Enabling Effective Human-Robot Collaboration via Trust-Driven Decision-Making

by

Shreyas Naveenchandran Bhat

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2026

Doctoral Committee:

Associate Professor Jessie Yang, Co-Chair
Professor Cong Shi, Co-Chair, University of Miami
Assistant Professor Patricia Alves-Oliveira
Professor Brian Denton
Dr. Joseph Lyons, Air Force Research Lab

Shreyas Naveenchandran Bhat
shreyasb@umich.edu
ORCID iD: 0000-0001-6052-3178

© Shreyas Naveenchandran Bhat 2026

DEDICATION

To my parents, loved ones, and friends.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my co-advisors, Dr. Jessie Yang and Dr. Cong Shi. Their complementary expertise, Dr. Yang's in human factors and Dr. Shi's in operations research, significantly shaped the trajectory of my research in human-robot interaction. Throughout the five years of my PhD, our weekly meetings provided me with invaluable opportunities to discuss ideas, analyze results, and delve into the intricate details of my work. Their guidance, encouragement, and commitment to my growth as a researcher were instrumental at every stage of my doctoral journey.

I am also deeply grateful to Dr. Joseph Lyons, with whom I met monthly. Your deep expertise in human psychology was invaluable during the design and analysis phases of my experiments. Our insightful discussions played a crucial role in interpreting the human-subjects data and in determining how and what conclusions could be responsibly drawn from the analyses.

I would also like to thank my committee members, Dr. Brian Denton and Dr. Patricia Alves-Oliveira, for their helpful suggestions and constructive feedback, which significantly improved the overall quality of my work. Thank you Dr. Yili Liu, Dr. Ruiwei Jiang, and Dr. Joyce Chai for being on my dissertation proposal committee and providing valuable insights early on in my research journey that helped shape my dissertation. I am deeply grateful for Dr. Marina Epelman and Matt Irelan for always keeping me informed about how to best navigate the doctoral journey.

Thank you to my labmates — Yaohui, Jin, Doowon, Hyesun, Patrik, Drew, Rongzhi, Yuanchen, Jennifer, Na, Ruikun, Taa, Jundi, Shaoze, Yutong, Catherine, Jose, Mann, Stacy — who always made me feel like a part of the family at the lab. Our many insightful discussions and collaborations not only enriched my research but also made the journey enjoyable and rewarding. I am grateful for their camaraderie, support, and the vibrant research community we shared. Thank you Catherine for helping me run human-subject experiments while I was away.

I am also thankful for all my friends in Ann Arbor who helped me relax and relieved the stress from research work. Without your companionship and encouragement, this experience would have been much more challenging.

Finally, I am truly grateful for the unwavering love and support of my family. Thank you to my mom and dad, without whom I never would have made it this far in life. To my girlfriend, Rucha Joshi, thank you for helping me keep sane, for regularly persuading me to take breaks from research, and for being my steadfast partner through it all. Without you, I would have certainly burnt out during the last phase of my research.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGMENTS iii

LIST OF FIGURES viii

LIST OF TABLES xi

LIST OF ACRONYMS xii

ABSTRACT xiii

CHAPTER

1 Introduction **1**

1.1 Dissertation Overview 1

2 Related Work **4**

2.1 Human-Autonomy Teaming 4
2.2 Qualitative and Snapshot Modeling of Trust 5
2.3 Dynamic Trust Modeling 6
2.4 Trust-Aware Planning 7
2.5 Value Alignment 9

3 Framework for Trust-Aware Decision-Making **11**

3.1 Introduction 12
3.2 Problem Formulation 13
3.2.1 Trust Aware Decision-Making 13
3.3 Experiment 18
3.3.1 Participants 18
3.3.2 Testbed 18
3.3.3 Measures 19
3.3.4 Experimental procedure 21
3.4 Results & Discussion 21
3.4.1 Immediate Actual Reward as a Performance Metric 21
3.4.2 Clustering of Trust Dynamics 22
3.4.3 Association between Personal Characteristics and Type of Trust Dynamics 24

