)	66
Table 11.	Pass/Fail Table of Means (Actual Condition).....	66
Table 12.	Pairwise Comparison of Competency by Round.....	68
Table 13.	Pairwise Comparison of Composite Trust by Round	69
Table 14.	Paired Samples T-Tests Between Actual Human and Bot Teammate Objective and Subjective Performance Metrics.....	71
Table 15.	Paired Samples T-Tests Between Actual and Perceived Human and Bot Teammate Subjective Performance Metrics	72
Table 16.	Perceived Human Teammate: Competency.....	73
Table 17.	Perceived Bot Teammate: Competency.....	73
Table 18.	Perceived Human Teammate: Trust.....	74
Table 19.	Perceived Bot Teammate: Trust.....	74
Table 20.	Actual Human Teammate: Competency	75
Table 21.	Actual Bot Teammate: Competency	76

Table 22.	Actual Human Teammate: Trust.....	76
Table 23.	Actual Bot Teammate: Trust.....	77

LIST OF ACRONYMS AND ABBREVIATIONS

ANOVA	Analysis of Variance
AI	Artificial Intelligence
CIV	Civilian
<i>d</i>	Cohen's-d
DOD	Department of Defense
DSS	Decision Support System
FN	Foreign National
GB	Gigabyte
GH	Gigahertz
HAI	Human Automation Interaction
HAT	Human Autonomy Teaming
Hz	Hertz
IBM	International Business Machines
JHU-APL	Johns Hopkins University Applied Physics Laboratory
<i>M</i>	Mean
n^2_p	Partial Eta Squared
NPS	Naval Postgraduate School
<i>p</i>	p-value
R^2	R-Squared
RAM	Random Access Memory
<i>SD</i>	Standard Deviation
SPSS	Statistical Package for Social Sciences
<i>t</i>	t-Statistic
UAS	Unmanned Aerial System
USA	United States Army
USAF	United States Air Force
USB	Universal Serial Bus
USCG	United States Coast Guard
USMC	United States Marine Corps
USN	United States Navy
VR-HMD	Virtual Reality-Head Mounted Display

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

This research would not have been possible without the assistance of Dr. Mollie McGuire, who offered endless guidance, direction, and support to ensure quality research and meaningful results. Special thanks to Mr. Steven Iatrou, Major Jacob Jones, and Major Erik Rastello, who generously provided outstanding feedback and recommendations throughout the entire thesis process. Most importantly, I want to recognize my wife, Stephanie. Her unwavering strength, support, and sacrifice provided motivation and the time required to see this project through to completion.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Trust is an essential element in the development of effective teams, and its characteristics apply to both human-human and human-machine teaming. Trust allows teams to work together smoothly and plays a vital role in system effectiveness in the areas of safety, performance, and use rate (Lee & See, 2004). The Department of Defense (DOD) has a long and successful history of partnering with machines. Dating back from the tracked remote-controlled explosive devices of World War I to the present-day use of unmanned aerial vehicles, the DOD has leveraged the advancements in technology to meet the needs of the increasingly complex modern battlefield. However, as autonomous system technology continues to advance, the DOD will have to adapt to the new human-autonomous systems dynamic, where machines are not just tools, but teammates (Laird et al., 2019). Blackhurst et al. (2011) suggested we are entering into a new era, where “Joint Warfare” means human-robot teams. What is not clear, however, is whether the trust of autonomous systems develops in the same way that trust develops between human-human teams. This change to the traditional team composition presents the DOD an opportunity to develop a better understanding of autonomy and apply increased effort towards designing systems for human-machine interdependence (Blackhurst et al., 2011).

A. PROBLEM STATEMENT

Current literature on the role of trust in the human-machine relationship tends to rely on subjective measures (Brzowski & Nathan-Roberts, 2019; Galster & Parasuraman, 2003; Lee & See, 2004). However, the inherent nature of subjective measures provides an incomplete set of metrics for understanding the development of trust among human-machine teammates. The problem is that there are too few studies that identify and test objective performance-based metrics of trust. More objective measures are needed that can be used in conjunction with subjectively derived metrics of trust to optimize human-machine team design and performance. While subjective measures of trust are important in studies, they can be influenced by how the responder wishes to be perceived rather than indicate their actual behavior and are sometimes criticized because they are open to

interpretation and opinion (Muckler & Seven, 1992). Objective measures on the other hand are not biased by the respondent's inability to accurately convey their thoughts and beliefs in retrospect. The lack of comprehensive research into and incorporating objective measures of trust in autonomy presents a challenge for the DOD in their effort to foster successful human-autonomy teams. As autonomous systems become more prevalent in the military and their roles begin to shift from being a mere tool to an actual teammate, the human operator's trust in such systems plays a crucial role in successful interactions and further use (Lee & See, 2004). Therefore, identifying and incorporating objective performance-based metrics of trust combined with subjective means is important because it can provide balance to research and contribute to proper validation of trust in the human-autonomous team dynamic.

B. PURPOSE STATEMENT

The primary purpose of this research is to identify and combine objective performance-based metrics of trust with subjective measures in order to better understand the development of trust between humans and autonomous systems in a teaming dynamic, especially during complex tasks that include risk. A secondary goal of this research is to explore the similarities and differences in the development of trust between human-human and human-autonomous teams. Research in this area is important to help the DOD better understand the development of trust between humans and autonomy. The knowledge gained through continued research can then be applied towards the design and acquisition of future autonomous systems that maximize efficiency and lethality in the human-autonomous team construct.

C. HYPOTHESES AND RESEARCH QUESTIONS

1. Hypotheses

This research will test the following hypotheses:

- H1: Trust scores are different depending on whether the teammate is a human or a bot (i.e., an autonomous robot teammate designed by the Johns

Hopkins University Applied Physics Laboratory (JHU-APL) for the purposes of this experiment)

- H2: Trust scores are different depending on whether it is believed the teammate is a human or a bot.

2. Research Questions

The study will also address the following research questions:

1. How do objective performance metrics compare to subjective trust scores when the teammate is a human or a bot?
2. How do objective performance metrics compare to subjective trust scores when the teammate is believed to be a human or a bot?

D. RESEARCH METHOD

Quantitative methods to address the two hypotheses and two research questions described above will be applied during this study. A laboratory experiment to measure trust will be conducted using volunteer Naval Postgraduate School students and staff as participants. A virtual reality game called ESCAPE was developed by the JHU-APL will be the platform used to measure trust scores between both homogeneous human teams and human-autonomous teams. Approximately 80 participants will be recruited for this proposed experiment. No preference will be given to Department of Defense service, rank, specialty, or Naval Postgraduate staff position for participation.

The experiment will require the participant to play four rounds of ESCAPE, utilizing the Oculus Rift S virtual reality head-mounted display (VR-HMD) and associated touch controllers to maneuver their avatar and coordinate with their bot teammate in the virtual environment. The intent for each round is to have the participant and their teammate, either the bot or a member of the research team (identities will be withheld to ensure that bias about performance expectations between humans and bots can be controlled) successfully escape together as a team. To be successful, the team will have to work

together to navigate through the virtual environment, overcome any obstacles, and exit the round together within the allotted time.

During the experiment, objective performance data will be collected after each round (i.e., round time, participants virtual reality [VR] deaths, and round pass/fail scores) in addition to using Schaffer-Lay and Likert Scale questionnaires to collect subjective data such as the participant's perception of their performance, the performance of their teammate, performance of the team, and overall trust. Together, these data will be used to measure trust between the human-human and human-autonomous teams.

E. PROPOSED DATA, OBSERVATION AND ANALYSIS METHODS

The principal phenomenon of analysis in the study will be trust. The objective performance metrics and the subjective survey results discussed above will be used to quantitatively address the two hypotheses and the two research questions. To test if trust scores are different depending on whether the teammate is a human or bot or believed to be a human or a bot will be conducted using ANOVA tests. Dependent variables in the ANOVA tests will be participant partner preference, round time, participants VR deaths, and round pass or fail scores. The independent variables used for analysis will be the consistent condition (i.e., the participant played with the teammate of their choice in Round 1), inconsistent condition (i.e., the participant played with the teammate opposite of their choice in Round 1), the participant's perception of teammate (i.e., a human or a bot), and the actual conditions. The two research questions will be addressed in a similar fashion.

F. THESIS ORGANIZATION

This thesis is divided into five chapters, including the current Chapter I. Chapter II consists of a comprehensive literature review of several works that explored the foundations of trust in automation and autonomous systems and the factors that influence it. It will also highlight critical aspects of automated and autonomous systems and explore the research conducted on reliance as it applies to these systems. The chapter will conclude with a discussion of the similarities and differences between human-human and human-autonomous trust and a critical review of previous research conducted in this field. Chapter III lays out the methodology applied in this research focusing on the experiment's

participants, design, materials utilized, and an explanation of the experiment's procedures and protocols derived from pilot testing. Chapter IV then provides a summary of the results from the experiment phase. Finally, Chapter V offers a discussion of the work performed along with thoughts and recommendations for future research in the area of trust and human-autonomous teaming.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. BACKGROUND

We are in a period of incredible technological flux. Advances in autonomy and in artificial intelligence and autonomous control systems and advanced computing and big data, and learning machines and intuitive graphic visualization tools, metamaterials, miniaturization -- they're leading us to a time of great human-machine collaboration. (Office of the Under Secretary of Defense, 2017, p. 17)

— Arati Prabhakar, Defense Advanced Research
Projects Agency Director (2012–2017)

The concept of humans using machines (in this thesis “machine” refers to computer-based machines) as a tool has been around for many decades, enabling humans to operate more efficiently and effectively. For example, human partnerships with machines have played an essential role in domains that encompass typically hazardous working conditions (e.g., automotive industry) where machines and automation serve to offset human efficiency factors such as stress, workload, fatigue, and distraction (Schaefer et al., 2016). As technology continues to improve, complex domains are now becoming inundated with advanced automation and decision aids supported by artificial intelligence (AI). However, Laird et al. (2019) suggested that although current intelligent systems, such as Siri and Alexa (i.e., virtual assistants) support and extend human capabilities, such machines are only tools and not yet real teammates. But continued advancements in automation, intelligence, and decision making make the potential for interdependent human-machine teaming a reality (Schaefer et al., 2016). The development and evolution of the machine teammate, in a practical sense, is intended to reduce operating costs, enhance performance, and improve safety in their application (Johnson & Vera, 2019). As such, Schaefer et al. (2016) surmised that as automation continues to develop into more advanced and sophisticated systems, the value in these systems to an organization is not in the complete replacement of human controllers, but the capacity for collaboration between the human and autonomous system.

While the premises of reducing costs, increasing performance, and enhancing safety through the use of automation and intelligent systems would undoubtedly be an attractive proposition for any civilian organization or the DOD, the promises of these benefits, in some cases, have been slow to materialize. For example, Blackhurst et al. (2011) described how DOD investments in autonomy have resulted in increased operating costs due to the “autonomy paradox.” The autonomy paradox suggests that the autonomous systems designed to reduce human inputs and workload require increases in human resources to support them (Blackhurst et al., 2011). Also, although automated support systems such as vehicle navigation systems and aircraft autopilots provide human operators numerous benefits, they are not perfect systems. These systems occasionally behave unpredictably, make wrong decisions, or provide inappropriate advice, which can degrade overall performance (Chavaillaz et al., 2016). Factors such as misuse or complete disuse of automated systems also result in less than desired performance, which in turn, negatively affects the proper integration between the human and automated system (Drnec & Metcalfe, 2016). Another contributing factor to poor integration is the lack of user acceptance of a system. Lack of acceptance was noted as being directly related to a human’s trust in automated systems (Drnec & Metcalfe, 2016).

Furthermore, as automation becomes more prevalent in the execution of complex tasks, underdeveloped partnerships between humans and automated systems can induce higher costs and can lead to catastrophic results (Lee & See, 2004). Hence, with further research on the subject of teaming, advanced technologies can result in making selected tasks less challenging, requiring less personnel training, expertise, and intervention with the systems (Johnson & Vera, 2019). That is an important point when considering teaming in future military contexts as warfighter interaction with autonomous systems is not uncommon. However, future warfighters will likely be required to interact regularly with various forms of automation and robots in stressful and dynamic combat environments (Chen & Terrence, 2009). For the man-machine team to accomplish their goal or mission, an essential factor is the human’s trust in the robot or like-intelligent system to protect the interests and welfare of the team, particularly in high-risk situations (Hancock et al., 2011).

Schaefer et al. (2016) postulated that if there is no trust in an automated system, it will not be used. As such, this leads to the assertion that trust will have a significant impact on the design of future autonomous systems (Schaefer et al., 2016). Therefore, factoring trust in the design, acquisition, and employment of future DOD intelligent systems should be an essential consideration.

As human interaction with technology is moving toward using machines and intelligent systems as competent teammates, a large body of literature has been written on the role that trust plays in the use and reliance on such systems and about enhancing human-machine interaction. However, although the literature on human-automation interaction is vast and dates back to the early seventies, trust in this context remains challenging to define and measure (Brzowski & Nathan-Roberts, 2019). Over the years, research conducted on the subject overwhelmingly employed subjective measures for analysis, using methods such as self-reporting on surveys that require participants to rate their levels of trust before, during, and after the interaction (Brzowski & Nathan-Roberts, 2019). Comparatively, few research studies employed objective measures for analysis or applied a combination of both (Brzowski & Nathan-Roberts, 2019). Therefore, Brzowski et al. (2019) suggested that to validate measures of trust in human-system interaction properly; researchers need to standardize both the definitions and methods to measure it effectively.

B. SYSTEMS AND TEAMING

1. Systems

What automated systems, autonomous systems, and AI all have in common is that they are all affiliated with machines, robots, or other software that allow humans to operate more efficiently and effectively (Hankiewicz, 2018). For context, however, the literature points to subtle differences between automation, autonomy, and AI. Parasuraman and Riley (1997) defined *automation* as “the execution by a machine agent (usually a computer) of a function that was previously carried out by a human” (p. 231). Lee et al. (2004) described automation as a technology that actively selects data, transfers information, makes decisions, or controls processes. Also, Schaefer et al. (2016) stated that “automated systems are designed to accomplish a specific set of largely deterministic steps (often in a repeated

pattern) in order to achieve one of an envisaged and finite set of pre-defined goals” (p. 380).

Autonomous systems, on the other hand, according to Schaefer et al. (2016), “learn and evolve through the input of operational and contextual information and thus their actions necessarily become more indeterminate across time” (p. 380). Also, de Visser et al. (2018) submitted that “in contrast to a conventional automated system designed to carry out a limited set of pre-programmed supervised tasks on behalf of the user, autonomy is designed to carry out a user’s goals, but that does not require supervision” (p. 1). Further, de Visser et al. (2018) described how autonomous systems could adapt and change, set goals, and respond to the environment utilizing sensors and input data.

AI, in general terms, as described by Shukla-Shubhendu and Vijay (2013), are “machines that respond to stimulation consistent with traditional responses from humans, given the human capacity for contemplation, judgment, and intention” (p. 28). Fittingly, AI are expert and intelligent systems, capable of making decisions that typically require a level of human expertise on the subject (Shukla Shubhendu & Vijay, 2013). Furthermore, the success of modern AI applications can be attributed to many factors. Such factors include improved algorithms, enormous computing power, and huge amounts of data, providing AI systems with human-level perception capabilities like image interpretation, speech-to-text, and text understanding that supports machine-learning (Rossi, 2018).

For this work, automation refers to the full range of capabilities inherent in programmable automated systems designed to control processes (Lee & See, 2004) and accomplish pre-defined goals (Schaefer et al., 2016) to autonomous systems capable of learning, evolving, and executing tasks by adapting to conditions sans supervision (de Visser et al., 2018; Schaefer et al., 2016). Given the current reality of technologically evolving autonomous systems and the applicability of AI, the DOD will need to continuously adapt to this emerging system dynamic, where machines are no longer mere tools, but teammates.

2. Teaming

Before getting into specifics about human-machine teaming, it would be prudent to cover the concept and makeup of a team and the importance of teamwork. Salas et al. (2004) defined a team as “a collective of interdependent individuals who work together, have shared objectives, mental models, and procedures that guide their perceptions, thinking, and behaviors toward a common goal.” In good teams, individuals support one another through rapport building and trust repair activities (Duhigg, 2016). In the same vein, the literature on the concept of teams supports the notion that teams, compared to individuals, make fewer mistakes when the responsibilities of each team member are known and understood (Baker et al., 2006). In order to be a capable team, however, team members must possess specific attributes that benefit the whole. Baker et al. (2006) referred to such attributes as possessing requisite knowledge, skills, and attitudes. The interaction of these team-related attributes contributes to the concept of teamwork. For reference, Table 1 describes the characteristics and behaviors of effective teams. Another vital aspect of team success is training. Whether training in a human-human or human-machine construct, practical training enables improved coordination, communications, error management, and decision making (Baker et al., 2006).

Table 1. The Characteristics of Effective Teams. Source: Baker et al. (2006).

<i>Team Knowledge, Skills, and Attitudes</i>	<i>Characteristics of Effective Teams (Salas, Sims, and Klein 2004)</i>
Team leadership	Have a clear common purpose Team member roles are clear but not overly rigid Involve the right people in decisions Conduct effective meetings Establish and revise team goals and plans Team members believe the leaders care about them Distribute and assign work thoughtfully
Backup behavior	Compensate for each other Manage conflict well-team members confront each other effectively Regularly provide feedback to each other, both individually and as a team ("debrief") "Deal" with poor performers Are self-correcting
Mutual performance monitoring	Effectively "span" boundaries with stakeholders outside the team Members understand each others' roles and how they fit together Examine and adjust the team's physical workplace Periodically diagnose team "effectiveness," including its results
Communication Adaptability	Communicate often "enough" Members anticipate each other Reallocate functions Recognize and adjust their strategy under stress Consciously integrate new team members.
Shared mental models	Coordinate without the need to communicate overtly
Mutual trust	Trust other team members' "intentions"
Team orientation	Select team members who value teamwork Strongly believe in the team's collective ability to succeed

Like trust, a conclusive definition of teamwork appears to be elusive. However, Salas et al. (2004) synthesized previous research and definitions of teamwork to provide a useful definition for this paper. Salas et al. (2004) termed teamwork as, "a set of flexible behaviors, cognitions, and attitudes that interact to facilitate task work and achieve mutually desired goals and adaptation to the changing internal and external environments." As noted in the definition above, team member behavior plays an essential role in teamwork. Salas et al. (2004) noted that certain behaviors comprise teamwork, such as leadership, backup behavior, mutual performance monitoring, communication, and adaptability. As such, teamwork will be an essential aspect of establishing and maintaining trust in the man-machine team dynamic.

While humans have used machines to gain and maintain a competitive advantage in seemingly every aspect of life, the concept of teaming with a machine that has the cognitive capabilities necessary to interact interdependently is becoming a reality with

every advancement in autonomous technology supported by AI. Indeed, these are exciting times, as we may be on the precipice of a new human-machine paradigm. Machine capabilities once thought only to exist in science fiction or in the future are developing at a very rapid pace. For a good reason, the DOD has recognized the importance of such emerging technologies. For example, in the United States Third Offset Strategy, five critical areas of technology were the subject of focus. It included autonomous learning systems, human-machine collaborative decision-making, assisted human operations, advanced manned-unmanned systems operations, and network-enabled autonomous weapons and high-speed projectiles (Ellman et al., 2017). Furthermore, the Unmanned Systems Integrated Roadmap for Fiscal Year 2017–2042 consolidated the advancements, challenges, and trends in technology from DOD, academia, and industry into future areas of interest that include: interoperability, autonomy, network security, and human-machine collaboration. Finally, the Defense Science Board Summer Study on Autonomy (2016) concluded that the ubiquitousness of autonomous capabilities make them accessible to both allies and adversaries, thereby surmising that “DOD must take immediate action to accelerate its exploitation of autonomy while also preparing to counter autonomy employed by adversaries” (p. iii).

The DOD has leveraged emergent technologies like the steam engine, aircraft, nuclear capabilities, stealth technology, and unmanned aerial vehicles to gain competitive advantages in warfare. Accordingly, the teaming of intelligent machines will also be an opportunity for the DOD to become more efficient and lethal in the future. However, we are not there yet. More work in this field is required to realize a real man-machine teaming construct. According to the DOD Future Directions in Human-Machine Teaming Workshop, Laird et al. (2019) explained that for machines to become real teammates, intelligent machines will need to become more flexible and adaptable to the task environment and to that of the human partner. More specifically, the intelligent machines will need to develop the ability to decipher their human partner’s intentions and capabilities in addition to applying learned experiences to novel situations (Laird et al., 2019). In order to fully understand the challenges of machine teaming, the workshop participants considered the breadth of research in the areas of AI and cognitive sciences, identifying

several challenges and opportunities for future research in this field (Laird et al., 2019). As a result, four key topic areas emerged for analysis. These areas include machine capabilities and AI, machine modeling of the human teammate, human cognitive capabilities in complex and dynamic situations, and human mental models of machines (Laird et al., 2019).

More research is necessary to understand what capabilities a machine must possess (e.g., motor skills, communication skills, perception, reasoning, and problem-solving skills) to be an effective teammate (Laird et al., 2019). Machine models of humans require research into what machines need to model to adequately understand the physical and cognitive capabilities of humans and how the machines represent and reason about the human teammate (Laird et al., 2019). Challenges in the area of human cognitive abilities entail understanding how humans learn from events to make predictions and generalizations about the future and how mental models, a theory of mind, and shared knowledge guide reasoning and deductive inferences (Laird et al., 2019). Finally, challenges associated with human models of machines include further research on what humans need to know about the machine's physical and intellectual capacities in order to support efficient interactions and to establish and maintain trust in the machine (Laird et al., 2019). Along those lines, for any team to be successful, whether in sports, business, or combat, an essential element of that team dynamic is trust. The following two sections will discuss the foundations of trust and the factors that influence trust in automated systems.

C. FOUNDATIONS OF TRUST

The interest in the concept of trust as a means to promote efficiency and cooperation in the cognitively complex areas of interpersonal interactions, organizations, and automation, has increased in recent years (Lee & See, 2004). Fittingly, the concept of trust has been studied across various disciplines, including economics, human factors, political science, philosophy, psychology, and sociology to understand its applications in the various fields (Hoff & Bashir, 2015). Early literature on trust in automation used human-human models of trust as the basis for comparing and contrasting trust with automation (de Visser et al., 2018; Madhavan & Wiegmann, 2007; Muir & Moray, 1996). Not surprisingly,

the literature suggests that trust has become a significant concern in the development and integration of complex autonomous systems (Schaefer et al., 2016). Furthermore, the concept of trust has generated many definitions across the varying disciplines of study. However, it appears that scholars are unable to come to unanimity on a single definition for it.

Some of the pioneering literature on interpersonal trust focused on attitudes or expectations between individuals. For example, early work by Rotter (1967) defined trust as “expectancy held by an individual that the word, promise, or written communication of another can be relied upon” (p. 651). Deutsch (1973), offered that trust is “confidence that one will find what is desired from another, rather than what is feared.” Yet another definition of trust was offered by Rempel, Holmes, and Zanna (1985), who stated that trust is the “the expectation related to subjective probability an individual assigns to the occurrence of some set of future events” (p. 96).

