



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

**THE EFFECT OF INTERMEDIATE TRUST
RATINGS ON AUTOMATION RELIANCE**

by

Linus Torner

December 2023

Dissertation Supervisor:

Mollie R. McGuire

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 December 2023	3. REPORT TYPE AND DATES COVERED Dissertation	
4. TITLE AND SUBTITLE THE EFFECT OF INTERMEDIATE TRUST RATINGS ON AUTOMATION RELIANCE			5. FUNDING NUMBERS	
6. AUTHOR(S) Linus Torner				
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) As automated systems are increasingly capable of augmenting human decision-makers, appropriate reliance on automation has the potential to increase safety and efficiency in several high-stake domains. To that end, a solid understanding of how and under what conditions people rely on automation is needed to design decision aids that allow people to rely on them appropriately. Previous studies have used regular trust ratings during human-automation interactions to examine how trust develops and evolves, but such intermediate judgments might affect subsequent reliance decisions. This dissertation addresses a knowledge gap by empirically exploring how intermediate trust ratings affect automation reliance in human-automation interactions. A laboratory experiment in which 118 participants, supported by automated decision aids, identified UAVs in images was conducted to determine whether trust rating frequency, automation reliability, and participant motivation affected participant reliance behavior. Findings show that intermediate trust ratings increased automation reliance and retrospective trust ratings but did not affect response time. This dissertation proposes an extended theoretical model that might help explain and predict automation reliance. Additionally, it suggests that intermediate trust ratings might be suitable for calibrating automation reliance but not for research that seeks to measure trust without influencing reliance behavior.				
14. SUBJECT TERMS Automation, reliance, appropriate reliance, trust, calibrated trust, trust assessment, decision aid, decision support, intermediate judgments			15. NUMBER OF PAGES 129	
			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	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**THE EFFECT OF INTERMEDIATE TRUST RATINGS ON AUTOMATION
RELIANCE**

Linus Torner
Major, Swedish Army
MCS, Chalmers University of Technology, 2003
Information Warfare Systems Engineering, Naval Postgraduate School, 2015

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN INFORMATION SCIENCES

from the

**NAVAL POSTGRADUATE SCHOOL
December 2023**

Approved by: Mollie R. McGuire
Department of
Information Sciences
Dissertation Supervisor

Alex Bordetsky
Department of
Information Sciences
Dissertation Chair

Dan C. Boger
Department of
Information Sciences

Rudolph P. Darken
Department of
Computer Science

Bonnie W. Johnson
Department of
Systems Engineering

Approved by: Alex Bordetsky
Chair, Department of Information Sciences

Joseph P. Hooper
Vice Provost of Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

As automated systems are increasingly capable of augmenting human decision-makers, appropriate reliance on automation has the potential to increase safety and efficiency in several high-stake domains. To that end, a solid understanding of how and under what conditions people rely on automation is needed to design decision aids that allow people to rely on them appropriately. Previous studies have used regular trust ratings during human–automation interactions to examine how trust develops and evolves, but such intermediate judgments might affect subsequent reliance decisions. This dissertation addresses a knowledge gap by empirically exploring how intermediate trust ratings affect automation reliance in human–automation interactions. A laboratory experiment in which 118 participants, supported by automated decision aids, identified UAVs in images was conducted to determine whether trust rating frequency, automation reliability, and participant motivation affected participant reliance behavior. Findings show that intermediate trust ratings increased automation reliance and retrospective trust ratings but did not affect response time. This dissertation proposes an extended theoretical model that might help explain and predict automation reliance. Additionally, it suggests that intermediate trust ratings might be suitable for calibrating automation reliance but not for research that seeks to measure trust without influencing reliance behavior.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	RESEARCH PROBLEM.....	2
	1. Research Purpose.....	3
	2. Research Scope and Boundaries.....	4
B.	MOTIVATION AND CONTRIBUTION.....	5
C.	ORGANIZATION	5
II.	LITERATURE REVIEW	7
A.	THEORETICAL FOUNDATIONS OF TRUST AND TECHNOLOGY USE.....	8
	1. Automation	9
	2. Trust in Automation	10
	3. Automation Reliance	11
B.	PREVIOUS STUDIES ON TRUST AND RELIANCE.....	11
	1. The Relationship between Trust and Reliance.....	11
	2. Antecedents of Trust.....	13
	3. Trust Measurement	13
	4. Initial and Dynamic Trust.....	15
	5. The Dynamics of Trust	17
C.	INTERMEDIATE JUDGMENTS.....	18
	1. Effects of Intermediate Judgments.....	18
	2. Articulation of Impressions as Intermediate Judgments	19
	3. Intermediate Judgments as Reminders	19
	4. Bridging the Gap.....	20
D.	POTENTIAL MODERATING VARIABLES	21
	1. Automation Reliability	21
	2. Motivation.....	22
E.	CURRENT STUDY	23
	1. Research Hypotheses	24
	2. Operational Definitions of Variables	24
F.	SUMMARY	27
III.	RESEARCH METHOD	29
A.	EXPERIMENTAL DESIGN	29
	1. Experimental Manipulation.....	30
	2. Measures	31

B.	PARTICIPANTS.....	32
C.	MATERIALS	33
1.	Laboratory Setup	33
2.	Technology Readiness Index 2.0.....	34
3.	UAV Identification Task	34
4.	Exit Questionnaire	36
D.	PROCEDURES	36
1.	Informed Consent Form.....	37
2.	Attitudes Toward Technology	37
3.	Introduction.....	37
4.	Practice.....	37
5.	First Decision Aid.....	38
6.	Second Decision Aid.....	38
7.	Exit Questionnaire	38
8.	Debrief.....	38
E.	SUMMARY	38
IV.	RESULTS	41
A.	MAIN ANALYSIS	41
1.	Automation Reliance	42
2.	Retrospective Trust.....	45
3.	Response Time.....	50
B.	EXPLORATORY ANALYSIS	52
1.	Trust Ratings.....	53
2.	Types of Reliance	57
3.	Accumulative Evidence of Reliance	59
4.	Correlations	60
5.	Participant Motivation	67
6.	Order Effects	68
V.	DISCUSSION	71
A.	INTERPRETATION OF RESULTS	71
1.	The Effect of Trust Rating Frequency	72
2.	The Effect of Automation Reliability	74
3.	The Effect of Human Characteristics	74
B.	CONTRIBUTIONS.....	74
C.	LIMITATIONS	75
D.	FUTURE RESEARCH.....	77
E.	CONCLUSION	78

APPENDIX A. TECHNOLOGY READINESS INDEX 2.0	79
APPENDIX B. EXIT QUESTIONNAIRE	83
A. DEMOGRAPHICS	83
B. EXPERIENCE WITH THE SAME OR SIMILAR TECHNOLOGY	84
C. ATTENTION AND MOTIVATION	86
D. RETROSPECTIVE TRUST	87
APPENDIX C. CORRELATIONS.....	89
A. DATA DISTRIBUTION OF VARIABLES.....	89
B. SCATTER PLOTS: AVERAGE RELIANCE	91
C. SCATTER PLOTS: AVERAGE RETROSPECTIVE TRUST.....	92
D. SCATTER PLOTS: AVERAGE RESPONSE TIME	93
LIST OF REFERENCES.....	95
INITIAL DISTRIBUTION LIST	109

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1.	Existing Literature and Knowledge Gap.....	7
Figure 2.	Levels of Automation. Adapted from Parasuraman et al. (2000).	10
Figure 3.	Trust and Reliance in the IMOI Model. Adapted from Ilgen et al. (2005).	12
Figure 4.	Trust and Reliance during Human–Automation Interaction. Adapted from Hoff & Bashir (2015).	16
Figure 5.	Trust Scale for Retrospective Trust Ratings. Adapted from Merritt, Unnerstall et al. (2015).	26
Figure 6.	$2 \times 3 \times 2$ Factorial Treatment Design	30
Figure 7.	Automation Reliance	32
Figure 8.	Example Image	35
Figure 9.	Participant Subgroups	41
Figure 10.	REL by Subgroup	44
Figure 11.	RTR by Subgroup	47
Figure 12.	Statistical Interaction by MOT Condition.....	49
Figure 13.	Statistical Interaction by AUR Condition.....	50
Figure 14.	RES by Subgroup.....	52
Figure 15.	Intermediate Trust Ratings by Decision Aid	54
Figure 16.	Intermediate Trust Ratings Outliers.....	55
Figure 17.	Accumulated Reliance	59
Figure 18.	Data Distribution Average REL.....	63
Figure 19.	Age by Subgroup	64
Figure 20.	Kruskal–Wallis Test of Age.....	65
Figure 21.	Data Distribution Average RTR	66

Figure 22.	Data Distribution Average RES.....	67
Figure 23.	REL by Decision Aid Order	68
Figure 24.	RTR by Decision Aid Order	69
Figure 25.	Proposed Theoretical Model	72
Figure 26.	Data Distribution Age	89
Figure 27.	Data Distribution Self-Performance	89
Figure 28.	Data Distribution Self-Confidence	90
Figure 29.	Data Distribution Team-Confidence.....	90
Figure 30.	Scatter Plot of Average REL by Age.....	91
Figure 31.	Scatter Plot of Average REL by Self-Performance	91
Figure 32.	Scatter Plot of Average REL by Self-Confidence	92
Figure 33.	Scatter plot of Average RTR by Team-Confidence.....	92
Figure 34.	Scatter Plot of Average RES by Age	93

LIST OF TABLES

Table 1.	Independent and Dependent Variables	25
Table 2.	Participant Demographics	33
Table 3.	REL Lower Reliability Decision Aid	43
Table 4.	REL Higher Reliability Decision Aid	43
Table 5.	RTR Lower Reliability Decision Aid	46
Table 6.	RTR Higher Reliability Decision Aid	46
Table 7.	RES Lower Reliability Decision Aid	51
Table 8.	RES Higher Reliability Decision Aid	51
Table 9.	Intermediate Trust Ratings by Subgroup	55
Table 10.	Intermediate Trust Ratings Descriptive Statistics	56
Table 11.	Categorization of Responses in Reliance Decisions	57
Table 12.	Reliance Decisions by Decision Aid	58
Table 13.	Reliance Decisions by TRF Condition	58
Table 14.	Correlations Average REL	61
Table 15.	Correlations Average RTR	62
Table 16.	Correlations Average RES	62
Table 17.	Subjective Motivation by MOT Condition	68

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
AUR	Automation Reliability
BCa	Bias-Corrected and Accelerated (Bootstrap Confidence Interval)
IMOI	Input-Mediator-Output-Input
ML	Machine Learning
MOT	Participant Motivation
NPS	Naval Postgraduate School
REL	Automation Reliance
RES	Response Time
RTR	Retrospective Trust
TRF	Trust Rating Frequency
TRI 2.0	Technology Readiness Index 2.0
UAV	Unmanned Aerial Vehicle

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First, I would like to thank my dissertation supervisor, Dr. Mollie McGuire, for her commitment, support, and patience during the dissertation. Second, I would like to thank my dissertation committee members for their valuable contributions and guidance throughout my Ph.D. journey. Third, I would like to thank the NPS students, faculty, and staff for all the help, discussions, and dedication during my time here. Finally, I thank my family and friends for their moral support. I would not have been able to do this without you all.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Technological advancements are increasing the capability and accessibility of automation across all personal and professional domains. Automated aids can support human decision-makers faced with an increasing amount of information at a higher pace. Furthermore, automation can perform dangerous or tedious tasks for humans. Hence, automation can increase safety and efficiency in several high-stake domains, such as military, healthcare, and transportation. For example, a faster tempo and more advanced threats create challenges in tactical warfare, and automated decision aids can help address these challenges (Johnson, 2019). Moreover, a recent medical study in Sweden indicated that artificial intelligence (AI) increased cancer detection rates and reduced workload for radiologists (Dembrower et al., 2023).

However, as automated aids that support control and decision-making are increasingly ubiquitous in society, an appropriate level of human reliance on automation is required to improve overall safety and efficiency (Lee & See, 2004). It is important to acknowledge that automation is imperfect and to understand its limitations. Using automation in the wrong contexts or outside its specified range might result in under- or overreliance.

On the one hand, under-reliance on automation might lead to inefficiency or accidents. For example, Zuboff (1988) found that a pulp mill failed to take advantage of the potential benefit of a new automated control interface because operators did not rely enough on automation. Also, during Southwest Airlines Flight 1455 from Las Vegas to Burbank in 2000, when the pilots ignored warnings of excessive airspeed and sink rate, the airplane overran the runway (Dismukes et al., 2007). Moreover, the ship captain's decision to manually steer the ship instead of relying on the navigation system might have contributed to the death of 32 passengers in the Costa Concordia cruise ship disaster in 2012 (Levs, 2012).

On the other hand, overreliance on automation might also lead to failures. In 1995, the cruise ship *Royal Majesty* went off course and ran aground when the officers

relied too heavily on the automated navigation system, which had failed (National Transportation Safety Board, 1997). Another incident occurred in 2009 when the Turkish Airlines Flight 1951 crashed and nine people were killed. The pilots' reliance on a malfunctioning autopilot might have contributed to the crash (CNN, 2009). A recent example of overreliance is when a self-driving Uber car struck and killed a pedestrian in 2018. The human driver was found to have been streaming video and was unprepared when the self-driving mode alerted and handed back control (Harris, 2019).

To achieve appropriate reliance, system designers need to understand what causes human decision-makers to rely on automation. This question is central in information science research, which aims to “explicate the conceptual and methodological foundations on which existing information systems rest and to develop new concepts on which improved systems can be based” (Borko, 1965, p. 5). Similarly, Debons (2008) defined *information science* as

that activity directed to the search for laws and principles that are part of the analysis, design, and evaluation of information systems. These systems are environments of people, technology, and procedures that serve to augment/extend human capabilities in dealing with the many states (decisions, problem solving) that are a part of day-to-day living. (p. 58)

Previous research has shown that trust plays an essential role in how automation is used (Dzindolet et al., 2003; Hussein et al., 2020; Körber et al., 2018). Therefore, a strong causal understanding of trust formation and dynamics helps system designers develop automated aids that allow human decision-makers to rely on them appropriately.

A. RESEARCH PROBLEM

Research in trust and automation reliance has explored and answered many questions, including identifying different antecedents of trust. Within this body of research, one important question is how the variability of trust affects reliance during human–automation interactions. Studies have shown that trust varies over time (Yang et al., 2023), but trust dynamics is still an underdeveloped research topic.

Even though most empirical trust research has measured trust after an interaction is over, some studies have tried to explore the dynamic aspects of trust. Sampling trust

regularly is one approach to capture adjustments of trust in moment-to-moment interactions. In recent studies (Bhat et al., 2022; Guo & Yang, 2020; Yang et al., 2017), participants were asked to rate their trust during interactions with automation to explore the variability of trust. The studies showed that trust varied over time, but they did not examine if the trust ratings affected participant reliance behavior. This dissertation suggests that trust ratings during human–automation interaction are *intermediate judgments* that might influence subsequent reliance decisions. Intermediate judgments are made before or between decisions (or choices) of interest.

The problem is that researchers have limited knowledge of the effect of intermediate trust ratings on automation reliance. No empirical studies have explored this effect, which reveals a knowledge gap. This is a problem because system designers need a strong causal understanding of human reliance behavior to develop and implement automation that allows humans to rely on them appropriately. Furthermore, intermediate trust ratings might affect retrospective trust ratings of automation, which could then be more useful as a basis for system evaluation and design. The next section deals specifically with how this dissertation addresses this knowledge gap.

1. Research Purpose

This dissertation explored the effect of intermediate trust ratings on automation reliance during moment-to-moment interactions between a human decision-maker and an automated decision aid. Furthermore, potential interactions between trust ratings, automation reliability, and participant motivation were examined. Finally, the study investigated correlations between human characteristics and automation reliance.

The goal of this research was to provide a theoretical contribution to enhance the understanding of automation reliance. Starting with existing assumptions of automation reliance and integrating them with theory on intermediate judgments, a set of hypotheses was developed and tested in a laboratory setting using an experimental approach.

2. Research Scope and Boundaries

The dissertation was conducted through a military lens; most participants were military students, and the task was military-relevant. Participants were students, faculty, and staff at the Naval Postgraduate School (NPS) in Monterey, California. The task was to detect unmanned aerial vehicles (UAVs) in images, supported by automated decision aids.

Automation can be utilized at different levels of the decision-making process. Parasuraman et al. (2000) described four dimensions that can be automated in human information processing: information acquisition, information analysis, decision and action selection, and action implementation. This dissertation considers automated aids in the first two dimensions, such as decision aids and diagnostic tools, in which automation processes information and gives recommendations. In these dimensions, human decision-makers select and implement actions. This level of automation is suitable in situations where full automation is undesirable for different reasons. Legal and ethical considerations might require humans to be responsible for the decisions made. Uncertainty, unanticipated situations, adversary interference, deception, and context change, which all are typical in military cases, might also warrant human control of decisions.

Furthermore, this dissertation considers a scenario where unreliability and imperfection should be expected in automation. In such contexts, the role of automation is to simplify the human task—for example, by processing data and providing recommendations. A small number of imperfections are tolerable, and potential failures are not catastrophic, in contrast to the failure of an autopilot.

Recent advances in machine learning (ML) and AI have increased automation capabilities but also raised specific questions regarding the trustworthiness of automation. This research takes an algorithm-agnostic approach where findings are independent of the underlying technology of the automation that augments the decision-making process. Specific AI-related considerations are outside of its scope.

Even though trust might spread in an organization and be influenced by how other people use automation (National Academies of Sciences, Engineering, and Medicine [NASEM], 2022), this dissertation focuses on dyadic interactions between one human decision-maker and one automated decision aid. Trust beyond the dyadic relationship lies beyond the scope of this dissertation.

B. MOTIVATION AND CONTRIBUTION

The importance of appropriate reliance on automation is easy to recognize, and this dissertation provides new knowledge that contributes to the understanding of automation reliance. The main contribution of this dissertation is to provide knowledge about how intermediate trust ratings affect automation reliance.

This dissertation provides an extended theoretical model of trust and reliance on automation, including the effect of intermediate trust ratings. A thorough theoretical understanding of how and under what conditions people rely on automation can help predict and guide human–automation interaction. It is impossible to test every single case, and this is where theoretical models can help predict and explain phenomena. Improved dynamic trust models are beneficial since they complement existing listings and categorizations of trust-related factors (NASEM, 2022).

Additionally, this dissertation contributes to methodology by showing how explicit trust ratings during human–automation interactions affect reliance. If researchers seek to measure trust without affecting reliance behavior, other ways to sample trust during interactions are recommended.

The dissertation also has practical contributions. An enhanced understanding of causal relationships between trust-related factors informs the design of automated aids that enable decision-makers to rely on them appropriately. Intermediate trust rating might be one approach to foster appropriate reliance in human–automation interactions.

C. ORGANIZATION

The dissertation is organized as follows. Chapter I introduces the research topic, problem, and contribution. Chapter II reviews existing literature to establish a theoretical

point of departure and derives the research hypotheses. Chapter III describes the research method. Chapter IV presents the results of the analysis. Chapter V provides a discussion, including interpretation of results, contributions, limitations, and recommendations for future research.

II. LITERATURE REVIEW

This chapter reviews existing research that provides this dissertation's theoretical underpinnings and point of departure. The relationship between trust and technology has been extensively studied in the past 40 years to explore why and how people use technology. The first section introduces the theoretical foundations within this body of literature, such as key terms and constructs and their relationships. The second section then reviews previous studies on automation reliance to align this dissertation with existing findings and outstanding questions. The third section discusses intermediate judgments and how they can extend existing theoretical models of automation reliance. The fourth section then reviews potential moderating variables of the relationship between intermediate trust ratings and automation reliance. Finally, the last section of the chapter describes this dissertation's research hypotheses and operational definitions of variables. Figure 1 provides an overview of the automation reliance literature and where this dissertation contributes.

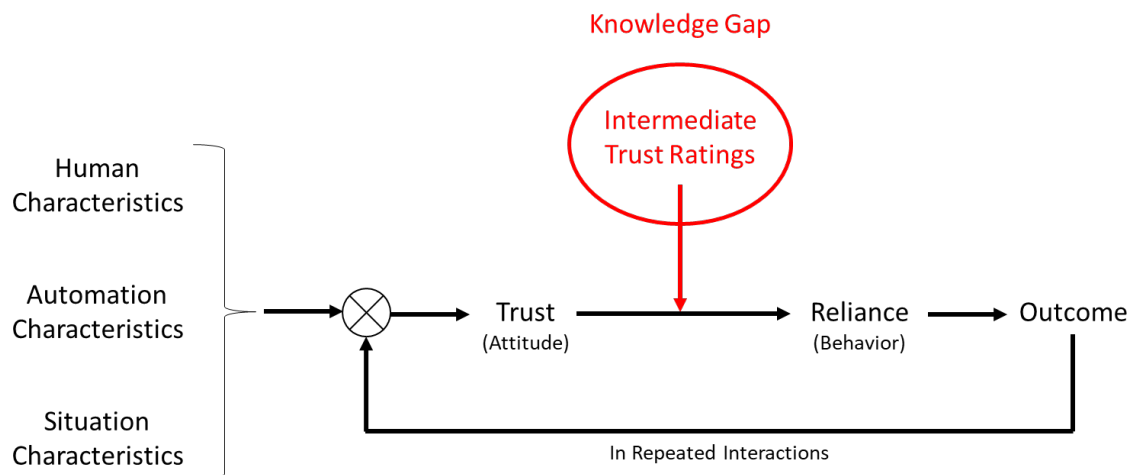


Figure 1. Existing Literature and Knowledge Gap

A. THEORETICAL FOUNDATIONS OF TRUST AND TECHNOLOGY USE

The complexity inherent in trust has led to different notions of its nature and uses across scientific disciplines, such as psychology, sociology, and philosophy (Marsh & Dibben, 2003). In general, there are three common notions of trust: a mental attitude toward a trustee, an intention to rely on a trustee, and behavior in relation to a trustee (Falcone & Castelfranchi, 2004; Lee & See, 2004; Lewis et al., 2018; McKnight, 2005). A trustor is the entity that trusts (or not), and a trustee is the entity that is being trusted (or not).