3.5 Conclusion . . . 26
4 Effect of Personalized Value Alignment . . . 28
4.1 Introduction . . . 29
4.2 Problem Formulation . . . 31
4.2.1 Human-Robot Teaming Task . . . 31
4.2.2 Partially Observable Markov Decision Process . . . 31
4.2.3 Bayesian Inverse Reinforcement Learning . . . 32
4.3 Simulation Study . . . 33
4.3.1 Threats and Threat Levels . . . 33
4.3.2 Simulating Human Decision and Feedback . . . 33
4.3.3 Results - Simulation Study . . . 34
4.3.4 Analyzing the Regions . . . 37
4.4 Experiment . . . 46
4.4.1 Testbed . . . 46
4.4.2 Participants . . . 46
4.4.3 Experiment Design . . . 47
4.4.4 Measures . . . 48
4.5 Results . . . 49
4.5.1 Trust, Agreement, and Reliance Intention . . . 50
4.5.2 Performance . . . 52
4.5.3 Workload . . . 53
4.6 Discussion and Conclusion . . . 54
4.6.1 Value Alignment with an Informed Prior . . . 55
4.6.2 Value Alignment with an Uninformed Prior . . . 56
4.6.3 Implications for a Broader Human-Robot Interaction (HRI) Context 57
4.6.4 Limitations and Future Research . . . 58
5 Refining the Reward Function: Adding State Dependence . . . 59
5.1 Introduction . . . 60
5.2 Mathematical Modeling . . . 61
5.2.1 Trust-Aware Markov Decision Process . . . 61
5.3 Study 1 - Learning the Reward Function . . . 63
5.3.1 Data Collection . . . 63
5.3.2 Data Cleaning . . . 64
5.3.3 Data Analysis . . . 64
5.4 Study 2 - Simulation Study . . . 65
5.4.1 Simulating the Human . . . 65
5.4.2 Setting Threats and Threat Levels . . . 66
5.4.3 Results - Simulation Study . . . 67
5.5 Study 3 - Studying the Effect of State Dependence of Reward Weights . . . 69
5.5.1 Testbed and Scenario . . . 69
5.5.2 Robot Interaction Strategies . . . 72
5.5.3 Active Threat Selection . . . 72
5.5.4 Experiment Design . . . 73

5.5.5 Data Collection 73
5.5.6 Measures 74
5.5.7 Results 75
5.6 Conclusion 79
5.6.1 Limitations and Future Work 81
6 Conclusion, Limitations, and Future Research Directions 83
6.1 Conclusion 83
6.2 Limitations and Future Research Directions 85
6.2.1 Further Personalizing Trust Dynamics Models 85
6.2.2 Tackling Diverse Human-Robot Collaborative Tasks 86
6.2.3 Tackling Multi-Action Scenarios 86
6.2.4 Shifting to the Multi-Robot Multi-Human Paradigm 87

APPENDIX 89

BIBLIOGRAPHY 94

LIST OF FIGURES

FIGURE

3.1 Graphical Representation of our POMDP 14
3.2 The testbed developed in the Unreal Engine 4 game engine 19
3.3 The four outcomes based on the presence of threat inside a site and the choice of action by the participant 20
3.4 Trust feedback with overlay of red and green triangles representing the rewards. Green triangles show that the participant would receive a higher reward by following the recommendation and red triangles show that the participant would receive a higher reward by not following the recommendation 22
3.5 Clustering of participants according to their trust dynamics. The features used are the root mean squared error between our predictions and their feedback and their average log trust. As evidenced by the elbow in the variance plot and the maximum in the silhouette scores plot, we chose $k = 3$ as the optimum number of clusters. There are 31 Bayesian Decision Makers, 5 Disbelievers and 9 Oscillators. 23
3.6 Three types of trust dynamics over an interaction period of 100 sites. The blue curve shows the reported trust feedback and the orange curve shows the model predicted trust. The blue points represent the sites at which reported feedback was used to train the model. Finally, the orange dots represent points sampled from the model's beta distribution. 24

4.1 The observed regions in end-of-mission trust as a function of the health weights of the human w_h^h and the robot w_h^r. The figure on the left is the simulation result when there is a relatively low chance of threat presence at any search site ($d = 0.3$). The figure on the right is when there is a higher chance of threat presence at any search site ($d = 0.7$). 34
4.2 The effect of the prior probability of threat presence in any house d on the end-of-mission trust, after fixing a trust region. 35
4.3 The effect of prior probability of threat presence in any house d on the end-of-mission trust when the human and the robot are both extremely risk-averse . . 36
4.4 Comparing the adaptive strategy with the non-adaptive strategy in the end-of-mission trust feedback given by the simulated human for two different levels of threat. The non-adaptive strategy sets the robot's health reward weight to 0.5 . 37
4.5 Comparing the adaptive strategy with the non-adaptive strategy in the end-of-mission trust feedback given by the simulated human for two different levels of threat. The non-adaptive strategy sets the robot's health reward weight to 0.5 . 38
4.6 Plot of the one-step expected rewards for the robot 40

4.7 A graph representing the penultimate stage of decision making. Here, the trust decreases as we go from top to bottom in either of the two stages. The transition probabilities are a function of the health weight of the human w_h^h and the action a. 41
4.8 The 2-stage decision-making problem at the i^{th} search site. Here, again, the trust decreases from top to bottom 43
4.9 The recommendation interface 47
4.10 The trust feedback slider used to get feedback from the participants after every search site. The mission timer is paused when the slider is shown to let the participants take their time in adjusting their trust. 48
4.11 Exp 1 – Post-mission reliance intentions 51
4.12 Exp 2– Average trust $\frac{1}{M}\sum_{i=1}^{M} t_i$ 52
4.13 Exp 2– End-of-mission trust t_M 53
4.14 Exp 2– Post-mission trust questionnaire 54
4.15 Exp 2– Number of agreements 55
4.16 Exp 2– Post-mission reliance intentions 56
4.17 Exp 2: Responses on the NASA TLX scale 57