Other research has characterized trust as intentions to behave or willingness to act accordingly (Lee & See, 2004). Using this approach, Johns (1996) defined trust as “willingness to place oneself in a relationship that establishes or increases vulnerability with the reliance upon someone or something to perform as expected” (p. 81). In this respect, Mayer, Davis, and Schoorman (1995) authored one of the most widely used definitions of trust in the mid-nineties. They offered that trust is “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 party” (Mayer et al., 1995, p. 712).

From the psychological perspective, Simpson (2007) researched the theoretical and empirical conceptualizations of trust. In this research, he presented two main historical approaches to conceptualizing trust. The first was the adoption of the dispositional (person-centric) view, which focused on someone’s belief that other people are likely to be reliable and helpful in experimental, game-like situations (Deutsch, 1973; Simpson, 2007). The second historical perspective focused on partners and relationships. Furthermore, the works of Rempel et al. (1985) and Holmes and Rempel (1989) presented a dyadic (interpersonal) perspective, where trust is a psychological state of an actor and the interdependence

between a partner where the actor requires cooperation from the partner to attain the desired outcome (Simpson, 2007).

Later work by Kramer and Carnevale (2001) argued that trust could be tied to the beliefs and expectations that someone has of a partner providing a unique or valuable outcome that is beneficial to one's self-interest. Kelley et al. (2003) offered an explanation of trust that was similar to Kramer and colleagues. They suggested that trust is assessable in certain interpersonal situations that involve high interdependence, a mix of rules for coordinating and sustaining interdependence, and similar interests (Kelley et al., 2003). From another perspective, Hoff and Bashir (2015) suggested that there were three commonalties of trust across the different fields of study on trust. They suggest that: 1) there must be a trustor and trustee to give and take trust when something is at risk, 2) there must be motivation for the trustee to do a task suggested by the trustor, and 3) a possibility must exist that the trustee may fail to execute a task, thereby creating risk and ambiguity (Hoff & Bashir, 2015).

Simpson's review of the literature on interpersonal trust highlighted four core principles of note. The first is that people gauge the trustworthiness of a partner by noting if they display proper action in a trust-diagnostic situation-meaning the partner puts the interests of the individual or relationship above their own self-interest in response to a given situation (Simpson, 2007). Second, trust-diagnostic situations happen by chance, occur naturally, and sometimes individuals produce such situations to determine if their trust in a partner is justified (Simpson, 2007). Third, the growth or decline of trust in a partner is affected by individual differences in working models of oneself (e.g. "attachment orientations, self-esteem, and self-differentiation" (Simpson, 2007, p.266)). Fourth, consideration must be given to the actions of the individual and partner to comprehend the level of trust in the relationship (Simpson, 2007). Taking the core principles of trust gleaned from the research, Simpson (2007) created the Dyadic Model of Trust in Relationships depicted in Figure 1.

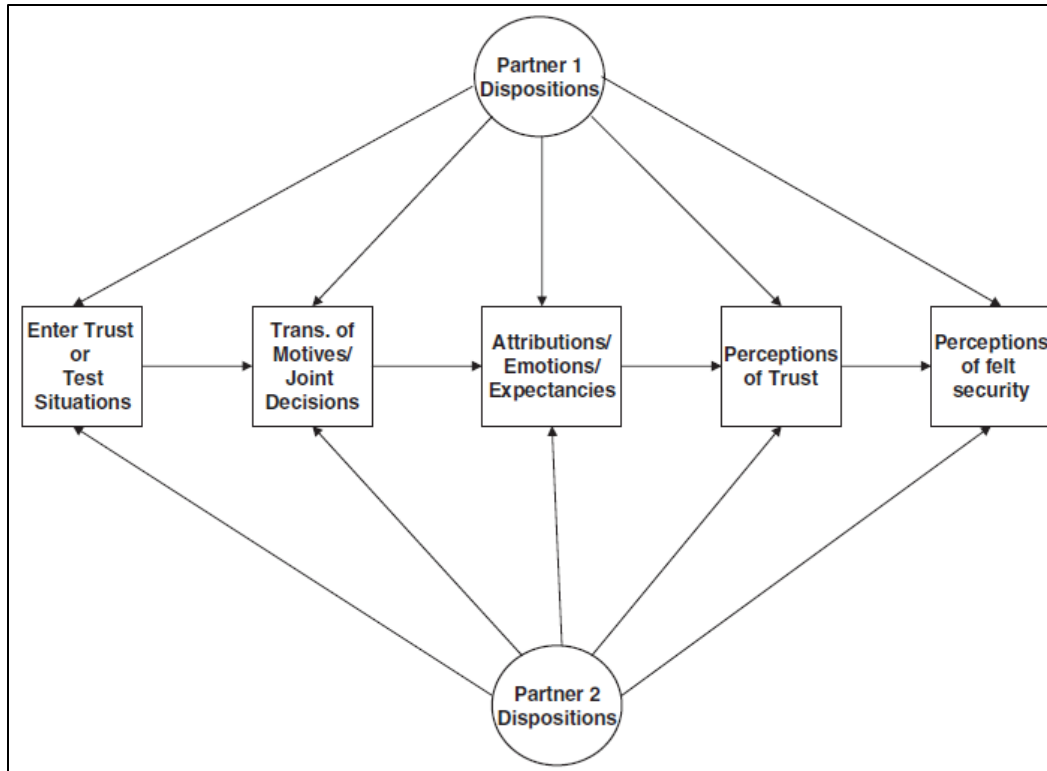


Figure 1. Dyadic Model of Trust in Relationships. Source: Simpson (2007).

The model includes both the normative and individual-difference components of trust. Individual-difference components are depicted as circles and represent the dispositions of each partner (working models) in the relationship and their attachments to the normative components at each stage of the model (Simpson, 2007). Normative components are shown as squares and represent the stages of the model; from entering, transforming, or creating trust-diagnostic situations on the left, to the perceptions of trust and security in a relationship on the right (Simpson, 2007). The model assumes that individuals with positive working models are more likely to enter trust-diagnostic situations, and each partner's working model in the relationship is impactful toward the outcome of each stage (Simpson, 2007). Feedback loops from the final stage back to the first for launching future trust interactions are assumed, but not depicted. The core principles of trust depicted in this model can serve as a way to conceptualize the foundations of trust and further inform how these foundations can apply to human-autonomous teaming.

Given the vast differences in interpretations of trust, for this research, trust will be defined in the context of automation that was authored by Lee and See (2004), describing trust as “the attitude that an agent will help achieve an individual’s goal in a situation characterized by uncertainty and vulnerability” (p. 51). Their definition is a distillation of the previous explanations of trust and best fits the purpose of this research.

D. FACTORS THAT INFLUENCE TRUST

1. Interpersonal Trust

Taking elements of earlier theoretical and operational definitions of trust, Rempel et al. (1985) offered a model of trust grounded in three components: predictability, dependability, and faith. The first influential component in the model, predictability, refers to the consistency of a partner’s behaviors and the ability to forecast behaviors in a social environment (Rempel et al., 1985). The next influential component of trust is dependability. Dependability, as described by Rempel et al. (1985), focuses less on a partner’s action and more on the internal qualities and characteristics of the partner. Finally, faith is the ability to trust in the absence of experiences or trusting without evidence that a partner’s behavior will be appropriate in the unknown future (Rempel et al., 1985).

The work of Lee et al. (2004) offered that trust does not develop on its own; instead, it evolves from individual, cultural, and organizational contexts. The individual context of trust includes an individual’s propensity to trust, and their history of interactions that lead to particular levels of trust (Lee & See, 2004). The organizational context of trust involves the interactions between people where the trustworthiness of others is determined (Lee & See, 2004). Similarly, the cultural context of trust can influence trust through people’s beliefs, norms, and expectations (Lee & See, 2004).

2. Trust in Technology

Turning toward technology, Corritore et al. (2003) proposed a model of online trust that considered such factors as credibility, ease of use, and risk. Perceived ease of use is a crucial variable identified in the Technology Acceptance Model, where it is described as “the degree to which the prospective user expects the target system to be free of effort”

(Davis et al., 1989, p. 985). Along with ease of use, reliability is another factor that influences trust. Parasuraman et al. (1997) suggested that when the reliability of an automated system was relatively high, the user's trust in the automation is not significantly reduced in response to occasional failures, but trust reduces if the failures are repeated or sustained. Another study on reliability was conducted by Madhavan, Weigmann, and Lacson (2006). They tested the hypothesis that a human's trust in automation is weakened by the automation's errors on tasks that can be easily performed by humans. The results of their experiment showed that automation errors that were perceived to be manageable by the human degraded the trust and reliance on such systems (Madhavan et al., 2006).

Hancock et al. (2011) conducted one of the first meta-analyses of available literature to quantify factors related to trust development in human-robot interactions. The studies included in their analysis were classified into three categories consisting of human-related factors, robot-related factors, and environmental-related factors of trust development. As depicted in Figure 2, human-related factors included characteristic and ability-based factors of the human, performance-based and attribute-based factors of the robot, and environmental factors categorized by team collaboration and tasking (Hancock et al., 2011). The results of their study indicated that the characteristics of the robot, specifically, performance-based factors (i.e., behavior, dependability, reliability, predictability, level of automation, failure rates, false alarms, and transparency) were the most influential on perceived trust in human-robot interaction, with environmental factors playing only a moderate role, and little evidence supporting effects of human-related factors (Hancock et al., 2011).

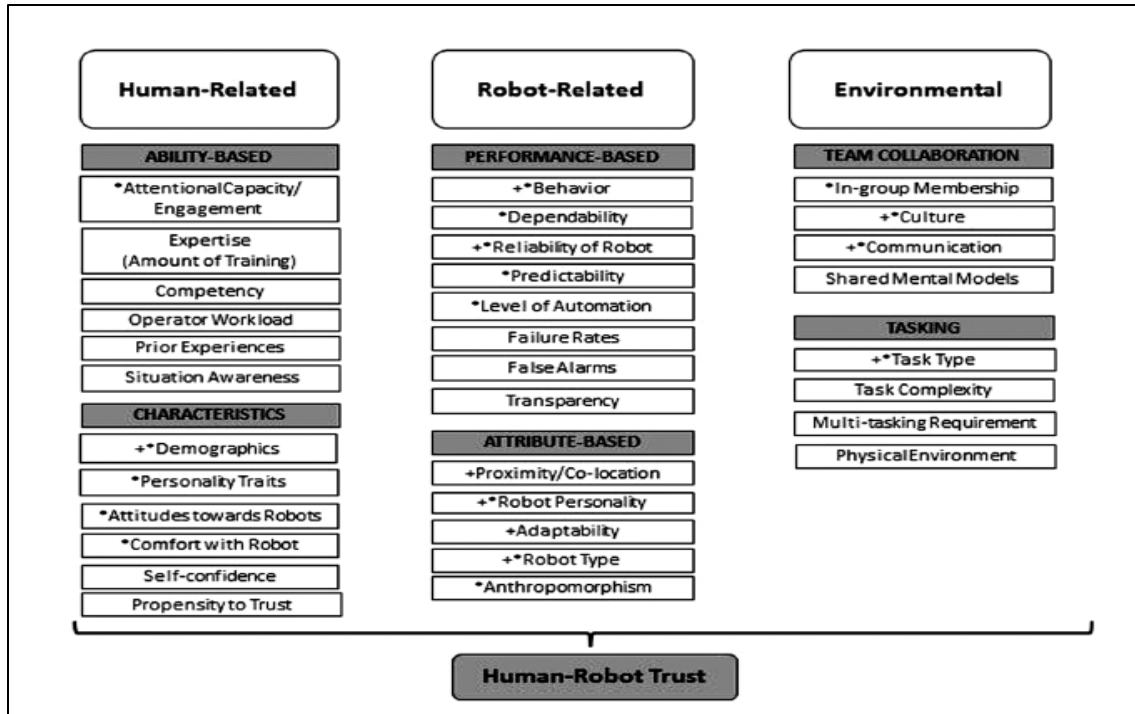


Figure 2. Factors of Trust Development in Human-Robot Interaction.
Source: Hancock et al. (2011).

In another study, Hoff et al. (2015) conducted a systematic review of empirical research on the factors that influence trust in automation. Their analysis revealed several factors that influence the trust formation process, most of which relate to the variability of the operator, environment, and the automated system (Hoff & Bashir, 2015). Drawing on the three layers of variability, Hoff et al. (2015) organized a conceptual model that reflected the dispositional, situational, and learned layers of trust that were derived by Marsh and Dibben (2003). For reference, the full model of the factors that influence trust in automation is depicted in Figure 3.

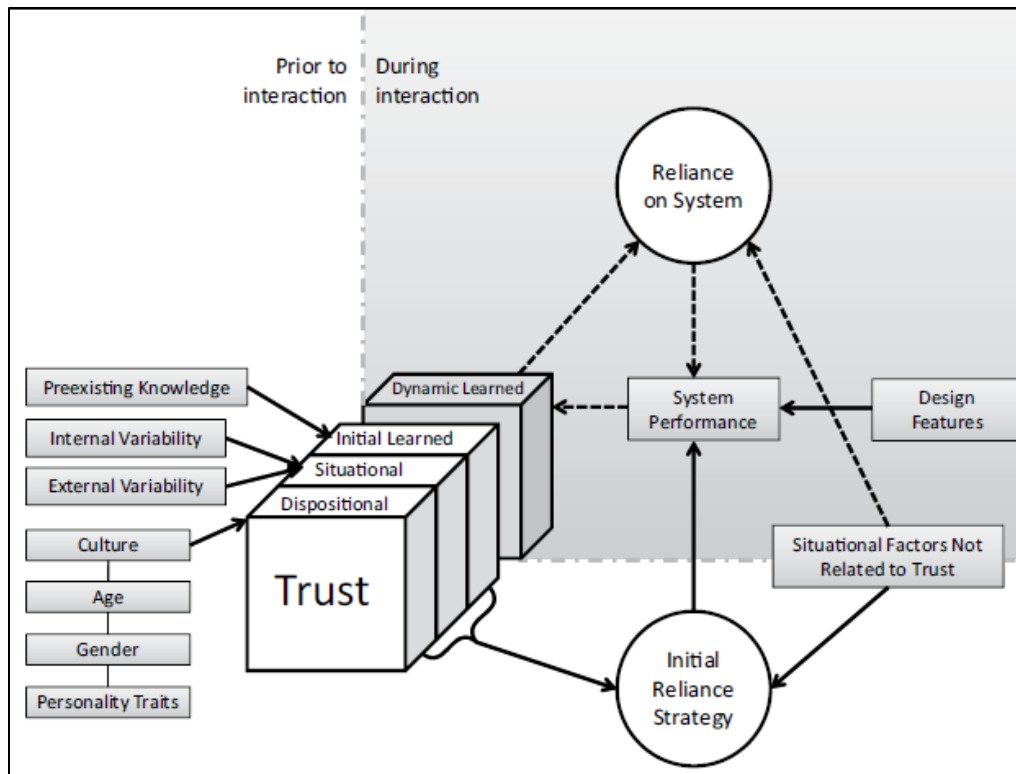


Figure 3. The Full Model of Factors That Influence Trust in Automation.
Source: Hoff and Bashir (2015).

In the model, dispositional trust relates to an individual's propensity to trust an automated system, independent of context or system specificity. It includes the operator's culture, age, gender, and personal traits (Hoff & Bashir, 2015). Situational trust encompasses external variability, which relates to the type of system and how complex it is, in addition to the internal variability of the operator's mental state (Hoff & Bashir, 2015). The third layer is the initial learned trust. This layer draws on the operator's preexisting knowledge of a system, which is based on past or current interactions, attitudes, and expectations when initially evaluating the automated system's trustworthiness (Hoff & Bashir, 2015). Initial learned trust correlates to an individual's trust before interaction with the automated system and is the differentiator between the dynamic learned trust layer-which is the trust that occurs during interaction with the system (Hoff & Bashir, 2015).

During the interaction, dynamic learned trust is affected by the system's design features and performance. It can change after a single interaction with the system, influencing the individual's reliance on the system. Design features and the appearance of the automation are essential considerations for influencing trust because they can indirectly influence an operator's perception of the system's performance (Hoff & Bashir, 2015). Hoff et al. (2015) also identified system design features to facilitate appropriate trust, such as ongoing feedback from the automated system about factors affecting its reliability and designing systems with anthropomorphism, transparency, politeness, and ease of use in mind (Hoff & Bashir, 2015). As an example, de Visser, Krueger, McKnight, Shied, Smith, Chalk, and Parasuraman (2012) investigated the effects of various types of automated agents on trust and performance by manipulating the humanness of the agent. De Visser et al. (2012) found that participants exhibited greater trust resilience as the anthropomorphism of the automated aid increased. System performance also plays a part in influencing trust. System performance factors include reliability, validity, predictability, dependability, the timing of errors, the difficulty of errors, type of errors, and system usefulness-where any or all of which can affect one's reliance on the system (Hoff & Bashir, 2015).

Hoff and Bashir (2015) concluded that many factors influence the three layers of trust, but a combination of human trusting tendencies, the situation, and perceptions of the automation drive trust in the system. Directions for future research in this area were suggested to focus on how an operator's past experiences influence trust and reliance on automation, researching how the automations aesthetics impacts trust, and how systematic manipulation of false alarms versus misses affects trust (Hoff & Bashir, 2015). Their model is a useful tool for visualizing the many factors that contribute to trust in automated systems. It also provides an excellent segue into the next discussion, which focuses on human reliance on automation.

E. RELIANCE ON AUTOMATION

The Cambridge English Dictionary (n.d.) defines reliance as "the state of depending on or trusting in something or someone." In regards to automation reliance, Madhavan et al. (2007) stated, "Automation reliance is based on the probability that an operator will use

the automation in the future and is influenced by the operator's level of trust in automation" (p. 291). Generally speaking, people are likely to rely more on trustworthy automation and tend to discard less trustworthy systems (Lee & See, 2004; Muir & Moray, 1996). However, Lee et al. (2004) suggested that human reliance on automation can be suboptimal, and a critical facet to automation use is appropriate reliance. On the subject, Lee et al. (2004) offered that trust influences reliance on automation but does not determine one's reliance on it. As a result, several factors, in addition to trust, affect how someone's intention to rely on automation transforms into actual reliance (Lee & See, 2004). Factor such as emotion, self-confidence in using the automation, perception of risk, subjective workload, effort to engage, time constraints, and system performance capabilities play a role in determining intention to rely on automation (Lee & See, 2004). Another factor that influences utilization and reliance is the operator's ability to perform tasks without the support of automation (Madhavan & Wiegmann, 2007). For example, when a human operator distrusts automation that has demonstrated poor reliability, the operator is likely to opt for manual operation (i.e., self-reliance) and ignore the automation (Dzindolet et al., 2003). The opposite of the above can also be exact. When an operator trusts in automation that is more reliable than manual operation, the operator is likely to rely on the automated aid over self-reliance (Dzindolet et al., 2003). In both examples, appropriate reliance on automation should occur (Dzindolet et al., 2003).

Work conducted by Parasuraman et al. (1997) discussed several factors and impacts of human reliance on automation. Mainly, they described inappropriate reliance in terms of misuse, disuse, and abuse. Misuse was defined as "overreliance on automation (e.g., using it when it should not be used, failing to monitor it effectively)" (Parasuraman & Riley, 1997, p. 233). Disuse was described as "underutilization of automation (e.g., ignoring or turning off automated alarms or safety systems)" (Parasuraman & Riley, 1997, p. 233). Furthermore, abuse was described as "inappropriate application of automation by designers or managers (e.g., automation that fails to consider the consequences for human performance in the resulting system)" (Parasuraman & Riley, 1997, p. 233). Factors contributing to misuse of automated systems can result from human errors such as decision biases (e.g., decision support automation reinforcing human decision heuristics) and failure

to monitor the automated systems accurately (Parasuraman & Riley, 1997). Overreliance in automation, specifically improper monitoring of automated aircraft systems, was noted as a contributing factor in several aircraft incidents (Parasuraman & Riley, 1997). Furthermore, Dzindolet et al. (2003) suggested that trust in an automated aid that is less reliable than manual operation or disuse of a system that is more capable than manual operation are examples of situations where inappropriate reliance on automation is likely to happen. Parasuraman and Riley (1997) emphasized that human-automation misuse, disuse, and abuse issues stem from differences between automation designers, managers, and operator's expectations of automation. Also, they highlighted the fact that human use of automation is a highly complex subject, encompassing its own set of patterns, characteristics, and influences. Hence, Parasuraman and Riley (1997) advocated for a better understanding of why automation is used, misused, disused, and abused in order to avoid human-automation errors, improve operator training, and overall system capability.

Situational factors can also influence one's reliance on automation. Hoff et al. (2015) identified situational factors that are not related to trust but which can guide reliance, such as alternatives to using automation, time constraints, and an individual's situational awareness, to name a few. Furthermore, attitudes, both positive and negative, can influence reliance on a system. Schaefer et al. (2016) asserted that positive attitudes can influence the liking of a system and may lead to over-reliance. Likewise, negative attitudes toward a system, typically stemming from difficulty interacting with or accessing information from a system, can lead to under reliance and disuse (Schaefer et al., 2016). Lastly, Hoff et al. (2015) concluded that "The strength of the relationship between trust and reliance depends on the complexity of automation, the novelty of the situation, the operators ability to compare manual to automated performance, and the operator's degree of decisional freedom" (p. 429).

F. SIMILARITIES AND DIFFERENCES BETWEEN HUMAN-HUMAN AND HUMAN AUTOMATION TRUST

Despite the extensive literature that addresses human-human trust and human-automation trust, significant comparative research has been sparsely reported between 1985 and 2016. However, some research was done to compare the two. One such study

was conducted by Lewandowsky et al. (2000) who investigated the dynamics of trust and self-confidence during a process control experiment, comparing human interactions with automation and human-human interactions under the same set of circumstances. The results of their study were informative in the comparison of the two forms of trust. Compared to the premise that interpersonal trust is resilient to momentary expectation violations (Rempel et al., 1985), they found there were moment-to-moment abrupt declines in trust between both the human-human and human-automation when faults occurred while attempting to complete tasks in the complex environment (Lewandowsky et al., 2000). This result led to a suggested implication that less than perfectly reliable automation will lead to a quick decline in trust in automation, and that the rapid declines in human-human trust imply that people expect others to conform to high standards as well (Lewandowsky et al., 2000). Hence, Lewandowsky et al. (2000) advocated that in some environments, moment-to-moment recalibration of trust in systems may be required.