Diverse notions of trust in various scientific fields have resulted in several different trust definitions. One of the most commonly cited definitions in the trust literature 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 other party. (Mayer et al., 1995, p. 712)

Another well-cited definition of *trust* is “the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another” (McAllister, 1995). Both of these definitions refer to interpersonal trust in organizational management.

Even though early studies of trust in technology come from the literature on interpersonal trust, not all elements of trust between humans are a good fit for trust in technology. The models and measures of human trust might require adjustment to accommodate nonhuman agents (Roberts et al., 2021). Previous studies have identified differences between interpersonal trust and trust in technology. Madhavan and Wiegmann (2007) suggested that people tend to consider information from automated systems more credible than information from humans, but mistakes by automated systems result in a more significant decrease in trust than mistakes by humans. McKnight (2005) stated that both trust in people and trust in information technology are based on perceived competence to perform a task, but that perceived benevolence and integrity also affect interpersonal trust. This dissertation considers *trust in automation* and how it affects *automation reliance*. The remainder of this theoretical introduction further

explains the three foundational constructs: automation, trust in automation, and automation reliance.

1. Automation

Various terms in the literature, such as automation, autonomy, and AI, refer to different types of autonomous systems. Automation is “technology that actively selects data, transforms information, makes decisions, or controls processes” (Lee & See, 2004, p. 50). Similarly, Moray et al. (2000) defined *automation* as “any sensing, detection, information processing, decision-making, or control action that could be performed by humans but is actually performed by machine” (p. 44). Technological advances allow more complex tasks to be automated. Parasuraman and Riley (1997) accounted for this evolution in their definition: “We define automation as the execution by a machine agent (usually a computer) of a function that was previously carried out by a human. What is considered automation will therefore change with time” (p. 231).

The difference between automation and autonomy lies within an autonomous system’s level of independence from human control. Autonomy refers to automation that is given authority to carry out human users’ goals without supervision (de Visser et al., 2018). For example, “Autonomy may be understood to involve a system operating without further human intervention after activation” (Department of State, 2023). Current developments in AI can potentially revolutionize the capability of autonomy (Schmidt et al., 2021; Shyu, 2022). The National Academies of Sciences, Engineering, and Medicine (2022) use the term *AI*

to describe a form of highly capable automation directed at highly perceptual and cognitive tasks. AI potentially improves upon previous forms of automation in its ability to sense and interpret situations, adapt to changes in conditions and the environment, prioritize and optimize based on changes in goals, and refine its abilities through learning. (p. 7)

The *level of automation* represents the degree of automation interdependence from human interventions. Figure 2 illustrates 10 levels of automation of decision and action selection described by Parasuraman et al. (2000), based on the work by Sheridan and Verplank (1978). For each level in the scale, automation is given more control. The

scale is often used in the human–automation literature to describe interactions for specific tasks and settings.

- HIGH 10. The computer decides everything, acts autonomously, ignoring the human.
9. informs the human only if it, the computer, decides to
8. informs the human only if asked, or
7. executes automatically, then necessarily informs the human, and
6. allows the human a restricted time to veto before automatic execution, or
5. executes that suggestion if the human approves, or
4. suggests one alternative
3. narrows the selection down to a few, or
2. The computer offers a complete set of decision/action alternatives, or
LOW 1. The computer offers no assistance: human must take all decisions and actions.

Figure 2. Levels of Automation. Adapted from Parasuraman et al. (2000).

This dissertation explores reliance on decision aids that process information and make recommendations to a human decision-maker, who then decides whether to rely on the recommendation. Using the scale in Figure 2, this corresponds to level four of automation. Of the terms described in this section, *automation* is the appropriate term for that level of automation, not *autonomy*. For this level of automation, even though some functions in the decision aids might be AI-enabled, a human decision-maker always selects and implements actions.

2. Trust in Automation

Trust specific to human–automation interactions has been labeled *trust in automation*, defined as “the attitude that one agent will achieve another agent’s goal in a situation where imperfect knowledge is given with uncertainty and vulnerability” (Lee & See, 2004, p. 51). This definition is widely adopted in the human–automation literature. This dissertation adopts that definition to ensure construct validity and align with the literature in this domain. The notion of trust in automation has been central in research about human–automation interaction and has been shown to influence automation reliance (e.g., Dzindolet et al., 2003; Hussein et al., 2020; Körber et al., 2018).

3. Automation Reliance

Researchers have long suggested that trust in automation impacts reliance decisions (Lee & Moray, 1992; Muir, 1994). In this body of literature, trust is conceived as an attitude that influences reliance, which is a behavior (Lee & See, 2004). However, factors not related to trust might affect reliance as well. For example, time pressure can cause humans to rely on technology despite a lack of trust (Rice & Keller, 2009).

What it means to rely on automation depends on the type of automation, the task the automation performs, and the situation. For automated decision aids, reliance can mean following recommendations. Yu et al. (2016) defined reliance as the rate of following recommendations made by an automated quality monitoring system. Similarly, Yuviler-Gavish and Gopher (2011) measured reliance as the rate at which participants followed decision aid recommendations. Automation reliance can be compared to self-reliance in situations where human decision-makers can rely on themselves or automation (Dzindolet et al., 2003). Since reliance is considered a behavior, it is more accessible to observe and measure than trust, which is an attitude.

B. PREVIOUS STUDIES ON TRUST AND RELIANCE

This section describes findings within the body of knowledge regarding trust and reliance in human–automation interaction, which provide the theoretical framework for this dissertation. Researchers still seek to better understand the dynamic nature of trust and reliance, which could help achieve appropriate reliance throughout human–automation interactions. The following sections cover the different elements in Figure 1.

1. The Relationship between Trust and Reliance

It is essential to clearly define the relationship between trust and reliance since the terms have different meanings and are sometimes used interchangeably in the human–automation literature. Previous research has suggested that trust is critical to whether and how automation is used (Lewandowsky et al., 2000; Muir & Moray, 1996; Sheridan, 2002). Falcone and Castelfranchi (2004) described trust as a set of mental components on which reliance behavior is based. Hence, trust precedes reliance.

The relationship between trust and reliance can be illustrated using the Input-Mediator-Output-Input (IMOI) model (Ilgen et al., 2005), which is a team performance framework. Humans and automation working together can be considered a human–automation team, where the respective strengths of team members complement each other to achieve higher performance than either alone (Calhoun, 2022; Cuevas et al., 2007). Figure 3 illustrates how the IMOI model describes trust as a mediator or an emergent state that is affected by inputs and, in turn, affects observable outputs. Multiple criteria can be examined as outputs, and relevant inputs should be considered. In the present model application, reliance is the output of interest, and inputs to trust are divided into human, automation, and situation characteristics. A feedback loop represents how previous outputs become inputs in a repeated interaction.

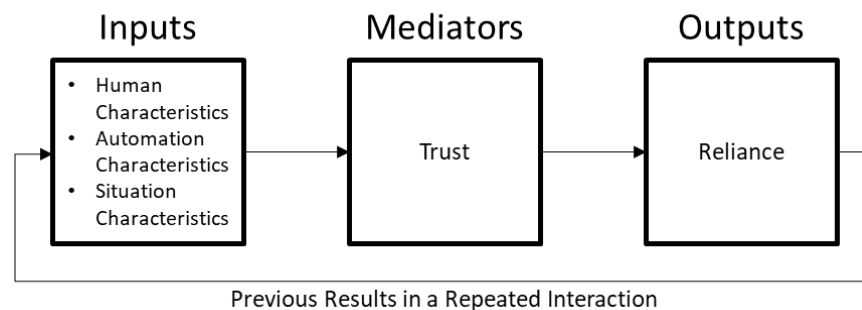


Figure 3. Trust and Reliance in the IMOI Model. Adapted from Ilgen et al. (2005).

Human trust should be calibrated to automation capabilities and performance to ensure appropriate reliance (Muir, 1987). A calibrated level of trust means that “trust matches system capabilities, leading to appropriate use” (Lee & See, 2004, p. 55). Over-trust leads to overreliance or *misuse*, and under-trust leads to underutilization or *disuse* (Parasuraman & Riley, 1997). Misuse can cause accidents when automation is relied upon in situations it was not designed for. Disuse can reduce efficiency or safety when tasks are done manually instead of taking advantage of automation capabilities. Inappropriate levels of trust might threaten Department of Defense directives and ethical considerations because they hamper human efforts to adequately direct and oversee

automated systems (NASEM, 2022). Therefore, it is essential to understand *antecedents* of trust, which means preceding events, conditions, or causes.

2. Antecedents of Trust

A vast amount of research has studied antecedents of trust in automation (Schaefer et al., 2016), trust in technology (Saßmannshausen et al., 2022), trust in robots (Hancock et al., 2011), and trust in AI (Glikson & Woolley, 2020). A general conclusion is that antecedents of trust in human–technology interaction can be categorized as human, technology, or situation characteristics (Endsley, 2017), as illustrated in Figure 1. This categorization aligns well with the Mayer et al. (1995) model of trust, in which trust depends on trustor, trustee, and context characteristics.

Previous studies have identified several significant antecedents of trust in each category. Examples of human characteristics include demographics (Merritt & Ilgen, 2008), attitudes toward technology (Merritt et al., 2013), domain expertise (Nourani et al., 2020), and personality (McBride et al., 2012). Examples of technology characteristics include reliability (Sanchez et al., 2004), transparency (Wang et al., 2011), and appearance (Pak et al., 2012). Examples of situational characteristics include task characteristics (Madhavan et al., 2006), risk (Perkins et al., 2010), and workload (Spain & Bliss, 2008). However, the method for measuring trust varies significantly between studies, which can limit the interpretation of the results.

3. Trust Measurement

Even though the importance of trust in automation is widely recognized in the literature, there is no standardized method for trust measurement. Measuring trust is critical for implementing mechanisms that keep trust calibrated and drive the appropriate use of technology (Tenhundfeld et al., 2022). However, the variety of applications of trust in automation has led to different approaches to trust measurement. The variability of trust measurements creates challenges in studying trust because the scientific comparison between various studies is difficult (Larasati et al., 2023). One reason for the variability of trust measurements is the disagreement over definitions of trust (Larasati et al., 2023). Another reason is the confusion of trust with related constructs such as

reliance, compliance, or confidence (Vereschak et al., 2021). Furthermore, nonexperts may have conceptions of trust that may dilute theoretical concepts of trust (NASEM, 2022). To avoid this dilution and to ensure that the correct construct is measured, research participants should be provided with a clear definition of trust (Vereschak et al., 2021).

Since trust is a construct that cannot be accessed directly, it has often been measured through subjective self-reporting after an interaction. However, trust behavior during an interaction might not be accurately described by a subjective trust assessment afterward (Desai et al., 2013; Hancock et al., 2011; Kohn et al., 2021). Although most existing studies considered a “snapshot” perspective of trust, either through static factors that influence trust or through single evaluations of trust after an experiment, research has shown that trust is dynamic and evolves throughout an interaction (Akash et al., 2017; Ekman et al., 2018; Kraus et al., 2020; Manzey et al., 2012; Yang et al., 2023). Therefore, continuous measures and evaluations of trust might provide more insights than a single retrospective evaluation.

Different measures of trust might be suitable in different situations. Trust is often measured on Likert-type ordinal or interval scales (Jian et al., 2000; Xu & Dudek, 2015). Malle and Ullman (2021) suggested that more comprehensive multi-item scales are preferred to single-item scales to capture different dimensions of trust. Furthermore, multi-item scales enable construct validity assessment, leading to more reliable trust measurements (Gillespie, 2015). Even though comprehensive scales might capture the richness and complexity of trust, they are more appropriate at the end of a task, whereas single-item measures might be more appropriate during interactions to reduce task interruption (Kohn et al., 2021; Vereschak et al., 2021). Wischniewski et al. (2023) reviewed 96 publications on trust calibration. They concluded that “the increased use of validated scales ensures the quality of the measurement” (p. 7) but that the use of ad hoc single-item trust measures “was often justified by increased usability and lowered intrusiveness” (p. 7).

4. Initial and Dynamic Trust

When studying trust in human–automation interactions, initial trust can be distinguished from dynamic trust. Figure 1 illustrates how trust is updated during repeated interactions. Initial trust is established before the interaction begins, and dynamic trust is established during the interaction. McKnight et al. (1998) defined initial trust as “when parties first meet or interact” (p. 473). Initial trust can help explain technology acceptance (i.e., if a person starts using technology). In contrast, dynamic trust influences whether a person continues to rely on technology during repeated interactions (Hoff & Bashir, 2015).

Li et al. (2008) studied initial trust formation in an organizational context as a new national identification system was introduced to the employees. The authors found that the following influenced the initial trust formation: (a) the participants’ perception that people important to them thought that they should use the system, (b) a rational cost–benefit analysis of using the system, (c) the reputation of the system, and (d) the perception that the organizational situation was normal.

Dynamic trust complements initial trust by accounting for how trust is updated during an interaction. Hoff and Bashir (2015) reviewed 127 empirical studies on trust in automation and created a theoretical model that captures the variability of trust and reliance through feedback mechanisms (see Figure 4). The model, which has been validated in subsequent studies (Bonneviot et al., 2021; Holthausen et al., 2020; Hsieh et al., 2022), illustrates through a cycle how trust affects reliance behavior, reliance behavior affects system performance, and system performance affects trust. This cyclical notion has often been ignored in previous theoretical models—for example, the Technology Acceptance Model (Ghazizadeh et al., 2012). The dotted lines in Figure 4 represent relationships that can vary during moment-to-moment interactions. The different parts of Hoff and Bashir’s (2015) model are briefly described next.



The model separates trust into three components previously described by Marsh and Dibben (2003): dispositional trust (depends on an individual's inclination to be trusting or not), situational trust (depends on the specific context), and learned trust (depends on previous experience of a particular technology). Dispositional trust has a more substantial influence on reliance behavior at the start of the interaction, but the influence of learned trust grows as experience builds after repeated interaction with the same technology (Merritt & Ilgen, 2008).

16

similar systems, and understanding of the system affect the initial learned trust established before interaction with a system. In contrast, dynamic learned trust is built during interactions and depends on system performance (e.g., reliability, predictability, errors, and usefulness).

b. Initial Reliance Strategy

Hoff and Bashir (2015) described how dispositional, situational, and initial learned trust help determine an initial reliance strategy together with situational factors not related to trust (e.g., time pressure). The initial reliance strategy might affect system performance, which in turn affects dynamic learned trust during the interaction.

c. Reliance on System: The Reliance Decision

Reliance on system in the model (Hoff & Bashir, 2015) represents a decision: whether to rely on automation. The relationship between dynamic learned trust and system reliance (red in Figure 4) illustrates how dynamic learned trust influences the reliance decision. Because trust is adjusted during human–automation interaction (Yang et al., 2023), reliance on automation might vary.

5. The Dynamics of Trust

The dynamics of trust refers to the evolution of trust over time (Jonker & Treur, 1999). Since trust depends on experience and other factors that might change over time, trust itself is a dynamic entity (Falcone & Castelfranchi, 2004). Dynamic models of trust evolution are an important research objective because trust evolves from repeated interactions between humans and automation and the outcomes of those interactions (NASEM, 2022).

Even though previous research has explored some aspects of trust dynamics, further research is needed. Madhavan and Wiegmann (2007) suggested that the type of automation failure affects how much trust decreases (e.g., immediacy and criticality of failures). Simple mistakes tend to have a more significant impact on trust. If a human understands the source of the failure, the effect of the failure might be modified (Wickens & Dixon, 2007). Yang et al. (2023) examined how trust is adjusted during human–

automation interaction and found that automation failures, compared to automation successes, had a more significant impact on trust adjustment. The same results were obtained by Yu et al. (2017), who also showed that the timing of automation failures and successes during an interaction affects the dynamics of trust. Trust loss caused by automation failures was less significant further into interactions, suggesting that familiarity with a specific system mitigates the effect of failures (Yu et al., 2017). As trust evolves over repeated interaction with the same automation, it tends to vary more initially and then stabilize over time (Yang et al., 2017; Yu et al., 2017).

Collecting trust ratings during interactions has been proposed in recent studies to examine micro-trust adjustments during interactions (Bhat et al., 2022; Clare et al., 2015; Guo & Yang, 2020; Holliday et al., 2016; Yang et al., 2017; Yu et al., 2017). The studies showed that trust develops throughout an interaction and that regularly measuring trust provides a more comprehensive picture than a single evaluation. However, intermediate trust ratings might affect subsequent reliance behavior. The following section reviews the literature on intermediate judgments to extend automation reliance theory, including the effect of intermediate trust ratings.

C. INTERMEDIATE JUDGMENTS

Intermediate judgments are made before a decision (or choice) of interest. An explicit rating before making a decision, such as rating trust in an automated aid before a reliance decision, is an example of an intermediate judgment. This section provides the rationale for the hypothesis that intermediate trust ratings affect automation reliance.

1. Effects of Intermediate Judgments

Empirical human–subject studies have shown that intermediate judgments might affect subsequent decisions. Croson (1999) showed that asking participants to guess what their opponents would do in an upcoming game before making a decision influenced the decisions. The study showed that the cooperation rate was significantly higher among participants in the no-guess treatment than in the guess treatment. Conte et al. (2009) asked participants to perform tests involving ambiguous images. One group performed only Test A, while another performed Test B before Test A. The findings showed that the

results of Test A differed between the groups. Wang and Busemeyer (2016) compared decisions between participants who made classifications before decisions and participants who did not make classifications. Their results indicated that making explicit classifications influenced subsequent decisions.

2. Articulation of Impressions as Intermediate Judgments

The effect of intermediate judgments is related to the cognitive impact of decision-making. Studies have shown that making a decision (i.e., choosing among alternatives) can alter the relative values of decision options (Sharot et al., 2010; Sherman, 1980). The studies showed that the preference for an alternative increased after it had been selected. In other words, even if decision-makers are hesitant about which option to choose, they might be more convinced of the selected alternative's merit after making the decision.

White et al. (2014) extended this theory to show that articulating an impression can affect subsequent decisions, even if the impression is unrelated to the following decision. The authors suggested that articulating an impression (i.e., making an intermediate judgment) before a decision might affect subsequent cognitive processing and alter the relative values of decision options. No additional information was provided to decision-makers. Instead, the act of articulating an impression seemed to alter the relative values of decision options. Their study suggests that asking for an explicit rating does not simply measure something; it has a constructive effect and adds something to the decision-making process. Drawing on White et al.'s (2014) theory, this dissertation suggests that explicit trust ratings during human–automation interaction are intermediate judgments that might affect subsequent reliance decisions.

3. Intermediate Judgments as Reminders

Furthermore, explicit trust ratings during human–automation interaction might remind decision-makers to reevaluate their trust in automation. Kahneman (2003) described how the accessibility of thoughts affects decision-making. Intuitions—thoughts and preferences that come easily to mind—are more accessible than other thoughts. Characteristics of stimuli that evoke a thought can change its accessibility (Kahneman,

2003). Inducing intermediate judgments by asking for regular trust ratings during human–automation interaction might increase the accessibility of trust and serve as a reminder to reflect on trust. Hoffman et al. (2013) suggested that people often use “default trust” instead of deliberating on whether or how they trust automation. Therefore, Hoffman et al. (2013) called for methods to allow systematic evaluation of the human–automation relationship.

In a related study, Buçinca et al. (2021) suggested that people often resort to heuristics rather than carefully considering each recommendation from AI-enabled decision aids. The authors conducted an experiment to test the effect of three different cognitive forcing interventions on reliance decisions. The three forcing interactions were (1) to have participants wait 30 seconds before the AI recommendation was shown, (2) to have participants make an initial decision before receiving the AI recommendation, and (3) to have the participants explicitly ask for an AI recommendation. Buçinca et al. (2021) found that the cognitive forcing functions significantly reduced overreliance on AI recommendations compared to only providing explanations for the recommendations. However, the more the intervention reduced overreliance, the worse participants rated the user interface design in terms of preference, trust, and difficulty.

4. Bridging the Gap

Existing research suggests that explicit trust ratings during human–automation interaction are intermediate judgments that might affect subsequent reliance decisions. However, this is currently a knowledge gap. Intermediate trust ratings might provide a way to mitigate inappropriate reliance and, if so, would be an important extension of existing theoretical frameworks. Also, intermediate judgments in the form of in-the-moment explicit trust ratings during human–automation interactions might provide additional benefits. Intermediate trust ratings during an interaction might keep trust accessible and affect retrospective trust ratings. If so, retrospective trust ratings might become more accurate (i.e., useful). However, intermediate judgments might have undesired impacts on decision-making (e.g., increasing response time). This dissertation

addresses the knowledge gap by exploring the effects of intermediate trust ratings on automation reliance.

D. POTENTIAL MODERATING VARIABLES

The effect of intermediate trust ratings on automation reliance might vary between different situations. In other words, moderating variables might affect the relationship between intermediate trust ratings and automation reliance. Understanding how various trust-related variables interact is a common research recommendation in the literature (Ekman et al., 2018; Kaplan et al., 2023; Li et al., 2008). Even though several potentially moderating variables exist, previous research suggests that automation reliability and participant motivation are two of the most relevant for this dissertation.

1. Automation Reliability

Automation reliability is one of the most significant antecedents of trust in automation (Bailey & Scerbo, 2007; de Visser & Parasuraman, 2011; Lee & Moray, 1992). Reliability means the degree to which automated agents are free from mistakes or errors (Lewis et al., 2018). Empirical studies have shown that automation reliability influences both trust in automation (de Vries et al., 2003; Dzindolet et al., 2003; Moray et al., 2000) and reliance on automation (de Visser et al., 2016; Hancock et al., 2011; Maltz & Shinar, 2003; Ross et al., 2008; Schaefer et al., 2016). Furthermore, automation reliability significantly affected response time in reliance decisions (Lu & Sarter, 2019). However, Yang et al. (2023) noted that “little research, if any, has examined the effects of automation reliability on moment-to-moment trust adjustment” (p. 865).