5.1 A sample query asked to an Amazon Mechanical Turk (MTurk) worker. The worker was shown this information and asked whether s/he would choose to use the Robotic Armored Rescue Vehicle (RARV) or not. 63
5.2 Sample logistic curve learned from the collected data 65
5.3 Heatmaps showing (a) raw data of learned health reward weights at each queried state and (b) the smoothed function for the state dependence of reward weights 66
5.4 Counts of states visited for the two robot strategies 67
5.5 Comparison of trust reported by the simulated human for the two robot strategies. Here, trust ranges from 0 to 1. 68
5.6 Comparison of the performance of the team, as measured by the sum of the health points and the time points 69
5.7 Comparison of the health and time points across the interaction for the two robot strategies 70
5.8 Interfaces to let the participants choose their (a) Initial Action Choice before receiving a recommendation from the intelligent agent and (b) Final Action Choice after receiving the recommendation. 71
5.9 Four possible outcomes seen by the participants depending on their choice of action and whether a threat was present inside the search site. 71
5.10 Interfaces to let the participants rate their trust and performance feedback on the intelligent agent's recommendations. The trust feedback slider converts to a value between 0 to 100 with a stepsize of 2. 72
5.11 Trust measures captured in the testbed: (a) Trust feedback given by the participants averaged over the 10 search sites, (b) Trust feedback given by the participants after searching the last search site. 76
5.12 Trust measured using a subjective trust questionnaire [75] after completing a mission. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability. 77

5.13 Reliance intentions measured using a 6-item 7-point Likert scale mentioned in Lyons et al. [63]. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability. 78
5.14 Agreement between the participants' initial choice of action with the robot's recommendation, averaged over the interaction period. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability. 79
5.15 Agreement between the participants' final choice of action with the robot's recommendation, averaged over the interaction period. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability. 80
5.16 Percent total points lost over the interaction period, taken with respect to the total points the soldier starts with (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability. 81
5.17 Subjective performance rating of the intelligent agent captured in the testbed, averaged over the interaction period (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability. 82

LIST OF TABLES

TABLE

3.1 Mean and Standard Deviation (SD) of personal characteristics between the three different trust dynamics (BDM = Bayesian Decision Maker) 24
3.2 Mean and Standard Deviation (SD) of post experiment metrics between the three different trust dynamics 25

4.1 Transition probabilities for the trust of the human depending on the health weight of the human w_h^h. Here, $\hat{d}_i$ is the probability of threat presence after scanning site i 39
4.2 Region numbers and their boundaries 44
4.3 Mean and standard deviation (SD) of dependent measures for the three interaction strategies in Experiments 1 and 2. NL - Non-learner, NAL - Non-Adaptive Learner, AL - Adaptive-learner. All values are Mean ± SD 50

5.1 Health and Time Loss 70

LIST OF ACRONYMS

MDP Markov Decision Process

POMDP Partially Observable Markov Decision Process

IRL Inverse Reinforcement Learning

CIRL Cooperative Inverse Reinforcement Learning

RARV Robotic Armored Rescue Vehicle

ISR Intelligence, Surveillance, and Reconnaissance

HRI Human-Robot Interaction

HAT Human-Autonomy Team(ing)

MTurk Amazon Mechanical Turk

DARPA Defense Advanced Research Projects Agency

ARMAV Auto-Regressive Moving Average Vector

OPTIMo Online Probabilistic Trust Inference Model

TACtiC Trust-Aware Conservative Control

VR Virtual Reality

ANOVA Analysis of Variance

AIC Akaike Information Criterion

ABSTRACT

With rapid advances in Artificial Intelligence and Robotics, we are quickly moving away from the notion of using automation as tools by human operators to the notion of teaming up with autonomous agents to perform collaborative tasks. For such teams to work efficiently, these agents need to be able to adapt their behavior according to the changing trust of their human counterparts. Many mathematical models of trust have been developed for predicting the dynamic trust of a human on the robot. However, the question of how to integrate these models into the decision-making system of a robot is an understudied problem. Most work in this domain has also been focused on a human-supervisor-robot-worker context in which the human can interfere with the robot's behavior if they lose trust in it. Our focus is on the robot-recommender-human-decision-maker context where the robot acts as a decision support system for the human. To provide this trust-based adaptability for robots, a trust-aware decision-making framework is proposed in Chapter 3 of this dissertation, wherein we model the human-robot interaction as a Partially Observable Markov Decision Process (POMDP) with trust as the hidden state variable. We demonstrate the framework in a simulated Intelligence, Surveillance, and Reconnaissance (ISR) mission with a human-subjects experiment. Equipped with a reverse-psychology model of human behavior, we add a trust-gaining reward term to ensure that the robot makes righteous recommendations.