Another finding from Lewandowsky et al. (2000) emphasized that during the experiment, operators were no more reluctant to delegate tasks to the automation than to the human partner. They implied that in certain environments, reluctance to delegate tasks is not necessarily driven by social mechanisms of human relations (e.g., doubt in a partner, unwillingness to share credit, or threat to self-esteem) (Lewandowsky et al., 2000). Instead, task allocation decisions are made by performance-based considerations of both the human partner and the automation (Lewandowsky et al., 2000). Finally, Lewandowsky et al. (2000) described the importance of calibrating perceived trustworthiness of the human partner and automation when allocating tasks, taking into account self-confidence in manual operations, partner reliability, and the shared responsibility of the outcome.

The work of Madhavan et al. (2007) noted that the research comparing human-human trust to human-automation trust points to humans responding socially to machines and automation and that trust is a critical psychological factor that influences acceptance or rejection of advice from decision support systems (DSS). However, there are contrasts in people's perceptions of human advisors compared to automated advisors. Madhavan et al. (2007) proposed a framework that synthesized and explained the process of trust development in humans versus automated aids, focusing in on the source of diagnostic

information, the source of credibility, and the source of reliability between the human advisor and the automated advisor.

In examining the sources of diagnostic information between human advisors and automated aids, Madhavan et al. (2007) noted several studies which support that information provided by automated systems are perceived to be more reliable and credible than the information provided by human advisors. These findings may be explainable by the attribution process, which is where a person makes inferences about an occurrence (i.e., interaction with automation) based on one's personality, disposition, or beliefs (Madhavan & Wiegmann, 2007). In the context of trust development in interpersonal relationships, trust often begins based on demonstrated performance or reliability and progresses based on dependability, and then further based on purpose or faith (Lee & See, 2004; Rempel et al., 1985). On the other hand, human trust in automation can develop in reverse of the above. In this pattern, initial trust in automation may be based on faith in the system, followed by system dependability and then predictability (Lee & See, 2004; Muir & Moray, 1996). However, as described previously, when automated system errors occur, user trust in such systems can rapidly decline. Dzindolet et al. (2001) offered that rapid drops in automation trust are the result of human expectations of automated systems to perform nearly perfectly, causing the operator to overly focus on the systems errors. Conversely, Madhavan et al. (2007) suggested that humans may have more realistic expectations of the information provided by human advisors based on familiarity with the human capacity of error. Since humans are not perfect, they may be more forgiving of incorrect information provided by the human agent as compared to incorrect information provided by the automated aid.

Following, Madhavan et al. (2007) explored the applicability of source credibility between the human and automated advisor. Studies have shown that source credibility is a significant influencer of information provided by an advisor (Madhavan & Wiegmann, 2007). In their study of credibility and trust in risk communications, Renn and Levine (1991) provided a definition of source credibility that is applicable for this study. They defined credibility as "the degree of shared and generalized confidence in a person or institution based on their perceived performance record of trustworthiness" (p. 179).

Similarly, according to the research of Madhavan et al. (2007), the most pertinent dimensions of credibility were related to perceptions of honesty, expertise, predictability, and reputation. In the context of automation, trust and credibility are frequently influenced by surface features of the source rather than the true capabilities of the system (Lee & See, 2004; Madhavan et al., 2006). These influences lead to interactions grounded in perceived expertise over the actual expertise of the system (Lee & See, 2004; Madhavan & Wiegmann, 2007).

Source reliability was also found to be another important factor in both human-human trust and human-automation trust. In the area of human relations, Mishra (1996) conducted interviews of 33 top-level managers in order to discern the nature of trust and its relationships to behavior during times of crisis. The results of the study identified that reliability was one of four key dimensions of trust between people, with the others being competency, concern, and openness (Mishra, 1996). Mirshra (1996) also noted that among people, one's reliability is salient when assessing trustworthiness. In other words, people generally trust others who demonstrate reliability. Human assessment of reliability when it comes to automation is similar in many respects. Some conceptual models of automation trust suggest that users of automation adjust their levels of trust to accommodate the level of system reliability. At the same time, other studies have shown that users tend to be either sensitive or insensitive to system reliability differences (Madhavan & Wiegmann, 2007). Additionally, other studies suggested that when an automated systems reliability is initially high, human trust and reliance on the system remains resilient to subsequent variations in reliability, but when the reliability is initially low, trust and reliance on the system tend to stay low, even if reliability improves (Lee & See, 2004). Madhavan et al. (2007) suggested that calibration of trust in DSS's can improve by changing human perceptions of automation to be closer to that of human perceptions of other humans, enabling similar levels of trust between automation that has differing levels of credibility and reliability. In order to advance the quality of interaction and human responses to automation, Madhavan et al. (2007) suggested improvements in automation design features are needed to invoke human responses to automation that reflect that of interpersonal interaction.

Finally, the study completed by de Visser et al. (2016) produced the first evidence that anthropomorphic features of the automation (i.e., human-like) can increase trust resiliency. Three experiments were designed to examine the opposing viewpoints of humans reacting to machines in specific ways compared to reacting to machines through the application of human social norms (de Visser et al., 2016). In their experiment, participants received advice that was gradually decreasing in reliability from either a human agent, a computer, or an avatar to explore the different effects of agent appearance and behavior on trust, compliance, and performance during decision-making tasks (de Visser et al., 2016). The results of the three experiments showed that anthropomorphic agents were associated with higher resistance to trust breakdowns and showed improved trust resiliency (de Visser et al., 2016). Another important finding was that high levels of uncertainty magnified the effects of the experiment and that the incorporation of human-like trust repair actions removed differences between the three agents (de Visser et al., 2016). Consequently, considerations for increasing anthropomorphism in the automation (e.g., apologizing for mistakes) may be an appropriate design consideration for increasing trust resilience when automation errors are sporadic or unique (de Visser et al., 2016). Furthermore, decreasing human-like features in automation design may be appropriate for reliable systems to appear more logical and to avoid human perceptions of the automation as being deceiving or lacking integrity (de Visser et al., 2016).

In sum, the research on the similarities and differences between human-human and human-automation trust suggests that humans have a natural inclination to react socially to machines and apply like response and filtering approaches in order to adjust their trust in both the human and automated teammate (Madhavan & Wiegmann, 2007).

G. EMPIRICAL RESEARCH

The foundations for the qualitative approach in this research were established by Jian et al. (2000). They conducted a three-phased experiment in order to develop an empirically based scale for trust measurements in automated systems. In their experiment, similarities and differences in the concept of trust, distrust, human-human trust, and human-machine trust were explored. The results of their experiment offered empirical

evidence that “trust and distrust are two opposite concepts, and that people’s perceptions of trust did not change when general trust was compared across human-human and human-machine relationships” (Abbass et al., 2016, p. 388). The 12-item scale they created is widely used in research as a subjective measurement of trust and is regarded as a staple for measuring trust in automation (Brzowski & Nathan-Roberts, 2019).

Building on the existing literature of trust in automation, Dzindolet et al. (2003) conducted three studies to explore the relationship between automation reliability, reliance, and trust. From the studies, Dzindolet et al. (2003) observed that participants considered the automation to be reliable and trustworthy initially. However, as automation errors occurred, even with reliable automation, participants distrusted the aid. Further, they also observed that when participants were provided an explanation as to why the automation failed, trust in the aid increased, even when the trust was unwarranted (Dzindolet et al., 2003). The results of the experiment demonstrated that trust was essential to understanding human-automation reliance decisions. Based on the findings, Dzindolet et al. (2003) concluded that optimization of automated aids would not successfully enhance human-machine performance. Instead, they suggested that understanding the processes humans use for dependency on automation will elicit better performance of the man-machine dynamic in the future.

Lee et al. (2004) built on the taxonomies of trust in automation of the early nineties. They highlighted appropriate trust and reliance on automation by considering trust from the organizational, sociological, interpersonal, psychological, and neurological perspectives. Their study honed in on the context, characteristics of the automation, and cognitive processes that affect the appropriateness of trust (Lee & See, 2004). In their research, Lee et al. (2004) also focused on the elements of performance, process, and purpose concerning automation trust. Lee et al. (2004) referred to performance as what the automation does in current and historic operating contexts, and it includes its competence, ability, and reliability. Next, process was described by how automation operates and the degree to which an automation’s algorithm meets the goals of an operator (Lee & See, 2004). Finally, Lee et al. (2004) noted purpose is why the automation was developed and the degree to which the system is being used for its desired purposes. The results of their

research brought to light gaps in trust and reliance on automation research-such as how automation interface features influence affect, the challenges of promoting appropriate trust, and how individual cultural differences influence trust and reliance (Lee & See, 2004).

Hancock et al. (2011) conducted a meta-analysis to identify the dimensions important for human trust in robots. In their meta-analysis, 21 empirical studies were used for comparison. For a study to be included in the analysis, trust had to be a direct measure of experimental manipulation, the examination of trust was focused on a robot, and the human participants engaged with a robot either visually, virtually, or through augmented means (Hancock et al., 2011). Their meta-analysis revealed that performance-based characteristics of the robot had the most substantial influence on perceived trust in human-robot interaction (Hancock et al., 2011). Suggestions for future research centered on human perception of robot intent, performance, and action through training methods aimed at aligning human perceptions more closely with that of the true capabilities of the robot. More research on the origins and sustainment of trust in robots using subjective and objective measurements was also recommended (Hancock et al., 2011).

Schaefer et al. (2016) conducted a meta-analysis of research on the development of trust in automation to inform the foundation on which autonomous systems could be built in the future. Their work is an expansion of the findings of Hancock, Billings, Schaefer, Chen, de Visser, and Parasuraman's (2011) three-factor model of human-robot trust to include all automation interaction. The analysis by Schaefer et al. (2016) explored how the human, automation, and the environment effects the development of trust. Their work yielded two important findings. The first finding identified human-automation interaction has an effect on trust, but the human-related antecedents of trust presented several research gaps (Schaefer et al., 2016). Notably, Schaefer et al. (2016) suggested further research was required into the dynamics of trust in automation concerning age across varying contexts, the relationship between traits, states, and trust, in addition to further studies of cognitive and emotive factors. Research into these characteristics will serve to enhance the overall knowledge and impact they have on the development of trust. A second significant finding by Schaefer et al. (2016), which supported the assessment of Hancock et al. (2011), was

the importance of the automation or robot's capabilities. Within their analysis, Schaefer et al. (2016) noted areas for additional research into cueing and feedback of more advanced autonomous systems in areas that involve interdependent operations, or human-autonomous teaming. Finally, Schaefer et al. (2016) assessed a lack of literature on the subjects of appearance and communication, which implies a need for more research into the features of automated systems.

Similarly, research conducted by Joe et al. (2014) was done to evaluate automation in relation to human teams. Their study was shaped by the existing literature on effective teams that work with complex systems. This was done to extrapolate general principles of teamwork. Then, based on those results, they examined the research on human-automation teams through the lenses of mixed agent teams and human-automation teamwork (Joe et al., 2014). The study revealed eight issues between human-human and human-machine teamwork. Challenges identified in the research included: 1) establishing shared mental models and understanding between human and machine agents, 2) anticipating intent, action, and team goals between the human and machine, 3) limitations in automation flexibility and adaptability in response to in new situations, 4) reductions in human-machine interaction with high levels of automation, 5) disruption of roles, responsibility, and teamwork when automation joins the human team, 6) the lack of responsibility, conflict resolution, or social acceptance on the part of the automation, 7) higher workloads when automation interaction requirements are high, and 8) communication issues between the human and non-human agents (Joe et al., 2014). The analysis by Joe et al. (2014) showed that principles of effective human teams do not necessarily correlate well to the automated agent due to the inherent differences between humans and automation. However, due to human propensity to interact similarly with automation in a human like way (Madhavan & Wiegmann, 2007), Joe et al. (2014) suggested greater emphasis is needed in research on human behavior and expectation calibration to facilitate genuine collaboration with the machine teammate.

McNeese et al. (2018) conducted one of the first experiments to better understand human-autonomy teaming (HAT) by pairing two human operators with an autonomous synthetic teammate during unmanned aerial vehicle (UAS) ground target photography

tasking. Team performance measurements were analyzed by efficiency in processing targets, situational awareness, and communication between the pilot, navigator, and the photographer (McNeese et al., 2018). The autonomous teammate in the mixed teams was able to perform UAS piloting tasks in addition to communicating textually but was lacking coordination skills (McNeese et al., 2018). Team composition included a synthetic team with an autonomous pilot, a control team with an inexperienced pilot, and an experimenter team with an experimenter acting as an experienced pilot. Ten of the three-team compositions participated in the experiment and were used as a comparison against all human teams (i.e., human pilot, navigator, and photographer) (McNeese et al., 2018).

Results of the experiment showed that the experimenter team performed better across all performance evaluation metrics compared to the control and synthetic teams. The synthetic team executed as well as the control team; however, it was less effective at processing targets (McNeese et al., 2018). This experiment demonstrated promise for autonomy working alongside humans, yet limitations in performance were observed. For instance, the synthetic team had difficulties efficiently coordinating together, and the autonomous teammate had difficulties understanding the informational needs of humans, suggesting further research into a better understanding of HAT, especially in more complex and dynamic environments (McNeese et al., 2018).

Finally, Brzowski et al. (2019) conducted a review of the empirical research done in the area of human-automation interaction (HAI) in order to provide a foundation for measuring trust in HAI. In addition, they aimed to provide a record of the research that utilized either subjective, objective or both types of measurements of trust in automation. As annotated in Table 2, 44 empirical studies met the criteria for inclusion in their research. Selection criteria required that the research articles 1) involved participants, 2) the article was open-source and available in full-text, 3) trust and other factors were measured, and 4) the measurement of trust involved human-automation, human-system, or human-computer interaction where some automation occurred in the test environment (Brzowski & Nathan-Roberts, 2019).

Table 2. Trust Measures by Type. Source: Brzowski and Nathan-Roberts (2019).

Measure	Articles
subjective only	33
objective only	5
both	6

Accepted articles for inclusion encompassed various forms of industry to include automotive, aviation, military, and augmented reality, to name a few. The study revealed that 75% of the accepted empirical research utilized only subjective measures of trust, 14% used a combination of subjective and objective measures, and only 11% used objective measurements of trust (Brzowski & Nathan-Roberts, 2019). Of note, according to Brzowski et al. (2019), six articles focused on the military, and all of them used subjective measures of trust in the research. Subjective measures of trust often utilized the Jian et al. (2000) Scale of Trust in Automation, utilizing limited focus questions, questionnaire forms, or Likert-type scales to assess and measure trust (Brzowski & Nathan-Roberts, 2019). Objective measurements were most often used in the automotive, aviation, computer science, and medicinal industries (Brzowski & Nathan-Roberts, 2019). These measurements were observed as acceptance rate, adoption rate, compliance, and glance rate, which was associated with automated control and navigation systems (Brzowski & Nathan-Roberts, 2019). Brzowski et al. (2019) concluded that subjective trust scales provide an accurate representation of a user's trust in automated systems. However, in order to ensure accurate HAI trust measurements, the use of a validated subjective scale combined with other system-specific measures may be necessary (Brzowski & Nathan-Roberts, 2019).

H. SUMMARY

Humans tend to react socially to machines (Madhavan & Wiegmann, 2007). However, calibration of human expectations of system capabilities, similar to human judgments of other people, is necessary for appropriate trust and reliance on the automation

(Joe et al., 2014; Lee & See, 2004). This review revealed gaps in the literature on trust in automation, and researcher recommendations for future work in specific areas were annotated.

Directions for follow-on studies included calibration of human expectations of intelligent systems (Joe et al., 2014), human-automation dependency (Dzindolet et al., 2003), cultural implications of trust in autonomous systems (Lee & See, 2004), the impact of automation aesthetics on trust (Hoff & Bashir, 2015), effects of anthropomorphic features on trust (de Visser et al., 2016), autonomous system communication (e.g., cueing) and feedback (McNeese et al., 2018; Schaefer et al., 2016), and further research on human perceptions of robot intent and performance (Hancock et al., 2011). The final empirical review explored trust measurement methods used in research. The work of Brzowski et al. (2019) noted a large amount of research on HAI was conducted using subjective measures of trust, while a relatively small percentage of studies used objective measures or a combination of both.

This points to a need for further research on the subject of man-machine teaming, much of which would be of interest to the DOD. Further research would provide valuable insight into the design and acquisition considerations of future autonomous systems. As such, this literature review has provided a solid foundation for examining the characteristics and development of trust between human and autonomous systems in a team dynamic. The results have identified an opportunity for further investigation into performance-based metrics of trust in man-machine teaming in an environment of risk and complexity using a combination of subjective and objective measurements.

III. METHODOLOGY

A. PARTICIPANTS AND DESIGN

1. Participants

During four months of testing, 51 volunteers from the NPS participated in the study. In accordance with the NPS Institutional Review Board, eligible participants included NPS students, staff, and faculty aged 18 and older. Participants were recruited through NPS bulk email solicitation, posted flyers, and word of mouth. Experiment sessions were self-scheduled by the participant using the www.signupgenius.com web application.

There were no risks associated with this study greater than that which participants are exposed to in normal daily life and gaming environments. However, extended time in virtual reality can lead to dizziness or nausea in some people. Also, mild fatigue or motion sickness is a possible adverse effect of playing games in virtual reality. Therefore, potential performers with known vestibular or motion sickness concerns or negative experiences in a virtual reality environment were discouraged from participating.

Of the 51 participants, 47 were male, and four were female. Participant ages ranged from 23 to 62 years old ($M_{age} = 32$, $SD = 8.10$). As depicted in Figure 4, the majority of participants were members of the United States Marine Corps (USMC) at 37%. However, service members of the United States Navy (USN), United States Army (USA), United States Air Force (USAF), and the United States Coast Guard (USCG) were represented as well. Also, nine NPS civilian (CIV) staff and faculty participated, along with four foreign national (FN) NPS students.

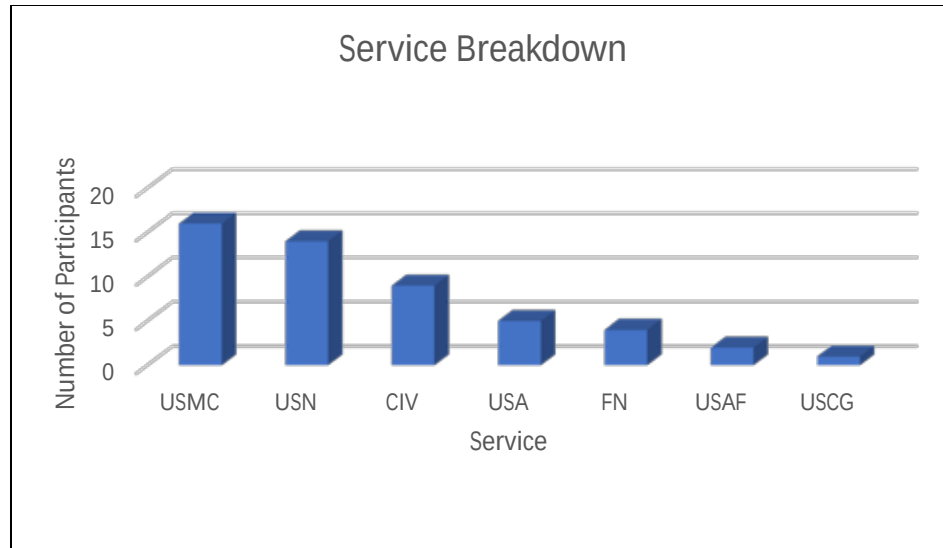


Figure 4. Participant Service Breakdown

Other participant demographic information such as average gaming experience (scaled 1–7) by age group, average VR gaming experience (scaled 1–7) by age group, and participant’s preferences for game genres were collected and are displayed in Figures 5–7.

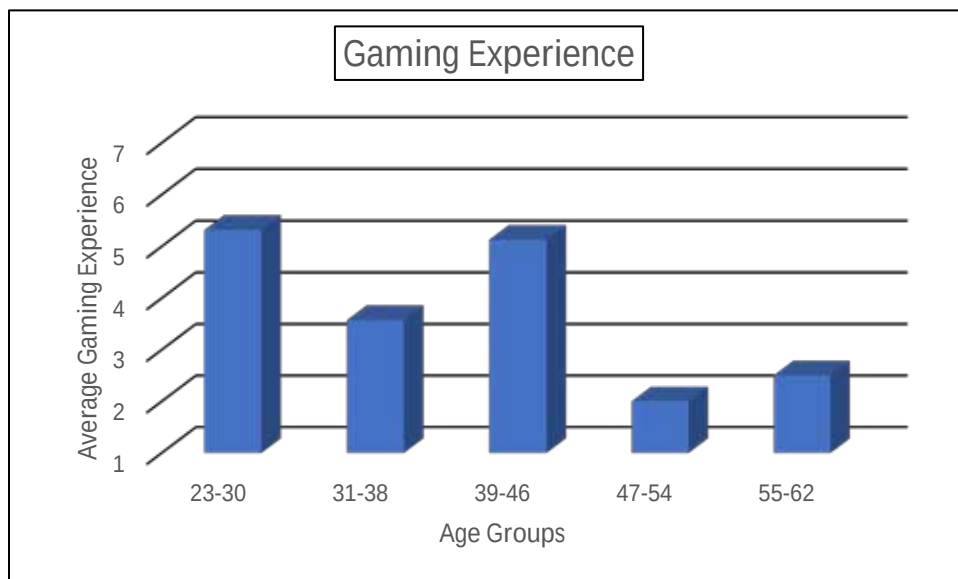


Figure 5. Participant Gaming Experience

The mean participant gaming experience was 4.39 out of 7, and the mean participant VR gaming experience was 2.35 out of 7. Of the nine game genres (e.g., Action, Action/Adventure, Adventure, Role Playing, Simulation, Strategy, Sports, Puzzles, and Idle), 61% of the participants preferred or played Action/Adventure games (e.g., The Legend of Zelda) compared to only 6% who typically played Idle games (i.e., clicking games).