Since reliability significantly impacts automation reliance, the effect of intermediate trust ratings on automation reliance might vary between different conditions of automation reliability. Studies have suggested both upper and lower “thresholds” of automation reliability, outside of which reliance almost becomes static at 100% or 0%. High reliability might contribute to automation bias, meaning that decision-makers accept recommendations from automated decision aids without thorough analysis (Parasuraman & Manzey, 2010). On the other hand, low reliability might cause decision-makers to ignore automation recommendations completely (Wickens & Dixon, 2007). Therefore,

the effect of intermediate trust ratings on automation reliance should be studied between these two extremes.

Previous research suggests that between 70% to 100% is a suitable range of automation reliability to explore the effect of intermediate trust ratings. Studies have shown that relying on automated decision aids with reliability lower than approximately 70% might result in worse performance than having no decision aid at all (Kantowitz et al., 1997; Rovira et al., 2007; Wickens et al., 2006). Other studies have shown that trust in decision aids significantly decreases when the reliability drops to around 70% (Yin et al., 2019; Yu et al., 2016; Zhang et al., 2020). Furthermore, there is no practical application for decision aids with around 40% to 60% reliability. One might as well flip a coin (assuming the decision space is binary). Therefore, automation reliability above 70% is considered in this study.

2. Motivation

Motivation is another potential moderating variable to consider when studying reliance on automation. Dzindolet et al. (2003) recognized the need to study trust and reliance under different motivation levels. One of Hoff and Bashir's (2015) recommendations for further research was to explore the effect of human task motivation on automation reliance. Van Dongen and van Maanen (2013) suggested that motivational processes affect the effort people are willing to put into a task, and they therefore help explain human reliance on decision aids. Their study showed that reliance on decision aids increased when participants felt more responsible for the outcome of decisions.

Both intrinsic and extrinsic motivation might affect human behavior. Intrinsic motivation refers to internal motivation processes, and extrinsic motivation comes from external stimuli such as rewards or punishments. Extensive research on how extrinsic rewards affect behavior has led to a broad acceptance of rewards as a motivational strategy (Deci et al., 1999). Punishments are negative extrinsic motivators that might be more effective than rewards but should be used carefully because they can negatively affect confidence, trust, and intrinsic motivation (Kinley & Ben-Hur, 2023).

Motivation affects information processing and decision-making because higher motivation can stimulate cognitive functioning. Several studies have shown that an increased motivational state improves cognitive task performance (Bansal et al., 2019; Chelazzi et al., 2013; Engelmann et al., 2009; Engelmann & Pessoa, 2007; Pochon et al., 2002; Sarter et al., 2006). Kouneiher et al. (2009) showed that motivation regulated cognitive control resources in decision-making. *Cognitive control* means “the ability to select thoughts or actions in relation to internal goals” (Kouneiher et al., 2009, p. 939). Their results indicated that both potential rewards and losses increased cognitive control resources. Paschke et al. (2015) suggested that motivation improves attentional control in cognitive tasks. Their study showed that both potential monetary gain and loss reduced response time and increased neural activity in the parts of the brain associated with attentional control, but through different mechanisms. The authors suggested that loss-based motivational strategies are less efficient because the benefits do not match the increased cognitive effort. Finally, Ballard et al. (2019) compared how potential rewards and punishments affected both the quality and quantity of information processing. Their results showed that potential punishments reduced both the quality and quantity of information processed, compared to potential rewards or no incentives.

E. CURRENT STUDY

Existing literature suggests several implications of intermediate trust ratings in human–automation interaction. Intermediate trust ratings might affect both automation reliance and response time in reliance decisions. Furthermore, intermediate trust ratings might affect retrospective trust ratings after an interaction.

Automation reliability is one of the most significant antecedents of trust in automation. Therefore, it is hypothesized that automation reliability moderates the effect of intermediate trust ratings on automation reliance. In other words, the impact of intermediate trust ratings on automation reliance varies between different automation reliability conditions.

Furthermore, since motivation affects cognitive functions and decision-making, it is hypothesized that participant motivation also moderates the effect of intermediate trust

ratings on automation reliance. Because motivation aids in higher-order cognitive functioning, the need for reminders might be lower. If so, reminders in the form of intermediate trust ratings may be more important when tasks are less motivational. The remainder of this section presents the research hypotheses and operational definitions of variables.

1. Research Hypotheses

Based on existing theory, four primary research hypotheses have been formulated:

H₁: Trust rating frequency affects automation reliance.

H₂: Trust rating frequency affects response time.

H₃: Trust rating frequency affects retrospective trust.

H₄: An interaction is predicted among trust rating frequency, automation reliability, and participant motivation on automation reliance.

To validate previous studies and to test the main effects of all independent variables on the primary dependent variable, two additional research hypotheses are included:

H₅: Higher automation reliability increases automation reliance.

H₆: Higher participant motivation increases automation reliance.

2. Operational Definitions of Variables

The operational definitions of this study's independent and dependent variables are grounded in existing literature. Table 1 lists all variables, their levels, and their measures. *Trust rating frequency* (TRF), *automation reliability* (AUR), and *participant motivation* (MOT) are independent variables, while *automation reliance* (REL) is the primary dependent variable. *Response time* (RES) and *retrospective trust* (RTR) are secondary dependent variables.

Table 1. Independent and Dependent Variables

Variable	Type	Level	Measure
Trust rating frequency (TRF)	Independent	Ordinal	Never, regularly, always
Automation reliability (AUR)	Independent	Nominal	Lower, higher
Participant motivation (MOT)	Independent	Nominal	No, Yes
Automation reliance (REL)	Dependent	Interval	0–1
Response time (RES)	Dependent	Ratio	Milliseconds
Retrospective trust (RTR)	Dependent	Interval	1–7

a. Trust Rating Frequency

This study implements intermediate judgments by requesting in-the-moment explicit trust ratings between reliance decisions. The independent variable *TRF* is operationally defined as how often participants are asked to provide explicit trust ratings during human–automation interaction.

b. Automation Reliability

The independent variable *AUR* is operationally defined in this dissertation as the percentage of correct answers provided by an automated decision aid. For example, 100% reliability means that every recommendation is correct, and 50% reliability means that half of the recommendations are correct. Previous studies have used similar definitions of reliability (Moray et al., 2000; van Dongen & van Maanen, 2013).

c. Participant Motivation

The independent variable *MOT* is operationally defined as the presence of extrinsic motivation through gift cards to the highest performers.

d. Automation Reliance

Reliance decisions were made any time the participant’s initial answer disagreed with the automation recommendation. *REL* is operationally defined as the percentage of reliance decisions in which the participant relied on the automation recommendation over their own initial answer. This operational definition is well aligned with the literature

concerning reliance on recommendations of decision aids (e.g., de Visser et al., 2020; Dzindolet et al., 2003; Kohn et al., 2021; van Dongen & van Maanen, 2013).

e. Response Time

RES is a secondary dependent variable, operationally defined as the time (in milliseconds) from when the automation provides a recommendation until the human makes a reliance decision.

f. Retrospective Trust

The final dependent variable, *RTR*, is operationalized as an overall trust rating after an interaction, using a 3-item scale (see Figure 5), suggested by Merritt et al. (2015). The scale is a subset of the 6-item scale developed by Merritt (2011), which has been used in several subsequent studies (Bobko et al., 2023; Drozdal et al., 2020; Kraus et al., 2021; Merritt, Unnerstall et al., 2015). An average is calculated across the three items to yield a number between 1 and 7.

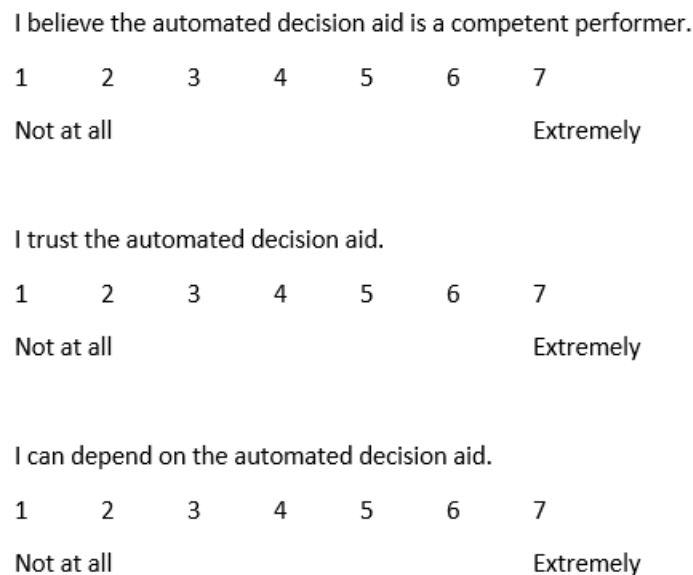


Figure 5. Trust Scale for Retrospective Trust Ratings. Adapted from Merritt, Unnerstall et al. (2015).

F. SUMMARY

Trust is a complex phenomenon to study, and this dissertation considers trust in automation specifically. Existing research show that trust in automation affects automation reliance and provides insights into trust formation. The dynamic nature of trust and reliance is still being explored to better understand trust adjustments during human–automation interaction, which might help achieve appropriate reliance. Intermediate trust ratings during interactions can capture trust dynamics but might affect subsequent reliance decisions. No empirical studies have explored this effect, which exposes a knowledge gap. Furthermore, existing literature suggests that automation reliability and participant motivation might moderate the effect of intermediate trust ratings on automation reliance. This dissertation extends existing theory on automation reliance to include the effects of intermediate trust ratings. The next chapter presents the research method to test six research hypotheses grounded in existing literature.

THIS PAGE INTENTIONALLY LEFT BLANK

III. RESEARCH METHOD

This dissertation suggests an extended theoretical model of automation reliance that accounts for the effects of intermediate trust ratings. Acknowledging that contextual factors might moderate the effect of intermediate trust ratings, this dissertation examined the moderating effect of automation reliability and participant motivation. Six research hypotheses, repeated below, were formulated to describe the hypothesized relationships. A $2 \times 3 \times 2$ factorial treatment design was then developed to test these hypotheses. This chapter describes the experimental design, participants, materials, and experimental procedures.

H₁: Trust rating frequency affects automation reliance.

H₂: Trust rating frequency affects response time.

H₃: Trust rating frequency affects retrospective trust.

H₄: An interaction is predicted among trust rating frequency, automation reliability, and participant motivation on automation reliance.

H₅: Higher automation reliability increases automation reliance.

H₆: Higher participant motivation increases automation reliance.

A. EXPERIMENTAL DESIGN

An experiment was conducted to test the research hypotheses. Figure 6 illustrates a factorial treatment design employed to test the effects of the independent variables. This $2(\text{AUR: lower, higher}) \times 3(\text{TRF: never, regularly, always}) \times 2(\text{MOT: no, yes})$ design yielded a total of 12 conditions, and the dependent variables REL, RES, and RTR were measured in each condition.

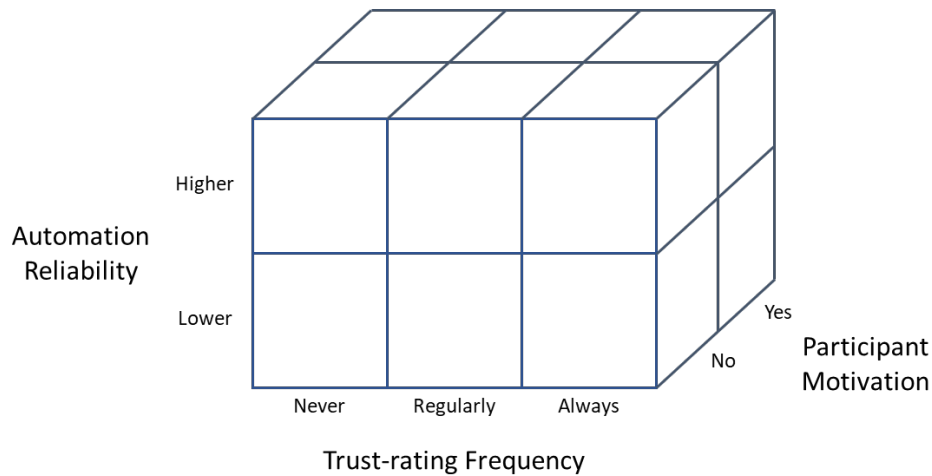


Figure 6. $2 \times 3 \times 2$ Factorial Treatment Design

Since various factors influence automation reliance, this dissertation attempted to control potential intervening variables and isolate the effect of manipulating independent variables. Therefore, a laboratory setting was selected to provide internal validity and ensure similar environments for all participants. Laboratory experiments permit more precise study of relationships between independent and dependent variables by reducing the impact of extraneous factors (Kerlinger & Lee, 2008). Even though automation reliance depends on contextual variables in the real world, laboratory experiments are suitable for exploring underlying cognitive functions in controlled environments (Kramer, 2018; Lyon et al., 2018). Hence, a laboratory environment was appropriate for this initial study of intermediate trust ratings' effect on automation reliance.

1. Experimental Manipulation

TRF was manipulated in three levels between subjects: never, regularly, and always. Participants were randomly assigned to one of three groups: TRF Never, TRF Regularly, and TRF Always. Participants in the TRF Never group never provided trust ratings during the interactions, which served as a control without intermediate trust ratings. Participants in the TRF Always group provided trust ratings after each trial. Even though trust ratings between every decision capture reactions directly after automation errors, trust ratings before every decision are constantly intrusive, time-consuming, and

potentially annoying in extended sequential decision-making. More efficient approaches must be developed for real-world scenarios. Therefore, participants in the TRF Regularly group provided trust ratings after every fifth trial. Trust ratings between every few reliance decisions might have a similar effect to ratings before every decision but could be more efficient and practical.

Participants rated their in-the-moment trust level in the automation on a sliding scale from 0% to 100%, where 0% equals no trust and 100% equals complete trust. A definition of trust was provided to avoid inconsistent participant interpretations of the concept (Lewis et al., 2018). The intermediate trust ratings were not intended to measure trust as a dependent variable. This study used explicit trust ratings to create intermediate judgments and observe the effects on reliance. Therefore, the study used a single-item scale to minimize interruptions during interaction, even though multi-item scales might capture a more comprehensive picture of trust.

AUR was manipulated within subjects by simulating two automated decision aids that all participants interacted with: one with 80% reliability (AUR Lower) and one with 90% reliability (AUR Higher). These two levels were selected between perfect automation (100%) and the critical point of 70%, where decision-makers start to ignore recommendations (see discussion in Section II.D.1).

MOT was manipulated between subjects in two levels: MOT Yes and MOT No. In the MOT Yes condition, participants were told that the three participants with the most correct answers would receive \$20 Amazon gift cards, even though all participants were eligible for the gift cards. Half of the participants in each TRF condition were randomly assigned to each condition of MOT.

2. Measures

REL was calculated for each participant by dividing the number of accepted recommendations by the number of provided recommendations (see Figure 7). The reliance rate ranged from 0 to 1, where 0 means that the decision-maker did not accept any recommendations, and 1 means that the decision-maker accepted all recommendations.

$$\text{Automation reliance} = \frac{\text{Accepted recommendations}}{\text{Provided recommendations}}$$

Figure 7. Automation Reliance

RES was measured for each reliance decision, and an average value was computed for each participant.

RTR scores for the two decision aids enable a comparison between retrospective trust and in-the-moment trust, as well as a comparison of RTR between experimental conditions. To prevent interference between conditions, RTR was measured directly after each interaction was concluded (Vereschak et al., 2021). In contrast to the intermediate trust ratings, a more reliable trust measurement was used for RTR. The intermediate trust assessments were used to create intermediate judgments, while RTR was a dependent variable.

B. PARTICIPANTS

The experiment was conducted between June and September of 2023 at the Naval Postgraduate School (NPS) in Monterey, California, with 120 students, faculty, or staff participants. Participants conducted the experiment individually in a single session. Two participants decided not to rely on recommendations from either decision aid for every decision. After the session, one of them said that they decided to see how well they could do without aid. Given this response it appears that the two participants might have misunderstood the task, therefore their data were excluded from the analysis. Table 2 summarizes participant demographics.

Table 2. Participant Demographics

Participants characteristics (N = 118)		
Age	Mean	36.68
	Median	35
	Standard deviation	8.894
Gender	Male	95
	Female	22
	Prefer not to answer	1
Ethnicity	Asian/Pacific Islander	20
	Black or African American	6
	Hispanic or Latino	10
	Native American or American Indian	1
	Other	3
	White	78
Employment	Military	100
	Civilian	18

C. MATERIALS

This section shows the materials and tasks used in the experiment. First, the laboratory setup is described. Then, the Technology Readiness Index 2.0 scale, which was used to collect participant attitudes toward technology, is presented. Next, the main experimental task is described. Finally, the exit questionnaire is presented.

1. Laboratory Setup

The experiment was conducted in an NPS research laboratory (Glasgow 103) using Presentation, a stimulus delivery and experiment control program developed by Neurobehavioral Systems, Inc. The software ran on a desktop connected to a 21.5” Dell E2214H monitor with 1920 × 1080 pixels resolution, which also was the image resolution. The images were taken from a previous NPS experiment in 2017 at Camp Roberts, CA (Jones, 2017). Participants used a wireless keyboard to provide inputs, which were recorded in the software along with time stamps.

2. Technology Readiness Index 2.0

Participant attitude toward technology was collected using a scale called the Technology Readiness Index 2.0 (TRI 2.0).¹ The scale, included in Appendix A, consist of 10 statements that participants were asked to rate their agreement with on a 5-point bipolar Likert scale (Strongly Disagree, Somewhat Disagree, Neutral, Somewhat Agree, Strongly Agree). Based on the answers, an overall score between 1 and 5 was calculated, with a higher score indicating a more positive attitude toward technology.

3. UAV Identification Task

Participants were tasked to determine the number of UAVs in a series of images with the help of an automated decision aid. The war in Ukraine has demonstrated the importance of UAVs. Participants were told that the study supported the development of automated decision aids, which are helpful in military contexts with fast information flow. The goal was to create a realistic and motivational scenario in which participants understood the importance of the task. Furthermore, the experiment was game-like, and participants could track their scores to increase attention and motivation (Vereschak et al., 2021; Yu et al., 2016). In summary, efforts were made to make participants feel that something was at stake and that their decisions mattered.

Participants were instructed to get as many correct answers as possible. First, they watched an image for a moment (1.5 seconds). Then, they were asked how many UAVs they saw in the image. If the decision aid disagreed, a recommendation was presented to the participants who chose to rely on either their initial answer or the decision aid recommendation. If the participant's initial answer matched the decision aid recommendation, that answer was automatically selected as the final answer. Finally, the correct answer was revealed. This procedure was repeated for 100 images. Figure 8 shows an example of the images that participants watched.

¹ The Technology Readiness Index 2.0 is copyrighted by A. Parasuraman and Rockbridge Associates, Inc., 2014. This scale may be duplicated only with written permission from the authors.



Figure 8. Example Image

The task was designed to require no domain knowledge or task expertise. Furthermore, to have the human decision-makers perform comparably to the automated decision aids, pilot tests were used to calibrate the time each image was shown. Decision-makers might be skewed toward reliance on decision aids if the decision aid performance is significantly higher than the decision-maker's (Zhang et al., 2020).

The experiment sessions were divided into two blocks, and the participants were told they were using a different decision aid for each block. Two decision aids with distinctive reliability were simulated. The lower reliability decision aid made correct recommendations for 80% of the images, and the higher reliability decision aid made correct recommendations for 90%. All participants interacted with both decision aids, and the order of the two decision aids was counterbalanced.

Each decision aid had a fixed series of recommendations, containing some errors at random occurrences. The decision aids were used for 50 images each, shown in the same order for all participants, with incorrect recommendations for the same images. The decision aids gave correct recommendations for images in which an error would be obvious because obvious automation errors tend to significantly hamper trust and quickly

lead to disuse (Dzindolet et al., 2003; Merritt et al., 2013). To further avoid participant detection of obvious automation errors, the incorrect recommendations were only off by one UAV. For example, if the correct number of UAVs in an image was three, an incorrect recommendation was two or four.

The two decision aids were assigned different names to help participants distinguish between them. To account for potential biases, *River* was always the first name and *Scout* the second, even though the order in which participants interacted with the higher and lower reliability decision aids was counterbalanced.

4. Exit Questionnaire

The experimental setup held situational and automation characteristics consistent throughout the experiment but could not account for human characteristics. Therefore, self-report measures in the exit questionnaire (see Appendix B) collected several human characteristics that previous studies have shown to impact automation reliance (e.g., demographics, trust propensity, and experience with similar technology). Those factors could then be assessed in data analysis.

Furthermore, self-confidence and the ability to perform a task without aid might affect human reliance on decision aids (Madhavan & Wiegmann, 2007), so participants were asked in the exit questionnaire how confident they were in solving the task without a decision aid. Also, since the participants provided their initial answer before receiving any decision aid recommendation, performance without aid could be recorded. To avoid a significant variance in performance without any aid, an experimental task was chosen that required no domain knowledge. However, previous experience with image recognition software or automated decision aids might affect reliance. Therefore, participants were also asked about such experiences in the exit questionnaire.

D. PROCEDURES

This section describes the chronological process of the experimental sessions. Each session included eight steps: informed consent form, attitudes toward technology,

introduction, practice, first decision aid, second decision aid, exit questionnaire, and debrief.

1. Informed Consent Form

At the start of each session, the participant was provided with an informed consent form, which explained the research purpose and procedures. Participants were allowed to read the consent form and ask questions before signing it. All participants were offered a copy of the informed consent form.

2. Attitudes Toward Technology

After the consent form was signed, participants were seated at the computer on which the experiment was conducted. Participants completed the TRI 2.0 scale before interaction with the decision aids started to prevent the experimental task from influencing the answers. The scale questions were presented in random order as per the usage guidelines.