The term value alignment in the context of HRI refers to the process of ensuring that robotic systems' behaviors, decisions, and goals are consistent with the values, preferences, and ethical principles of the humans with whom they interact. In most prior literature, this has been operationalized as the alignment of reward functions between team members. It is theorized that aligning the reward functions will benefit trust in the robot and team performance, but has rarely been empirically verified. In Chapter 4 of this dissertation, we use Bayesian Inverse Reinforcement Learning (IRL) to enable the robot to learn the human's reward function during interaction. We also propose the Bounded Rationality Disuse model of human behavior allowing us to apply Bayesian IRL to our problem. We empirically show that using such an "adaptive" reward function results in higher trust and team performance when the robot starts with an uninformed prior on the human's reward function.

Although we utilized IRL to learn the reward functions during interaction, we made a couple of rather big assumptions in Chapter 4. Firstly, we assumed that the reward function remains constant throughout the interaction. However, this may not be the case, since humans may change their preferences based on the context of interaction. Secondly, we always start the ISR mission from the same "state". This leads to a limited exploration of the state space. Within this limited exploration, we found that the "non-adaptive" reward function may perform well in terms of trust and team performance, if the robot starts with a good prior. However, these results may not hold when more of the state space is explored. We address these limitations in Chapter 5, where we learn a context-dependent reward function. We then compare a robot using this reward function with a robot using a constant reward function under different starting states. Our results indicate that the context-dependent reward function performs better in terms of trust and team performance in this case.

CHAPTER 1

Introduction

Human Autonomy Teams (HATs) [68] are becoming more and more common in a variety of domains like search and rescue, transportation, package delivery, and healthcare due to their ability to leverage the computational capabilities of autonomous agents and the higher-level understanding of the task from the human teammates. Trust between teammates has been identified as a crucial factor to ensure smooth and effective collaboration [52]. In human-human teams, trust naturally develops between its members and the members are able to utilize their trust on others to make decisions about depending on the other, delegating tasks to the other, and weighing their inputs. In a similar vein, autonomous agents need a mechanism to estimate their human partner's trust on them, and to predict whether or not the human partner will rely on it. Utilizing this ability, the autonomous agent can then adapt its behavior to encourage trust building, discourage overtrust and undertrust [80], and take steps to align its behavior with the values of the human.

Toward this, a number of dynamic models of trust have been developed in recent years [98, 38, 105]. In Chapter 3 of this dissertation, we introduce the trust-aware Markov Decision Process (MDP) framework which can use these models as the transition function to give autonomous agents the ability to predict the trust of their human teammate during interaction. We also use a human trust-behavior model that provides a way for the robot to predict the human's reliance behavior based on their trust. The framework also includes a reward function that the agent tries to maximize during interaction, which in turn, drives the behavior of the agent. In the rest of the chapters, we analyze various human trust-behavior models, and reward functions, and test their effects on human trust and team performance through simulations, mathematical analyses, and human-subjects studies.

1.1 Dissertation Overview

The dissertation takes steps toward trust-aware HRI through the following chapters:

- **Chapter 1 Introduction:** We discuss the motivation behind pursuing this research work. It also provides the structure of the rest of the dissertation.

- **Chapter 2 Related Work:** We discuss the areas of work that are related to research done in this dissertation. We discuss five related areas of Human-Autonomy Teaming, Qualitative and Snapshot Modeling of Trust, Dynamic Trust Modeling, Trust-Aware Planning, and Value Alignment. We provide summaries of seminal works and recent developments in these domains.

- **Chapter 3 Framework for Trust-Aware Decision-Making:** We propose a framework for trust-aware decision-making in a human-autonomy teaming context. We model the interaction as a trust-aware Markov Decision Process with trust as the underlying state variable, actions that can be implemented by the human and recommended by the robot, a dynamic trust model as the transition function, a reward function that encourages gaining trust, and a reverse-psychology based human trust-behavior model. We demonstrate the effectiveness of our approach through a human-subjects experiment conducted using a custom-built high-fidelity 3D game environment in Unreal Engine 4. We propose a reward-based performance metric for the dynamic trust model in a sequential decision-making context and show its effectiveness in capturing the loss/gain in trust. Finally, we find three different types of trust dynamics exhibited by people interacting with decision support agents - Bayesian Decision Makers whose trust varies according to the reward-based performance of the agent, Oscillators whose moment-to-moment trust varies greatly, making it difficult for our model to predict, and Disbelievers, whose trust remains low irrespective of the performance of the agent. We also associate personal characteristics like personality traits and trust propensity to the type of trust dynamics a person may exhibit.