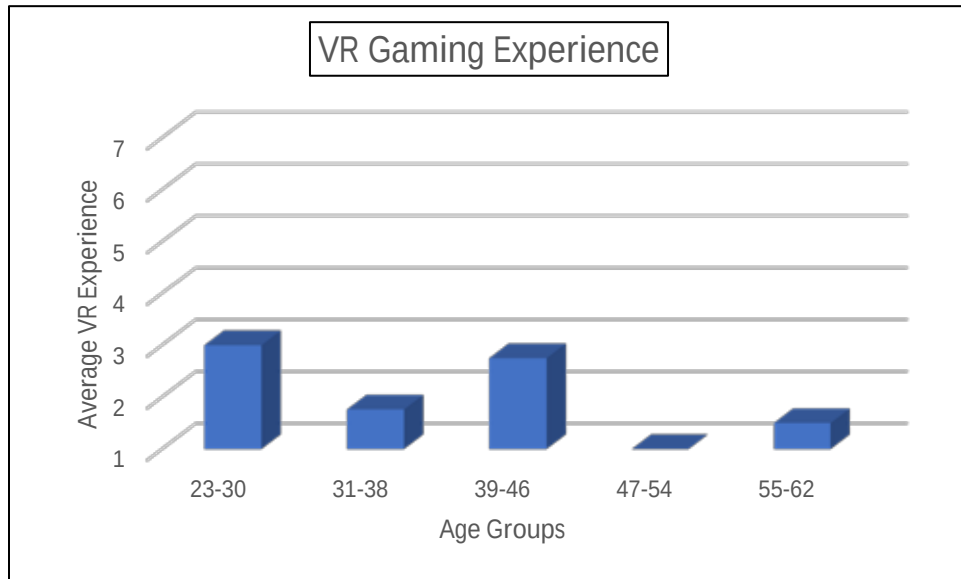


Figure 6. Participant VR Gaming Experience

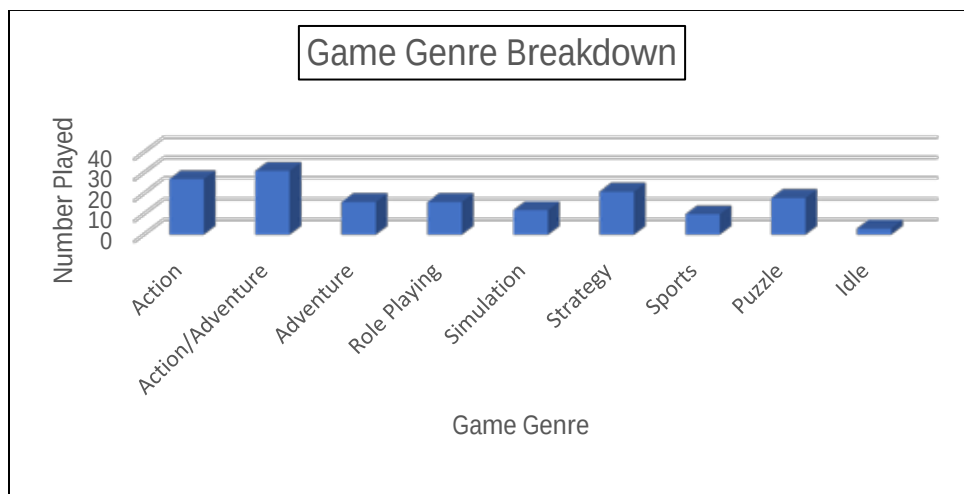


Figure 7. Participant Game Genre Preference

2. Experiment Design

a. Background

Participants played four rounds of a VR spatial puzzle type game. The game, called ESCAPE, was designed by the JHU-APL specifically for this experiment as the platform to measure trust scores between both homogeneous human teams and human-autonomous teams. The participant's teammate was either a member of the research team who played the game from a separate enclosed room in the laboratory and followed a script to act as a thinking teammate to limit giving away the solution to the puzzle. Alternatively, the participant's teammate was an autonomous bot that was developed by JHU-APL to play each round of the game. Deception was employed by withholding the true identity of the teammate for each round in addition to telling the participants their human partner was playing remotely from the JHU-APL on the east coast. This aspect of deception was done to ensure that bias about performance expectations between humans and bots could be controlled.

The intent for each of the increasingly difficult four rounds of Escape was to have the participant and their teammate complete (i.e., escape) each round together as a team. In order to be successful, the team had to work together to navigate through the virtual environment, overcome any obstacles they faced, and exit the round together within the allotted time of seven minutes per round.

b. Laboratory and Experiment Setup

The laboratory contains two rooms separated by an internal door. During gameplay, the participant was alone in the primary laboratory room (Room A) while the researcher controlled the human or bot teammate avatar from the control room (Room B). Room A is shown in Figure 8.



Figure 8. Room A

As depicted, Room A contained the main computer which ran the ESCAPE program, received the games output data, and supported the Oculus Rift-S VR-HMD and associated Touch Controllers. The Oculus Rift-S VR-HMD was connected to the main computer via a universal serial bus (USB) cable and was long enough to permit the Oculus Rift-S VR-HMD to be hung from the ceiling. Hanging the device provided participants plenty of maneuver space to walk, twist, bend, and reach during gameplay within the approximate 5' x 7' Guardian virtual play area established by the Oculus Rift-S software. In addition, hanging the Oculus Rift-S VR-HMD negated the cable from wrapping around the participant's legs during gameplay.

Room B, as shown in Figure 9, contained the secondary computer, which served as the console for the researcher to play as either the human or bot teammate. In order to maneuver the avatar teammate in the game, the researcher utilized the secondary computer's keyboard, mouse, and monitor. Enabling the autonomous bot mode for any level was accomplished by toggling the (p) key on the keyboard. A secondary monitor was installed in the room and mirrored what the participant saw in VR. This addition allowed

the researcher to observe what the participant was doing when obstacles or walls obscured them.



Figure 9. Room B

c. Picking Teams

When the participant arrived at the laboratory, a simple Python script was used to generate a random number (up to four digits) to serve as the volunteer's participation number for questionnaire tracking and data input purposes. A second Python script was used to generate a random number consisting of a 1 or 2, serving as the random assignment to the consistent or inconsistent condition. Participants in the consistent condition played the initial round (Round 1) with the partner of their choice (i.e., human or bot). Participants in the inconsistent condition played the initial round with the partner opposite of their choice. Participants were permitted to choose their partner, either a human or the bot, before the first round when filling out the pre-Round 1 questionnaire and after playing the second round and completing the associated end of round questionnaire (i.e., for their third-round partner). Teammate selection is where one of the aspects of deception comes into play as participants were led to believe they were playing the initial round with the partner

of their choice. Participants were then told that they would play with the opposite partner in Round 2, and they were switched to the opposite partner that they played with in Round 1. Before Round 3, they were asked their preference again. If they received their preference in Round 1, then they would receive the opposite of their preference in Round 3, and vice versa for those that initially received the opposite of their preference in Round 1. Finally, the partner in Round 4 was the opposite partner in Round 3. This setup for teammate pairing ensured that every participant played two rounds with a human and two rounds with the bot, one of each that was consistent with expectation, and one of each that was inconsistent with expectation.

d. Gameplay Scenarios

There were eight scenarios possible in this experiment. Each scenario was based on the random reference number provided to the participant prior to the start of the experiment, the participant's choice of partner before the first round, and their choice of partner for the third round, as described above. Figure 10 shows the experiment's scenario sync matrix used for initial teammate pairing and scenario flow. Scenarios were identified as: 1A2C, 1A2D, 1B2C, 1B2D, 2A1C, 2A1D, 2B1C, and 2B1D. This naming convention was chosen in order to differentiate which rounds the participants were playing with the actual teammate they selected and whom they believed they were playing with during the round.

1 st Half					
Ref#	(1) Consistent		Ref#	(2) Inconsistent	
	1A	1B		2A	2B
Rd1	H/H	B/B	Rd1	H/B	B/H
Rd2	B	H	Rd2	H	B
2 nd Half					
Ref#	(2) Inconsistent		Ref#	(1) Consistent	
	2C	2D		3C	3D
Rd3	H/B	B/H	Rd3	H/H	B/B
Rd4	H	B	Rd4	B	H

Figure 10. Scenario Sync Matrix

The scenario sync matrix is divided into four quadrants, with the top two horizontal quadrants representing the first half of the experiment (i.e., Rounds 1 and 2) and the bottom two representing the other two. In addition, the top two horizontal quadrants identify the reference number given to the participant at the beginning of the experiment. This number signified whether the teammate the participant plays with would be consistent or inconsistent with their choice for Round 1. Thus, the bottom two quadrants are the opposite. For example, if a participant draws a reference number of 1, they would begin in the top left “consistent” quadrant. If the participant chooses to play with human in Round 1, they would actually play with a human (annotated as H/H) in the 1A column. After playing Round 2, the participant has the option to choose their teammate for Round 3, and the scenario can branch off to either 2C or 2D, both inconsistent with their actual choice. If the participant chooses to play with a bot in Round 3, they would actually play with a human (annotated as B/H) in the 2D column. The overall scenario, in this case, would be 1A2D, and the same methodology applied to the other seven possible scenarios.

B. MATERIALS

1. Oculus Rift-S VR-HMD

The Oculus Rift-S VR-HMD was developed by Oculus VR and manufactured by Lenovo. The Oculus Rift-S display is a fast-switch liquid crystal display panel with a resolution of 2560 x 1440 (1280 x 1440 per eye) at an 80 hertz (Hz) refresh rate. It uses an inside-out tracking system consisting of a constellation of five cameras (i.e., one looking upward, one on the left and right, and two in the front) in addition to accelerometers located in both the headset and controllers to track, predict, and display (with the use of AI) the movement of the controllers, even when out of view of the headset's cameras. The Oculus Rift-S also has a speaker system integrated into the halo type adjustable headband for audio and comfort; however, there was no capability for voice communication during gameplay. In-game communication was conducted using the Oculus Touch controllers and will be described below. The Oculus Rift-S VR-HMD is shown in Figure 11.



Figure 11. Oculus Rift-S Headset. Source: Oculus (n.d.-a).

Oculus Rift-S manufacturer recommended (minimum) personal computer specifications include: an NVIDIA GTX 1050TI or AMD Radeon RX 470 or greater graphics card, an Intel i3-6100 or AMD Ryzen 3 1200, FX4350 or greater central processing unit, 8 gigabytes (GB) or more of random access memory (RAM), a DisplayPort 1.2 for video output, one USB 3.0 port, and the Windows 10 operating system. The computer used to support ESCAPE (main game configuration) in Room A was a Hewlett Packard Elite Desk 800 G4 Workstation Edition desktop. The processor was an Intel i3-6100 (@ 3.20 gigahertz (GHz)), with 16.0GB of RAM. Its operating system was the Windows 10 Pro edition. In Room B, the computer that supported the partner configuration was an MSI Trident 3 mini desktop personal computer. This computer had an Intel i7-8700 processor (@3.20GHz), with 16.0GB of RAM and the Windows 10 Home Edition operating system. Both computer systems met the minimum specification requirements to support the Oculus Rift-S and the ESCAPE program. Finally, all three computer monitors (i.e., one in Room A and two in Room B) were the acer Predator 24" XB241H type monitors.

2. Oculus Touch Controllers

Each Oculus Touch controller is ergonomically designed for comfort and allows for the virtual representation of both hands and hand gestures for gameplay. Both controllers come equipped with a vertical ring that is embedded with infrared light emitting diodes for controller tracking and display with the headset. Also, each controller has two action buttons (A) and (B), one side grip button, a moveable analog stick, one index finger trigger button, and a menu button. For reference, the Oculus Touch controller is depicted in Figure 12.



Figure 12. Oculus Touch Controllers. Source: Oculus (n.d.-b).

In ESCAPE, the index trigger button served two purposes. The first function simulated a laser pointer (when not holding the tractor beam gun) and was the primary means of communicating with the human operator or the bot. The laser pointer allowed the participant to illuminate an object of interest for action and also served to illuminate spots on the ground where the participant wished the teammate to move. The other function of the index trigger button was to shoot the tractor beam gun (i.e., the participant could pull or push objects or their teammate across gaps too far to jump across) when it was held using the side grip buttons. The (B) action button was used to display a virtual five-second countdown timer for partner coordination and communication. This function could be used by the participant to command the teammate to actuate platform buttons, pick up and move boxes, or move somewhere in the game as desired. Using the (B) timer button for commanding teammate movement was optional as the bot (and human operator) would usually recognize the command without the timer and move accordingly. An example of using this function would be after the participant illuminated a platform actuation button with their index finger trigger button (an action icon would be displayed above the actuation button), they would then press the (B) button to display the five-second

countdown timer. The timer would cue the teammate to press the button at time zero for platform actuation. This type of communication and action was practiced during the ESCAPE training session, which will be described later. The (A) action button on the right-hand controller simply enabled the avatar to jump. The side grip button on both controller's inner hand grip allowed the participant to virtually grab, hold, and move with objects such as boxes or the tractor beam gun. A moveable analog stick (used on the left controller only) enabled movement of the avatar (i.e., forward, back, left, and right). However, to yaw left or right, either standing or while moving, the participant would have to move their body and the VR-HMD in the direction they wished to view or move. A complete list of ESCAPE objects and obstacles the team encountered can be viewed in Appendix A.

3. ESCAPE

ESCAPE was created at JHU-APL using the Unity game engine. The Unity game engine provides game creators the ability to program 2D, 3D, and VR games and supports the Oculus Rift-S platform. Further, the game was designed to enable multiplayer gaming from separate computers via Wi-Fi connection. The main ESCAPE game configuration (for use with the Oculus Rift-S VR-HMD) was used in Room A. The partner configuration (for use by the human partner via keyboard and mouse) was employed in Room B. As described previously, the ESCAPE software recorded and saved gameplay performance data within a results folder in the games directory of the primary computer.

Round 1 of ESCAPE is the easiest of the four rounds. In this round, as depicted in Figure 13, the team was required to coordinate who will press the platform activation buttons that enable the retractable platforms, and who will jump across those platforms to get to the other side to unlock the exit door. In order to execute this task, one partner has to synchronize pushing the two platform activation buttons, and the other has to time their jump across the platforms to avoid falling into the “green goo” and dying. The mean time for completion was 94.53 out of 420 seconds, and 100% of participants successfully passed the round.

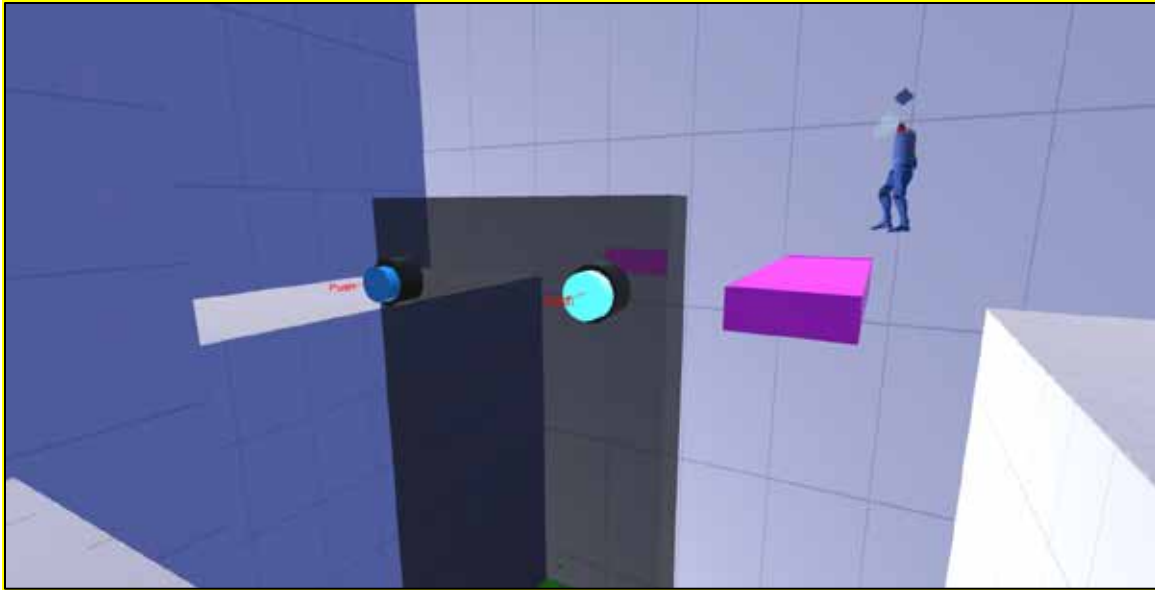


Figure 13. ESCAPE Round 1

Round 2 of ESCAPE is slightly more challenging than Round 1. As shown in Figure 14, this round featured a pressure plate on the floor, which, when depressed, activated retractable platforms which were required to jump across to the exit. The pressure plate also activated deadly green goo rising from the floor below. The challenge in this round was for the team to coordinate who was to retrieve the box (not shown in Figure 14) which was required to maintain pressure on the plate in order to allow both teammates to jump across the platforms to the exit before being killed by the rising green goo. The mean time for completion was 211.98 out of 420 seconds, and 80% of participants successfully passed the round.



Figure 14. ESCAPE Round 2

Round 3 added another layer of difficulty for the team. In this round, the participant must use a tractor beam gun to shoot a box across the room in order to activate a pressure plate that is on a retractable platform (activated by the teammate via an activation button). Once the pressure plate was activated, retractable platforms extended to allow the teammate to jump down to the mid-level platform below. The task was challenging because the participant must coordinate the timing of when the teammate presses the platform activation button in order to correctly time when they were to shoot the box across to engage the pressure plate. Also, once the pressure plate was engaged (the platform remains in the out position), a laser fence is enabled, which comes out of the wall and tracks toward the participant. The participant must maintain pressure on the plate (via the box and tractor beam gun) while simultaneously moving toward the edge of the platform long enough for the platforms to remain extended so the teammate can jump down to the mid-level platform. Once the teammate is at the mid-level platform, the participant must release pressure on the plate to allow the box to drop and disengage the moving laser fence (also retracts the platforms). The actions required for the participant to get the teammate to the mid-level platform is shown in Figure 15.



Figure 15. ESCAPE Round 3

From there, the participant must shoot another box down to their teammate on the mid-level platform where there is another pressure plate located on the floor that functions in the same manner as the other one. The participant must then coordinate to have the teammate pick up the box and put it on the pressure plate (or the teammate can stand on it), so the participant can jump down to the mid-level platform. The key to this round is placing the box on the mid-level platform pressure plate before the teammate jumps down to the bottom level where the exit is located. This is a crucial step to accomplish because the exit door is guarded by another laser fence, which is only deactivated when the pressure plate is depressed (i.e., by a box or the teammate). The mean time for completion was 287.96 out of 420 seconds, and 75% of participants successfully passed the round.

The fourth and final round of ESCAPE was the most challenging. In this round, the team must coordinate in order to activate two sets of actuation buttons, one set of buttons to activate two retractable platforms, and the other set of buttons located across the gap in the room to activate the platform connector cables needed to lock the platforms in-place. Locking the platforms in-place was required to form one half of the “ladder” needed to climb to the top of the wall located on the left side of Figure 16. The most difficult aspect of this round is getting either teammate across the gap in the room to activate the platform

connector cables. Most participants did not realize they could use the tractor beam gun to shoot their teammate across the gap to the other side. Once either teammate made it across the gap to the other side of the room, a coordinated effort was required to lock the platforms in place. After that was completed, additional teamwork was required to get the box on the other side of the room, and the teammate back across in order to build the second half of the ladder needed to get to the top of the wall. When both teammates made it to the top of the wall, they were required to use the main tractor beam (i.e., the red object in the upper right side of Figure 16) to be pulled across to the exit. The mean time for completion was 405.38 out of 420 seconds, and 14% of participants successfully passed the round.



Figure 16. ESCAPE Round 4

a. Limitations

ESCAPE was a prototype game created for this experiment. As such, the game had a few software issues that were revealed during pilot testing. All but one of the issues was disclosed to the participants during the in-brief in order to manage expectations during gameplay. The one issue that was not disclosed, however, was how the bot occasionally got out of sequence with the participant's commands or its autonomy protocols. Withholding this information was done to ensure there were no biases about the bot's

performance before gameplay began. Toggling control from bot mode to human control for intervention, as described below, was effective at mitigating this issue, and no participant reported observing teammate anomalies in that temporary condition.

Another issue that occasionally occurred was the participant's avatar would lock-up or freeze after virtually dying in the game. When this happened, participants were instructed to physically move forward, back, left, and right (while also moving the analog stick) and press the (A) button to jump out of the frozen state. This method occasionally worked to free the avatar, and gameplay would continue. However, if the game was permanently locked-up or frozen, there was no way to reset the game at the specific point in time the error occurred. For example, if the game froze during play in Round 3, the game would need to be reset and started at Round 1. When a restart was required, participants were offered the opportunity to replay the rounds with a member of the research team (not bot) or have the research team 'administratively' progress them to the point where the error occurred. An administrative reset consisted of one research team member playing the game in Room A using the VR-HMD and the other member playing as the teammate in Room B. Most participants chose to replay the previous rounds. In either case, repositioning the participant to where the error occurred usually took no more than 5 minutes. Participants were not penalized for the error, as the data from the game prior to the error was saved, and the recording of game performance continued once reset.

4. Questionnaires

In order to collect subjective performance metrics, five questionnaires were provided to all participants. The questionnaires for this experiment were created at the JHU-APL, with slight modifications made by the research team to account for recommendations provided during pilot testing, in addition to adding a demographic section at the end of the Round 4 questionnaire. The first questionnaire was administered after the training session during the break before the start of Round 1. This questionnaire contained three questions, two about prior collaborative gaming experience, and the other solicited the participant's preference for teammate during Round 1, given the team selection criteria described previously. There were three sections in the follow-on

questionnaires, each with identical questions. The first section consisted of three questions asking about overall personal performance, overall teammate performance, and the overall team performance during that round. The second section was a 20 question 1–7 Likert Scale and covered general aspects of game strategy, difficulty, and team chemistry. The final section consisted of 26 questions eliciting responses as to what percentage of the time (i.e., 10–100%) the teammate exhibited trustworthy qualities such as competence, dependability, reliability, or predictability. The end of the Round 4 questionnaire contained an additional six questions following the third section. These questions focused on determining whom the participant believed they were playing with, either a human or the bot during each round, and also provided an opportunity to do a final ordinal ranking of the performance of the perceived human and bot teammate. Finally, a six-question demographic section created by the research team solicited information about age, gender, service, gaming experience, VR experience, and preferences for game genres. For reference, each questionnaire is shown in full in Appendix B.

C. PILOT TESTING

Pilot testing for the experiment phase was conducted during October and November of 2019. The goal of pilot testing was to observe volunteers play the ESCAPE training session and the subsequent four rounds in order to gauge the pace and time requirement for each round and questionnaire session, test the viability of split-room multiplayer gaming over the NPS Wi-Fi, and to practice human and bot partner protocols for each round.

Before pilot testing, the research team played multiple iterations of each round of ESCAPE and worked with the game’s programmer at the JHU-APL in order to understand the bot’s autonomous actions and to learn how to collaboratively solve each puzzle in both the human-human and human-bot scenarios. From this iterative approach, the research team identified instances where operator intervention was required to ensure that the bot performed as originally programmed in certain scenarios. As a result, the research team developed human-bot protocols (see Appendix C).

When playing in the human-bot composition, and intervention was necessary by the researcher, the protocol enabled taking control of the bot to simulate autonomy in any

scenario when necessary. For example, if the bot's autonomy protocol got out of sequence during play, the researcher could manually toggle control from bot to human mode (utilizing the (p) keyboard button) and position the bot in the appropriate position in order to re-sync the bot with the current scenario. After this was completed, toggling control back to bot mode would resume the bot's autonomous action. Through practice, this enabled the researcher to fluidly manage software errors without cuing the participant as to the state of the teammate.

Three NPS students volunteered to participate in three pilot testing sessions. Volunteer feedback and research team observations led to several procedural modifications that improved the efficiency of the experiment sessions. During each pilot test, the various components of the experiment were timed in order to gauge overall time requirements. The overall time requirement for each session was found to take between 1 and 1.5 hours. This time estimation was used for NPS bulk email solicitation purposes and to support scheduling multiple participants during a workday.