3. Introduction

Before the main task started, participants were provided with an introduction to the study. They were told that NPS was developing a decision support system to help identify UAVs in images. Participants were also told that human decision-makers need appropriate trust in automated decision aids. To prevent confusion about the construct, a definition of *trust* was provided: “The attitude that one agent will achieve another agent’s goal in a situation where imperfect knowledge is given with uncertainty and vulnerability” (Lee & See, 2004, p. 51). Then, participants were told that two different decision aids had been tested in isolation and were now being evaluated with a human decision-maker. Finally, the introduction described the main task and instructed participants to get as many correct answers as possible.

4. Practice

Participants conducted three practice trials to get familiarized with the experimental task and the user interface.

5. First Decision Aid

Participants were randomly selected to interact with either the lower or the higher reliability decision aid first. Each interaction lasted for 50 trials. When the interaction was completed, the participants were asked to provide an overall trust rating of the decision aid, defined as the RTR. Participants were also asked to assess the overall reliability of the decision aid on a sliding scale from 0% to 100%.

6. Second Decision Aid

Next, participants interacted with the second decision aid for another 50 trials. RTR and a reliability assessment were again collected directly after the interaction ended.

7. Exit Questionnaire

When both interactions were completed, participants filled out an exit questionnaire where demographics, previous experience, and subjective ratings of motivation and engagement during the experiment were collected. The exit questionnaire also included an attention check by asking participants to describe their assigned main task. The participant data were included in the analysis if the response was satisfactory.

8. Debrief

Experimental conditions were not revealed to participants until a debrief to avoid bias in participant behavior. After each session, the researcher revealed that NPS was not developing a decision support system and that the deception was used to create a realistic and motivational scenario. Participants were asked if they still consented to participate in the study after the extent and purpose of the deception were explained.

E. SUMMARY

A human subject laboratory experiment was conducted to test the research hypotheses. A $2 \times 3 \times 2$ factorial treatment design was employed in which automation reliability was manipulated within subjects, whereas trust rating frequency and participant motivation were manipulated between subjects. Automation reliance, response time, and retrospective trust were measured. Participants in the experiment included 118

NPS students, faculty, and staff. Participants were tasked to correctly identify the number of UAVs in a series of images supported by two automated decision aids. Attitudes towards technology were collected at the start of the sessions, and participants completed an exit questionnaire at the end.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. RESULTS

Data analysis was performed in two steps. First, the main analysis was based on statistical hypothesis testing. Analysis of variance (ANOVA) and *t* tests were used to test for statistically significant differences in dependent variables between treatment groups. Then, exploratory data analysis was conducted to further examine the data. The exploratory analysis revealed additional patterns in the data and helped to interpret results from the main analysis. The analysis was conducted on IBM SPSS and ROV BizStats (Real Options Valuation, n.d.).

A. MAIN ANALYSIS

All participants interacted with both the lower reliability (AUR Lower) and the higher reliability (AUR Higher) decision aids. There were three levels of TRF and two levels of MOT. Hence, there were six subgroups illustrated in Figure 9. Participants were randomly assigned to one of the subgroups.

		Trust Rating Frequency (TRF)		
		Never	Regularly	Always
Participant Motivation (MOT)	No	TRF:never MOT:no	TRF:regularly MOT:no	TRF:always MOT:no
	Yes	TRF:never MOT:yes	TRF:regularly MOT:yes	TRF:always MOT:yes

Figure 9. Participant Subgroups

First, descriptive statistics were generated and interpreted. Histograms and box and whisker plots helped in visually interpreting the data. Both the Shapiro-Wilk test and the Lilliefors test were used to test the data for normality. Moreover, Box's test for equality of covariance and Levene's test for homogeneity of variances were used to verify that the assumptions of parametric tests were met. Finally, ANOVA tests were

used to test for statistically significant differences between treatment groups, and pairwise *t* tests were used to further explore the differences.

The main analysis was based on three ANOVA tests, one for each dependent variable. Three-way mixed ANOVA tests (between-between-within) were performed to simultaneously test the effect of all three independent variables and their interactions. This section describes the statistical tests for each dependent variable.

1. Automation Reliance

ANOVA tests assume data are approximately normally distributed. Numerical methods were used to assess if the data were normally distributed for each combination of the independent variable levels. Both the Shapiro–Wilk and Lilliefors tests for normality indicated that the REL scores were statistically normal ($p > .05$). These tests are two of the most powerful tests for normality, especially on small sample sizes (Mendes & Pala, 2003).

Basic descriptive statistics on REL in each subgroup are presented in Table 3 for the lower reliability decision aid and Table 4 for the higher reliability decision aid. Figure 10 visualizes the REL scores in all subgroups, using box and whisker plots.

Table 3. REL Lower Reliability Decision Aid

MOT	Statistic	TRF Never	TRF Regularly	TRF Always
No	N	20	20	20
	Mean	.3610	.4456	.4385
	Median	.3810	.4440	.4310
	Std dev	.1702	.1583	.1439
	Skewness	.049	.105	.796
	Kurtosis	-1.478	-.105	1.217
	95% confidence interval for Mean	.2813-.4407	.3715-.5197	.3711-.5058
Yes	N	19	19	20
	Mean	.3258	.3924	.4498
	Median	.3080	.4210	.4855
	Std dev	.1556	.2002	.2211
	Skewness	.889	-.376	-.195
	Kurtosis	.342	-.805	-.366
	95% confidence interval for Mean	.2508-.4008	.2959-.4890	.3463-.5533

Table 4. REL Higher Reliability Decision Aid

MOT	Statistic	TRF Never	TRF Regularly	TRF Always
No	N	20	20	20
	Mean	.5286	.5536	.5624
	Median	.4620	.5810	.5535
	Std dev	.2618	.1762	.1476
	Skewness	.280	-.512	.684
	Kurtosis	-1.343	-.039	.767
	95% confidence interval for Mean	.4056-.6516	.4711-.6361	.4933-.6314
Yes	N	19	19	20
	Mean	.4364	.5888	.6434
	Median	.4440	.6470	.6290
	Std dev	.2291	.2289	.1879
	Skewness	.210	-.086	.264
	Kurtosis	-.312	-1.320	-.904
	95% confidence interval for Mean	.3260-.5469	.4785-.6992	.5554-.7313

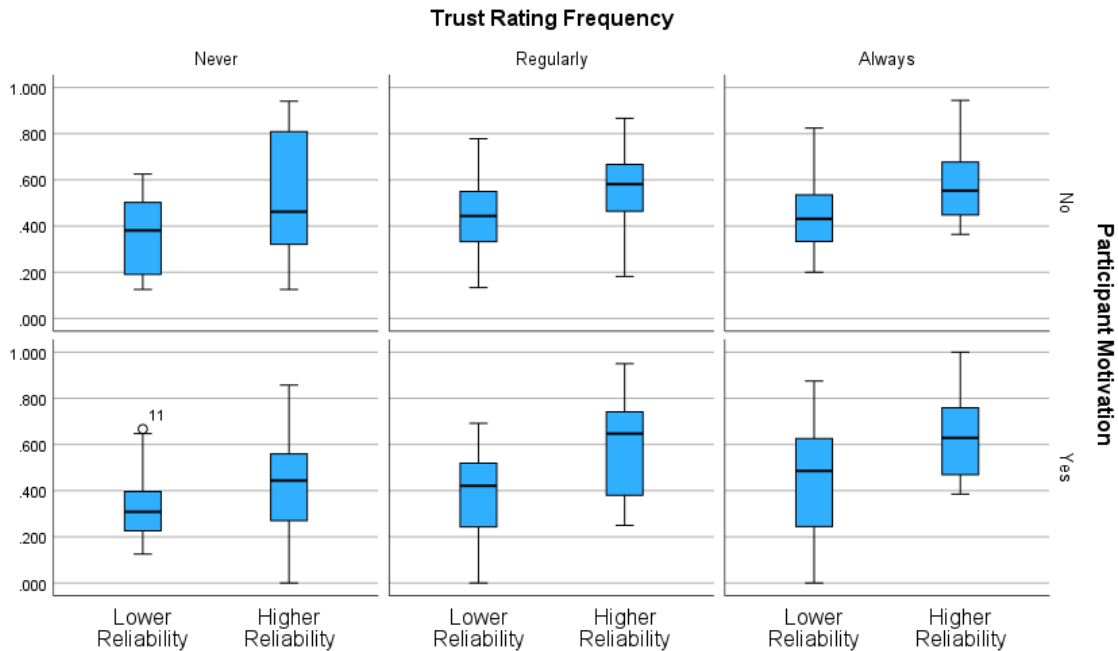


Figure 10. REL by Subgroup

Assessment of boxplots revealed no extreme outliers in the data. Since the within-factor only had two levels, the assumption of sphericity was automatically met. Box's test for equality of covariance matrices indicated that covariance matrices of the dependent variables were approximately equal across groups ($p = .243$). Levene's test for homogeneity of variances did not show a difference in variance between the subgroups, for either the lower reliability decision aid ($p = .384$) or the higher reliability decision aid ($p = .093$).

The ANOVA test revealed that there was a main effect of TRF on REL, $F(2,112) = 4.367$, $p = .015$, $\eta_p^2 = .072$. Follow-up t tests (independent samples) showed that REL was lower in the TRF Never condition ($M = .414$, $SD = .18634$, $CI_{95} [.3534, .4742]$) than in the TRF Regularly condition ($M = .495$, $SD = .166$, $CI_{95} [.4413, .5491]$), $t(76) = -2.037$, $p = .045$, $d = -.461$, and in the TRF Always condition ($M = .524$, $SD = .161$, $CI_{95} [.4719, .5750]$), $t(77) = -2.801$, $p = .006$, $d = -.630$. There was no statistically significant difference in REL between the TRF Regularly and TRF Always conditions ($p = .446$).

Additionally, there was a main effect of AUR on REL, $F(1,112) = 85.504$, $p < .001$, $\eta_p^2 = .433$. Follow-up t tests (paired samples) showed that REL was higher on the higher reliability aid ($M = .553$, $SD = .214$, $CI_{95} [.514, .592]$) than on the lower ($M = .403$, $SD = .179$, $CI_{95} [.370, .436]$), $t(117) = 9.235$, $p < .001$, $d = .850$.

Finally, MOT had no statistically significant main effect on REL ($p = .781$), and there were no statistically significant interactions between TRF*MOT, TRF*AUR, MOT*AUR, or TRF*MOT*AUR ($p > .05$).

2. Retrospective Trust

Basic descriptive statistics on RTR per subgroup are presented in Tables 5 and 6. Figure 11 visualizes RTR in all subgroups using box and whisker plots. The RTR scores were not normally distributed in all subgroups. Since the scores were strongly negatively skewed, a reflect and logarithmic transformation was used to improve normality (Osborne, 2008). Each score was multiplied by -1, and a constant was added to get a minimum score of 1. Since the range of the values was not extreme, log base e was used rather than log base 10 to preserve data resolution (Cleveland, 1984). After transformation, the RTR scores were statistically normal ($p > .05$).

Table 5. RTR Lower Reliability Decision Aid

MOT	Statistic	TRF Never	TRF Regularly	TRF Always
No	N	20	20	20
	Mean	3.850	4.883	4.533
	Median	3.834	5.000	5.000
	Std dev	1.254	.9568	1.535
	Skewness	-.578	-.484	-1.206
	Kurtosis	.760	.109	.496
	95% confidence interval for Mean	3.263–4.437	4.436–5.331	3.815–5.252
Yes	N	19	19	20
	Mean	3.544	3.947	4.517
	Median	3.667	4.000	4.834
	Std dev	1.344	1.675	1.441
	Skewness	-.450	-.560	-.758
	Kurtosis	-1.010	-.798	.772
	95% confidence interval for Mean	2.896–4.191	3.140–4.755	3.842–5.191

Table 6. RTR Higher Reliability Decision Aid

MOT	Statistic	TRF Never	TRF Regularly	TRF Always
No	N	20	20	20
	Mean	4.783	5.100	5.583
	Median	4.667	5.667	6.000
	Std dev	1.294	1.125	1.070
	Skewness	-.372	-.517	-2.179
	Kurtosis	.466	-.732	6.218
	95% confidence interval for Mean	4.178–5.389	4.574–5.626	5.083–6.084
Yes	N	19	19	20
	Mean	4.719	5.298	5.283
	Median	5.000	6.000	5.330
	Std dev	1.297	1.328	1.044
	Skewness	-1.461	-.962	-.724
	Kurtosis	1.735	.571	.699
	95% confidence interval for Mean	4.094–5.344	4.658–5.939	4.795–5.772

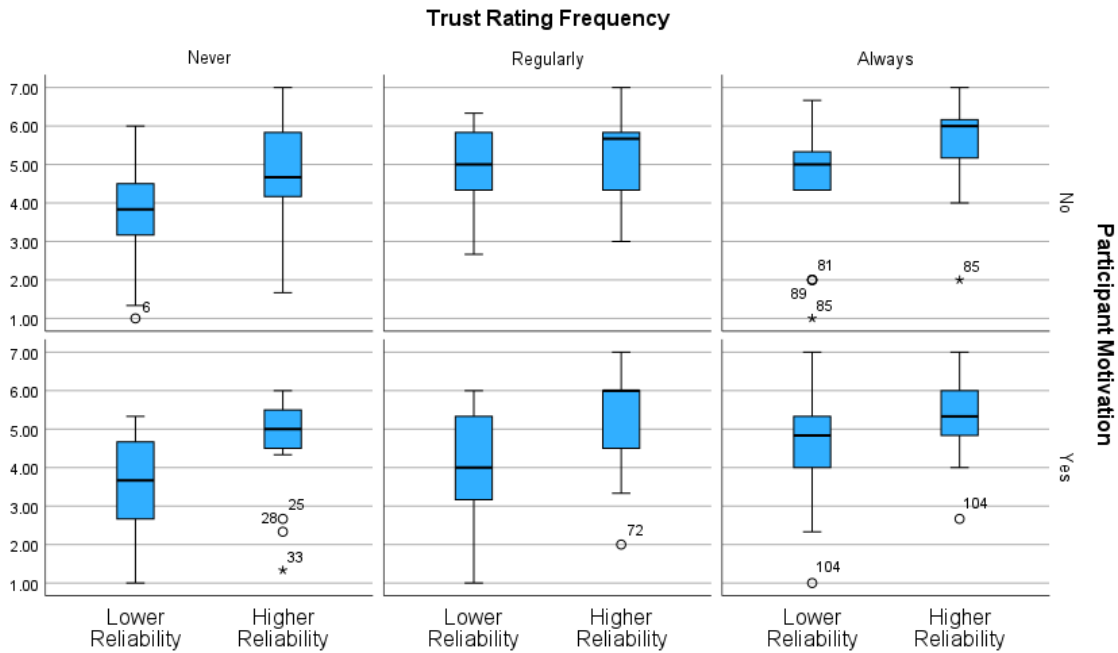


Figure 11. RTR by Subgroup

For RTR, the ANOVA test was run on the reflect and logarithmic (log base e) transformed data. Mean values, standard deviations, and confidence intervals are reported on the untransformed data, but the analysis was done on the transformed data. Analysis of boxplots revealed no extreme outliers. Again, the assumption of sphericity was automatically met since the within-factor only had two levels. Box's test for equality of covariance matrices indicated that covariance matrices of the dependent variables were approximately equal across groups ($p = .530$). Levene's test for homogeneity of variances showed no statistically significant difference in variance between the subgroups for either the lower reliability decision aid ($p = .645$) or the higher ($p = .770$).

There was a main effect of TRF on RTR, $F(2,112) = 5.106$, $p = .008$, $\eta_p^2 = .084$. Follow-up t tests (independent samples) showed that RTR were lower in the TRF Never condition ($M = 4.227$, $SD = 1.138$, $CI_{95}[3.858, 4.595]$) than in the TRF Regularly condition ($M = 4.812$, $SD = 1.197$, $CI_{95}[4.424, 5.200]$), $t(76) = -2.214$, $p = .030$, $d = -.501$ and in the TRF Always condition ($M = 4.979$, $SD = 1.166$, $CI_{95}[4.606, 5.352]$),

$t(77) = -2.903, p = .005, d = -.653$. There was no statistically significant difference in RTR between the TRF Regularly and TRF Always conditions ($p = .531$).

Additionally, there was a main effect of AUR on RTR, $F(1,112) = 84.840, p < .001, \eta_p^2 = .431$. Follow-up t tests (paired samples) showed that RTR was higher for the higher reliability aid ($M = 5.130, SD = 1.209, CI_{95}[4.910, 5.350]$) than for the lower ($M = 4.220, SD = 1.432, CI_{95}[3.959, 4.481]$), $t(117) = 8.841, p < .001, d = .814$.

However, there was no statistically significant main effect of MOT on RTR, $F(1,112) = .818, p = .368$. This result indicates that competing for gift cards did not affect RTR.

The ANOVA test revealed a statistically significant three-way interaction between TRF*MOT*AUR, $F(2,112) = 4.443, p = .014, \eta_p^2 = .074$. Specifically, this interaction was found in the MOT No condition, where RTR for the lower reliability decision aid in the TRF Regularly condition were higher than in any other condition. This anomaly caused some divergence from the previously reported main effects.

In comparison of RTR between AUR Lower and AUR Higher, in the MOT No condition, while there was no difference in RTR between AUR conditions in the TRF Regularly condition ($p = .103$), there was a significant difference in RTR between AUR in both the TRF Never ($p = .005$) and TRF Always conditions ($p < .001$). Additionally, in the MOT Yes condition, there was a significant difference in RTR between AUR lower and AUR Higher in all TRF conditions: TRF Never ($p < .001$), TRF Regularly ($p < .001$), and TRF Always ($p = .003$). Figure 12 illustrates how the high RTR scores in the AUR Lower–TRF Regularly–MOT No condition violates the main effects.

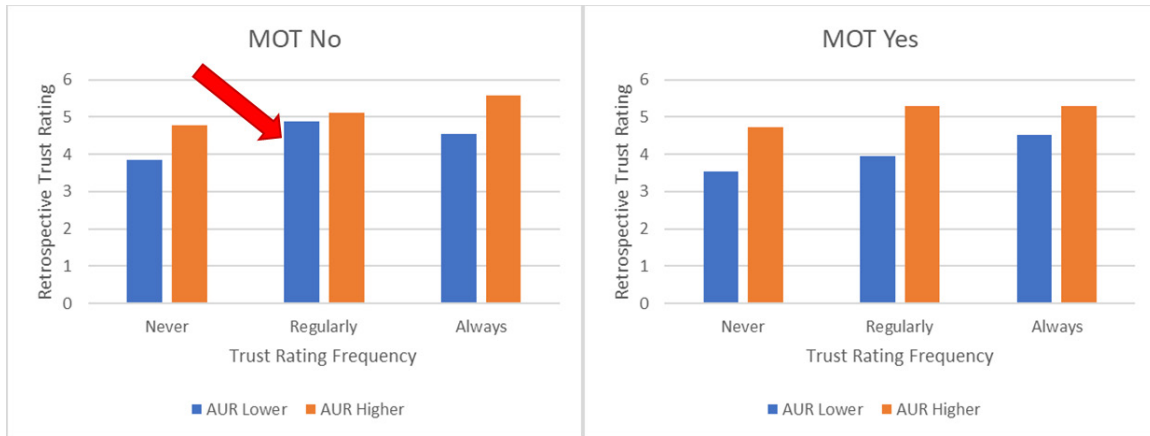


Figure 12. Statistical Interaction by MOT Condition

Furthermore, in the comparison in RTR between TRF Never and TRF Regularly, while there was a significant difference for the lower reliability decision aid in the MOT No condition ($p = .006$), there was neither a difference in RTR between TRF Never and TRF Regularly for the lower reliability decision aid in the MOT Yes condition ($p = .265$), nor for the higher reliability decision aid in either MOT condition: MOT No ($p = .518$) or MOT Yes ($p = .087$).

Finally, in the comparison in RTR between TRF Never and TRF Always, while there was a significant difference for the lower reliability decision aid in the MOT Yes condition ($p = .023$) and for the higher reliability decision aid in the MOT No condition ($p = .048$), there was neither a difference for the lower reliability decision aid in the MOT No condition ($p = .066$), nor for the higher reliability decision aid in the MOT Yes condition ($p = .108$). Figure 13 illustrates differences between TRF conditions for both decision aids and MOT conditions.

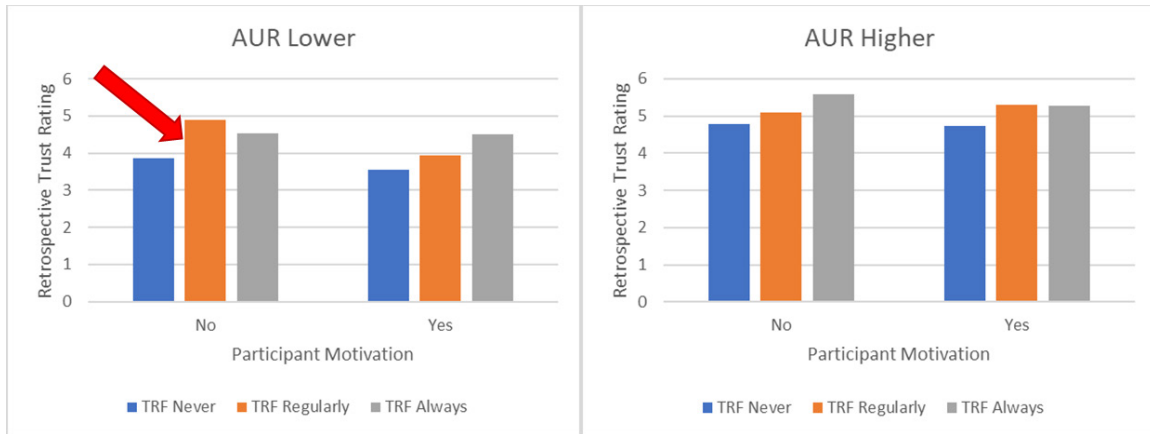


Figure 13. Statistical Interaction by AUR Condition

3. Response Time

Basic descriptive statistics on RES per subgroup are presented in Tables 7 and 8. Figure 14 visualizes the RES scores in all subgroups using box and whisker plots. The RES scores were not normally distributed but strongly positively skewed. A square root transformation was not powerful enough to improve normality. Instead, a logarithmic transformation with log base e made RES scores statistically normal ($p > .05$), except for one subgroup. In the TRF Always–MOT No condition for the higher reliability decision aid, the Shapiro–Wilk test score was $p = .048$, just under the $p = .05$ threshold. The data were assumed close enough to normally distributed, since ANOVA is robust to violations of normality, especially under homogeneity of variance (Blanca et al., 2017, 2023).