- **Chapter 4 Effects of Personalized Value Alignment:** We study the effect of learning personalized reward functions within our trust-aware MDP framework on trust and team performance. Toward this, we first propose the Bounded Rationality Disuse model of human behavior, which gives probabilities of human action choice based on trust and personalized preferences. We show that by inverting this model in a Bayesian sense, we are able to get insights into the underlying preferences of the human during interaction. We also show that while using this behavior model, the robot inherently prefers high trust states, removing the need to add a trust-gaining reward term in our reward function. This serves two purposes - makes our reward function simpler and makes the robot focus on maximizing the task reward, thus improving performance. Through human-subjects studies, we compare three strategies for the robot behavior

- the non-learner that assumes that the human shares its own reward function, the non-adaptive learner that learns the human's reward function and uses it to predict human behavior but still optimizes its own reward function and the adaptive learner that learns the human's reward function and adopts it as its own, while also using it to predict human behavior. We compare these strategies across trust and team performance in two scenarios - one in which the robot starts with an uninformed prior on the human rewards and the other in which it starts with a data-driven prior. Our results indicate that the adaptive learner strategy results in the highest trust and team performance in the uninformed prior case. In the informed prior case, all three strategies perform similarly. This allows us to conclude that being adaptive and personalized is important for trust and team performance if the robot does not have a good prior on human preferences (rewards). If the robot does have a good prior, further personalization may not be necessary.

- **Chapter 5 Refining the Reward Function: Adding State Dependence:** We postulate that a human's preferences may change depending on the context of interaction. We provide a way to learn a context-dependent reward function in the ISR mission scenario. We learn the reward function through data collected from MTurk and show that it matches the intuitive expectations. We compare two interaction strategies for the robot - one using the constant reward function that worked well in the previous chapter and the other using the context-dependent reward function. As a first step, we do a simulation study to analyze the comparison of these strategies on trust and team performance. We see that interacting with the robot using the context-dependent strategy results in higher trust and better team performance. We empirically verify these findings through a human subjects study. We add another factor of vulnerability in the design of this experiment. Our experimental results align with the simulation results, especially in the high vulnerability context. This shows the importance of learning more fine-grained reward functions when interacting with humans to capture their context-dependent preferences.

- **Chapter 6 Conclusion, Limitations, and Future Research Directions:** This chapter concludes the dissertation and discusses certain limitations of our work, which give rise to promising ideas for future research.

CHAPTER 2

Related Work

This chapter discusses bodies of work that motivate the work in this dissertation. Researchers first started studying human trust in robots qualitatively, analyzing factors that influence it [45, 41]. A multitude of questionnaires [93, 47, 71, 89, 63, 15] were designed to be administered post-interaction to measure trust of the human on the robot. This view, however, seemed limiting since it only gave a snapshot of the human's trust at the end of interaction. In order to ensure effective usage of the autonomous system, an appropriate level of trust in the system is necessary [80, 55] since trust controls the reliance behavior of humans on the systems [80, 30]. This motivated researchers to develop dynamic models of trust capable of estimating the moment-to-moment trust of the human on the system [98, 38, 37]. Using these models, researchers then developed frameworks for trust-aware planning [17, 7] where trust is explicitly modeled within the decision-making system of a robot, enabling the robot to change it's behavior to foster appropriate trust. These trust-aware frameworks typically use a reward function that the robot tries to optimize. A natural question arises - how should we design this reward function? This question gave rise to the field of value alignment [40, 33, 14] wherein the robot uses inverse reinforcement learning [78, 84, 107] to learn this reward function through human demonstrations and preferences. We discuss these bodies of work in detail below.

2.1 Human-Autonomy Teaming

Human-Autonomy Team(ing) (HAT) [68, 77] has attracted significant attention across robotics, artificial intelligence, and human factors domains, motivated by the potential of combining human flexibility and judgment with the efficiency and scalability of autonomous systems. Studies have explored various aspects of HAT, including trust calibration [1], decision support [7], communication protocols [48], and shared situation awareness [16].

Robinson et al. [85] tested human-robot teams vs robot only teams in 16 real-world

missions based on the Defense Advanced Research Projects Agency (DARPA) Subterranean Challenge. Their results indicated that human-robot teams performed significantly better in the presence of errors or unforeseen circumstances that lead to the robot failing without intervention from a human operator. Perez D'Arpino et al. [82] experimentally verified the effectiveness of a human-robot teaming approach versus a direct control approach for teleoperating a robotic arm. Zhao et al. [106] argue for the need for adaptivity of autonomy to individuals within a group and discuss the challenges for collaborative autonomous systems that will interact with various individuals. O'Neill et al. [79] present an excellent review of empirical research conducted in the area of HAT.

2.2 Qualitative and Snapshot Modeling of Trust

Trust is defined as the "attitude that the automated system will help achieve one's goals in a situation characterized by uncertainty and vulnerability" [55]. Researchers studying trust in human-automation and human-robot collaboration started by finding the antecedents on human trust on robots. A central question that researchers tried to answer is "what are the factors that affect human trust in automation and autonomous systems" [45, 90, 41].

Hancock et al. [42] divided human trust on robots into three broad categories - human-related factors, robot-related factors, and environment-related factors. Human-related factors include things like the human's expertise, performance, self-confidence, and skills. Robot-related factors included performance and attributes like anthropomorphism and physical appearance. Environmental factors included the constitution of the team and task features. The authors later extended their work [41] by including newer studies and provided empirical evidence for the inclusion of many of the factors listed in their original qualitative model.