Through volunteer feedback, unclear questionnaire items were re-written or removed to ensure clarity. Volunteer feedback concerning VR sickness led to testing and approval for participants to play the game in a wheeled desk chair to mitigate sickness. In the event actual VR sickness occurred during gameplay, participants were asked to stop movement, remove the VR-HMD, and sit in a chair and relax until they were able to continue. Also, during pilot testing, it was observed that the participant's legs were becoming entangled with the VR-HMD's USB cord. The remedy to this problem was to install a ring in the ceiling to allow the VR-HMD to hang from a position which enabled full movement within the laboratory's working area and inside the Oculus Rift-S' Guardian virtual play area (i.e., approximately 5'x7'). Finally, pilot testing revealed the need for another monitor in Room B to observe what the participant was viewing through the VR-HMD in order to improve team coordination.

D. PROCEDURE

1. Experiment

The experiment consisted of the following components: participants self-scheduling using the signup genius web application, reading and signing an informed consent form, a verbal pre-brief of the event (see Appendices D and E), and demonstration of the Oculus Rift-S VR-HMD and Oculus Touch controllers. After the demonstration, participants played one round of the ESCAPE training session followed by playing the four rounds of the game and completing the associated end of round questionnaires. The experiment concluded with a final debrief and re-signing of the informed consent.

After participants self-scheduled using the signup genius application and arrived at the laboratory in Glasgow Hall Room 103, they were asked to read and sign an informed consent form. The informed consent form explained the purpose of the experiment, informed the participant of the potential adverse effects of playing in VR, and provided an overview of the experiment's remaining flow (i.e., training session, four game rounds, and debrief session). During this time, the primary investigator used the python script to generate a participant number for data tracking purposes and a reference number (1 or 2) for initial teammate pairing.

The verbal pre-brief explained the structure of the experiment and covered instructions on ESCAPE gameplay. Also, during this time, the primary investigator disclosed known Escape software issues in order to manage participant expectations of the game and gameplay. After the verbal briefing, the primary investigator explained the safe movement areas in Room A as defined by the Oculus Rift-S Guardian virtual play area (see Figure 17) to ensure participant safety. Finally, participants were introduced to the Oculus Touch controller's buttons/moveable analog stick control and were shown how to fit and wear the VR-HMD during gameplay.

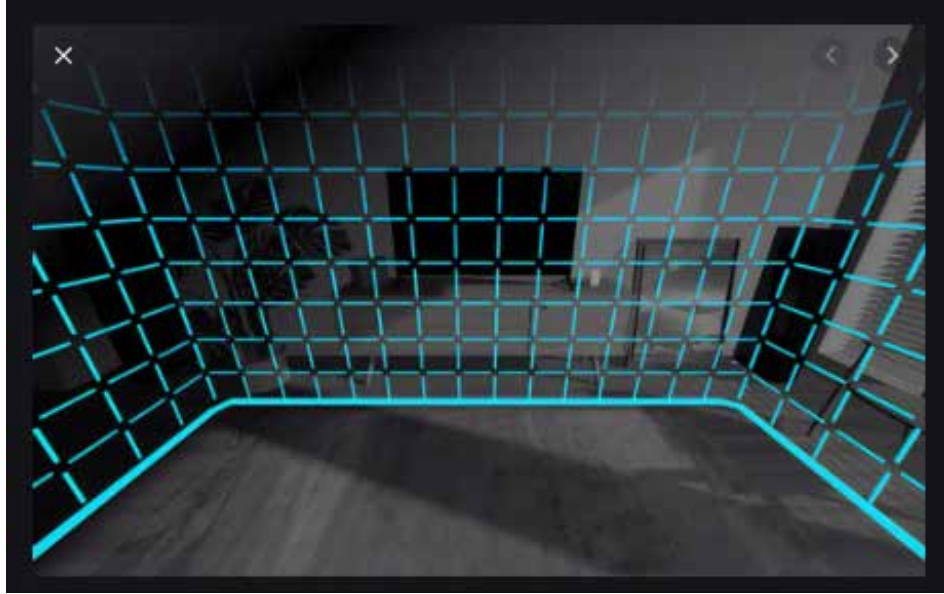


Figure 17. Example Oculus Rift-S Guardian Virtual Play Area. Source: Baker (2020).

Following the pre-brief, participants played one Escape training session (with the assistance of the primary investigator), as shown in Figure 18. The training session allowed the participants to become familiar with the Oculus Rift-S VR-HMD and Oculus Touch controllers usage, adapt to the Escape virtual environment and gain familiarity with the obstacles they would encounter and the objects they could manipulate during gameplay. Once the training session was complete, participants were asked to fill out the pre-Round 1 questionnaire in order to choose their partner, either a human or the bot, for the first round. However, they may or may not have received their partner preference depending on the randomly assigned condition described previously.



Figure 18. ESCAPE Training Session

After the training session and choosing their partner for Round 1, participants began gameplay with either a human or bot teammate. The human-human or human-bot teams had seven minutes to complete each round. After completing a round, the participant's avatar would enter a virtual waiting room. Of note, at the end of seven minutes, and the team did not successfully solve the puzzle, the participant's avatar was slewed to the virtual waiting room, and the round was logged as a failure. The virtual waiting room is an empty room with a single button on the wall, which, when pressed, started the next round. Upon entry in the virtual waiting room, participants were asked to take off the VR-HMD and Oculus Touch controllers and sit at a table in order to fill out the associated end of round questionnaire. In addition to collecting subjective performance data, sitting and filling out the questionnaires was also used to provide the participants with a break from the VR environment in order to mitigate the occurrence of VR sickness. Once questionnaires were completed, participants were asked if they felt up to playing another round, and if so, they would don the VR-HMD, press the virtual button on the waiting room wall, and begin play in the next round. If the previous round was Round 4, the participants were complete with VR play, and after the final questionnaire was completed, the experiment debrief commenced.

The debrief session reiterated the purpose of the experiment and disclosed the use of deception by providing the identity of the teammate they played within each round. Also, if a team failed to complete a round, the primary investigator would provide the solution for solving that round. This time was also used to obtain a second signature on the informed consent form to allow the research team to use their gameplay data. The debrief was designed to take up as much of the planned 10 minutes as possible in order to ensure participants were not feeling any effects of VR sickness. Before departing, participants were kindly asked not to disclose any information about their session in order to preserve the integrity of the experiment.

2. Data and Collection

During the experiment, objective performance data was recorded by the ESCAPE software saved as a Microsoft Notepad text editor file in the ESCAPE file directory of the main computer. Each Notepad text editor file was renamed according to the date, time, and experiment scenario. The performance data included the date, time, puzzle round, success or failure, time in seconds (up to 420), subject number of deaths, and partner number of deaths. Subjective performance data was obtained using questionnaires provided to the participant before Round 1, and between each successive round.

After each experiment session, the primary investigator input the participant's performance data generated by the Escape program and participant questionnaire data into Microsoft Excel for organizational purposes. At the completion of the experimentation phase, the research team transferred the Microsoft Excel data to the International Business Machines (IBM) Statistical Package for Social Sciences (SPSS) Statistics Premium Grad Pack version 26 for statistical analysis. The next chapter will discuss the results of the experiment phase of this research.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. EXPERIMENT RESULTS

Results from the collected objective and subjective measures of trust between the human-human and human-autonomous teammates were analyzed using Microsoft Excel for initial data organization and for producing necessary statistical information. Further analysis of the data was done using the IBM SPSS Premium Grad Pack version 26.

The following analysis is broken down into four sections. Section A covers the analysis of the objective measures of trust, where the variables, conditions, and test results will be explained. Section B examines the analysis of the subjective measures of trust following the same format as above. Section C covers the results from paired sample tests, and Section D will highlight the exploratory analysis conducted during the research.

A. OBJECTIVE MEASURES OF TRUST

Objective performance metrics that may indicate trust and those that are important in the development of trust included the participant's partner preference (obtained from the pre-Round 1 and pre-Round 3 questionnaires), in addition to the metrics obtained from the ESCAPE output data such as round time, the number of VR deaths, and whether each round was scored as a pass or a fail. The performance metrics listed above served as the dependent variables during analysis. The independent variables used for analysis included the consistent condition (i.e., the participant played with the teammate of their choice in Round 1), inconsistent condition (i.e., the participant played with the teammate opposite of their choice in Round 1), the participant's perception of teammate (i.e., a human or a bot), and the actual condition.

1. Consistent and Inconsistent Conditions

a. Partner Preference

The first objective measurement was whether there was an overall change in partner preference between the pre-Round 1 preference and the pre-Round 3 preference. The preferences were coded so that 0 = preference for a human partner, and 1 = preference for a bot partner. A paired t-test was conducted, and there was no difference in preference for

the human or bot between the pre-Round 1 ($M = .53$, $SD = .50$) and the pre-Round 3 ($M = .49$, $SD = .51$) choice; $t(50) = .41$, $p = .69$.

Additionally, an analysis was run to test if there was a difference in the change in partner preference between the consistent and inconsistent conditions. An independent samples t-test showed that there was no difference in change in partner preference between the consistent ($M = .12$, $SD = .73$) and inconsistent ($M = -.039$, $SD = .66$) conditions; $t(49) = .82$, $p = .42$.

Finally, a test was done to determine if there was any significance between the participant's perception of whether their teammate was a bot or a human between the consistent and inconsistent conditions. The analysis showed that there was no main effect of correct partner choice, $F(1, 47) = .63$, $p = .60$, $\eta_p^2 = .039$, however, the interaction between rounds was found to be significant, $F(1, 47) = 4.14$, $p = .011$, $\eta_p^2 = .21$. Further, t -testing revealed a significant difference between the consistent ($M = .60$, $SD = .50$) and inconsistent ($M = .15$, $SD = .37$) conditions for Round 1; $t(49) = 3.64$, $p = .001$. The results of these data suggest that in Round 1 participants in the inconsistent condition were successfully led to believe they were playing with their chosen teammate when in fact they were not, but it was not the case for the subsequent rounds.

b. Round Time

Round time is of interest in trust as it may indicate how much the participant relied on or worked with their partner, as more time might be indicative of trying to figure out how to escape without working collaboratively. A 4 (round time) x 2 (consistency condition) repeated measures ANOVA was run to assess whether there was a difference between the consistency conditions and round time. There was a main effect for time per round of $F(3, 47) = 325.92$, $p < .001$, $\eta_p^2 = 1.00$. There was no interaction found between rounds and the consistency condition, $F(3,47) = .85$, $p = .47$, $\eta_p^2 = .47$. As shown in Table 3, all rounds were found to be significantly different from each other, meaning as the rounds progressed, the team round time increased.

Table 3. Pairwise Comparison of Time by Round

Pairwise Comparisons						
Measure: MEASURE_1						
(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	-117.267 [*]	19.985	.000	-157.428	-77.107
	3	-193.574 [*]	17.186	.000	-228.110	-159.037
	4	-311.095 [*]	10.081	.000	-331.354	-290.836
2	1	117.267 [*]	19.985	.000	77.107	157.428
	3	-76.307 [*]	19.162	.000	-114.814	-37.799
	4	-193.827 [*]	18.304	.000	-230.610	-157.044
3	1	193.574 [*]	17.186	.000	159.037	228.110
	2	76.307 [*]	19.162	.000	37.799	114.814
	4	-117.521 [*]	14.721	.000	-147.104	-87.937
4	1	311.095 [*]	10.081	.000	290.836	331.354
	2	193.827 [*]	18.304	.000	157.044	230.610
	3	117.521 [*]	14.721	.000	87.937	147.104
Based on estimated marginal means						
*. The mean difference is significant at the .05 level.						
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).						

c. VR Deaths

The number of participant VR deaths is of interest in trust as more VR deaths may be indicative of the level of teamwork used to meet the needs of a task (e.g., synchronizing the pressing of platform activation buttons as the partner jumps across the green goo), as well as the ongoing development of trust. A 4 (participant deaths per round) x 2 (consistency condition) repeated measures ANOVA was run to determine whether there was a difference between the consistency conditions and participant deaths. There was a main effect for participant deaths per round of $F(3, 47) = 16.3$, $p < .001$, $\eta_p^2 = .51$. Furthermore, there was no interaction between rounds and subject deaths, $F(3, 47) = .60$, $p = .62$, $\eta_p^2 = .04$. As shown in Table 4, all rounds except Round 2 and Round 3 were significantly different from the others.

Table 4. Pairwise Comparison of VR Deaths by Round

Pairwise Comparisons						
Measure: MEASURE_1						
(I) Sub_Deaths	(J) Sub_Deaths	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	-3.322 ^a	.522	.000	-4.372	-2.273
	3	-2.436 ^a	.509	.000	-3.459	-1.413
	4	-.825 ^a	.301	.009	-1.431	-.220
2	1	3.322 ^a	.522	.000	2.273	4.372
	3	.886	.626	.163	-.371	2.143
	4	2.497 ^a	.544	.000	1.404	3.590
3	1	2.436 ^a	.509	.000	1.413	3.459
	2	-.886	.626	.163	-2.143	.371
	4	1.611 ^a	.489	.002	.628	2.593
4	1	.825 ^a	.301	.009	.220	1.431
	2	-2.497 ^a	.544	.000	-3.590	-1.404
	3	-1.611 ^a	.409	.002	-2.593	-.620

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

d. Pass/Fail

Finally, whether the team successfully passed a round or failed is of interest because repeated failures may impact trust between teammates. A 4 (pass/fail per round) x 2 (consistency condition) repeated measures ANOVA was run to determine whether there was a difference between the consistency conditions and round pass/fail. There was a main effect for subject pass/fail per round of $F(3, 47) = 98.78, p < .001, \eta_p^2 = .86$. However, there was no interaction found between rounds and pass/fail, $F(3, 47) = 1.19, p = .33, \eta_p^2 = .07$. Similar to the participant deaths per round analysis, Table 5 reveals the pass or fail score for all rounds except Round 2, and Round 3 was significantly different from the others.

Table 5. Pairwise Comparison of Pass/Fail by Round

Pairwise Comparisons						
Measure: MEASURE_1						
(I) PF	(J) PF	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	.195*	.056	.001	.083	.307
	3	.256*	.062	.000	.132	.380
	4	.863*	.049	.000	.764	.962
2	1	-.195*	.056	.001	-.307	-.083
	3	.062	.074	.412	-.088	.211
	4	.668*	.066	.000	.536	.801
3	1	-.256*	.062	.000	-.380	-.132
	2	-.062	.074	.412	-.211	.088
	4	.607*	.069	.000	.467	.746
4	1	-.863*	.049	.000	-.962	-.764
	2	-.668*	.066	.000	-.801	-.536
	3	-.607*	.069	.000	-.746	-.467
Based on estimated marginal means						
*. The mean difference is significant at the .05 level.						
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).						

2. Perception Conditions

The next three objective measurements were conducted using the same dependent variables, only this time with comparison between the participant's perception of their teammate being either a human or the bot. Perception of teammate data was collected in the end of Round 4 questionnaire where participants were asked to annotate who they thought they played with for each of the four rounds (i.e., human or the bot). The perception of teammate was coded so that 0 = human, and 1 = bot.

a. Round Time

Differences between round time and the perception of teammate as human or bot was examined by a one-way ANOVA testing of each round. Analysis revealed no difference in round time and perception of teammate: (a) Round 1, $F(1, 49) = .09$, $p = .77$, $d = .08$; (b) Round 2, $F(1, 49) = .03$, $p = .87$, $d = .05$; (c) Round 3, $F(1, 49) = .00$, $p = .99$, d

= -.004; and (d) Round 4, $F(1, 49) = .28$, $p = .60$, $d = -.15$. Table 6. displays the round time table of means for the perception condition.

Table 6. Round Time Table of Means (Perception Condition)

Perception of Teammate	RD 1	RD 2	RD 3	RD 4
(0) Human	97.41 (76.48)	214.97 (136.47)	287.70 (111.60)	402.40 (43.31)
(1) Bot	91.99 (53.37)	208.63 (130.26)	288.23 (105.62)	408.26 (35.70)
<i>M(SD)</i>				

b. VR Deaths

The differences between participants VR deaths per round and the perception of teammate as either human or bot was assessed with a one-way ANOVA. Analysis showed the following: (a) Round 1, $F(1, 49) = .02$, $p = .89$, $d = .04$; (b) Round 2, $F(1, 49) = .07$, $p = .80$, $d = .07$; (c) Round 3, $F(1, 49) = .05$, $p = .82$, $d = .06$; and (d) Round 4, $F(1, 49) = 5.9$, $p = .02$, $d = -.69$. Results from Round 4 ($p = .02$) was considered significant and will be explored in-depth in the next chapter. Table 7. shows the VR deaths table of means for the perception condition.

Table 7. VR Deaths Table of Means (Perception Condition)

Perception of Teammate	RD 1	RD 2	RD 3	RD 4
(0) Human	.50 (1.40)	3.93 (3.43)	3.00 (3.43)	.72 (.89)
(1) Bot	.44 (1.42)	3.67 (3.76)	2.80 (2.81)	1.85 (2.15)
<i>M(SD)</i>				

c. Pass/Fail

The differences between the participants pass or fail score and the perception of teammate as human or bot was assessed in a one-way ANOVA. Results revealed no differences in any of the rounds: (a) Round 1, all participants in all conditions passed; (b)

Round 2, $F(1, 49) = .24, p = .63, d = -.12$; (c) Round 3, $F(1, 49) = 2.34, p = .13, d = -.44$; and (d) Round 4, $F(1, 49) = .21, p = .65, d = .11$. Table 8. displays the Pass/Fail table of means for the perception condition.

Table 8. Pass/Fail Table of Means (Perception Condition)

Perception of Teammate	RD 1	RD 2	RD 3	RD 4
(0) Human	1.0 (0.00)	.78 (.42)	.65 (.49)	.16 (.37)
(1) Bot	1.0 (0.00)	.83 (.38)	.84 (.37)	.12 (.33)
<i>M(SD)</i>				

3. Actual Conditions

The last three objective measurements were conducted using the same dependent variables, this time for comparison between when the participant's teammate was an actual human or bot. Actual teammate data per round was annotated via the scenario synch matrix described in Chapter III.

a. Round Time

Differences between round time and the actual teammate being a human or bot was examined by a one-way ANOVA testing of each round. Analysis revealed the following: (a) Round 1, $F(1, 49) = .26, p = .61, d = -.14$; (b) Round 2, $F(1, 49) = 41.40, p < .001, d = -1.80$; (c) Round 3, $F(1, 49) = 1.65, p = .21, d = -.36$; and (d) Round 4, $F(1, 49) = 6.41, p = .02, d = -.75$. Table 9. displays the round time table of means for the actual condition.

Table 9. Round Time Table of Means (Actual Condition)

Actual Teammate	RD 1	RD 2	RD 3	RD 4
Human	89.57 (60.66)	128.42 (84.50)	266.73 (103.12)	393.39 (50.32)
Bot	98.95 (68.83)	306.01 (112.03)	305.41 (109.93)	420.00 (0.00)
<i>M(SD)</i>				

b. VR Deaths

The differences between participants deaths per round and the actual teammate being a human or bot was assessed with a one-way ANOVA. Analysis showed the following: (a) Round 1, $F(1, 49) = 1.70, p = .18, d = .38$; (b) Round 2, $F(1, 49) = 24.50, p = < .001, d = -1.37$; (c) Round 3, $F(1, 49) = 5.95, p = .02, d = -.70$; and (d) Round 4, $F(1, 49) = .04, p = .84, d = -.06$. Table 10. displays the VR deaths table of means for the actual condition.

Table 10. VR Deaths Table of Means (Actual Condition)

Actual Teammate	RD 1	RD 2	RD 3	RD 4
Human	.75 (1.87)	1.89 (2.10)	1.78 (2.10)	1.25 (1.40)
Bot	.22 (.70)	5.96 (3.65)	3.82 (3.53)	1.35 (2.15)
<i>M(SD)</i>				

c. Pass/Fail

Finally, the differences between the participants pass or fail score and the actual teammate being a human or bot was assessed in a one-way ANOVA. Results revealed the following: (a) Round 1, all participants in all conditions passed; (b) Round 2, $F(1, 49) = 10.79, p = .002, d = .87$; (c) Round 3, $F(1, 49) = .30, p = .59, d = .16$; and (d) Round 4, $F(1, 49) = 7.37, p = .01, d = .80$. Table 11. displays the Pass/Fail table of means for the actual condition.

Table 11. Pass/Fail Table of Means (Actual Condition)

Actual Teammate	RD 1	RD 2	RD 3	RD 4
Human	1.0 (0.00)	.96 (.19)	.78 (.42)	.25 (.44)
Bot	1.0 (0.00)	.63 (.50)	.71 (.46)	.00 (0.00)
<i>M(SD)</i>				

B. SUBJECTIVE MEASURES OF TRUST

Subjective metrics of trust were determined from the end of round survey questionnaires. Specifically, section two of the questionnaires related to the participant's perception of partner competency and section three of the questionnaires related to the participant's perception of partner trustworthiness. A factor analysis was conducted in order to determine which questions from sections two and three of the questionnaires most accurately described the competency and trustworthiness variables. The factor analysis revealed that questions 1–6, 9, 12–15, and 20 from section two of the questionnaires represented the underlying variable of partner competency. Questions 1, 3, 4–5, 7, 12, 14, 20, 23, and 26 from section three of the questionnaires represented the underlying variable of trustworthiness. Composite scores for teammate competency and trustworthiness were then produced. The following subjective measurements of trust used ANOVA tests to explore the main effect and interaction between the competency and trustworthiness variables by the consistent and inconsistent conditions, and the participant perception of teammate condition, and the actual condition by round.

1. Consistent and Inconsistent Conditions

a. Competency

The first subjective measurement was between the composite score for competency by round and the consistent and inconsistent conditions. Competency is of interest because a higher perception of competency in the teammate may be indicative of greater trust. A 4 (composite competency score per round) x 2 (consistency condition) repeated measures ANOVA was run to determine if there was a difference between the consistency conditions and competency scores. There was a main effect for competency per round of $F(3, 47) = 35.8, p < .001, \eta_p^2 = .7$. Also, there was an interaction between rounds and competency at $F(3, 47) = 3.60, p = .02, \eta_p^2 = .19$.

A pairwise comparison shown in Table 12 revealed that Round 1 was significant compared to the other rounds ($p < .001$). Round 2, compared to Round 4, had a significance of $p = .006$. Further, a follow-up independent samples t-test was run to find out the nature of interaction; in Round 3 there was significant difference in competency between

consistent ($M = 4.00$, $SD = 1.69$) and inconsistent ($M = 5.34$, $SD = 1.44$) conditions; $t(49) = -3.05$, $p = .004$.