Table 7. RES Lower Reliability Decision Aid

MOT	Statistic	TRF Never	TRF Regularly	TRF Always
No	N	20	20	20
	Mean	3452	4005	4002
	Median	3264	3737	3814
	Std dev	1166	1472	1272
	Skewness	.333	1.222	.431
	Kurtosis	-.856	1.778	-1.119
	95% confidence interval for Mean	2906–3998	3316–4694	3406–4597
Yes	N	19	19	20
	Mean	3183	3648	3907
	Median	3248	3449	3414
	Std dev	843.2	1477	1489
	Skewness	.364	1.862	.669
	Kurtosis	-.383	4.929	-.177
	95% confidence interval for Mean	2777–3589	2936–4360	3210–4604

Table 8. RES Higher Reliability Decision Aid

MOT	Statistic	TRF Never	TRF Regularly	TRF Always
No	N	20	20	20
	Mean	3594	4092	4107
	Median	3377	3998	3648
	Std dev	1440	1471	1659
	Skewness	1.633	.726	.187
	Kurtosis	2.912	.218	-1.638
	95% confidence interval for Mean	2920–4268	3404–4780	3331–4883
Yes	N	19	19	20
	Mean	3491	3850	3816
	Median	3436	3437	3713
	Std dev	1067	1822	1593
	Skewness	.215	1.381	1.348
	Kurtosis	-.372	1.427	1.949
	95% confidence interval for Mean	2977–4006	2972–4728	3071–4562

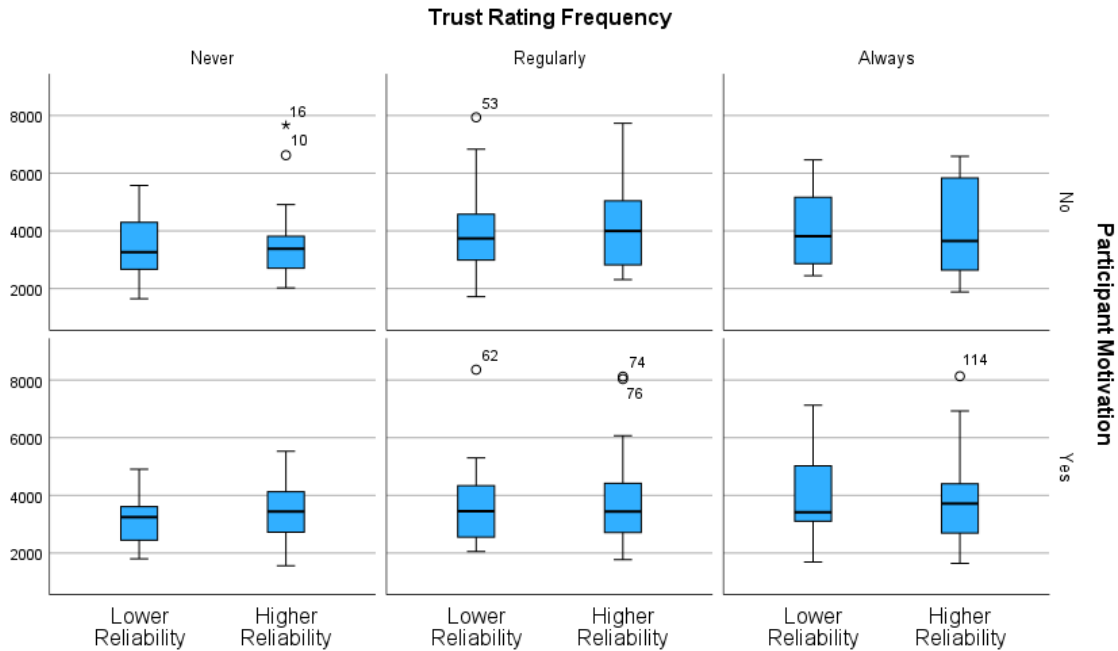


Figure 14. RES by Subgroup

For RES, the ANOVA test was run on the logarithmic (log base e) transformed data. Analysis of boxplots revealed no extreme outliers. Again, the assumption of sphericity was automatically met since the within-factor only had two levels. Box's test for equality of covariance matrices indicated that covariance matrices of the dependent variables were approximately equal across groups ($p = .927$). Levene's test for homogeneity of variances showed no statistically significant difference in variance between the subgroups for either the lower reliability decision aid ($p = .817$) or the higher ($p = .468$).

There were no main effects of either AUR ($p = .631$), TRF ($p = .099$), or MOT ($p = .233$). Furthermore, there were no statistically significant interactions between TRF*MOT, TRF*AUR, MOT*AUR, or TRF*MOT*AUR ($p > .05$).

B. EXPLORATORY ANALYSIS

This section describes exploratory data analysis (EDA) used to further examine the data. EDA complements confirmatory data analysis (CDA), such as statistical hypothesis testing (Tukey, 1980). For example, EDA can help “discovering patterns to

foster hypotheses development and refinement [and] interpret the results of CDA and reveal unexpected or misleading patterns in the data” (Behrens, 1997, p. 131).

1. Trust Ratings

This section compares trust ratings between the TRF Regularly and TRF Always conditions. Figure 15 illustrates average trust ratings in both TRF conditions for each decision aid. The red vertical lines mark trials where the decision aids made incorrect recommendations. As expected, trust ratings varied with automation performance. The TRF Always curve shows clear reactions to automation mistakes. Trust ratings drop faster after mistakes than they grow with successes, which aligns with previous research (Yang et al., 2023; Yu et al., 2017). Since trust ratings are sampled more seldom in the TRF Regularly condition, that curve does not have the same fidelity as the TRF Always curve. However, it roughly follows the same pattern. Interestingly, for both decision aids, the trust ratings in the TRF Always condition were consistently higher than in the TRF Regularly condition.



Figure 15. Intermediate Trust Ratings by Decision Aid

An average trust rating score was calculated per participant and decision aid to examine if there was a statistically significant difference in trust ratings between the two TRF conditions. The data had a strong positive kurtosis and a strong negative skewness because two participants rated both decision aids consistently lower than the other participants (see Table 9 and Figure 16). Removing the two outliers that differed considerably from the primary trend in the data fixed the issues with skewness and kurtosis. Keeping the outliers would draw the model towards them, but predict them poorly (Field, 2017). Basic descriptive statistics on trust ratings with the two outliers removed are presented in Table 10.

Table 9. Intermediate Trust Ratings by Subgroup

AUR	TRF	N	Skewness	Kurtosis
Lower	Regularly	40	-1.082	.622
	Always	40	-1.272	1.830
Higher	Regularly	40	-1.289	1.646
	Always	40	-1.274	1.804

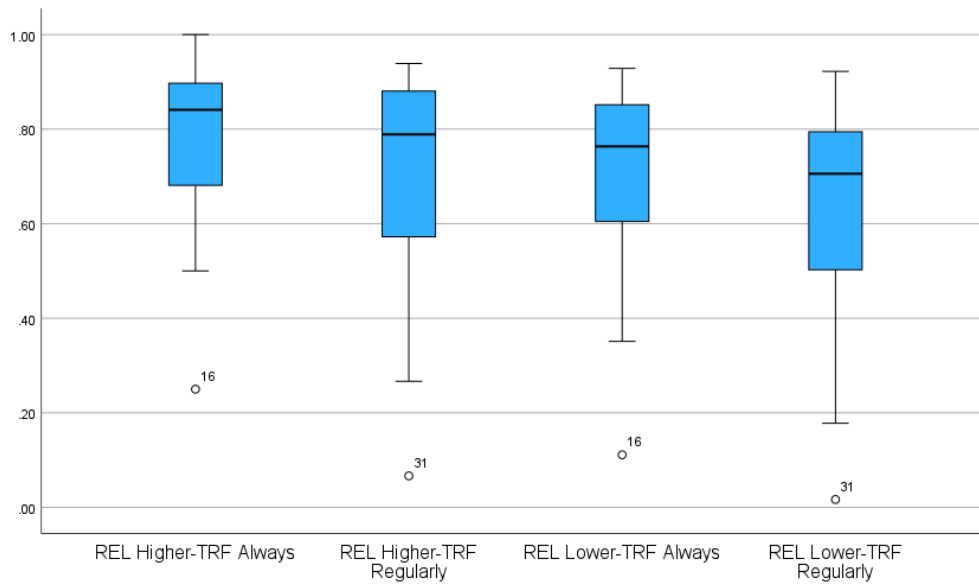


Figure 16. Intermediate Trust Ratings Outliers

Table 10. Intermediate Trust Ratings Descriptive Statistics

AUR	TRF Regularly	TRF Always
Lower	N: 39	N: 39
	Mean: .6514	Mean: .7333
	Median: .7056	Median: .7653
	Std dev: .1957	Std dev: .1558
	Range: .74	Range: .58
	Skewness: -.910	Skewness: -.782
	Kurtosis: .048	Kurtosis: -.200
	95% confidence interval for Mean: .5880–.7148	95% confidence interval for Mean: .6828–.7838
	Shapiro–Wilk: .004	Shapiro–Wilk: .014
Higher	N: 39	N: 39
	Mean: .7362	Mean: .7990
	Median: .7944	Median: .8449
	Std dev: .1716	Std dev: .1365
	Range: .67	Range: .50
	Skewness: -.872	Skewness: -.765
	Kurtosis: -.140	Kurtosis: -.381
	95% confidence interval for Mean: .6806–.7918	95% confidence interval for Mean: .7547–.8432
	Shapiro–Wilk: .001	Shapiro–Wilk: .009

The data in each subgroup were approximately normally distributed, according to a Shapiro–Wilk test ($p < .05$). Independent sample t tests showed that for the higher reliability aid, the trust ratings in the TRF Always condition ($M: .799$, $SD: .136$, $CI_{95} [.755, .843]$) were statistically significantly higher than in the TRF Regularly condition ($M: .736$, $SD: .172$, $CI_{95} [.681, .792]$), $t(76) = 1.788$, $p = .039$, $d = .405$. Similarly, for the lower reliability decision aid, trust ratings in the TRF Always condition ($M: .733$, $SD: .156$, $CI_{95} [.683, .784]$) were statistically significantly higher than in the TRF Regularly condition ($M: .651$, $SD: .196$, $CI_{95} [.588, .715]$), $t(76) = 2.045$, $p = .022$, $d = .463$. These results indicate that participants who rated their trust more often provided higher trust ratings.

2. Types of Reliance

The outcome of each reliance decision might be placed in one of four possible categories. Participants could either accept or reject automation recommendations, and the recommendations could either be correct or incorrect. This dissertation defines appropriate reliance as accepting automation recommendations when the automation is correct and rejecting recommendations when the automation is incorrect. Conversely, accepting recommendations when the automation is incorrect and rejecting recommendations when the automation is correct is considered inappropriate reliance. The four categories are illustrated in Table 11.

Table 11. Categorization of Responses in Reliance Decisions

	Accept Recommendation	Reject Recommendation
Automation Correct	Appropriate Reliance	Inappropriate Reliance
Automation Incorrect	Inappropriate Reliance	Appropriate Reliance

Reliance decisions for each decision aid were compared using the structure of Table 11. Recall that reliance decisions were made any time a participant's answer differed from the automation recommendation. A total of 1,730 reliance decisions were made on the lower reliability decision aid, and 1,749 were made on the higher reliability decision aid. The ratios of each category of responses are presented in Table 12. Appropriate reliance is shown in green, and inappropriate reliance is shown in red. Summarized, 56% of the reliance decisions on the lower reliability decision aid and 53% on the higher are categorized as appropriate reliance. Even though those two numbers are similar, there are differences between the decision aids. Most appropriate reliance on the lower reliability decision aid comes from correctly rejecting recommendations. In contrast, most appropriate reliance on the higher reliability decision aid comes from correctly accepting recommendations. This discrepancy is intuitive, since the higher reliability decision aid performed better than the lower (i.e., its recommendations were correct more often).

Table 12. Reliance Decisions by Decision Aid

	Accept Recommendation	Reject Recommendation
Automation Correct	Lower Reliability: .19 Higher Reliability: .40	Lower Reliability: .22 Higher Reliability: .30
Automation Incorrect	Lower Reliability: .22 Higher Reliability: .17	Lower Reliability: .37 Higher Reliability: .13

The categorization of decisions was compared between the TRF conditions by breaking down reliance decisions per condition. The ratios of each category of reliance decisions are presented in Table 13. Appropriate reliance is shown in green, and inappropriate reliance is shown in red. For the lower reliability decision aid, appropriate reliance was 55% at TRF Never, 58% at TRF Regularly, and 55% at TRF Always. For the higher reliability decision aid, appropriate reliance was 52% at TRF Never, 54% at TRF Regularly, and 53% at TRF Always. Even though the amount of appropriate reliance was similar between the groups, intermediate trust ratings moved cases of appropriate reliance from correctly rejecting recommendations to correctly accepting recommendations. Similarly, inappropriate reliance moved from incorrectly rejecting recommendations to incorrectly accepting recommendations. These findings further support previous results from hypothesis testing that trust ratings increase automation reliance. However, significantly, intermediate trust ratings do not necessarily increase appropriate reliance.

Table 13. Reliance Decisions by TRF Condition

	Accept Recommendation			Reject Recommendation		
Automation Correct		Lower	Higher		Lower	Higher
	TRF Never	.15	.37	TRF Never	.26	.35
	TRF Regularly	.20	.41	TRF Regularly	.19	.28
	TRF Always	.22	.42	TRF Always	.21	.28
Automation Incorrect		Lower	Higher		Lower	Higher
	TRF Never	.20	.13	TRF Never	.40	.15
	TRF Regularly	.22	.18	TRF Regularly	.38	.13
	TRF Always	.24	.19	TRF Always	.33	.11

3. Accumulative Evidence of Reliance

This section explores how reliance behavior evolved during the interactions. Accumulating evidence over time is an emerging approach to understanding human decision-making (Markkula et al., 2021). For example, drift-diffusion models can use sequential sampling to model sequential effects of stimuli over trials (Ratcliff et al., 2016). In this dissertation, to accumulate reliance over repeated trials, starting at a value of 0, for each trial, a value of 1 was added if a participant relied on the automation recommendation, or a value of 1 was subtracted if a participant relied on their own answer. No changes were made to the accumulated value if there was no reliance decision (i.e., if the automation and the participant answers agreed). An accumulated value higher than 0 means that the participants relied more often on the automation than themselves, and vice versa. An average value was calculated across participants for each trial for both decision aids (lower and higher reliability). Figure 17 illustrates the average accumulated reliance over trials.

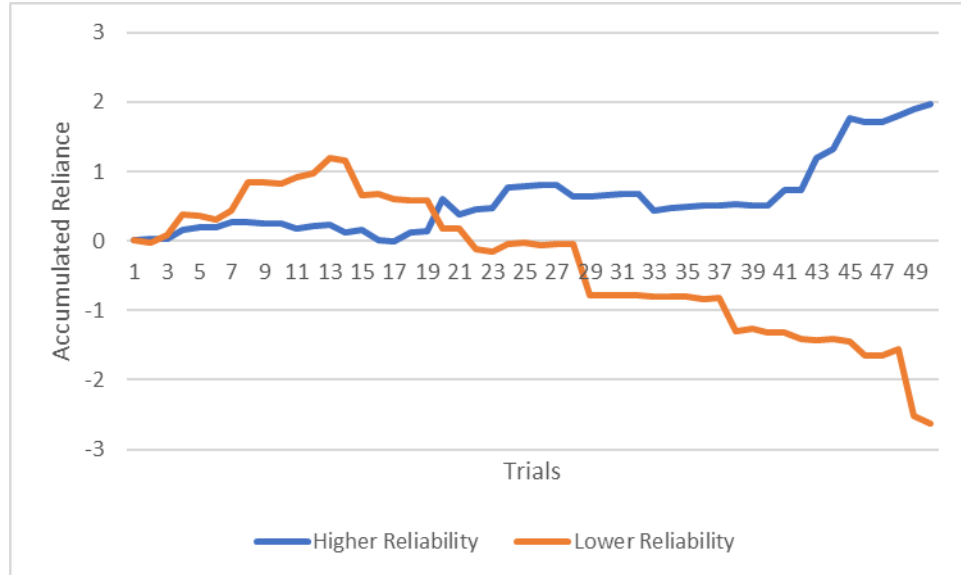


Figure 17. Accumulated Reliance

As seen in Figure 17, the curve for the higher reliability decision aid trends upward, and the curve for the lower reliability decision aid trends downward. In other

words, participants relied more on the decision aid than themselves when interacting with the higher reliability decision aid. In contrast, participants relied more on themselves than the decision aid when interacting with the lower reliability decision aid.

4. Correlations

This section examines correlations between variables to discover general trends in the data across all subgroups. Average scores were calculated across the two decision aids for the three dependent variables. Each of the average scores were correlated with the following variables:

- **Age:** Participant age provided in the exit questionnaire.
- **Attitudes:** Attitudes toward technology collected with the TRI 2.0 scale.
- **Self-performance:** The number of correct participant answers before receiving automation recommendations.
- **Familiarity DA:** Familiarity with decision aids. Subjective self-report provided on a 1–7 Likert scale in the exit questionnaire.
- **Familiarity IR:** Familiarity with image recognition software. Subjective self-report provided on a 1–7 Likert scale in the exit questionnaire.
- **Self-confidence:** The participant’s confidence in completing the task without a decision aid. Subjective self-report provided on a 1–7 Likert scale in the exit questionnaire.
- **Team-confidence:** How confident the participant was in completing the task with the decision aids. Subjective self-report provided on a 1–7 Likert scale in the exit questionnaire.
- **Motivated:** How motivated the participant was to participate in the experiment. Subjective self-report provided on a 1–7 Likert scale in the exit questionnaire.

- **Engaged:** How engaged the participant was during the experiment. Subjective self-report provided on a 1–7 Likert scale in the exit questionnaire.

Several assumptions about effect size, desired power, significance level, and required sample sizes were made. This dissertation follows Cohen’s (1988) suggestions that a correlation coefficient $r = .10$ represents a small effect size, $r = .30$ represents a medium effect size, and $r = .50$ represents a large effect size.

Since most variables did not meet the assumptions of normality and linearity, the nonparametric Spearman’s rank order correlation was used. In addition to not requiring the assumptions of parametric methods, Spearman’s correlation coefficient reduces the effect of outliers (Field, 2017). Tables 14 through 16 present Spearman’s correlation coefficients r_s and p values for each dependent variable. Significant medium correlations detected, highlighted in green, are discussed in more detail. Scatterplots and distributions are included in Appendix C.

Table 14. Correlations Average REL

Variable	r_s	p value (2-tailed)
Age	.379	<.001
Attitudes		.644
Self-performance	-.382	<.001
Familiarity DA	.219	.017
Familiarity IR		.128
Self-confidence	-.376	<.001
Team-confidence		.745
Motivated		.106
Engaged		.054

Table 15. Correlations Average RTR

Variable	r_s	p value (2-tailed)
Age		.308
Attitudes		.862
Self-performance	-.194	.036
Familiarity DA	.201	.029
Familiarity IR	.257	.005
Self-confidence		.335
Team-confidence	.466	< .001
Motivated		.678
Engaged		.580

Table 16. Correlations Average RES

Variable	r_s	p value (2-tailed)
Age	.438	< .001
Attitudes		.425
Self-performance		.556
Familiarity DA		.076
Familiarity IR		.277
Self-confidence		.676
Team-confidence		.791
Motivated		.183
Engaged		.077

a. Average Reliance

The average reliance score was calculated by taking the average value of REL across the two decision aids for each participant. The histogram in Figure 18 shows the shape of the data distribution. The average REL scores were approximately normally distributed, according to a Shapiro–Wilk test ($p = .319$), with a skewness of .159 and a kurtosis of -.590.

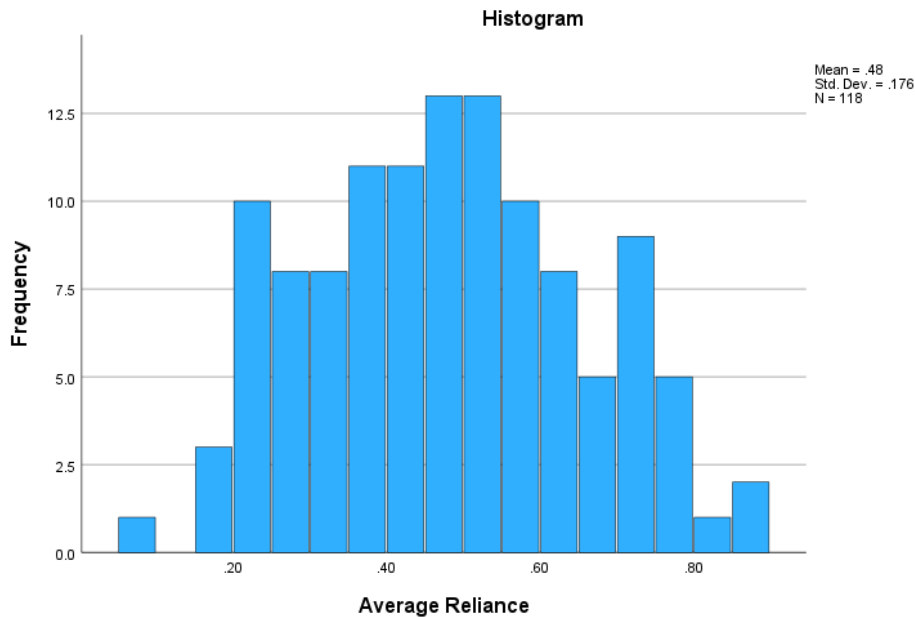


Figure 18. Data Distribution Average REL

(1) Age and Self-Performance

The variable age was not normally distributed. Five outliers contributed to a skewness of 1.457 and a kurtosis of 3.004. Self-performance was also not normally distributed. Outliers contributed to a skewness of - 1.023 and a kurtosis of 1.497. Since the original data points are of research interest, outliers were not removed, and data were not transformed to improve normality. Instead, the nonparametric Spearman's correlation coefficient was used. Inspection of scatterplots indicated monotonic relationships both between age and average REL and between self-performance and average REL.