Hoff and Bashir [45] present a systematic review of literature on factors that affect trust. They propose a three-layered model for trust, namely dispositional trust, situational trust, and learned trust. According to their model, dispositional trust represents a person's inherent tendency to trust automation, independent of the context of interaction. Factors that influence dispositional trust include culture, age, gender, and personality traits. Situational trust, on the other hand, explicitly depends on the context of interaction. It depends on factors internal to the human, such as self-confidence and expertise, and also on factors external to the human such as the task difficulty and system complexity. Finally, learned trust represents an operator's evaluations of the specific system through direct interactions with it. Pre-existing knowledge about the system, its performance during interaction, and certain design features influence learned trust in the system.

Although these models of trust provide an understanding of the various factors that affect trust, it is necessary to be able to quantitatively measure trust. Toward this, researchers have developed various scales to measure the subjective level of trust of an individual on automated systems. These questionnaires can be administered to the participants post-interaction to get a value for their level of trust. Muir [75] developed a 4-item scale with slider type answers ranging from a value of 0 (do not trust the system at all) to a value of 100 (trust the system completely). Jian et al. [47] created the "Scale of Trust in Automated Systems" as a 12-item scale with each item being a 7-Point Likert type question. The 20-item Complacency Potential Rating scale developed by Singh et al. [93] measures whether the operator of a system will become complacent while depending on the system.

These scales give an insight into the human's trust on the system after they have interacted with it. However, in order to give the system an insight into the trust, and therefore, into behavior of the human during interaction, researchers felt the need to develop dynamic trust estimation algorithms. We talk about these models in the next subsection.

2.3 Dynamic Trust Modeling

In their pioneering work, Lee et al. [54] proposed an Auto-Regressive Moving Average Vector (ARMAV) model to estimate trust at a given time step based on its value at the previous time step, task performance, and whether an automation failure occurred. The idea of trust being a function of the performance of the automation was later used in many other mathematical models of trust.

Kaniarasu et al. [49] use the number of warnings issued by the system and the number of times the human operator intervenes with the operation of the system as indicators of loss in trust. Later, Desai et al. [26] extended this model to also evaluate increases in trust, by having the operators report their feedback at regular intervals.

Xu et al. [98] presented the Online Probabilistic Trust Inference Model (OPTIMo) that uses a dynamic Bayesian Network to constantly update trust in asymmetric human-robot collaboration. They study a human supervisor, robot worker relationship. The model updates its estimation of trust based on the history of observed interaction experiences.

Floyd et al. [34] propose an "inverse trust metric" using the number of task completions, task failures, and interruptions as the basis for estimating the human's trust on a robot.

Nam et al. [76] study the problem of estimating human trust on a swarm of drones. They find that rather than performance, features such as physical characteristics of the swarm are more important for trust in this context.

Chen et al. [17] present a performance-based trust dynamics model in a human su-

pervisor, robot worker context. Mathematically, their model relates performance to trust causally through a linear Gaussian system. They learn the parameters of the model through human-subject experiments. Their model predicts that trust gradually increases on observing successful actions by the robot, but remains same or decreases slightly if interventions occur.

Akash et al. [2] formulate the human-robot interaction problem as a POMDP with trust as one of the partially observed states. The transition probabilities from low to high trust and vice versa depend on the performance of the recommendations of the robot and the transparency of its interface. Their results indicate that higher transparency is more likely to increase trust when the trust is already low but is also more likely to decrease trust when it is already high. The authors present a similar analysis in the context of human-Autonomous vehicle interactions [1].

Soh et al. [94] propose a Bayesian Gaussian Process based trust model that is capable to predicting human trust across tasks. They conclude that a task-dependent functional trust model can accurately represent human trust across tasks. Trust transfer across tasks can be predicted by using a shared task space representation and update process.

Guo et al. [38, 100] introduce the beta distribution based dynamic trust model. Personalized parameters can be learned for the model through a few training sessions with the robot. The parameters of the beta distribution are adjusted based on the performance of the robot, shifting it to the right on a success and shifting it to the left on a failure. The authors also extend this model to a multi-human, multi-robot context [37]. They consider both direct experiences with a robot and indirect experiences with a robot through their human teammate to update their trust estimates. [25] extend this model to tasks at continuous time scales by using inverse reinforcement learning to learn reward functions and performance metrics.

Rodriguez Rodriguez et al. [86] provide an excellent review of the various mathematical models of human trust in automation. Liu [59] also provides a survey of modeling trust from a Bayesian perspective.

2.4 Trust-Aware Planning

Dynamic trust models allow robots to peek into the real-time trust level of the human. This allows the robot to be able to adapt its behavior to the changing trust of the human. Researchers have put a lot of effort in this area of work in recent years. This section discusses this work in detail.