Table 12. Pairwise Comparison of Competency by Round

Pairwise Comparisons						
Measure: MEASURE_1						
(I) Competency	(J) Competency	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	.961 [*]	.215	.000	.530	1.392
	3	1.353 [*]	.200	.000	.935	1.771
	4	1.736 [*]	.209	.000	1.316	2.155
2	1	-.961 [*]	.215	.000	-1.392	-.530
	3	.392	.260	.138	-.131	.915
	4	.775 [*]	.268	.006	.237	1.313
3	1	-1.353 [*]	.200	.000	-1.771	-.935
	2	-.392	.260	.138	-.915	.131
	4	.383	.293	.197	-.205	.971
4	1	-1.736 [*]	.209	.000	-2.155	-1.316
	2	-.775 [*]	.268	.006	-1.313	-.237
	3	.383	.293	.197	-.971	.205
Based on estimated marginal means						
*. The mean difference is significant at the .05 level.						
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).						

b. Trust

The second subjective measurement was done to determine if there was a difference between the composite score for trust by round and the consistent and inconsistent conditions. A 4 (composite trust score per round) x 2 (consistency condition) repeated measures ANOVA was run, and there was a main effect for trust per round of $F(3, 47) = 8.74$, $p < .001$, $\eta_p^2 = .36$. There was no interaction found between rounds and composite trust, $F(3, 47) = 1.45$, $p = .24$, $\eta_p^2 = .09$.

As depicted in Table 13, a Pairwise comparison found significance between Round 1 and Round 3 at $p = .002$, Round 1 and Round 4 at $p < .001$, Round 2, and Round 3 at $p = .049$, and between Round 2 and Round 4 at $p = .002$.

Table 13. Pairwise Comparison of Composite Trust by Round

Pairwise Comparisons						
Measure: MEASURE_1						
(I) Trust	(J) Trust	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	3.885	2.433	.117	-1.006	8.775
	3	8.705*	2.655	.002	3.370	14.039
	4	12.437*	2.675	.000	7.061	17.812
2	1	-3.885	2.433	.117	-8.775	1.006
	3	4.820*	2.387	.049	.023	9.617
	4	8.552*	2.547	.002	3.434	13.670
3	1	-8.705*	2.655	.002	-14.039	-3.370
	2	-4.820*	2.387	.049	-9.617	-.023
	4	3.732	2.935	.209	-2.165	9.630
4	1	-12.437*	2.675	.000	-17.812	-7.061
	2	-8.552*	2.547	.002	-13.670	-3.434
	3	-3.732	2.935	.209	-9.630	2.165
Based on estimated marginal means						
*. The mean difference is significant at the .05 level.						
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).						

2. Perception Conditions

a. Competency

The third subjective measurement was done to determine if there was a difference on the teammate competency score among the consistency and perception conditions. A 2 (perceived teammate human or bot) x 2 (consistency condition) repeated measures ANOVA was run, and there was no main effect for overall perception of teammate as human or bot, $F(1, 49) = .8, p = .38, \eta_p^2 = .02$. However, there was an interaction between perception and consistency of $F(1, 49) = 6.58, p = .013, \eta_p^2 = .12$. Follow-up t -tests showed

there was a difference between the consistent ($M = 4.64$, $SD = 1.50$) and inconsistent ($M = 5.66$, $SD = 1.04$) conditions on teammate competency when the teammate was perceived to be a bot; $t(49) = -2.83$, $p = .007$.

b. Trust

The fourth subjective measurement was done to determine if there was a difference on the teammate trust score among the consistency and perception conditions. A 2 (perceived teammate human or bot) x 2 (consistency condition) repeated measures ANOVA was run, and there was no main effect for overall perception of teammate as human or bot, $F(1, 49) = 2.76$, $p = .10$, $\eta_p^2 = .05$ and there was non-significant interaction between perception and consistency $F(1, 49) = 2.03$, $p = .16$, $\eta_p^2 = .04$.

3. Actual Conditions

a. Competency

The next subjective measurement was done to determine if there was a difference on the teammate competency score among the consistency and actual conditions. A 2 (actual teammate human or bot) x 2 (consistency condition) repeated measures ANOVA was run, and there was a main effect for actual teammate at $F(1, 49) = 17.21$, $p < .001$, $\eta_p^2 = .26$. However, there was no interaction between actual teammate and consistency on competency, $F(1, 49) = 2.78$, $p = .10$, $\eta_p^2 = .05$.

b. Trust

The final subjective measurement was done to determine if there was a difference on the teammate trust score among the consistency and actual conditions. A 2 (actual teammate) x 2 (consistency condition) repeated measures ANOVA was run, and there was a main effect for actual teammate at $F(1, 49) = 11.63$, $p = .001$, $\eta_p^2 = .19$ and there was non-significant interaction between actual teammate and consistency on trust $F(1, 49) = 1.84$, $p = .18$, $\eta_p^2 = .07$.

C. PAIRED SAMPLES TESTING

Two paired samples t-tests were conducted to determine if there were differences between actual and perceived teammates and objective and subjective performance metrics. Table 14 depicts the differences in objective performance metrics (i.e., round time, participant VR deaths, and pass/fail scores) and subjective performance metrics (i.e., composite competency and trust scores) between the actual human and bot teammate. The results of the paired samples t-tests revealed that all pairings were considered significant.

Table 14. Paired Samples T-Tests Between Actual Human and Bot Teammate Objective and Subjective Performance Metrics

Paired Samples Test								
		Paired Differences						
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df
					Lower	Upper		Sig. (2-tailed)
Pair 1	Actual Human Partner Time - Actual Bot Partner Time	-53.53863	109.61623	15.34934	-84.36869	-22.70857	-3.488	50
Pair 2	Actual Human Partner Sub_Deaths - Actual Bot Partner Sub_Deaths	-1.39216	3.16751	.44354	-2.28303	-.50128	-3.139	50
Pair 3	Actual Human Partner P/F - Actual Bot Partner P/F	.12745	.41040	.05747	-.01202	.24288	2.218	50
Pair 4	Actual Human Partner_Competency Score - Actual Bot Partner_Competency Score	.67496	1.19160	.16686	-.33982	1.01010	4.045	50
Pair 5	Actual Human Partner_Trust Score - Actual Bot Partner_Trust Score	6.39706	13.80994	1.90577	2.56920	10.22492	3.357	50

Table 15 depicts the differences in subjective performance metrics (i.e., composite competency and trust scores) between the actual human or bot teammate and the perceived human or bot teammate. The results of the paired samples t-tests revealed that all pairings were highly significant.

Table 15. Paired Samples T-Tests Between Actual and Perceived Human and Bot Teammate Subjective Performance Metrics

Paired Samples Test								
		Paired Differences				t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower Upper			
Pair 1	Actual Human Partner_Competency Score - Perceived human component score competency questionnaire	.41478	.96066	.13452	.14459 .68497	3.083	50	.003
Pair 2	Actual Bot Partner_Competency Score - Perceived bot component score competency questionnaire	-.41478	.96066	.13452	-.68497 -.14459	-3.083	50	.003
Pair 3	Actual Human Partner_Trust Score - Perceived human component score trust questionnaire	4.44118	10.12060	1.41717	1.59471 7.28764	3.134	50	.003
Pair 4	Actual Bot Partner_Trust Score - Perceived bot component score trust questionnaire	-4.44118	10.12060	1.41717	-7.28764 -1.59471	-3.134	50	.003

D. EXPLORATORY FINDINGS

Additional analysis was done to explore how certain participant demographic data and select objective measures related to teammate competency and trust. Participant age, gaming experience, VR deaths, and overall pass/fail data were used in eight standard regression analyses to predict perceived human and bot teammate competency and trust, in addition to predicting actual human and bot teammate competency and trust. The model summaries of each exploratory analysis are shown in Tables 16–23.

Table 16. Perceived Human Teammate: Competency

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.483a	.233	.217	1.06229	.233	14.876	1	49	.000
2	.488b	.238	.207	1.06944	.006	.347	1	48	.558
3	.492c	.242	.194	1.07790	.004	.250	1	47	.620
4	.506d	.256	.192	1.07952	.014	.859	1	46	.359
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths									
c. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age									
d. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age, Gaming Experience									

As shown in Table 16, Model 1, whether or not the participant passed or failed the round, was the best predictor of perceived human teammate competency score, $F(1, 49) = 14.88$, $p < .001$, $R^2 = .233$ and accounted for 23.3% of the variance of perceived human teammate competency.

Table 17. Perceived Bot Teammate: Competency

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.542a	.294	.280	1.16673	.294	20.412	1	49	.000
2	.596b	.355	.328	1.12662	.061	4.551	1	48	.038
3	.598c	.357	.316	1.13673	.002	.150	1	47	.700
4	.612d	.374	.320	1.13396	.017	1.230	1	46	.273
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Age									
c. Predictors: (Constant), Overall_PF, Age, Gaming Experience									
d. Predictors: (Constant), Overall_PF, Age, Gaming Experience, Overall_Sub_Deaths									

As depicted in Table 17, Model 2 was the best predictor for perceived bot teammate competency score, $F(1, 48) = 4.55$, $p = .038$, $R^2 = .355$ and accounted for 35.5% of the

variance of perceived bot teammate competency. The largest predictor in the model was overall pass/fail scores, accounting for 29.4% of the variance, with participant age adding 6% more variance to the model.

Table 18. Perceived Human Teammate: Trust

Model	Model Summary								
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.386a	.149	.131	13.38298	.149	8.558	1	49	.005
2	.388b	.151	.115	13.50464	.002	.121	1	48	.729
3	.411c	.169	.116	13.49984	.018	1.034	1	47	.314
4	.444d	.197	.127	13.41387	.028	1.604	1	46	.212
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Age									
c. Predictors: (Constant), Overall_PF, Age, Gaming Experience									
d. Predictors: (Constant), Overall_PF, Age, Gaming Experience, Overall_Sub_Deaths									

As noted in Table 18, Model 1, whether or not the participant passed or failed the round, was the best predictor of perceived human teammate trust score, $F(1, 49) = 8.56$, $p = .005$, $R^2 = .149$ and accounted for 14.9% of the variance of perceived human teammate trust.

Table 19. Perceived Bot Teammate: Trust

Model	Model Summary								
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.413a	.171	.154	11.48832	.171	10.093	1	49	.003
2	.537b	.288	.259	10.75349	.118	7.926	1	48	.007
3	.569c	.324	.281	10.59137	.036	2.481	1	47	.122
4	.595d	.354	.298	10.46453	.030	2.146	1	46	.150
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths									
c. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age									
d. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age, Gaming Experience									

As shown in Table 19, Model 2 was the best predictor for perceived bot teammate trust score, $F(1, 48) = 7.93$, $p = < .001$, $R^2 = .288$ and accounted for 28.8% of the variance of perceived bot teammate trust. The largest predictor in the model was overall pass/fail scores, accounting for 17.1% of the variance, with overall participant VR deaths adding 12% more variance to the model.

Table 20. Actual Human Teammate: Competency

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.597a	.356	.343	.90333	.356	27.112	1	49	.000
2	.598b	.357	.330	.91195	.001	.078	1	48	.782
3	.612c	.375	.335	.90875	.018	1.338	1	47	.253
4	.625d	.391	.338	.90668	.016	1.215	1	46	.276
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths									
c. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age									
d. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age, Gaming Experience									

As depicted in Table 20, Model 1, whether the participant passed or failed the round, was the best predictor of actual human teammate competency score, $F(1, 49) = 27.11$, $p = < .001$, $R^2 = .356$ and accounted for 35.6% of the variance of actual human teammate competency.

Table 21. Actual Bot Teammate: Competency

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.462a	.213	.197	1.27920	.213	13.298	1	49	.001
2	.513b	.263	.232	1.25101	.050	3.233	1	48	.078
3	.518c	.268	.222	1.25974	.005	.337	1	47	.564
4	.521d	.271	.208	1.27080	.003	.186	1	46	.669
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths									
c. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age									
d. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age, Gaming Experience									

As shown in Table 21, Model 1, whether the participant passed or failed the round, was the best predictor of actual bot teammate competency score, $F(1, 49) = 13.30$, $p = .001$, $R^2 = .213$ and accounted for 21.3% of the variance of actual bot teammate competency.

Table 22. Actual Human Teammate: Trust

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.380a	.144	.127	10.05720	.144	8.250	1	49	.006
2	.418b	.175	.140	9.97848	.031	1.776	1	48	.189
3	.423c	.179	.127	10.05661	.004	.257	1	47	.615
4	.508d	.258	.193	9.66629	.079	4.872	1	46	.032
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths									
c. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age									
d. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age, Gaming Experience									

As depicted in Table 22, Model 1, whether the participant passed or failed the round, was the best predictor of actual human teammate trust, $F(1, 49) = 8.25$, $p = .006$, $R^2 = .144$ and accounted for 14.4% of the variance of actual human teammate trust.

Finally, Table 23 shows Model 2 was the best predictor for actual bot teammate trust score, $F(1, 48) = 4.81$, $p = .002$, $R^2 = .232$ and accounted for 23.2% of the variance of actual bot teammate trust. The largest predictor in the model was overall pass/fail scores, accounting for 15.5% of the variance, with overall participant VR deaths adding 8% more variance to the model.

Table 23. Actual Bot Teammate: Trust

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.393a	.155	.137	15.61609	.155	8.966	1	49	.004
2	.481b	.232	.200	15.04207	.077	4.811	1	48	.033
3	.498c	.248	.200	15.03705	.017	1.032	1	47	.315
4	.504d	.254	.189	15.13993	.006	.363	1	46	.550
a. Predictors: (Constant), Overall_PF									
b. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths									
c. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age									
d. Predictors: (Constant), Overall_PF, Overall_Sub_Deaths, Age, Gaming Experience									

Chapter V will cover the findings from this chapter which were found to be significant and will discuss how they address the stated hypotheses and research questions annotated in Chapter I. Additionally, the following chapter will discuss the conclusions drawn from this research and provide recommendations for future work in the area of trust and human-autonomous teaming.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION AND THOUGHTS FOR FUTURE RESEARCH

This experiment was designed to measure the development of trust between human and autonomous teammates in a virtual environment that presents an aspect of risk. The virtual environment was a four-level escape room type game developed by the JHU-APL. Participants played two rounds with a human teammate and two rounds with an autonomous teammate. Additionally, half of the participants started Round 1 with the teammate they chose, and half started play with the teammate opposite their choice, however all participants thought they started playing Round 1 with the partner of their choice. Various aspects of teamwork were then analyzed between the human-human and human-autonomous teams. The results are discussed below in relation to the study's hypotheses and research questions.

A. DISCUSSION

1. Hypotheses

Of primary interest was to measure the development of trust between human and autonomous teammates. The two hypotheses concern trust development when the actual teammate was a human or bot (H1) and whether the teammate was believed to be a human or bot (H2).

Testing of the first hypothesis: *Trust scores are different depending on whether the teammate is a human or a bot*, was done to examine if there were differences in trust when the participant was playing the game with either the actual human teammate or the bot (i.e., actual conditions). In the actual conditions, the findings show that the objective performance metrics that were of interest to trust development (i.e., round time, participant VR deaths, and pass/fail scores) were better overall when participants were teamed with the human compared to the bot. One-way ANOVA testing of the differences between the actual conditions and objective performance metrics revealed significance for Rounds 2 and 4, respectively. Round 2 showed the most effect for round time and examination of the mean round times revealed that it took participants an average of 177.59 seconds longer to complete Round 2 when playing with the bot compared to a human. However, this result

may have been influenced by the researcher's proficiency at playing the round as the human teammate, enabling completion of the round in less time. In Round 4, when participants were playing with the bot, no participants passed the round, resulting in a round timeout at 420 seconds.

For VR deaths, Round 2 returned the most significance in the difference between participants VR deaths per round when the teammate was an actual human or bot. Participants died (virtually) approximately 3.5 times more often while playing with the bot teammate in Round 2. Round 2 required the most amount of participant and teammate coordination for jumping between the set and retractable platforms to avoid the rising green goo, therefore the performance difference between the bot and human teammate was more evident.

Finally, for pass/fail scores, Round 1 was the easiest of the four rounds, therefore all teams passed. However, Round 2 revealed the most significance in the difference between participant pass/fail scores and the actual condition. Table of means data shows that 96% of participants passed Round 2 when teamed with a human, while only 63% of participants passed Round 2 when teamed with a bot. As noted previously, no participants passed Round 4 when playing with a bot teammate and only 25% of the participants passed the round while playing with a human. This result may speak to the overall difficulty of the final round, but it may also be indicative of a breakdown of trust and teamwork. For example, if participants gain greater proficiency over time and their confidence in the reliability of the autonomy was shaken based on observed performance, they may begin to distrust the teammate and opt for self-reliance over continued team coordination, when in fact, trust and teamwork was needed to be successful (Dzindolet et al., 2003).

Subjective measurements were analyzed through 2 x 2 repeated measures ANOVA tests to determine if there was a difference on 1) composite competency score, and 2) composite trust score among the consistency (i.e., participant played either Round 1 or 3 with the actual teammate of their choice) and the actual conditions. For the first test, it was revealed that there was a main effect for the actual teammate on competency, however, there was no interaction between the actual teammate and consistency on competency. The

second test showed that there was a main effect for actual teammate on trust, but there was no interaction between actual teammate and consistency on trust.

Furthermore, paired samples t-tests were done with the objective and subjective measures to examine the differences between: 1) actual human teammate round time and actual bot teammate round time, 2) actual human teammate participant VR deaths and actual bot teammate participant VR deaths, 3) actual human teammate pass/fail scores and actual bot teammate pass/fail score, 4) actual human teammate composite competency score and actual bot teammate composite competency score, and 5) actual human teammate composite trust score and actual bot teammate composite trust score. The results of these pairing showed that there were significant differences on all comparisons between the teammate being a human or a bot. Therefore, for this study, round time, VR deaths, pass/fail score, and subjective competency score were different between the actual teammate being a human or a bot, with scores favoring the human teammate.

Testing of the second hypothesis: *Trust scores are different depending on whether it is believed the teammate is a human or a bot*, was conducted to determine if there were differences in trust when the participant was playing the game with a teammate they perceived to be a human or bot (i.e., perception conditions), regardless of whether their teammate was actually a human or bot. Examination of the objective performance metrics in the perception conditions highlighted negligible differences when the participant correctly or incorrectly perceived who their actual teammate was during gameplay. Analysis of the findings showed that there were no differences between the perception conditions and round time or pass/fail scores. However, one-way ANOVA testing of the differences between participant VR deaths and the perception conditions was shown to be significant for Round 4. Participant VR deaths table of means for the perception conditions showed that when participants who were led to believe their teammate was a bot in Round 4, they died (virtually) 2.57 times more than when playing with the bot teammate compared to the human. This result may be explainable by the participant losing trust in the bot by the fourth (most difficult) round; possibly choosing not to work with it based on prior performance, leading to decreases in their own performance and more deaths.

Subjective measurements were analyzed through 2 x 2 repeated measures ANOVA tests to determine if there was a difference on 1) composite competency score, and 2) composite trust score among the consistency and the perception conditions. The first test showed that there was no main effect for overall perception of teammate as a human or bot on competency. However, there was interaction between the perception and consistency conditions on competency. Further t-tests revealed there was a significant difference between the consistent and inconsistent conditions on teammate competency when the teammate was perceived to be a bot. For the second test, the results revealed that there was no main effect for overall perception of teammate as human or bot on trust and there was also non-significant interaction between perception and consistency on trust.

Moreover, a second paired samples t-tests was done to examine the differences between: 1) actual human teammate composite competency score and perceived human teammate composite competency score, 2) actual bot teammate composite competency score and perceived bot teammate composite competency score, 3) actual human teammate composite trust score and perceived human teammate composite trust score, and 4) actual bot teammate composite trust score and perceived bot teammate composite trust score. Each pairing in this test revealed significant differences between the actual and perceived human and bot teammate. Regarding the human teammate, both the competency and trust scores were higher in the actual than perceived human condition. The opposite pattern was found with the bot, such that both the competency and trust scores were higher in the perceived than actual bot condition.

The discussions on H1 and H2 point to trust scores being different depending on both the actual and perceived conditions. What this could mean is that the level of trust attributed to the teammate, whether it was actually or perceived to be a human or a bot, may not have been completely based on performance. Further analysis of the actual and perception conditions by round suggests that objective performance measurements and subjective trust scores are a function of round rather than teammate ability. Put another way, as the rounds progressed in difficulty, overall objective and subjective trust scores for the teammate decreased linearly.

Decreases in objective performance metrics and subjective ratings of trust as the rounds increased in difficulty may be attributed to a lack of bot design features, such as ongoing feedback (Hoff & Bashir, 2015) and communication abilities (other than subtle cues) (Joe et al. 2014; Laird et al., 2019; Schaefer et al., 2016). Also, a lack of anthropomorphism which, de Visser et al. (2012) and de Visser et al. (2018) asserted is needed to increase trust resiliency in environments of uncertainty, may have contributed to the lower trust levels as well.

2. Research Questions

The research questions of this study are concerned with how objective performance metrics compare to subjective trust scores when the teammate is: a) a human or a bot, and b) believed to be a human or a bot. Regressions with objective measures were used as predictors of subjective measures to see how well performance predicted subjective scores. Additionally, participant age and gaming experience were of interest and included in each regression, along with participant VR deaths, and overall pass/fail scores.

a. Teammate is a Human or a Bot

The first four regressions addressed competency and trust when the teammate was an actual human or an actual bot. To begin, the first model was run to predict competency when the actual teammate was a human. The results showed that whether the participant passed or failed the round was the best predictor for actual human teammate competency, accounting for over one third of the variance in the model. The second model was also run to predict competency, only this time when the actual teammate was a bot. Results revealed that the pass/fail scores variable was again the best predictor for actual bot teammate competency, accounting for over one fifth of the variance in the model. The third model was then done to predict trust when the actual teammate was a human. The results of this test showed that the pass/fail scores were the best predictor for actual human teammate trust, accounting for one seventh of the variance. Finally, the fourth model was run to predict trust when the actual teammate was a bot, and the results revealed that the model containing both overall pass/fail scores and overall participant VR deaths was the best predictor for actual bot teammate trust. The largest predictor for actual bot teammate trust

in this case was overall pass/fail scores, with overall participant VR deaths marginally contributing to the model's overall variance.

b. Teammate is Believed to be a Human or a Bot

The next four regressions addressed competency and trust when the teammate was believed to be a human or a bot. The first model was run to predict competency when the perceived teammate was a human. In this model, the results showed that whether the participant passed or failed the round was the best predictor for perceived human teammate competency, accounting for a little less than one quarter of the variance in the model. The second model was run to predict competency when the perceived teammate was a bot. Results highlighted that the model containing both the overall pass/fail scores and participant age was the better predictor for perceived bot teammate competency, with overall pass/fail scores being the largest predictor and age contributing only slightly to the overall variance of the model. The third model was then run to predict trust when the perceived teammate was a human. These results revealed that whether or not the participant passed or failed the round was the greatest predictor for perceived human teammate trust and accounted for a little more than one seventh of the model's variance. Finally, the fourth model was run to predict trust when the perceived teammate was a bot. Results showed that the model containing overall pass/fail scores and overall participant VR deaths was the best predictor for perceived bot teammate trust, with overall pass/fail scores being the greatest predictor and participant VR deaths contributing almost equally to the overall variance of the model.