There was a statistically significant medium positive correlation between age and average REL, $r_s(116) = .379$, $p < .001$, with a 95% percentile bias-corrected and accelerated (BCa) bootstrap confidence interval of (.209, .537). This result indicates that older participants relied more on the decision aids' recommendations than younger participants.

Furthermore, there was a statistically significant medium negative correlation between self-performance and average REL, $r_s(116) = -.382$, $p < .001$, with a 95% percentile BCa bootstrap confidence interval of (-.534, -.183). This result indicates that

participants who were better at performing the task without a decision aid tended to rely less on the decision aids' recommendations than those worse at performing the task without a decision aid.

The positive correlation between age and reliance might partly be explained by the fact that younger participants were better at performing the task without a decision aid. There was a statistically significant medium negative correlation between self-performance and age, $r_s(116) = -.373$, $p < .001$, with a 95% percentile BCa bootstrap confidence interval of $(-.522, -.205)$. This result indicates that younger participants were better at performing the task unassisted than older participants.

Since age was correlated with self-performance and reliance, any significant age differences between subgroups might have biased the previously reported results of hypothesis testing. Figure 19 presents descriptive statistics on age for each of the six subgroups.

Descriptives								
Age								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	20	35.85	6.869	1.536	32.64	39.06	24	53
2	19	36.32	11.373	2.609	30.83	41.80	25	70
3	20	37.00	10.296	2.302	32.18	41.82	26	70
4	19	39.53	10.292	2.361	34.57	44.49	27	66
5	20	36.20	6.469	1.447	33.17	39.23	26	46
6	20	35.30	8.007	1.790	31.55	39.05	26	60
Total	118	36.68	8.951	.824	35.05	38.31	24	70

Figure 19. Age by Subgroup

Since age was not normally distributed in all subgroups, a nonparametric Kruskal–Wallis test was used to examine potential age differences between the groups. The test showed no statistically significant age differences between subgroups (see Figure 20). Hence, uneven distributions of age do not threaten the validity of the results.

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig. ^{a,b}	Decision
1	The distribution of Age is the same across categories of sg.	Independent-Samples Kruskal-Wallis Test	.736	Retain the null hypothesis.

a. The significance level is .050.
b. Asymptotic significance is displayed.

Independent-Samples Kruskal-Wallis Test Summary	
Total N	118
Test Statistic	2.763 ^a
Degree Of Freedom	5
Asymptotic Sig.(2-sided test)	.736

a. The test statistic is adjusted for ties.

Figure 20. Kruskal–Wallis Test of Age

(2) Self-confidence

The variable self-confidence was not normally distributed, according to a Shapiro-Wilks test ($p < .001$), with a skewness of $-.938$ and a kurtosis of $.861$. Inspection of a scatterplot indicated a monotonic relationship between self-confidence and average REL.

There was a statistically significant medium negative correlation between confidence in solving the task without a decision aid and average REL, $r_s(116) = -.376$, $p < .001$, with a 95% percentile BCa bootstrap confidence interval of $(-.531, -.200)$. This result indicates that participants who were more confident in solving the task without a decision aid relied less on the decision aids than those who were less confident in solving the task without a decision aid.

b. Average Retrospective Trust

The average retrospective trust score was calculated by taking the average value of RTR across the two decision aids for each participant. The histogram in Figure 21 shows the shape of the data distribution. The data were not normally distributed, according to a Shapiro–Wilk test ($p < .001$), with a skewness of $-.867$ and a kurtosis of $.627$.

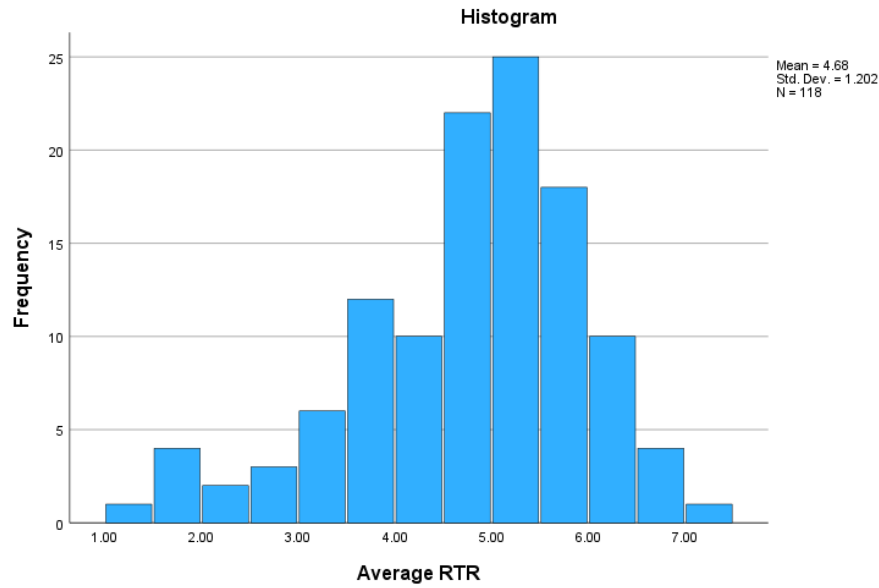


Figure 21. Data Distribution Average RTR

The variable team-confidence was not normally distributed, according to a Shapiro–Wilk test ($p < .001$), with a skewness of -1.272 and a kurtosis of 2.214. Inspection of a scatterplot indicated a monotonic relationship between team-confidence and average RTR.

There was a statistically significant medium positive correlation between confidence in solving the task with the decision aid and average RTR, $r_s(116) = .466$, $p < .001$, with a 95% percentile BCa bootstrap confidence interval of (.289, .603). This result indicates that participants who were more confident in solving the task with the decision aids gave higher retrospective trust ratings of the decision aids than those who were less confident in solving the task with the decision aids.

c. *Average Response Time*

The average response time score was calculated by taking the average RES value across the two decision aids for each participant. The histogram in Figure 22 shows the shape of the data distribution. The average RES scores were not normally distributed, according to a Shapiro–Wilk test ($p < .001$), with a skewness of .939 and a kurtosis of .751.

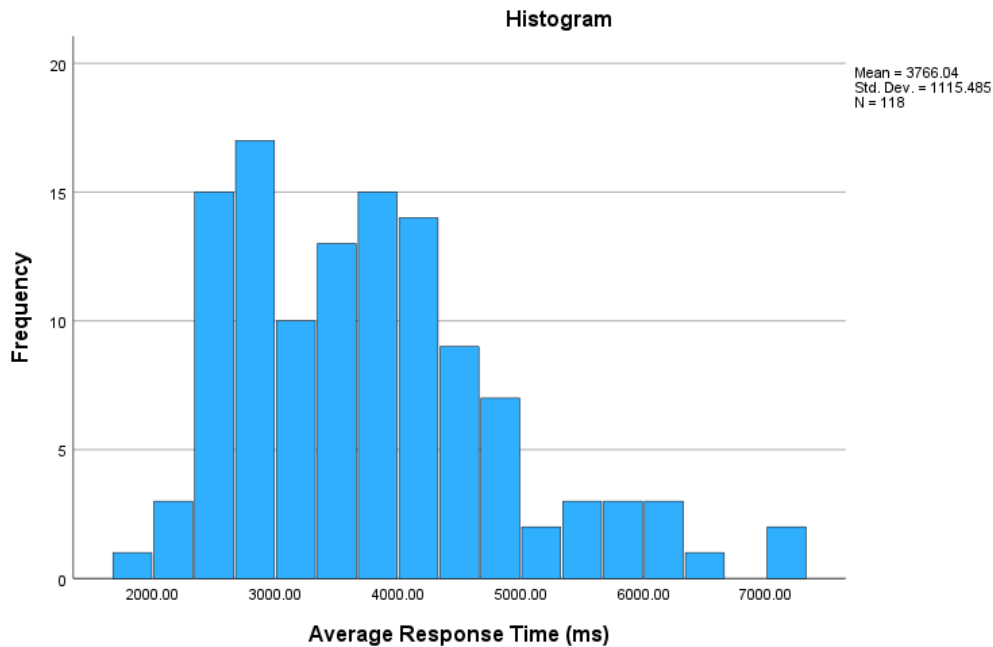


Figure 22. Data Distribution Average RES

According to the Spearman's ranked order test, age was the only variable with a statistically significant correlation with average RES. There was a statistically significant medium positive correlation between age and average RES, $r_s(116) = .438, p < .001$, with a 95% percentile BCa bootstrap confidence interval of (.293, .567). This result indicates that older participants took longer to make reliance decisions than younger participants.

5. Participant Motivation

The previously reported ANOVA tests showed no statistically significant effect of MOT on any dependent variables. In the exit questionnaire, participants were asked to rate their motivation to participate in the experiment. To further explore if the manipulation of MOT had an effect, Table 17 compares the subjectively reported motivation between the two conditions of MOT.

Table 17. Subjective Motivation by MOT Condition

MOT	N	Mean	Std Dev
No	60	6.42	1.169
Yes	58	6.60	.528

An independent samples *t* test showed no significant statistical difference between the groups ($p = .268$). This result suggests that the chance to win a \$20 gift card based on performance did not significantly increase MOT.

6. Order Effects

This section explores potential order effects between the two decision aids. Figure 23 compares mean values between REL for the lower reliability decision aid, the higher reliability decision aid, and average REL across the two decision aids. Order 1 means that the lower reliability decision aid was used first, and Order 2 means that the higher reliability decision aid was used first. Mean values were higher for both the lower and the higher reliability decision aids when they were used first. The mean value for average REL was higher when the higher reliability decision aid was used first.

Group Statistics					
	order	N	Mean	Std. Deviation	Std. Error Mean
REL Lower Reliability Aid	1	58	.41769	.157931	.020737
	2	60	.38862	.198163	.025583
REL Higher Reliability Aid	1	58	.50738	.201596	.026471
	2	60	.59683	.217149	.028034
Average Reliance	1	58	.4625	.16064	.02109
	2	60	.4927	.19051	.02459

Figure 23. REL by Decision Aid Order

Independent samples *t* tests were used to test if the differences were statistically significant. There was a statistically significant difference in reliance on the higher reliability decision aid between participants who used it first ($M: .597$, $SD: .217$, $CI_{95} [.541, .653]$) and those who used it last ($M: .507$, $SD: .202$, $CI_{95} [.454, .560]$), $t(116) =$

-2.317, $p = .022$, $d = -.427$. There were no statistically significant differences in reliance on the lower reliability decision aid ($p = .381$) or average REL ($p = .355$) between the participants who used the lower or higher reliability decision aid first.

A corresponding examination of RTR gave similar results. Figure 24 compares the mean values of RTR for both decision aids and an average rating across the decision aids. RTR for both decision aids appeared higher if participants used them first, and the average RTR score appeared higher when participants interacted with the higher reliability decision aid first.

Group Statistics					
	order	N	Mean	Std. Deviation	Std. Error Mean
RTR Lower Reliability Aid	1	58	4.3333	1.30382	.17120
	2	60	4.1111	1.54852	.19991
RTR Higher Reliability Aid	1	58	4.8678	1.24087	.16293
	2	60	5.3833	1.13136	.14606
Average RTR	1	58	4.6006	1.20954	.15882
	2	60	4.7472	1.19967	.15488

Figure 24. RTR by Decision Aid Order

Independent samples t tests were used to test for statistically significant differences. There was a statistically significant difference in RTR for the higher reliability decision aid between participants who used it first ($M: 5.383$, $SD: 1.131$, $CI_{95}[5.091, 5.676]$) and those who used it last ($M: 4.868$, $SD: 1.241$, $CI_{95}[4.542, 5.194]$), $t(116) = -2.360$, $p = .020$, $d = -.435$. There were no statistically significant differences in RTR for the lower reliability decision aid ($p = .402$) or average RTR ($p = .510$) between the participants who used the lower or higher reliability decision aid first.

The tests indicate that the order in which participants interacted with the two decision aids affected both REL and RTR. In this dissertation, the order in which participants interacted with the two decision aids was counterbalanced to mitigate potential order effects.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

This dissertation explored the effects of intermediate trust ratings on automation reliance, retrospective trust, and response time in human–automation interactions. In addition, the effects of automation reliability and participant motivation were of interest and manipulated along with trust rating frequency. An experiment was conducted to test six research hypotheses that were grounded in existing literature. In the experiment, research participants were asked to determine the number of UAVs in a series of images, with the help of an automated decision aid. Overall, results showed that intermediate trust ratings increased automation reliance and retrospective trust assessments of automation. This final chapter presents the interpretation of results, contributions, study limitations, and recommendations for future research.

A. INTERPRETATION OF RESULTS

The essence of this dissertation is captured in Figure 25. The red text represents the contribution of this dissertation: that intermediate trust ratings affect automation reliance. Antecedents of trust in automation have previously been broken down into human, automation, and situation characteristics by several authors (Endsley, 2017; Schaefer et al., 2016). Also, previous work has suggested that trust and reliance are updated in repeated interactions (Hoff & Bashir, 2015; Yang et al., 2023; Yu et al., 2017). The proposed theoretical model was developed by integrating theories about automation reliance and intermediate judgments. Research hypotheses were formulated to test the theorized relationships, and results from analysis provided initial validation of the model.

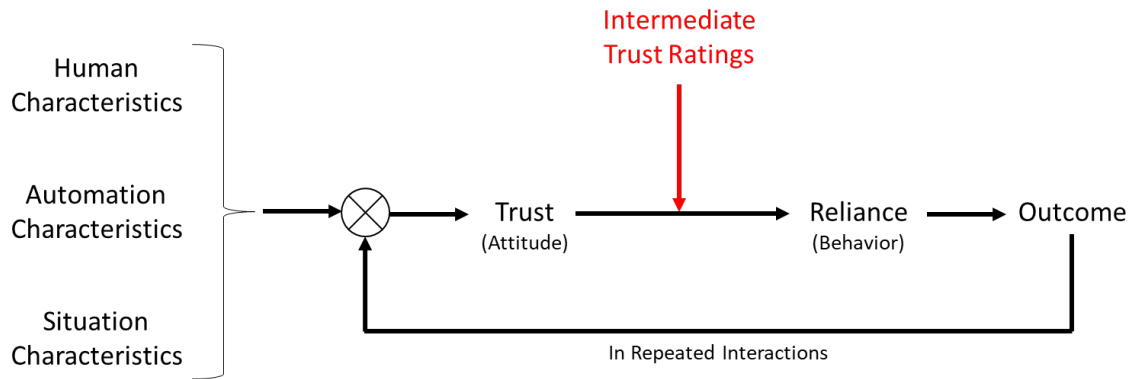


Figure 25. Proposed Theoretical Model

The six research hypotheses tested in this dissertation are repeated here for reference:

H₁: Trust rating frequency affects automation reliance.

H₂: Trust rating frequency affects response time.

H₃: Trust rating frequency affects retrospective trust.

H₄: An interaction is predicted among trust rating frequency, automation reliability, and participant motivation on automation reliance.

H₅: Higher automation reliability increases automation reliance.

H₆: Higher participant motivation increases automation reliance.

The main analysis showed a main effect of TRF on both REL and RTR, which supports H₁ and H₃. However, TRF did not affect RES, which does not support H₂. Furthermore, the tests showed no statistically significant interactions between TRF, AUR, and MOT on REL, which does not support H₄. Additionally, higher AUR did increase REL, which supports H₅. Finally, there was no evidence that higher MOT increased REL, which does not support H₆.

1. The Effect of Trust Rating Frequency

This dissertation provided empirical evidence that intermediate trust ratings during human–automation interaction increased automation reliance, as shown by the

ANOVA test, which compared REL between participants who made intermediate trust ratings and those who did not. Furthermore, the intermediate trust ratings were higher in the group that made trust ratings between every reliance decision than in the group that made trust ratings between every five decisions (see Figure 15). This study prompted participants to reevaluate their trust by asking them to make intermediate trust ratings. Previous research suggests that humans often use “default trust” and that mechanisms to trigger reevaluation of trust are needed (Hoffman et al., 2013). Trust might be lower without frequent reevaluations, since automation mistakes affect trust more than automation successes (Yang et al., 2023). Trust repair functions have been suggested to keep perceived trust calibrated to actual trustworthiness (de Visser et al., 2020). Findings in this dissertation indicate that intermediate trust ratings might be used for trust repair in situations when humans under-rely on automation.

In terms of retrospective trust, intermediate trust ratings increased RTR. These results align with the previously described findings that intermediate trust ratings increased automation reliance. However, a three-way interaction between TRF, AUR, and MOT was detected. It is unclear why RTR was relatively higher for the lower reliability decision aid in the TRF Regularly–MOT No condition compared to other combinations of MOT and TRF. This anomaly caused divergence from the main effects; for example, AUR had a significant effect on RTR in all combinations of TRF and MOT, except for the TRF Regularly–MOT No condition. Since the results were inconclusive, further research is needed to explain this interaction.

Finally, the variability in both intermediate trust ratings and RTR demonstrates how differently people might rate the same technology. One reason for the discrepancies is that human characteristics, such as age, experience, and attitudes, all affect trust in automation (Endsley, 2017; Hoff & Bashir, 2015; Schaefer et al., 2016). Another reason is the existence of diverse notions of trust among the participants (Larasati et al., 2023; Vereschak et al., 2021). However, the effect of intermediate trust ratings on automation reliance was significant, despite the variability of trust ratings.

2. The Effect of Automation Reliability

As predicted, AUR had a significant effect on REL. The experimental data suggest that participants were able to observe the decision aid performance and adjust their reliance accordingly. The fact that reliance is dynamically adjusted during interactions based on previous outcomes is a central idea in the automation–reliance literature (Hoff & Bashir, 2015).

Exploratory analysis showed that reliance on the higher reliability decision aid increased as the interactions progressed, whereas reliance on the lower reliability decision aid decreased (see Figure 17). These findings further support that participants adjusted their reliance to the observed performance of the two decision aids during the interactions. Hence, participants started to rely more on the higher reliability decision aid and less on the lower reliability decision aid as the interaction progressed.

3. The Effect of Human Characteristics

Variability in all dependent variables among participants in the same treatment groups suggest that human characteristics matter, since automation and situational characteristics were kept consistent in the experiment. For example, self-performance was negatively correlated with REL, and younger participants performed the experimental task better than older participants. These correlations might have contributed to why younger participants relied less on the automated decision aids. However, further analysis showed that participant age was approximately equal between the treatment groups, indicating that age differences between groups did not bias the results.

B. CONTRIBUTIONS

The findings in this dissertation contribute to theory, methodology, and practice. The theoretical contribution consists of a proposed model of automation reliance that extends existing models of automation reliance to include the effect of intermediate trust ratings (see Figure 25). The proposed model was tested empirically, and results showed an effect of intermediate trust ratings on reliance during human–automation interaction.

The model represents a contribution to information science by identifying causal relationships that govern how people and technology make sense of information to improve decision making. A stronger theoretical understanding of the phenomenon helps explain and predict reliance behavior. Empirical evidence also supported previous results that automation reliability is a significant predictor of automation reliance.

The findings also have methodological implications. Researchers who seek to capture trust dynamics during human–automation interactions should consider potential effects of intermediate trust ratings on reliance behavior. If they seek to measure trust without affecting reliance behavior, other, less intrusive methods than explicit trust ratings might be preferred.

Practical applications of the findings include informing the design of decision aids that allow people to rely on them appropriately. Intermediate trust ratings can then be one method to reevaluate trust during human–automation interactions to foster appropriate reliance, which might improve efficiency and safety in several high-stake domains.

Moreover, intermediate trust ratings might be a method to normalize and stabilize reliance on automation in a training setting before deploying it in a real situation. Since the variability of trust is higher before people get familiar with the automation (Yang et al., 2017; Yu et al., 2017), a training period is recommended before using automation in real situations. Furthermore, there might not be time for intermediate trust ratings in a real-world scenario. Therefore, intermediate trust ratings during the training period might be a way to calibrate trust and reliance.

C. LIMITATIONS

The suggested model has been validated in one specific situation. It is possible that the model could have a broader application in human technology interactions. However, since trust depends on specific technology and situational characteristics, further research is necessary to enhance generalizability across settings and populations.

One limitation is that the experiment was conducted in a laboratory environment. Since trust depends on situational characteristics, generalization to real-world scenarios is

limited. However, this was an initial study of the effect of intermediate trust ratings on automation reliance. Therefore, a laboratory experiment was selected to show causal relationships with internal validity. Future tests conducted outside of the laboratory will help provide more external validity (Kramer, 2018). Furthermore, the experimental interactions were relatively short and only once per decision aid. More extended studies would provide further insight into how intermediate trust ratings affect automation reliance over extended periods.

Another limitation is that participants were not domain experts. Experts provide more reliable evaluations of trust in automation since they have the appropriate knowledge and experience (Hoffman et al., 2013). Because recruiting enough participants with specific expertise can be challenging, using laypeople to increase the number of participants might be appropriate when studying more straightforward tasks (Doshi-Velez & Been, 2017). Therefore, this dissertation involved a task that required no specific domain knowledge. Furthermore, a training session was provided to familiarize participants with the task.