Xu and Dudek [97] propose a utility-based trust model to estimate the trust of a human as they supervise a robot worker. They evaluate a strategy that locally retrains their robot

planner in case of a small loss in trust and globally retrains it in case of a high loss in trust. Their experimental results indicate that this adaptive strategy results in better performance over a non-trust driven strategy. The authors later extend this algorithm to the Trust-Aware Conservative Control (TACtiC) [99] strategy, enabling robots to engage in trust-seeking behaviors. Their results indicate that robots using this strategy and being mildly conservative result in higher collaboration efficiency.

Floyd et al. [34] use their proposed inverse trust estimation model to adapt robot behavior to elicit a higher level of trust from humans. They demonstrate it in a simulated robotics environment, showing that a robot using their algorithm converges to trustworthy behaviors.

Chen et al. [17, 18] propose a trust-aware framework for human-robot collaboration in the context of human supervising the robot worker. They model the interaction as a POMDP with trust as the partially observed variable. The solution to this POMDP in a pick-and-place task results in behaviors such as the robot picking up a low risk item in order to build trust and then moving on to a high risk item. This results in lower interruptions by the human supervisor when compared with a trustless robot strategy.

Akash et al. [3] discuss a solution to their trust-workload-POMDP [2] by defining intuitive reward functions. They computed near-optimal policies for increasing trust and reducing workload. Their results point to the fact that it may not always be beneficial to increase trust, and decision support systems should rather try to maximize some context-dependent performance metric.

Guo et al. [36] present the use of the Beta distribution trust model [38] in a sequential decision-making scenario for a human-robot team. They use this model to define a trust-aware Markov Decision Process, which gives optimal actions for the robot considering the human's trust and subsequent behavior. Their simulation results indicate the need for a trust-gaining reward term that encourages the robot to suggest righteous actions.

Zahedi et al. [105] present a trust dynamics model and a meta-MDP framework that chooses a robot's behavior depending on the level of trust of the human. They analyze the case where the human's model of the environment may be false leading to sub-optimal trustworthy policies and untrustworthy optimal policies. Their framework generates an optimal policy for the robot that chooses the trustworthy sub-optimal policies when trust is low in order to increase it and chooses the untrustworthy optimal policy when trust is high enough to improve the team's performance.

Liao and Wang [58] develop a trust-based variable impedance controller for human-robot co-manipulation tasks. Through extensive human-subjects testing, the authors show the efficacy of their trust-aware controller in terms of efficiency, agreement, safety, and physical and social HRI factors.

Dubois and Le Ny [29] propose an MDP based framework for adaptive task allocations in human-machine teams based on the human's trust and cognitive workload. They provide simulation results that indicate the emergence of automation behaviors like trust seeking, trust repair after error, and managing the human's workload.

Li et al. [57] propose a linear Gaussian process based trust model for estimating a human's trust in an indoor patrolling task scenario. They employ the Monte Carlo Tree Search method to determine the robot's optimal decision-making. Their Virtual Reality (VR) based experiments confirm that the robot using their policy prioritizes gaining the human's trust to mitigate the long-term costs of collaboration.

2.5 Value Alignment

Value alignment is an emerging and critical concept in the field of HRI. At its core, value alignment seeks to ensure that autonomous robotic systems act in accordance with the values, intentions, and expectations of their human users and broader society.

Hadfield-Menell et al. [40] formally define the value alignment problem as Cooperative Inverse Reinforcement Learning (CIRL). They show that the more traditional framework of apprenticeship learning can be formulated as a CIRL game. Their results indicate that the human acting optimally in isolation may not be an effective way to teach the robot. They show that under this formulation, behaviors such as active teaching, active learning, and communicative actions become optimal.

Fisac et al. [33] discuss a solution to the CIRL game based on established models of cognition and theory of mind. Under this solution, the human thinks pedagogically, choosing actions that give the most information to the robot about the underlying reward function. The robot, in turn, expects this behavior and acts pragmatically on the human's behavior. This enables the robot to learn the reward function quickly and efficiently.

Milli et al. [73] compare a robot that completely abides by the human's literal order with a robot that instead behaves according to its estimate of the human's underlying preferences. They use simulations to compare how much more reward the human would get if the robot directly followed the human's orders vs if the robot used an estimate of the human's preferences. Their results indicate that 1) when a human is not rational, a robot should not directly obey their commands, 2) The optimal robot obeys only optimal commands from the human, and uses the estimate of the reward features to drive its behavior otherwise.

Liu et al. [60] present a Bayesian approach to goal inference through the human's motion patterns in a human-robot collaborative task. The robot then re-plans its own motion according to these predictions. Their results indicate that the combination of goal inference

and dynamic task planning significantly improves both objective and perceived performance of the human-robot team.

A bidirectional value alignment problem is studied by Yuan et al. [104]. In their study, the human knows the true reward function and behaves accordingly while interacting as a supervisor to a group of worker robots. The robots try to learn this true reward function through correctional inputs to their behavior from the human. The human, on the other hand, tries to update her belief on the robot's belief of the true reward function and inputs corrections to their behavior accordingly. They compare the degree of alignment of human estimates of the robot's value function and the degree of alignment of the robot's value function to the true value function. Their results reveal evidence of a bidirectional value learning behavior from the human and the robot.