From these data points, it is clear to see that whether the participant passed or failed the round was a major contributor for predicting trust in both the actual and perceived conditions. As such, these results could suggest that task accomplishment (e.g., passing the round) or self-error rate (i.e., VR deaths) may contribute to increases or decreases in overall trust and be a consideration for trust development between human and autonomous teammates.

3. Limitations

This study had a several shortcomings that were a result of the ESCAPE game design in addition to an unforeseeable worldwide pandemic. A considerable limitation with the ESCAPE game design that may have affected the results of the experiment was that the bot teammate was unable to adequately communicate. The game's design did not provide for the bot teammate to communicate via voice, hand or arm signals, or using the laser pointer. The researcher, when acting as the human teammate, and in some cases simulating the bot when required, had the ability to use the laser pointer for communication (like the participant), but in order to remain consistent between the human teammate and the autonomy, only the participant used the laser pointer to communicate. Had the game's design facilitated the ability for the bot to communicate with the laser pointer, and subsequently the human teammate, objective performance metrics and subjective measures of trust may have been different. Another limitation of the game's design was the increasingly difficult rounds. The lack of randomness in round difficulty made the game predictable and most likely led the objective performance measurements and subjective trust scores to be function of round rather than teammate ability. Furthermore, unpredictable software glitches that caused the bot to get out of sequence with the participants commands was also a shortcoming of the game's design. This situation required the researcher to toggle control from autonomous bot mode to human mode and position the bot in the appropriate position in order to re-sync the bot with the current scenario. While this transition between modes was well rehearsed, there was a possibility that the participant was aware of the momentary change and it may have influenced their perception of the teammate's true identity or its performance. Another shortcoming of the experiment that was not related to the game's design was an inadequate number of participants representing all age groups, of particular interest would be more participants older than 50. Having a larger sample size of older participants may have provided a better representation of trust development. Lastly, the target sample size of 80 participants need for the experiment was not achieved. The experiment phase of this research was cut short due to the novel coronavirus disease pandemic. Consequently, only 51 participants were able to complete the experiment and the reduced sample size may have limited the results.

B. CONCLUSIONS AND THOUGHTS FOR FUTURE RESEARCH

1. Conclusions

As autonomous systems become more prevalent within the DOD and their roles begin to shift from being a mere tool to an actual teammate, the human operator's trust in such systems plays a crucial role in successful interactions and further use (Lee & See, 2004). This research aimed to identify and combine objective performance-based metrics of trust with subjective measures in order to better understand the development of trust between humans and autonomous systems in a teaming dynamic. Three objective performance metrics (i.e., round time, participant VR deaths, and pass/fail scores) were identified that may indicate trust or are important to the development of trust within the context of this experiment.

Specifically, the round time metric, especially in a time sensitive situation, could be indicative how much or how little a person relied on or worked with their teammate to solve a problem or complete a task. In the context of this study, longer round times may signal less trust in the teammate if more intra-team coordination is required. Conversely, shorter round times could imply enhanced intra-team coordination and greater trust in the teammate.

The number of participant VR deaths (i.e., performance degradation due to reliance on autonomy) could also be of interest to trust due the partner being reliant on the teammate to accomplish a task, that if not done properly, could lead to disastrous results (e.g., the participant relying on the teammate to stand on a pressure plate long enough for them to jump across the retractable platforms). If someone cannot rely on the autonomous teammate to accomplish a task that may result in peril, trust in the teammate may rapidly decrease. While a solid inference cannot be made without further investigation, the inverse of this condition may contribute to trust development if the automation consistently demonstrates high levels of reliability. However, when the autonomy's reliability comes into question, or if there is an egregious violation of trust, just how rapidly the trust decreases and whether that trust is repairable is a current topic of study (de Visser et al., 2018).

Finally, pass/fail scores (i.e., task completion) could also be indicative of trust development. In this experiment, as round difficulty increased, overall levels of reported trust decreased in proportion to the round failures in the later rounds. It can be inferred from the results that when human and autonomous teammates repeatedly fail a task, or in the case of this study, fail to pass a round, human trust in autonomy may begin to erode. Although a strong conclusion cannot be derived without further research, perhaps the opposite of this condition may be true where repeated human-autonomy team successes may enhance trust development and become more resilient to trust breakdowns when risk and adversity is introduced.

Overall, the objective performance metrics identified in this thesis were not indicative of overall trust per se but provided statistical significance in the various conditions by round and may be worthy of consideration in the study of trust development between man and machine. Moreover, the results show that objective measures of trust, such as the ones presented here may be used in conjunction with subjectively derived metrics of trust to provide a more complete set of measures that can contribute to a greater understanding of the development of trust between human and autonomous teammates.

2. Thoughts for Future Research

A secondary goal of this research was to explore the similarities and differences in the development of trust between human-human and human-autonomous teams. The literature on the subject suggests that humans tend to react socially to machines (Madhavan & Wiegmann, 2007). However, calibration of human expectations of system capabilities, similar to human judgements of other people, is necessary for appropriate trust and reliance on the automation (Joe et al., 2014; Lee & See, 2004).

Accordingly, the autonomous teammate in this experiment lacked certain design features that would enable basic social interactions, such as the ability to communicate directly and provide constant feedback, two capabilities that are important to successful teamwork and trust. This observation was supported by low scores returned for questions related to teammate communication abilities between both human and bot teammate within each end of round questionnaire. If researchers desire to perform an experiment similar to

this in the future, enhancing the autonomy's level of anthropomorphism and communicable abilities could better calibrate human expectations and may yield more realistic objective and subjective measures of trust between the human and autonomy.

Furthermore, future adaptations of this research should also consider randomizing the rounds and their associated level of difficulty in order to explore how trust evolves in a less predictable environment. The linear escalation of difficulty in this experiment may have induced unrealistic performance expectations of the autonomy early on in the game. Round randomization may preclude this situation and provide a better gauge of trust development between the teammates.

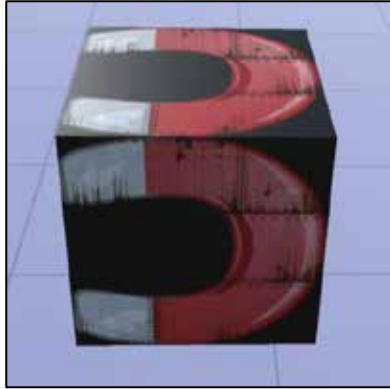
The exploratory tests revealed that pass/fail scores were a consistent and major predictor of trust. Future research should consider extending the total number of rounds played to enable more teaming with the autonomous teammate in order to investigate the impact of team success rate on trust the longer the participant plays with the autonomy. Additionally, the exploratory tests provided ideas for future testing of the effect of age and gender on trust development. In this study, age was shown to contribute to the variance in predicting competency in the perceived bot teammate condition. In follow-on studies, it would be interesting to observe how age factors into trust development in the man-machine teaming dynamic. In the present study, the participant age range was between 23 to 63 years old ($M_{age} = 32$, $SD = 8.10$) with only three participants being 50 years old or older. It is recommended that later studies recruit a sufficient sample of older participants (e.g., over 50 years old) in order to explore how they interact with and rate trust between the human and autonomous teammates. Similarly, this study only had four females volunteer for participation. With a greater sample size of female participants, it would also be interesting to investigate how trust development between humans and autonomy is impacted by gender differences.

In closing, while this research was able to identify and combine objective performance metrics with subjectively derived metrics of trust, it has also produced some questions on the subject worthy of consideration. This study also offered a few areas to explore during future investigations into trust between man and machine. It is hoped that the results of this humble experiment provokes thought and generates ideas to further

explore the development of trust between human-autonomous teammates. The knowledge gained through continued research in this area can then be applied towards the design and acquisition of future DOD autonomous systems that maximize efficiency and lethality in the human-autonomous team construct.

THIS PAGE INTENTIONALLY LEFT BLANK

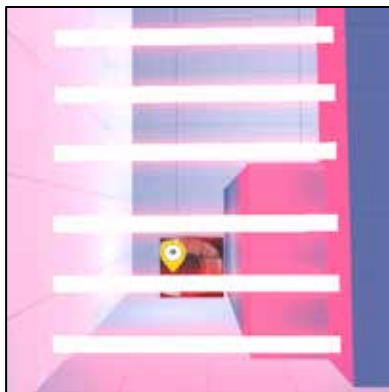
APPENDIX A. ESCAPE OBJECTS AND OBSTACLES



Moveable Box



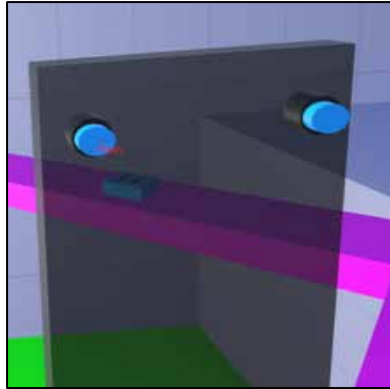
Pressure Plate



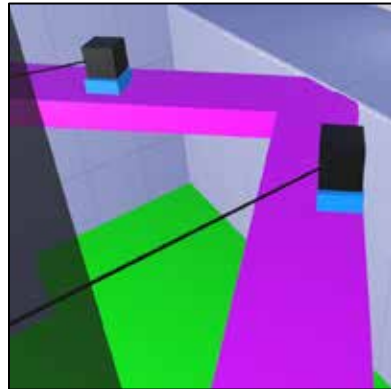
Laser Wall



Tractor-beam Gun



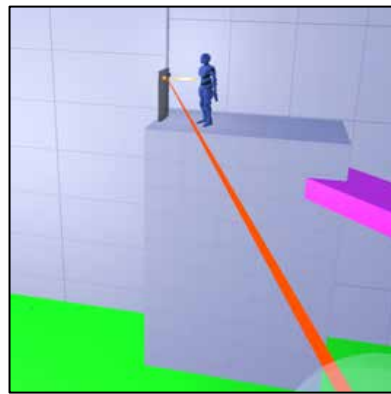
Activation Panel



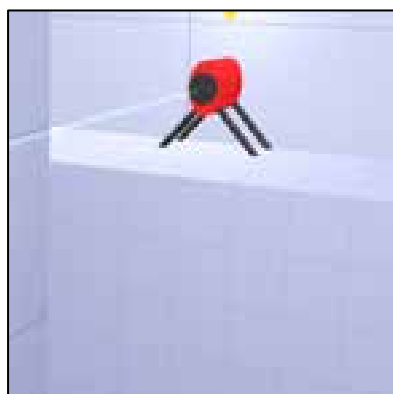
Cable/Connectors



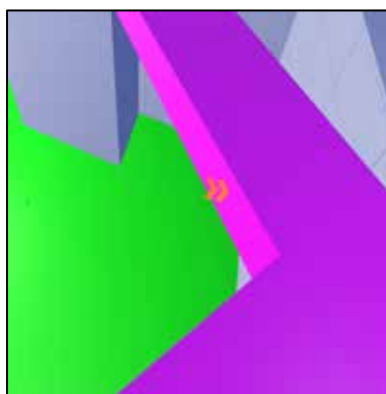
Movement Command /Timer



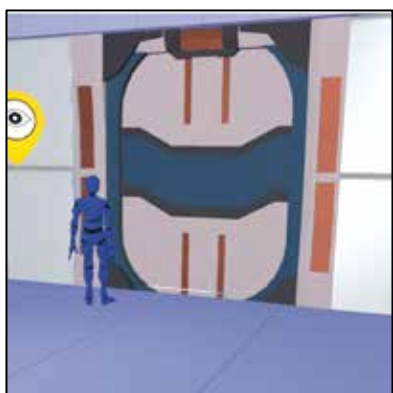
Action Command



Tractor-beam



Platform



Round Exit

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. QUESTIONNAIRES

A. PRE-ROUND 1 QUESTIONNAIRE

Waiting room prior to Round 1.

1. Have you ever played a video game that required collaboration with a teammate?
 - A. Yes
 - B. No
2. If yes, did you play with another person, a bot, or both?
 - A. Person
 - B. Bot
 - C. Both
 - D. N/A
3. ESCAPE is a collaborative, spatial puzzle game. You will need to collaborate with your teammate to find the solution to exit the level, and to overcome the obstacles in the room. Would you prefer to play this round with another person or a bot?
 - A. Person
 - B. Bot

B. END OF ROUND 1 QUESTIONNAIRE

On a scale of 1 to 7, where 1 is poor and 7 is awesome, rate the following:

1. My performance on this round:
1 2 3 4 5 6 7
2. My teammate's performance on this round:
1 2 3 4 5 6 7
3. Our team's performance on this round
1 2 3 4 5 6 7

On a scale of 1 to 7, where 1 is disagree and 7 is completely agree, please rate the following:

1. My teammate actively assisted in solving the puzzle
1 2 3 4 5 6 7
2. My teammate understood the goal at each step of the puzzle
1 2 3 4 5 6 7
3. My teammate performed actions that were difficult (because of timing requirements or complexity) very well
1 2 3 4 5 6 7
4. My teammate was important to successfully solving (figuring out what actions needed to be performed to escape) the puzzle
1 2 3 4 5 6 7

5. My teammate was important to successfully escaping the level (performing the required actions to escape)

1 2 3 4 5 6 7

6. I believe my teammate was performing at the top of their ability

1 2 3 4 5 6 7

7. I could have performed the actions better

1 2 3 4 5 6 7

8. I easily understood how to escape the level

1 2 3 4 5 6 7

9. My teammate is capable of performing the actions needed to escape

1 2 3 4 5 6 7

10. My teammate was good at communicating what actions needed to be performed

1 2 3 4 5 6 7

11. I was good at communicating the actions that needed to be performed

1 2 3 4 5 6 7

12. My teammate understood what I was trying to do

1 2 3 4 5 6 7

13. I understood what my teammate was trying to do

1 2 3 4 5 6 7

14. I would play another round with the teammate

1 2 3 4 5 6 7

15. If this teammate and I were to play another round together, we would perform much better

1 2 3 4 5 6 7

16. This puzzle was difficult to figure out what actions were needed to escape

1 2 3 4 5 6 7

17. The actions required to escape from this room were difficult to perform

1 2 3 4 5 6 7

18. When looking down from edges, I felt like I was on a cliff

1 2 3 4 5 6 7

19. It was important to me to succeed at the game

1 2 3 4 5 6 7

20. My teammate and I worked together well

1 2 3 4 5 6 7

What % of time was the other player

1. Considered part of the team

10 20 30 40 50 60 70 80 90 100

2. Incompetent

10 20 30 40 50 60 70 80 90 100

3. Dependable

10 20 30 40 50 60 70 80 90 100

4. Reliable

10 20 30 40 50 60 70 80 90 100

5. Unresponsive

10 20 30 40 50 60 70 80 90 100

6. Autonomous

10 20 30 40 50 60 70 80 90 100

7. Predictable

10 20 30 40 50 60 70 80 90 100

8. Conscious

10 20 30 40 50 60 70 80 90 100

9. Lifelike

10 20 30 40 50 60 70 80 90 100

10. A good teammate

10 20 30 40 50 60 70 80 90 100

11. Led astray by unexpected changes in the environment

10 20 30 40 50 60 70 80 90 100

12. Act consistently

10 20 30 40 50 60 70 80 90 100

13. Act as part of the team

10 20 30 40 50 60 70 80 90 100

14. Function successfully

	10	20	30	40	50	60	70	80	90	100
15. Malfunction										
	10	20	30	40	50	60	70	80	90	100
16. Have errors										
	10	20	30	40	50	60	70	80	90	100
17. Perform a task better than a novice										
	10	20	30	40	50	60	70	80	90	100
18. Provide feedback										
	10	20	30	40	50	60	70	80	90	100
19. Possess adequate decision-making capability										
	10	20	30	40	50	60	70	80	90	100
20. Meet the needs of the task										
	10	20	30	40	50	60	70	80	90	100
21. Provide appropriate information										
	10	20	30	40	50	60	70	80	90	100
22. Work best with a team										
	10	20	30	40	50	60	70	80	90	100
23. Perform exactly as instructed										
	10	20	30	40	50	60	70	80	90	100
24. Work in close proximity with people										
	10	20	30	40	50	60	70	80	90	100
25. Perform many functions at one time										

10 20 30 40 50 60 70 80 90 100

26. Follow directions

10 20 30 40 50 60 70 80 90 100

C. **END OF ROUND 2 QUESTIONNAIRE**

On a scale of 1 to 7, where 1 is poor and 7 is awesome, rate the following:

1. My performance on this round:

1 2 3 4 5 6 7

2. My teammate's performance on this round:

1 2 3 4 5 6 7

3. Our team's performance on this round

1 2 3 4 5 6 7

On a scale of 1 to 7, where 1 is disagree and 7 is completely agree, please rate the following:

1. My teammate actively assisted in solving the puzzle

1 2 3 4 5 6 7

2. My teammate understood the goal at each step of the puzzle

1 2 3 4 5 6 7

3. My teammate performed actions that were difficult (because of timing requirements or complexity) very well

1 2 3 4 5 6 7

4. My teammate was important to successfully solving (figuring out what actions needed to be performed to escape) the puzzle

1 2 3 4 5 6 7

5. My teammate was important to successfully escaping the level (performing the required actions to escape)

1 2 3 4 5 6 7

6. I believe my teammate was performing at the top of their ability

1 2 3 4 5 6 7

7. I could have performed the actions better

1 2 3 4 5 6 7

8. I easily understood how to escape the level

1 2 3 4 5 6 7

9. My teammate is capable of performing the actions needed to escape

1 2 3 4 5 6 7

10. My teammate was good at communicating what actions needed to be performed

1 2 3 4 5 6 7

11. I was good at communicating the actions that needed to be performed

1 2 3 4 5 6 7

12. My teammate understood what I was trying to do

- 1 2 3 4 5 6 7
13. I understood what my teammate was trying to do
- 1 2 3 4 5 6 7
14. I would play another round with the teammate
- 1 2 3 4 5 6 7
15. If this teammate and I were to play another round together, we would perform much better
- 1 2 3 4 5 6 7
16. This puzzle was difficult to figure out what actions were needed to escape
- 1 2 3 4 5 6 7
17. The actions required to escape from this room were difficult to perform
- 1 2 3 4 5 6 7
18. When looking down from edges, I felt like I was on a cliff
- 1 2 3 4 5 6 7
19. It was important to me to succeed at the game
- 1 2 3 4 5 6 7
20. My teammate and I worked together well

What % of time was the other player

1. Considered part of the team

10 20 30 40 50 60 70 80 90 100

2. Incompetent

10 20 30 40 50 60 70 80 90 100

3. Dependable

10 20 30 40 50 60 70 80 90 100

4. Reliable

10 20 30 40 50 60 70 80 90 100

5. Unresponsive

10 20 30 40 50 60 70 80 90 100

6. Autonomous

10 20 30 40 50 60 70 80 90 100

7. Predictable

10 20 30 40 50 60 70 80 90 100

8. Conscious

10 20 30 40 50 60 70 80 90 100

9. Lifelike

10 20 30 40 50 60 70 80 90 100

10. A good teammate

10 20 30 40 50 60 70 80 90 100

11. Led astray by unexpected changes in the environment

10 20 30 40 50 60 70 80 90 100

12. Act consistently

10 20 30 40 50 60 70 80 90 100

13. Act as part of the team

10 20 30 40 50 60 70 80 90 100

14. Function successfully

10 20 30 40 50 60 70 80 90 100

15. Malfunction

10 20 30 40 50 60 70 80 90 100

16. Have errors

10 20 30 40 50 60 70 80 90 100

17. Perform a task better than a novice

10 20 30 40 50 60 70 80 90 100

18. Provide feedback

	10	20	30	40	50	60	70	80	90	100
19. Possess adequate decision-making capability										
	10	20	30	40	50	60	70	80	90	100
20. Meet the needs of the task										
	10	20	30	40	50	60	70	80	90	100
21. Provide appropriate information										
	10	20	30	40	50	60	70	80	90	100
22. Work best with a team										
	10	20	30	40	50	60	70	80	90	100
23. Perform exactly as instructed										
	10	20	30	40	50	60	70	80	90	100
24. Make sensible decisions										
	10	20	30	40	50	60	70	80	90	100
25. Perform many functions at one time										
	10	20	30	40	50	60	70	80	90	100
26. Follow directions										
	10	20	30	40	50	60	70	80	90	100

You have played with both the person and the bot. In Round 3, would you like to play with a **person** or the **bot**?