Intermediate trust ratings might not always be feasible in real-world scenarios. Trust ratings would constitute an additional task, which might increase decision time and add workload for the decision-maker. There might not be time for intermediate trust ratings in rapid decision-making. Furthermore, the perceived extra work might cause decision-makers to disapprove of the approach. To mitigate these concerns, the timing of intermediate trust ratings could be adaptive and adjusted to when the workload permits.

Finally, the chance to earn gift cards based on performance was used to increase participant motivation in this dissertation. Half of the participants were informed about the gift cards, and half were not. However, no statistically significant differences were observed between the two groups. The gift card value was \$20, which was the highest amount that the institutional review board would allow. That value might have been too low to affect motivation.

D. FUTURE RESEARCH

There are several opportunities for future research. First, the effect of intermediate trust ratings should be explored outside of the laboratory setting—for example, in a field experiment with a real-world task over an extended period. Experimental scenarios with higher stakes, in which the consequences of automation reliance are more significant, might be necessary to enhance generalizability to high-stake domains, such as the military. In general, before implementing mechanisms for trust calibration, extensive testing is required to assess their impacts (Wischnewski et al., 2023).

Second, to extend the applicability of the proposed model, further research is recommended on other populations and with different types of automation. Also, this dissertation explored explicit trust ratings, but there might be other types of intermediate judgments that affect reliance behavior in human–automation interactions. Therefore, future research should also test the effects of different types of intermediate judgments.

Third, future research should investigate how adaptive intervals of intermediate trust ratings can be used to maintain appropriate reliance with minimal interruption. Both situational and individual characteristics should be monitored to determine when intermediate trust ratings are warranted. Some individuals might benefit from more frequent trust ratings, and some situations might not allow for trust ratings as often.

Fourth, future research should explore if intermediate trust ratings is a viable method to help normalize and stabilize reliance on automation in a training setting before deploying it in a real situation. This might require more longitudinal studies, and reliable methods to verify that reliance matches system capabilities.

Finally, future research is recommended to further explore how human characteristics affect automation reliance—for example, how different human characteristics interact and how human characteristics might moderate the effect of intermediate trust ratings on automation reliance.

E. CONCLUSION

As automation becomes increasingly ubiquitous and capable, appropriate reliance is required to increase safety and efficiency in several high-stake domains. A strong understanding of how trust develops and varies during human–automation interactions helps explain and predict reliance behavior, but trust dynamics remains an underdeveloped topic in the automation–reliance literature. This dissertation addressed a knowledge gap by exploring how intermediate trust ratings affected reliance behavior in an experiment. Intermediate trust ratings have been used in previous studies to capture trust variations during human–automation interactions, but their effect on reliance decisions had not been empirically tested. Findings showed that intermediate trust ratings increased both automation reliance and retrospective trust. A theoretical model supported by these findings was presented to help explain and predict automation reliance. Additionally, findings from this dissertation have practical implications for both system designers and researchers. However, trust in automation is a complex topic, and further research is needed to enhance the generalizability of this dissertation’s findings.

APPENDIX A. TECHNOLOGY READINESS INDEX 2.0

We are interested in your views on how technology influences your life. Please indicate how much you agree with the following statements.

Technology gives me more freedom of mobility.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

Technology makes me more productive in my personal life.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

Other people come to me for advice on new technologies.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

In general, I am among the first in my circle of friends to acquire new technology when it appears.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

I keep up with the latest technological developments in my areas of interest.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

Technical support lines are not helpful because they don't explain things in terms I understand.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

Sometimes, I think that technology systems are not designed for use by ordinary people.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

People are too dependent on technology to do things for them.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

Too much technology distracts people to a point that is harmful.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

Technology lowers the quality of relationships by reducing personal interaction.

- Strongly disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Strongly agree

Note: These questions comprise the Technology Readiness Index 2.0, which is copyrighted by A. Parasuraman and Rockbridge Associates, Inc., 2014. This scale may be duplicated only with written permission from the authors.

APPENDIX B. EXIT QUESTIONNAIRE

A. DEMOGRAPHICS

Please enter your age.

Please enter your gender.

- ☐ Female
- ☐ Male
- ☐ I prefer not to answer

Please enter your ethnicity.

- ☐ Asian / Pacific Islander
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Native American or American Indian
- ☐ White
- ☐ Other
- ☐ I prefer not to answer

Are you a military or civilian?

- ☐ Military
- ☐ Civilian
- ☐ I prefer not to answer

If applicable, please enter your branch.

To your knowledge, do you have any type of color blindness?

- ☐ Yes
- ☐ No
- ☐ I prefer not to answer

B. EXPERIENCE WITH THE SAME OR SIMILAR TECHNOLOGY

How well do you agree with the following statements? Please select one option.

I am familiar with using automated decision aids.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Somewhat disagree
- ☐ Neutral
- ☐ Somewhat agree
- ☐ Agree
- ☐ Strongly agree

In my experience, automated decision aids perform well.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Somewhat disagree

- Neutral
- Somewhat agree
- Agree
- Strongly agree

I am familiar with using image recognition software.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

In my experience, image recognition software performs well.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

1	2	3	4	5	6	7
Not at all			Extremely			

1	2	3	4	5	6	7
Not at all			Extremely			

How well do you agree with the following statements? Please select one option.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

- I was motivated to participate in this experiment.
- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

Please describe the main task you were assigned in this experiment.

D. RETROSPECTIVE TRUST

Please rate the automated decision aid you just interacted with.

I believe the automated decision aid is a competent performer.

1	2	3	4	5	6	7
Not at all					Extremely	

I trust the automated decision aid.

1	2	3	4	5	6	7
Not at all					Extremely	

I can depend on the automated decision aid.

1 2 3 4 5 6 7

Not at all

Extremely

What is the overall reliability of the automated decision aid?

0% -----100%

APPENDIX C. CORRELATIONS

A. DATA DISTRIBUTION OF VARIABLES

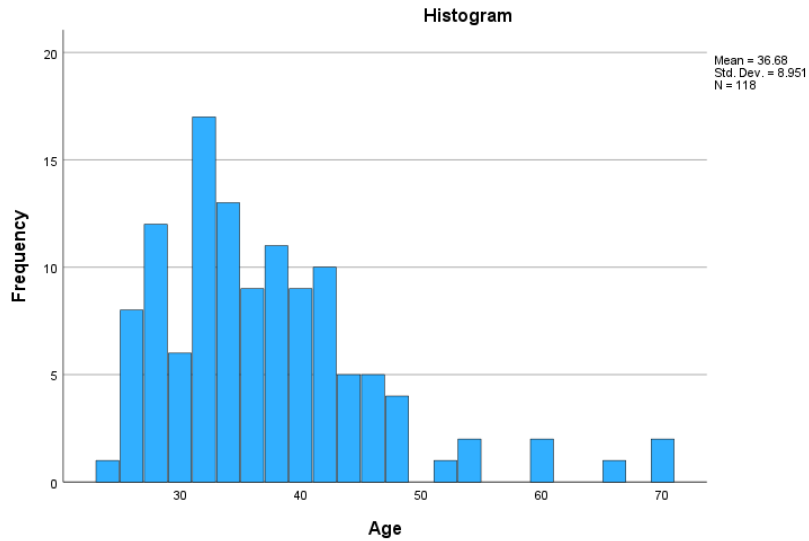


Figure 26. Data Distribution Age

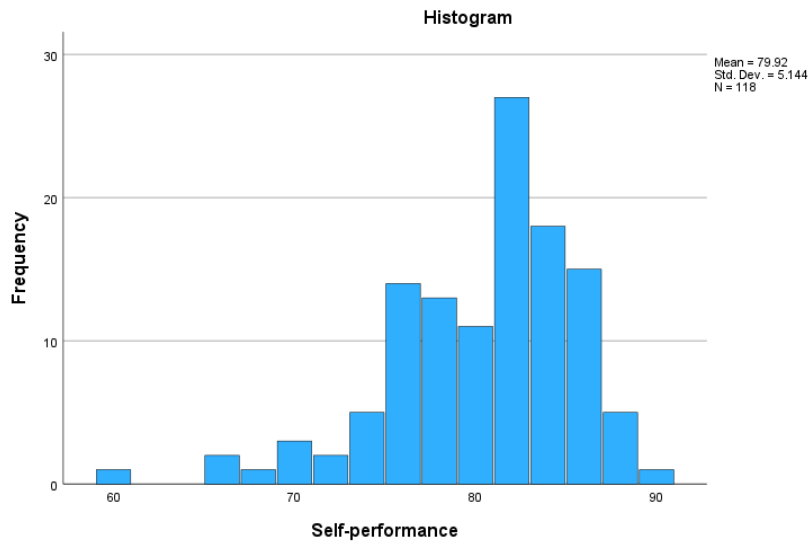


Figure 27. Data Distribution Self-Performance

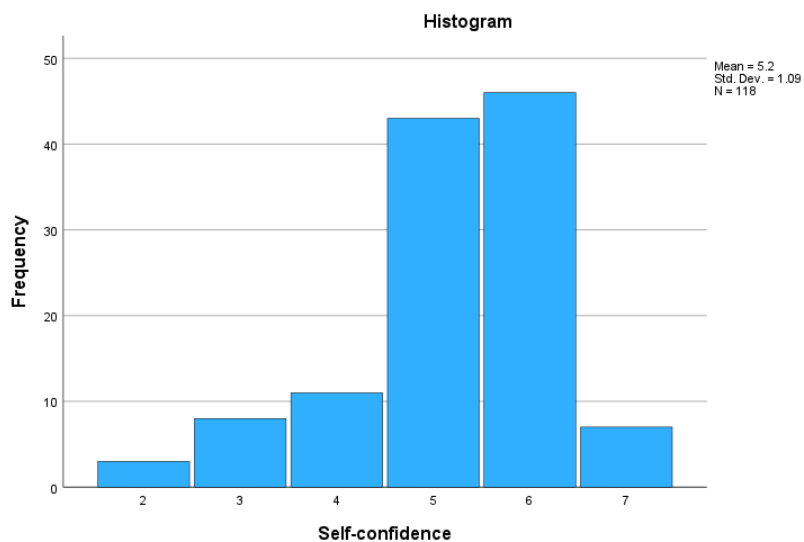


Figure 28. Data Distribution Self-Confidence

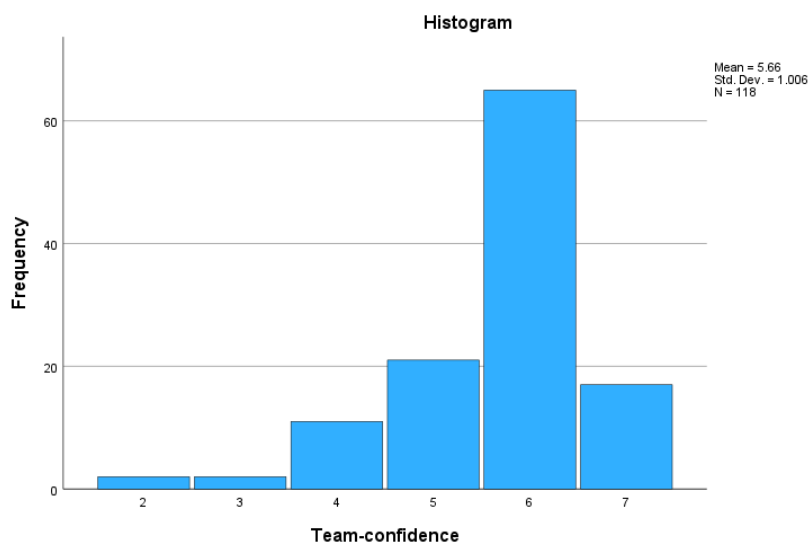


Figure 29. Data Distribution Team-Confidence

B. SCATTER PLOTS: AVERAGE RELIANCE

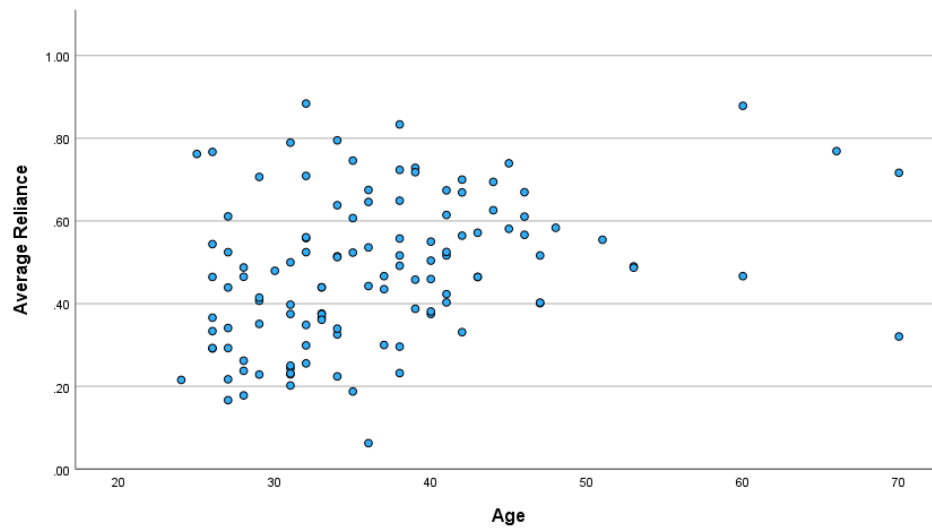


Figure 30. Scatter Plot of Average REL by Age

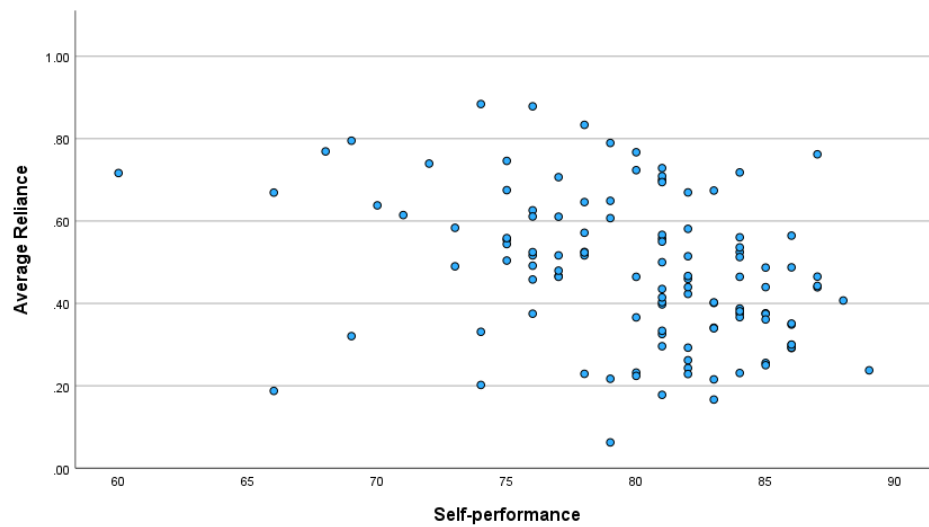


Figure 31. Scatter Plot of Average REL by Self-Performance

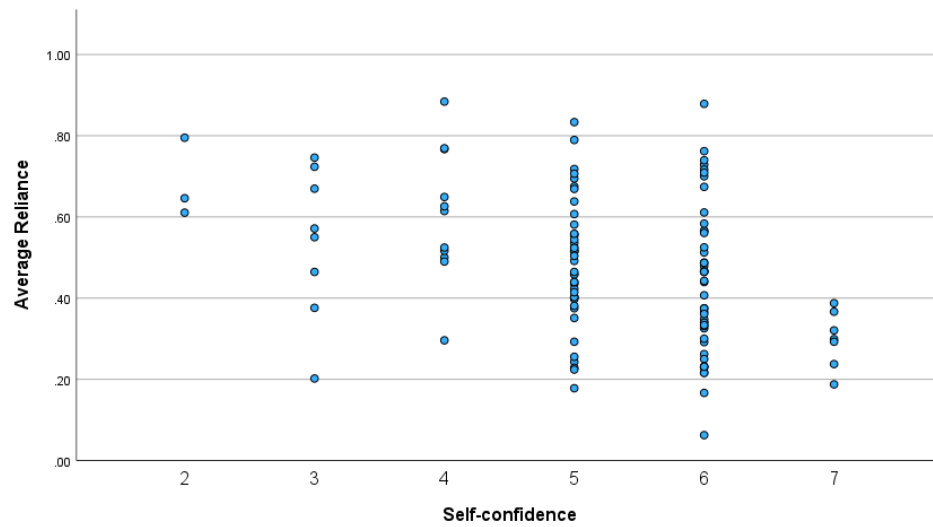


Figure 32. Scatter Plot of Average REL by Self-Confidence

C. SCATTER PLOTS: AVERAGE RETROSPECTIVE TRUST

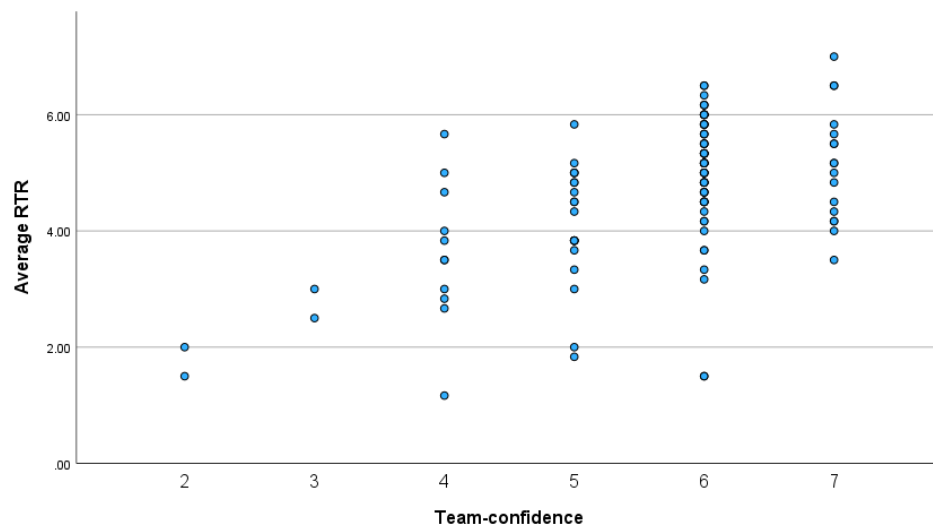


Figure 33. Scatter plot of Average RTR by Team-Confidence

D. SCATTER PLOTS: AVERAGE RESPONSE TIME

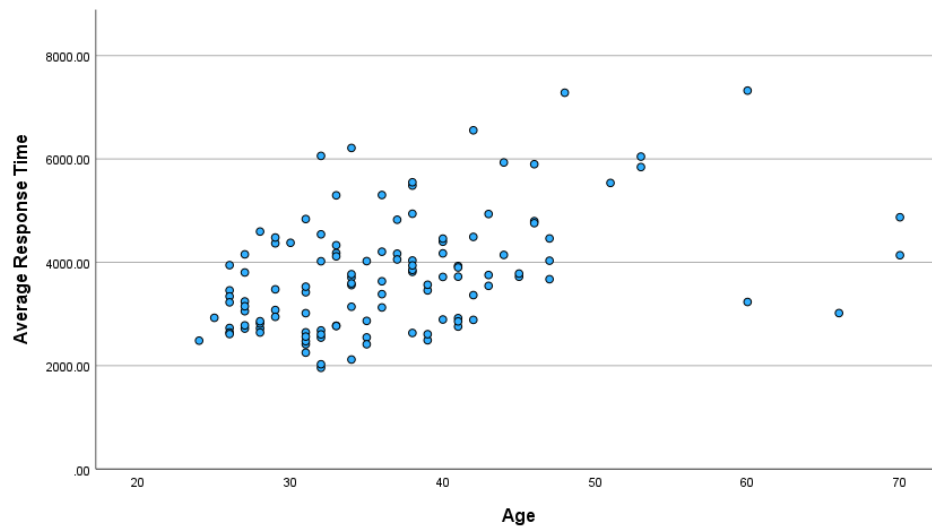


Figure 34. Scatter Plot of Average RES by Age

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- Akash, K., Hu, W.-L., Reid, T., & Jain, N. (2017). Dynamic modeling of trust in human-machine interactions. *2017 American Control Conference (ACC)*, 1542–1548. <https://doi.org/10.23919/ACC.2017.7963172>
- Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, 8(4), 321–348. <https://doi.org/10.1080/14639220500535301>
- Ballard, T., Sewell, D. K., Cosgrove, D., & Neal, A. (2019). Information processing under reward versus under punishment. *Psychological Science*, 30(5), 757–764. <https://doi.org/10.1177/0956797619835462>
- Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., & Horvitz, E. (2019). Updates in human-AI teams: Understanding and addressing the performance/compatibility tradeoff. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 2429–2437.
- Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), 131–160. <https://doi.org/10.1037/1082-989X.2.2.131>
- Bhat, S., Lyons, J. B., Shi, C., & Yang, X. J. (2022). Clustering trust dynamics in a human-robot sequential decision-making task. *IEEE Robotics and Automation Letters*, 7(4), 8815–8822. <https://doi.org/10.1109/LRA.2022.3188902>
- Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, 29(4), 552–557. <https://doi.org/10.7334/psicothema2016.383>
- Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal data in repeated measures ANOVA: Impact on type I error and power. *Psicothema*, 35(1), 21–29. <https://doi.org/10.7334/psicothema2022.292>
- Bobko, P., Hirshfield, L., Eloy, L., Spencer, C., Doherty, E., Driscoll, J., & Obolsky, H. (2023). Human-agent teaming and trust calibration: A theoretical framework, configurable testbed, empirical illustration, and implications for the development of adaptive systems. *Theoretical Issues in Ergonomics Science*, 24(3), 310–334. <https://nps.primo.exlibrisgroup.com>
- Bonneviot, F., Coeugnet, S., & Brangier, E. (2021). Pedestrians-automated vehicles interaction: Toward a specific trust model? *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, 221, 568–574.

- Borko, H. (1965). *The conceptual foundations of information systems*. System Development Corporation.
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–21. <https://doi.org/10.1145/3449287>
- Calhoun, G. (2022). Adaptable (not adaptive) automation: Forefront of human–automation teaming. *Human Factors*, 64(2), 269–277. <https://doi.org/10.1177/00187208211037457>
- Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Research (Oxford)*, 85, 58–72. <https://doi.org/10.1016/j.visres.2012.12.005>
- Clare, A. S., Cummings, M. L., & Repenning, N. P. (2015). Influencing trust for human–automation collaborative scheduling of multiple unmanned vehicles. *Human Factors*, 57(7), 1208–1218. <https://doi.org/10.1177/0018720815587803>
- Cleveland, W. S. (1984). Graphical methods for data presentation: Full scale breaks, dot charts, and multibased logging. *The American Statistician*, 38(4), 270–280. <https://doi.org/10.2307/2683401>
- CNN. (2009, March 4). *Faulty reading helped cause Dutch plane crash*. <http://www.cnn.com/2009/WORLD/europe/03/04/plane.crash/index.html>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Conte, E., Khrennikov, A. Y., Todarello, O., Federici, A., Mendolicchio, L., & Zbilut, J. P. (2009). Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems & Information Dynamics*, 16(1), 85–100. <https://doi.org/10.1142/S1230161209000074>
- Croson, R. T. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes*, 80(2), 118–133. <https://doi.org/10.1006/obhd.1999.2846>
- Cuevas, H. M., Fiore, S. M., Caldwell, B. S., & Strater, L. (2007). Augmenting team cognition in human–automation teams performing in complex operational environments. *Aviation, Space, and Environmental Medicine*, 78(5 Suppl), B63–B70.

- de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331–349. <https://doi.org/10.1037/xap0000092>
- 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>
- de Visser, E. J., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, 5(2), 209–231. <https://doi.org/10.1177/1555343411410160>
- de Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerinx, M. A. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics*, 12(2), 459–478. <https://doi.org/10.1007/s12369-019-00596-x>
- de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human–Computer Studies*, 58(6), 719–735. [https://doi.org/10.1016/S1071-5819\(03\)00039-9](https://doi.org/10.1016/S1071-5819(03)00039-9)
- Debons, A. (2008). *Information science 101*. Scarecrow Press.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668. <https://doi.org/10.1037/0033-2909.125.6.627>
- Dembrower, K., Crippa, A., Colón, E., Eklund, M., Strand, F., & ScreenTrustCAD Trial Consortium. (2023). Artificial intelligence for breast cancer detection in screening mammography in Sweden: A prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health*, 5(10), e707–e711. [https://doi.org/10.1016/S2589-7500\(23\)00153-X](https://doi.org/10.1016/S2589-7500(23)00153-X)
- Department of State. (2023, February 16). *Political declaration on responsible military use of artificial intelligence and autonomy*. <https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/>
- Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., & Yanco, H. (2013). Impact of robot failures and feedback on real-time trust. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 251–258. <https://doi.org/10.1109/HRI.2013.6483596>

- Dismukes, K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents*. Ashgate Publishing. <https://doi.org/10.4324/9781315238654>
- Doshi-Velez, F., & Been, K. (2017). Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608*. <https://doi.org/10.48550/arxiv.1702.08608>
- Drozdal, J., Weisz, J., Wang, D., Dass, G., Yao, B., Zhao, C., Muller, M., Lin, J., & Su, H. (2020). Trust in AutoML: Exploring information needs for establishing trust in automated machine learning systems. *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 297–307. <https://doi.org/10.1145/3377325.3377501>
- 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)
- Ekman, F., Johansson, M., & Sochor, J. (2018). Creating appropriate trust in automated vehicle systems: A framework for HMI design. *IEEE Transactions on Human-Machine Systems*, 48(1), 95–101. <https://doi.org/10.1109/THMS.2017.2776209>
- Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, 59(1), 5–27. <https://doi.org/10.1177/0018720816681350>
- Engelmann, J. B., Damaraju, E., Padmala, S., & Pessoa, L. (2009). Combined effects of attention and motivation on visual task performance: Transient and sustained motivational effects. *Frontiers in Human Neuroscience*, 3, 4. <https://doi.org/10.3389/neuro.09.004.2009>
- Engelmann, J. B., & Pessoa, L. (2007). Motivation sharpens exogenous spatial attention. *Emotion (Washington, DC)*, 7(3), 668–674. <https://doi.org/10.1037/1528-3542.7.3.668>
- Falcone, R., & Castelfranchi, C. (2004). Trust dynamics: How trust is influenced by direct experiences and by trust itself. *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, 2, 740–747. <https://ieeexplore.ieee.org/document/1373544>
- Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, 14(1), 39–49. <https://doi.org/10.1007/s10111-011-0194-3>

- Gillespie, N. (2015). Survey measures of trust in organizational contexts: An overview. In F. Lyon, G. Möllering, & M. N. K. Saunders (Eds.), *Handbook of research methods on trust* (pp. 225–240). Edward Elgar Publishing.
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Guo, Y., & Yang, X. J. (2020). Modeling and predicting trust dynamics in human–robot teaming: A Bayesian inference approach. *International Journal of Social Robotics*, 13(8), 1899–1909. <https://doi.org/10.1007/s12369-020-00703-3>
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., 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>
- Harris, M. (2019, November 7). NTSB investigation into deadly Uber self-driving car crash reveals lax attitude toward safety. *IEEE Spectrum*. <https://spectrum.ieee.org/ntsb-investigation-into-deadly-uber-selfdriving-car-crash-reveals-lax-attitude-toward-safety>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in automation. *IEEE Intelligent Systems*, 28(1), 84–88. <https://doi.org/10.1109/MIS.2013.24>
- Holliday, D., Wilson, S., & Stumpf, S. (2016). User trust in intelligent systems: A journey over time. *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 164–168. <https://doi.org/10.1145/2856767.2856811>
- Holthausen, B. E., Wintersberger, P., Walker, B. N., & Riener, A. (2020). Situational trust scale for automated driving (STS-AD): Development and initial validation. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20)*, 40–47. <https://doi.org/10.1145/3409120.3410637>
- Hsieh, S.-J., Wang, A. R., Madison, A., Tossell, C., & de Visser, E. (2022). Adaptive driving assistant model (ADAM) for advising drivers of autonomous vehicles. *ACM Transactions on Interactive Intelligent Systems*, 12(3), 1–28. <https://doi.org/10.1145/3545994>

- Hussein, A., Elsayah, S., & Abbass, H. A. (2020). Trust mediating reliability–reliance relationship in supervisory control of human–swarm interactions. *Human Factors*, 62(8), 1237–1248. <https://doi.org/10.1177/0018720819879273>
- Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, 56(1), 517–543. <https://doi.org/10.1146/annurev.psych.56.091103.070250>
- 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
- Johnson, B. (2019). Artificial intelligence—An enabler of naval tactical decision superiority. *The AI Magazine*, 40(1), 63–78. <https://doi.org/10.1609/aimag.v40i1.2852>
- Jones, K. (2017, February 17). ARSENL FX29 highlights reel [Video]. YouTube. <https://www.youtube.com/watch?v=6LyUe7Y4uxc>
- Jonker, C. M., & Treur, J. (1999). Formal analysis of models for the dynamics of trust based on experiences. In F. J. Garijo & M. Boman (Eds.), *Multi-agent system engineering: 9th European workshop on modelling autonomous agents in a multi-agent world* (Vol. 1647, pp. 221–232). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-48437-X_18
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American Psychologist*, 58(9), 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>
- Kantowitz, B. H., Hanowski, R. J., & Kantowitz, S. C. (1997). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors*, 39(2), 164–176. <https://doi.org/10.1518/001872097778543831>
- Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, 65(2), 337–359. <https://doi.org/10.1177/00187208211013988>
- Kerlinger, F. N., & Lee, H. B. (2008). *Foundations of behavioral research* (5th ed.). Wadsworth Publishing.
- Kinley, N., & Ben-Hur, S. (2023). Extrinsic motivation: Using reward and punishment. In *Changing employee behavior: How to drive performance by bringing out the best in people* (2nd ed., pp. 59–84). Palgrave Macmillan.

- Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y.-C., & Shaw, T. H. (2021). Measurement of trust in automation: A narrative review and reference guide. *Frontiers in Psychology*, 12, 604977–604977. <https://doi.org/10.3389/fpsyg.2021.604977>
- Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, 66, 18–31. <https://doi.org/10.1016/j.apergo.2017.07.006>
- Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 12(7), 939–945. <https://doi.org/10.1038/nn.2321>
- Kramer, R. M. (2018). Pursuing ecological validity in trust research: Merits of multi-method research. In F. Lyon, G. Möllering, & M. N. K. Saunders (Eds.), *Handbook of research methods on trust* (pp. 25–35). Edward Elgar Publishing.
- Kraus, J., Scholz, D., & Baumann, M. (2021). What’s driving me? Exploration and validation of a hierarchical personality model for trust in automated driving. *Human Factors*, 63(6), 1076–1105. <https://doi.org/10.1177/0018720820922653>
- Kraus, J., Scholz, D., Stiegemeier, D., & Baumann, M. (2020). The more you know: Trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. *Human Factors*, 62(5), 718–736. <https://doi.org/10.1177/0018720819853686>
- Larasati, R., De Liddo, A., & Motta, E. (2023, April 23). *Human and AI trust: Trust attitude measurement instrument development*. Workshop on Trust and Reliance in AI-Assisted Tasks (TRAIT) at CHI 2023, Hamburg, Germany.
- Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human–machine systems. *Ergonomics*, 35(10), 1243–1270. <https://doi.org/10.1080/00140139208967392>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Levs, J. (2012, January 16). *What caused the cruise ship disaster?* CNN. <https://www.cnn.com/2012/01/15/world/europe/italy-cruise-questions/index.html>
- 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>
- Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In J. Scholz, H. Abbass, & D. Reid (Eds.), *Foundations of trusted autonomy* (Vol. 117, pp. 135–159). Springer Nature.

- Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, 17(1), 39–71. <https://doi.org/10.1016/j.jsis.2008.01.001>
- Lu, Y., & Sarter, N. (2019). Eye tracking: A process-oriented method for inferring trust in automation as a function of priming and system reliability. *IEEE Transactions on Human-Machine Systems*, 49(6), 560–568. <https://doi.org/10.1109/THMS.2019.2930980>
- Lyon, F., Möllering, G., & Saunders, M. N. K. (2018). Researching trust: The ongoing challenge of matching objectives and methods. In *Handbook of research methods on trust* (2nd ed., pp. 1–22). Edward Elgar Publishing.
- 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/10.1518/001872006777724408>
- Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human–robot trust. In C. S. Nam & J. B. Lyons (Eds.), *Trust in human–robot interaction* (pp. 3–25). Academic Press.
- Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors*, 45(2), 281–295. <https://doi.org/10.1518/hfes.45.2.281.27239>
- Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, 6(1), 57–87. <https://doi.org/10.1177/1555343411433844>
- Markkula, G., Uludag, Z., Wilkie, R. M., & Billington, J. (2021). Accumulation of continuously time-varying sensory evidence constrains neural and behavioral responses in human collision threat detection. *PLoS Computational Biology*, 17(7), e1009096. <https://doi.org/10.1371/journal.pcbi.1009096>
- 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. *The Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.2307/258792>

- McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24–59. <https://doi.org/10.5465/256727>
- McBride, M., Carter, L., & Ntuen, C. (2012). The impact of personality on nurses' bias towards automated decision aid acceptance. *International Journal of Information Systems and Change Management*, 6(2), 132–146. <https://doi.org/10.1504/IJISCM.2012.051148>
- McKnight, D. H. (2005). Trust in information technology. In *The Blackwell encyclopedia of management. Vol. 7 Management information systems* (pp. 329–331). Blackwell.
- McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *The Academy of Management Review*, 23(3), 473–490. <https://doi.org/10.5465/amr.1998.926622>
- Mendes, M., & Pala, A. (2003). Type I error rate and power of three normality tests. *Pakistan Journal of Information and Technology*, 2(2), 135–139.
- Merritt, S. M. (2011). Affective processes in human–automation interactions. *Human Factors*, 53(4), 356–370. <https://doi.org/10.1177/0018720811411912>
- Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*, 55(3), 520–534. <https://doi.org/10.1177/0018720812465081>
- Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human–automation interactions. *Human Factors*, 50(2), 194–210. <https://doi.org/10.1518/001872008X288574>
- Merritt, S. M., Lee, D., Unnerstall, J. L., & Huber, K. (2015). Are well-calibrated users effective users? Associations between calibration of trust and performance on an automation-aided task. *Human Factors*, 57(1), 34–47. <https://doi.org/10.1177/0018720814561675>
- Merritt, S. M., Unnerstall, J. L., Lee, D., & Huber, K. (2015). Measuring individual differences in the perfect automation schema. *Human Factors*, 57(5), 740–753. <https://doi.org/10.1177/0018720815581247>
- Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, 6(1), 44–58. <https://doi.org/10.1037/1076-898X.6.1.44>

- Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5), 527–539. [https://doi.org/10.1016/S0020-7373\(87\)80013-5](https://doi.org/10.1016/S0020-7373(87)80013-5)
- Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11), 1905–1922. <https://doi.org/10.1080/00140139408964957>
- 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>
- National Academies of Sciences, Engineering, and Medicine. (2022). *Human–AI teaming: State-of-the-art and research needs*. The National Academies Press.
- National Transportation Safety Board. (1997). *Marine accident report—Grounding of the Panamanian passenger ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts, June 10, 1995* (NTSB/MAR-97/01). <https://www.nts.gov/investigations/AccidentReports/Reports/mar9701.pdf>
- Nourani, M., King, J. T., & Ragan, E. D. (2020). The role of domain expertise in user trust and the impact of first impressions with intelligent systems. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 8, 112–121. <https://doi.org/10.48550/arxiv.2008.09100>
- Osborne, J. (2008). Best practices in data transformation: The overlooked effect of minimum values. In *Best practices in quantitative methods* (pp. 197–204). SAGE Publications. <https://doi.org/10.4135/9781412995627.d17>
- Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059–1072. <https://doi.org/10.1080/00140139.2012.691554>
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410. <https://doi.org/10.1177/0018720810376055>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics. Part A, Systems and Humans*, 30(3), 286–297. <https://doi.org/10.1109/3468.844354>

- Paschke, L. M., Walter, H., Steimke, R., Ludwig, V. U., Gaschler, R., Schubert, T., & Stelzel, C. (2015). Motivation by potential gains and losses affects control processes via different mechanisms in the attentional network. *NeuroImage (Orlando, Fla.)*, 111, 549–561. <https://doi.org/10.1016/j.neuroimage.2015.02.047>
- Perkins, L., Miller, J. E., Hashemi, A., & Burns, G. (2010). Designing for human-centered systems: Situational risk as a factor of trust in automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(25), 2130–2134. <https://doi.org/10.1177/154193121005402502>
- Pochon, J. B., Levy, R., Fossati, P., Lehericy, S., Poline, J. B., Pillon, B., Le Bihan, D., & Dubois, B. (2002). The neural system that bridges reward and cognition in humans: An fMRI study. *Proceedings of the National Academy of Sciences*, 99(8), 5669–5674. <https://doi.org/10.1073/pnas.082111099>
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4), 260–281. <https://doi.org/10.1016/j.tics.2016.01.007>
- Real Options Valuation. (n.d.). *Welcome to Real Options Valuation*. Retrieved November 21, 2023, from <https://www.realoptionsvaluation.com/>
- Rice, S., & Keller, D. (2009). Automation reliance under time pressure. *Cognitive Technology*, 14(1), 36–44.
- Roberts, A., Webster, L., Salmon, P., Flin, R., Salas, E., Cooke, N., Read, G., & Stanton, N. (2021). State of science: Models and methods for understanding and enhancing teams and teamwork in complex sociotechnical systems. *Ergonomics*, 1–45. <https://doi.org/10.1080/00140139.2021.2000043>
- Ross, J. M., Szalma, J. L., Hancock, P. A., Barnett, J. S., & Taylor, G. (2008). The effect of automation reliability on user automation trust and reliance in a search-and-rescue scenario. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(19), 1340–1344. <https://doi.org/10.1177/154193120805201908>
- Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, 49(1), 76–87. <https://doi.org/10.1518/001872007779598082>
- Sanchez, J., Fisk, A. D., & Rogers, W. A. (2004). Reliability and age-related effects on trust and reliance of a decision support aid. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 586–589. <https://doi.org/10.1177/154193120404800366>

- Sarter, M., Gehring, W. J., & Kozak, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*, 51(2), 145–160. <https://doi.org/10.1016/j.brainresrev.2005.11.002>
- Saßmannshausen, T., Burggräf, P., Hassenzahl, M., & Wagner, J. (2022). Human trust in otherware—A systematic literature review bringing all antecedents together. *Ergonomics*, 1–23. <https://doi.org/10.1080/00140139.2022.2120634>
- Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, 50(3), 377–400.
- Schmidt, E., Work, R., Catz, S., Horvitz, E., Chien, S., Jassy, A., Clyburn, M., Louie, G., Darby, C., Mark, W., Ford, K., Matheny, J., Griffiths, J.-M., McFarland, K., & Moore, A. (2021). *National Security Commission on artificial intelligence final report*. National Security Commission on Artificial Intelligence.
- Sharot, T., Velasquez, C. M., & Dolan, R. J. (2010). Do decisions shape preference? Evidence from blind choice. *Psychological Science*, 21(9), 1231–1235. <https://doi.org/10.1177/0956797610379235>
- Sheridan, T. B. (2002). *Humans and automation: System design and research issues*. Human Factors and Ergonomics Society.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Man-Machine Systems Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology. <https://apps.dtic.mil/sti/citations/ADA057655>
- Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, 39(2), 211–221. <https://doi.org/10.1037/0022-3514.39.2.211>
- Shyu, H. (2022). *USD (R&E) technology vision for an era of competition*. U.S. Under Secretary of Defense for Research & Engineering. https://www.cto.mil/wp-content/uploads/2022/02/usdre_strategic_vision_critical_tech_areas.pdf
- Spain, R. D., & Bliss, J. P. (2008). The effect of sonification display pulse rate and reliability on operator trust and perceived workload during a simulated patient monitoring task. *Ergonomics*, 51(9), 1320–1337. <https://doi.org/10.1080/00140130802120234>
- Tenhundfeld, N., Demir, M., & de Visser, E. (2022). Assessment of trust in automation in the “real world”: Requirements for new trust in automation measurement techniques for use by practitioners. *Journal of Cognitive Engineering and Decision Making*, 16(2), 101–118. <https://doi.org/10.1177/15553434221096261>

- Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician*, 34(1), 23–25. <https://doi.org/10.1080/00031305.1980.10482706>
- van Dongen, K., & van Maanen, P.-P. (2013). A framework for explaining reliance on decision aids. *International Journal of Human-Computer Studies*, 71(4), 410–424. <https://doi.org/10.1016/j.ijhcs.2012.10.018>
- Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to evaluate trust in AI-assisted decision making? A survey of empirical methodologies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–39. <https://doi.org/10.1145/3476068>
- Wang, L., Jamieson, G. A., & Hollands, J. G. (2011). The effects of design features on users' trust in and reliance on a combat identification system. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1), 375–379. <https://doi.org/10.1177/1071181311551077>
- Wang, Z., & Busemeyer, J. R. (2016). Interference effects of categorization on decision making. *Cognition*, 150, 133–149. <https://doi.org/10.1016/j.cognition.2016.01.019>
- White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: The constructive role of articulating impressions. *Cognition*, 133(1), 48–64. <https://doi.org/10.1016/j.cognition.2014.05.015>
- Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201–212. <https://doi.org/10.1080/14639220500370105>
- Wickens, C. D., Dixon, S. R., & Ambinder, M. S. (2006). Workload and automation reliability in unmanned air vehicles. In N. J. Cooke, H. L. Pringle, H. K. Pedersen, & O. Connor (Eds.), *Human factors of remotely operated vehicles* (Vol. 7, pp. 209–222). Emerald Group Publishing.
- Wischnewski, M., Krämer, N., & Müller, E. (2023). Measuring and understanding trust calibrations for automated systems: A survey of the state-of-the-art and future directions. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. <https://doi.org/10.1145/3544548.3581197>
- Xu, A., & Dudek, G. (2015). OPTIMo: Online probabilistic trust inference model for asymmetric human-robot collaborations. *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 221–228. <https://doi.org/10.1145/2696454.2696492>

- Yang, X. J., Schemanske, C., & Searle, C. (2023). Toward quantifying trust dynamics: How people adjust their trust after moment-to-moment interaction with automation. *Human Factors*, 65(2), 862–878. <https://doi.org/10.1177/00187208211034716>
- Yang, X. J., Unhelkar, V., Li, K., & Shah, J. (2017). Evaluating effects of user experience and system transparency on trust in automation. *Proceedings of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*, 408–416. <https://doi.org/10.1145/2909824.3020230>
- Yin, M., Wortman Vaughan, J., & Wallach, H. (2019). Understanding the effect of accuracy on trust in machine learning models. *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3290605.3300509>
- Yu, K., Berkovsky, S., Conway, D., Taib, R., Zhou, J., & Chen, F. (2016). Trust and reliance based on system accuracy. *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 223–227. <http://dx.doi.org/10.1145/2930238.2930290>
- Yu, K., Berkovsky, S., Taib, R., Conway, D., Zhou, J., & Chen, F. (2017). User trust dynamics: An investigation driven by differences in system performance. *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, 307–317. <https://doi.org/10.1145/3025171.3025219>
- Yuviler-Gavish, N., & Gopher, D. (2011). Effect of descriptive information and experience on automation reliance. *Human Factors*, 53(3), 230–244. <https://doi.org/10.1177/0018720811406725>
- Zhang, Y., Liao, Q., & Bellamy, R. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295–305. <https://doi.org/10.1145/3351095.3372852>
- Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. Basic Books. <https://dl.acm.org/doi/abs/10.5555/47303>

INITIAL DISTRIBUTION LIST

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



DUDLEY KNOX LIBRARY

NAVAL POSTGRADUATE SCHOOL

WWW.NPS.EDU

WHERE SCIENCE MEETS THE ART OF WARFARE