D. END OF ROUND 3 QUESTIONNAIRE

On a scale of 1 to 7, where 1 is poor and 7 is awesome, rate the following:

- My performance on this round:

1	2	3	4	5	6	7
---	---	---	---	---	---	---
- My teammate's performance on this round:

1	2	3	4	5	6	7
---	---	---	---	---	---	---

3. Our team's performance on this round

1 2 3 4 5 6 7

On a scale of 1 to 7, where 1 is disagree and 7 is completely agree, please rate the following:

1. My teammate actively assisted in solving the puzzle
1 2 3 4 5 6 7
2. My teammate understood the goal at each step of the puzzle
1 2 3 4 5 6 7
3. My teammate performed actions that were difficult (because of timing requirements or complexity) very well
1 2 3 4 5 6 7
4. My teammate was important to successfully solving (figuring out what actions needed to be performed to escape) the puzzle
1 2 3 4 5 6 7
5. My teammate was important to successfully escaping the level (performing the required actions to escape)
1 2 3 4 5 6 7
6. I believe my teammate was performing at the top of their ability
1 2 3 4 5 6 7
7. I could have performed the actions better
1 2 3 4 5 6 7
8. I easily understood how to escape the level
1 2 3 4 5 6 7
9. My teammate is capable of performing the actions needed to escape
1 2 3 4 5 6 7
10. My teammate was good at communicating what actions needed to be performed
1 2 3 4 5 6 7
11. I was good at communicating the actions that needed to be performed
1 2 3 4 5 6 7
12. My teammate understood what I was trying to do
1 2 3 4 5 6 7
13. I understood what my teammate was trying to do
1 2 3 4 5 6 7
14. I would play another round with the teammate
1 2 3 4 5 6 7
15. If this teammate and I were to play another round together, we would perform much better
1 2 3 4 5 6 7
16. This puzzle was difficult to figure out what actions were needed to escape
1 2 3 4 5 6 7
17. The actions required to escape from this room were difficult to perform
1 2 3 4 5 6 7
18. When looking down from edges, I felt like I was on a cliff
1 2 3 4 5 6 7

19. It was important to me to succeed at the game

1 2 3 4 5 6 7

20. My teammate and I worked together well

1 2 3 4 5 6 7

What % of time was the other player

1. Considered part of the team

10 20 30 40 50 60 70 80 90 100

2. Incompetent

10 20 30 40 50 60 70 80 90 100

3. Dependable

10 20 30 40 50 60 70 80 90 100

4. Reliable

10 20 30 40 50 60 70 80 90 100

5. Unresponsive

10 20 30 40 50 60 70 80 90 100

6. Autonomous

10 20 30 40 50 60 70 80 90 100

7. Predictable

10 20 30 40 50 60 70 80 90 100

8. Conscious

10 20 30 40 50 60 70 80 90 100

9. Lifelike

10 20 30 40 50 60 70 80 90 100

10. A good teammate

	10	20	30	40	50	60	70	80	90	100
11. Led astray by unexpected changes in the environment										
	10	20	30	40	50	60	70	80	90	100
12. Act consistently										
	10	20	30	40	50	60	70	80	90	100
13. Act as part of the team										
	10	20	30	40	50	60	70	80	90	100
14. Function successfully										
	10	20	30	40	50	60	70	80	90	100
15. Malfunction										
	10	20	30	40	50	60	70	80	90	100
16. Have errors										
	10	20	30	40	50	60	70	80	90	100
17. Perform a task better than a novice										
	10	20	30	40	50	60	70	80	90	100
18. Provide feedback										
	10	20	30	40	50	60	70	80	90	100
19. Possess adequate decision-making capability										
	10	20	30	40	50	60	70	80	90	100
20. Meet the needs of the task										
	10	20	30	40	50	60	70	80	90	100
21. Provide appropriate information										
	10	20	30	40	50	60	70	80	90	100

22. Work best with a team

10 20 30 40 50 60 70 80 90 100

23. Perform exactly as instructed

10 20 30 40 50 60 70 80 90 100

24. Make sensible decisions

10 20 30 40 50 60 70 80 90 100

25. Perform many functions at one time

10 20 30 40 50 60 70 80 90 100

26. Follow directions

10 20 30 40 50 60 70 80 90 100

E. END OF ROUND 4 QUESTIONNAIRE

On a scale of 1 to 7, where 1 is poor and 7 is awesome, rate the following:

4. My performance on this round:

1 2 3 4 5 6 7

5. My teammate's performance on this round:

1 2 3 4 5 6 7

6. Our team's performance on this round

1 2 3 4 5 6 7

On a scale of 1 to 7, where 1 is disagree and 7 is completely agree, please rate the following:

1. My teammate actively assisted in solving the puzzle

1 2 3 4 5 6 7

2. My teammate understood the goal at each step of the puzzle

1 2 3 4 5 6 7

3. My teammate performed actions that were difficult (because of timing requirements or complexity) very well

1 2 3 4 5 6 7

4. My teammate was important to successfully solving (figuring out what actions needed to be performed to escape) the puzzle

1 2 3 4 5 6 7

5. My teammate was important to successfully escaping the level (performing the required actions to escape)

1 2 3 4 5 6 7

6. I believe my teammate was performing at the top of their ability

1 2 3 4 5 6 7

7. I could have performed the actions better

1 2 3 4 5 6 7

8. I easily understood how to escape the level

1 2 3 4 5 6 7

9. My teammate is capable of performing the actions needed to escape

1 2 3 4 5 6 7

10. My teammate was good at communicating what actions needed to be performed

1 2 3 4 5 6 7

11. I was good at communicating the actions that needed to be performed

1 2 3 4 5 6 7

12. My teammate understood what I was trying to do

1 2 3 4 5 6 7

13. I understood what my teammate was trying to do

1 2 3 4 5 6 7

14. I would play another round with the teammate

1 2 3 4 5 6 7

15. If this teammate and I were to play another round together, we would perform much better

1 2 3 4 5 6 7

16. This puzzle was difficult to figure out what actions were needed to escape

1 2 3 4 5 6 7

17. The actions required to escape from this room were difficult to perform

1 2 3 4 5 6 7

18. When looking down from edges, I felt like I was on a cliff

1 2 3 4 5 6 7

19. It was important to me to succeed at the game

1 2 3 4 5 6 7

20. My teammate and I worked together well

1 2 3 4 5 6 7

What % of time was the other player

1. Considered part of the team

10 20 30 40 50 60 70 80 90 100

2. Incompetent

10 20 30 40 50 60 70 80 90 100

3. Dependable

10 20 30 40 50 60 70 80 90 100

4. Reliable

10 20 30 40 50 60 70 80 90 100

5. Unresponsive

10 20 30 40 50 60 70 80 90 100

6. Autonomous

10 20 30 40 50 60 70 80 90 100

7. Predictable

10 20 30 40 50 60 70 80 90 100

8. Conscious

10 20 30 40 50 60 70 80 90 100

9. Lifelike

10 20 30 40 50 60 70 80 90 100

10. A good teammate

10 20 30 40 50 60 70 80 90 100

11. Led astray by unexpected changes in the environment

10 20 30 40 50 60 70 80 90 100

12. Act consistently

10 20 30 40 50 60 70 80 90 100

13. Act as part of the team

10 20 30 40 50 60 70 80 90 100

14. Function successfully

	10	20	30	40	50	60	70	80	90	100
15. Malfunction										
	10	20	30	40	50	60	70	80	90	100
16. Have errors										
	10	20	30	40	50	60	70	80	90	100
17. Perform a task better than a novice										
	10	20	30	40	50	60	70	80	90	100
18. Provide feedback										
	10	20	30	40	50	60	70	80	90	100
19. Possess adequate decision-making capability										
	10	20	30	40	50	60	70	80	90	100
20. Meet the needs of the task										
	10	20	30	40	50	60	70	80	90	100
21. Provide appropriate information										
	10	20	30	40	50	60	70	80	90	100
22. Work best with a team										
	10	20	30	40	50	60	70	80	90	100
23. Perform exactly as instructed										
	10	20	30	40	50	60	70	80	90	100
24. Make sensible decisions										
	10	20	30	40	50	60	70	80	90	100
25. Perform many functions at one time										

10 20 30 40 50 60 70 80 90 100

26. Follow directions

10 20 30 40 50 60 70 80 90 100

27. Did you play *more* than one round with the same player or bot?

- A) Yes
- B) No

28. If you thought you played rounds with the same teammates, which rounds were they?

- A) Each round was a different teammate
- B) 1 and 2
- C) 1 and 3
- D) 1 and 4
- E) 2 and 3
- F) 2 and 4
- G) 3 and 4
- H) 1, 2, and 3
- I) 1, 2, and 4
- J) 2, 3, and 4
- K) I played all 4 rounds with the same teammate
- L) 1, 3, and 4

29. Which round or rounds did you play with bots?

- A) 1
- B) 2
- C) 3
- D) 4
- E) 1 and 2
- F) 1 and 3
- G) 1 and 4
- H) 2 and 3
- I) 2 and 4
- J) 3 and 4
- K) 1, 2, and 3
- L) 1, 2, and 4
- M) 2, 3, and 4
- N) I did not play with a bot
- O) 1, 3, and 4
- P) 1, 2, 3, and 4

30. Which round or rounds did you play with a human?

- A) 1
- B) 2
- C) 3

- D) 4
- E) 1 and 2
- F) 1 and 3
- G) 1 and 4
- H) 2 and 3
- I) 2 and 4
- J) 3 and 4
- K) 1, 2, and 3
- L) 1, 2, and 4
- M) 2, 3, and 4
- N) I did not play with another person
- O) 1, 3, and 4
- P) 1, 2, 3, and 4

31. Overall, how good were the human players? (1 is terrible, 7 is great)

1 2 3 4 5 6 7

32. Overall, how good were the bots? (1 is terrible, 7 is great)

1 2 3 4 5 6 7

Demographic Information:

1. Age: _____

2. Gender: _____

3. Service: _____

4. What is your overall gaming experience? (1 is little, 7 is extensive)

1 2 3 4 5 6 7

5. What is your prior virtual reality gaming experience? (1 is little, 7 is extensive)

1 2 3 4 5 6 7

6. If you have or currently play games, which genre did you/do you typically play?

Circle all that apply.

- 1) Action 2) Action/Adventure 3) Adventure 4) Roleplaying 5) Simulation
- 6) Strategy 7) Sports 8) Puzzle 9) Idle

Thank you! Survey questions are complete

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. HUMAN-BOT PROTOCOLS

Round	Human	Bot
1	Walk toward laser wall, then move to the buttons and press. After, move to edge facing the platforms and wait for direction *Or go opposite of where they go (edge or buttons). (If player is at the console) to time the jump better, line up against the wall on the 4th block facing the platforms and press 'P'	Line up against the wall on the 4th block facing the platforms and press 'P'
2	Jump down to other platform, walk around and wait for direction. *When box is on plate, jump across.	Press 'P'
3	Walk to buttons and press until teammate figures out the gun. Once steps come out, go to bottom and wait for direction *If teammate shoots box over, put box on plate and jump down	Press 'P' *May need to intervene if bot gets out of sequence.
4	Put box facing platforms, jump up and off, then press buttons, then go to edge and look around, wait.	*Same, then follow directions.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. INFORMED CONSENT FORM

Naval Postgraduate School Consent to Participate in Research

Introduction. You are invited to participate in a research study entitled Exploring the Development of Trust Between Human and Autonomous Teammates. The purpose of the research is explore how people can effectively team with robots. This study is being done to explore how people can team with robots. We aim to define objective, performance bases metrics of trust for human-autonomous system teams, explore the similarities and differences in these performance metrics between teams made of humans and teams with mixes of humans and robots, and assess how problem complexity and risk affect team performance. Understanding these interactions will allow us to better define the information requirements for human-autonomous systems, and define how robot behaviors should vary based on the how the robot must interact with the human. Participations is completely voluntary and confidential.

Procedures. You will be asked to play a spatial puzzle video game using a virtual reality head mounted display. The display is an Oculus Rift. Some people can get a bit motion sick using virtual reality displays. You will be asked to wear the Oculus Rift and Touch hand controllers. You will have one training session to allow you to become familiar with the head mounted display, adapt to the virtual environment, and become familiar with the use of the handheld controls and the kinds of obstacles you will encounter. At the end of the training session, you will be asked if you would like to play 4 rounds of the game. Participation is expected to last no more than 1 hour. This experiment will require between 70–90 participants. Performance metrics will be collected, however, they will not be associated with an individual; all data is anonymized. Participation will last approximately an hour; no more than an hour and a half.

Exclusionary Criteria. Those with vestibular, balance, or motion sickness are excluded from participation. Participation is restricted to those above 18 years of age.

Location. The experiment will take place in Glasgow Hall, room 103.

Cost. There is no cost to participate in this research study.

Voluntary Nature of the Study. Your participation in this study is strictly voluntary. If you choose to participate you can change your mind at any time and withdraw from the study. You will not be penalized in any way or lose any benefits to which you would otherwise be entitled if you choose not to participate in this study or to withdraw. The alternative to participating in the research is to not participate in the research.

Potential Risks and Discomforts. The potential risks of participating in this study include slight discomfort from virtual reality. Some people experience virtual reality sickness.

Anticipated Benefits. There will be no direct benefit for individual participation. However, this research will contribute to the understanding of trust in automation.

Confidentiality & Privacy Act. Any information that is obtained during this study will be kept confidential to the full extent permitted by law. All efforts, within reason, will be made to keep your personal information in your research record confidential but total confidentiality cannot be guaranteed. All identifiable information will be locked in a file and kept in a private office.

Points of Contact. If you have any questions or comments about the research, or you experience an injury or have questions about any discomforts that you experience while taking part in this study please contact the Principal Investigator, Dr. Mollie McGuire, 831-656-2995, mrmcguir@nps.edu. Questions about your rights as a research subject or any other concerns may be addressed to the Navy Postgraduate School IRB Chair, Dr. Larry Shattuck, 831-656-2473, lgshattu@nps.edu.

Statement of Consent. I have read the information provided above. I have been given the opportunity to ask questions and all the questions have been answered to my satisfaction. I have been provided a copy of this form for my records and I agree to participate in this study. I understand that by agreeing to participate in this research and signing this form, I do not waive any of my legal rights.

Participant's Signature

Date

To be completed after completion of research session:

☐ I consent to have my data used in this study

☐ I do not consent to have my data used in this study

Participant's Signature

Date

APPENDIX E. EVENT BRIEF

ESCAPE In-Brief / Gameplay Instructions

Introduction

- The purpose of the research, entitled Exploring the Development of Trust Between Human and Autonomous Teammates, is to explore how people can effectively team with robots.
- We aim to define objective, performance based metrics of trust for human-autonomous system teams, explore the similarities and differences in these performance metrics between teams made of humans and teams with mixes of humans and robots, and assess how problem complexity and risk affect team performance.
- Understanding these interactions will allow us to better define the information requirements for human-autonomous systems, and define how robot behaviors should vary based on how the robot must interact with the human.

Overview

- You will be asked to play four rounds of a virtual reality spatial puzzle game developed by Johns Hopkins University called *Escape*.
- The game will be played using the Oculus Rift-S virtual reality head mounted display (VR-HMD) and Oculus Touch hand controllers.
- You will play the game with another player, either a human or a bot developed by Johns Hopkins University specifically for this game. The human player will be a member of the research team playing remotely from Johns Hopkins University.
- Participation is expected to last no more than 1.5 hours. Performance metrics will be collected, however, they will not be associated with an individual; all data is anonymized.
- Participation in this experiment is voluntary and confidential.

Procedures

- You will be asked to sign an informed consent before beginning the experiment.
- A member of the research team will be available to assist you with fitting the head mounted display, wear and use of the Touch hand controllers, and answer basic questions about gameplay.

- You will play one training session of *Escape* to allow you to become familiar with the VR-HMD, adapt to the virtual environment, and become familiar with the use of the handheld controls and the kinds of obstacles you will encounter.
*If at any time you feel discomfort using the VR-HMD, please notify a member of the research team.
- At the end of the training session, you will be asked if you would like to play the four rounds of *Escape*.
- If you choose to play, you will be asked to fill out a pre-round one questionnaire and state your preference for your partner in Round 1 (Human or Bot). At the end of Round 2, you will be asked to state your preferred partner for Round 3.
- During gameplay, you will be the leader of your team and be responsible for directing your partner, either the human or the bot using the Touch controllers as demonstrated in the training session. As a reminder, the human partner will be playing remotely from Johns Hopkins University and there will be no verbal communication.
- You will have approximately 7 minutes to complete each round. The timer for each round (elapsed time) can be viewed by looking at your left wrist in the VR-HMD. In Rounds 3 and 4, a session timer (starting at 0:00) is displayed on the wall for added situational awareness.
- After completing a round, you will enter a waiting room chamber. Once inside the chamber, **DO NOT PRESS THE GO BUTTON** on the far wall.
- After completing a round, you will be asked to take off the VR-HMD and Touch controllers and be handed an end of round questionnaire to fill out. Once complete with the questionnaire and you and the research team are ready to proceed, you may press the go button and begin the next round.
- After the fourth and final round, you will be asked to fill out a final questionnaire and then be debriefed on your performance.

Disclaimer

- *Escape* is a developmental game created for this experiment. With any first generation software prototype, there are going to be bugs. The following are a some of the bugs encountered by the research team and remedial actions to be taken:

- **Locking or freezing up.** If the game locks or freezes, there is no way to reset the game at a specific point in time (the game will need to be restarted from the beginning). The research team will conduct a reset of the game and ‘administratively’ progress you to the point where the error occurred. An administrative reset should take no more than 5 minutes. You will not be penalized for the error, as the data from the game prior to the error is saved and recording of game performance will continue once reset.
- **Hint button on left wrist does not work.** The hint button does not work, do not waste valuable time pressing the button and looking for a hint.
- **Timer on left wrist is elapsed time.** The timer on your left wrist is elapsed time. You will have approximately 7 minutes to complete each round. After you complete a round and answer the post-round questionnaires, the timer on your wrist keeps counting-up. When you begin the next round, take note of the time and add 7 minutes to it for your time to go for the current round. In Rounds 3 and 4, a session timer (starting at 0:00) is displayed on the wall for added situational awareness.
- **Round time-out.** If you do not successfully complete a round within 7 minutes, you will be transported to the waiting room chamber and the level will be recorded as unsuccessful. Although your progress will be recorded as unsuccessful, you will still progress to the next round.
- Reminder, **DO NOT PRESS THE GO BUTTON** on the far wall once in the chamber after an unsuccessful round. You will be asked to take off the VR-HMD and Touch controllers and be handed an end of round questionnaire to fill out. Once complete with the questionnaire and you and the research team are ready to proceed, you may press the go button and begin the next round.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- Abbass, H., Petraki, E., Merrick, K., Harvey, J., & Barlow, M. (2016). Trusted autonomy and cognitive cyber symbiosis: Open challenges. *Cognitive Computation*, 8(3), 385–408. <https://doi.org/10.1007/s12559-015-9365-5>
- Baker, D. P., Day, R., & Salas, E. (2006). Teamwork as an essential component of high-reliability organizations. *Health Services Research*, 41(4p2), 1576–1598.
- Baker, H. (2020). How to disable the Oculus Quest Guardian Boundary and tracking system. Upload. <https://uploadvr.com/how-to-disable-quest-guardian/>
- Blackhurst, J. L., Gresham, J. S., & Stone, M. O. (2011). The autonomy paradox: Why “unmanned systems” don’t shrink manpower needs. *Armed Forces Journal*. <http://armedforcesjournal.com/the-autonomy-paradox/>
- Brzowski, M., & Nathan-Roberts, D. (2019). Trust measurement in human-automation interaction: A systematic review. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1), 1595–1599. <https://doi.org/10.1177/1071181319631462>
- Cambridge English Dictionary. (n.d.). Reliance. In *Cambridge English Dictionary*. Retrieved May 13, 2020, from <https://dictionary.cambridge.org/us/dictionary/english/reliance>
- Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. *Applied Ergonomics*, 52, 333–342. <https://doi.org/10.1016/j.apergo.2015.07.012>
- Chen, J. Y. C., & Terrence, P. I. (2009). Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics*, 52(8), 907–920. <https://doi.org/10.1080/00140130802680773>
- Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, 58(6), 737–758. [https://doi.org/10.1016/S1071-5819\(03\)00041-7](https://doi.org/10.1016/S1071-5819(03)00041-7)
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>

- de Visser, E. J., Krueger, F., McKnight, P., Scheid, S., Smith, M., Chalk, S., & Parasuraman, R. (2012). The world is not enough: Trust in cognitive agents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), 263–267. <https://doi.org/10.1177/1071181312561062>
- de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., 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>
- de Visser, E. J., 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>
- Deutsch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. Yale University Press.
- Drnec, K., & Metcalfe, J. S. (2016). Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. *International Conference on Augmented Cognition*, 157–167. https://link.springer.com/content/pdf/10.1007%2F978-3-319-39952-2_16.pdf
- Duhigg, C. (2016). *Smarter faster better: The transformative power of real productivity*. Random House.
- 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)
- Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, 13(3), 147–164. https://doi.org/10.1207/S15327876MP1303_2
- Ellman, J., Samp, L. S., & Coll, G. (2017). *Assessing the third offset strategy*. Center for Strategic & International Studies.
- Galster, S. M., & Parasuraman, R. (2003). The concordance of subjective measures and military pilot performance in a study of human-interaction with automation in a complex simulated environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 47(1), 212–216. <https://doi.org/10.1177/154193120304700145>
- Hancock, Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5), 517–527. <https://doi.org/10.1177/0018720811417254>

- Hankiewicz, K. (2018). *What is the real difference between automation and AI?* Medium. <https://becominghuman.ai/what-is-the-real-difference-between-automation-and-ai-366513e0c910>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Holmes, J. G., & Rempel, J. K. (1989). Trust in close relationships. In *Close Relationships* (pp. 187–220). Sage Publications, Inc.
- Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04
- Joe, J. C., O'Hara, J., Medema, H. D., & Oxstrand, J. H. (2014). *Identifying requirements for effective human-automation teamwork*. Idaho National Laboratory (INL).
- Johns, J. L. (1996). A concept analysis of trust. *Journal of Advanced Nursing*, 24(1), 76–83. <https://doi.org/10.1046/j.1365-2648.1996.16310.x>
- Johnson, M., & Vera, A. H. (2019). No AI is an island: The case for teaming intelligence. *AI Magazine*, 40(1), 16–28. <https://doi.org/10.1609/aimag.v40i1.2842>
- Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. (2003). *An atlas of interpersonal situations*. Cambridge University Press.
- Kramer, R., & Carnevale, P. (2001). Trust and intergroup negotiation. In *Blackwell handbook of social psychology: Intergroup processes* (pp. 431–450). Blackwell Publishers Ltd.
- Laird, J., Ranganath, C., & Gershman, S. (2019). *Future Directions in Human Machine Teaming Workshop*. Department of Defense. <https://basicresearch.defense.gov/Portals/61/Future%20Directions%20in%20Human%20Machine%20Teaming%20Workshop%20report%20%20%28for%20public%20release%29.pdf>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80. <https://doi.org/10.1518/hfes.46.1.50.30392>
- Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: comparing humans to automation. *Journal of Experimental Psychology: Applied*, 6(2), 104–123. <https://doi.org/10.1037/1076-898X.6.2.104>

- Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, 8(4), 277–301. <https://doi.org/10.1080/14639220500337708>
- Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors*, 48(2), 241–256. <https://doi.org/10.1518/00187200677772440>
- Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, 37(1), 465–498. <https://doi.org/10.1002/aris.1440370111>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
- McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors*, 60(2), 262–273. <https://doi.org/10.1177/0018720817743223>
- Mishra, A. K. (1996). Organizational responses to crisis. Trust in organizations: *Frontiers of Theory and Research*, 3(5), 261–287.
- Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures: “objective” versus “subjective” measurement. *Human Factors*, 34(4), 441–455. <https://doi.org/10.1177/001872089203400406>
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429–460. <https://doi.org/10.1080/00140139608964474>
- Oculus. (n.d.-a). [Oculus Rift-S headset]. https://www.oculus.com/rift-s/?locale=en_US
- Oculus. (n.d.-b). [Oculus Touch controllers]. https://www.oculus.com/rift-s/?locale=en_US
- Office of the Under Secretary of Defense. (2017). *Unmanned systems integrated roadmap 2017–2042*. Department of Defense. https://www.defensedaily.com/wp-content/uploads/post_attachment/206477.pdf
- Office of the Under Secretary of Defense, & for Acquisition, Technology and Logistics. (2016). *Report of the Defense Science Board Summer Study on Autonomy*. Department of Defense. <http://edocs.nps.edu/2016/October/Report%20of%20the%20Defense%20Science%20Board%20Summer%20Study%20on%20Automony.pdf>

- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1), 95.
- Renn, O., & Levine, D. (1991). Credibility and trust in risk communication. In *Communicating risks to the public* (pp. 175–217). Springer.
- Rossi, F. (2018). Building trust in artificial intelligence. *Journal of International Affairs*, 72(1), 127–134. <https://doi.org/10.2307/26588348>
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust 1. *Journal of Personality*, 35(4), 651–665.
- Salas, E., Sims, D. E., & Klein, C. (2004). Cooperation at work. *Encyclopedia of Applied Psychology*, 1, 497–505.
- Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 58(3), 377–400. <https://doi.org/10.1177/0018720816634228>
- Shukla 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. https://pdfs.semanticscholar.org/2480/a71ef5e5a2b1f4a9217a0432c0c974c6c28c.pdf?_ga=2.12673783.11415393.1598322365-2028994908.1588293046
- Simpson, J. A. (2007). Psychological foundations of trust. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, 16(5), 264–268. <https://doi.org/10.1111/j.1467-8721.2007.00517.x>

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