Christiano et al. [20] propose an algorithm to learn from human preferences and show that it can be used to "solve" reinforcement learning tasks in which the robot's goal is to minimize cost to reach a goal state. They show that this can be done for complex tasks within an hour of a human's time. Additionally, they show that by incorporating human preferences, the robot can learn more efficiently than using traditional deep reinforcement learning methods.

Another related area of research is that of active learning [31] which focuses on the robot's ability to proactively query or interact with humans to improve its understanding of the human's underlying preferences. In contrast to passive data collection, active learning empowers robotic systems to selectively seek information that maximally reduces uncertainty in their models, thereby accelerating learning with fewer examples.

Sadigh et al. [87] present an algorithm to actively generate queries to understand the underlying preferences of humans for how a dynamical system should act. They then learn the reward by maximally removing the volume of possible reward weights based on the answers to these queries. Their results indicate that their algorithm converges faster to the desired reward compared to approaches that are not active or that do not synthesize queries in an autonomous driving domain and also empirically validate their experiments in a user study.

Basu et al. [6] generate active hierarchical queries for efficiently eliciting human preferences in the HRI domain. Their approach allows for learning reward dynamics, instead of just focusing on time invariant reward functions. Human-subject experiments show that their learned reward dynamics match human preferences accurately.

Biyik et al. [12] present batch active preference-based methods to efficiently learn human preferences from a small number of queries. They also focus on balancing the query generation time and the number of queries needed to learn the preferences.

CHAPTER 3

Framework for Trust-Aware Decision-Making

Trust has been identified as a crucial factor to ensure smooth and efficient collaboration in teams. In human-human teams, trust evolves naturally and the team members are able to effectively gauge and adjust their trust on others. In HAT, however, the autonomous agent needs an explicit model of human trust in order to gain an insight into the human's behavior. Toward this, a number of quantitative models of trust have been developed in recent research. In this chapter, we provide a framework for integrating these trust models into the decision-making framework of an intelligent agent. We model the human-autonomy interaction as a trust-aware POMDP with the trust of the human on the agent as the state variable. Therefore, the trust dynamics model acts as the state transition function. In order for the robot to plan around the human's behavior through their trust, we also add a human trust-behavior model into our formulation. In this chapter, we use the reverse psychology model of human behavior that states that the human is likely to accept a recommendation from the intelligent agent with a probability equal to their current level of trust on the intelligent agent. With the remaining probability, the human does the opposite of what was recommended (thus expressing reverse psychology). We add a trust-gaining reward term to the reward function in our trust-aware POMDP framework for encouraging the robot to make righteous recommendations. We also introduce a reward-based performance metric to judge the performance of the intelligent agent's recommendations.

We demonstrate our framework in an ISR mission context through a human-subjects study. From the collected data, we find three distinct clusters of trust dynamics types and categorize them as Bayesian Decision Makers, Oscillators, and Disbelievers. We associate certain personal characteristics (like personality traits, perfect automation schema, etc.) to the type of trust dynamics a person would likely exhibit.

This chapter is based on the work published in [7]. The remainder of the chapter is organized as follows. Section 3.1 introduces the research area and provides some background; Section 3.2 formulates the trust-aware decision making problem in an POMDP framework

and describes the trust-behavior model as well as the trust-seeking reward function; Section 3.3 introduces a reconnaissance mission as a relevant human-autonomy teaming use case to examine how different settings affect the interaction; Section 3.4 reports the results and our observations; Section 3.5 summarizes our findings and discusses the limitations and future directions.

3.1 Introduction

Human-robot collaborations are becoming more prevalent in a range of fields including package delivery, warehouse management, search and rescue, transportation, and healthcare. To facilitate effective human-robot collaboration, trust has been identified as a key factor. In order to enable trustworthy human-robot interaction, substantial research efforts have been devoted to identifying factors that influence humans' trust in robots [41], developing computational models for trust estimation [98, 38], and developing trust-aware decision making [18, 36].

Despite these studies, existing research on trust in human-robot interaction is subject to several major limitations. A majority of prior studies that focused on trust in automation adopt a *snapshot* view of trust [101]. Trust is measured via questionnaires, typically administered at the end of the experiment. As it is challenging to repeatedly administer trust surveys during the normal functioning of the autonomy, this snapshot view of trust remains inadequate in providing the robotic agent with the moment-to-moment trust of the human.

More recent developments have tried addressing this research gap via computational models of trust, capable of estimating moment-to-moment changes in trust as a human repeatedly interacts with an autonomous agent [98, 38]. The models in these studies require the definition of a clear binary performance measure, i.e., the autonomy is either correct or wrong. Due to this requirement, these models are used in episodic decision-making scenarios, for example, in a search and rescue scenario where the autonomy detects if a victim is present or not at a site independently of other search sites. In such scenarios, at every site (independent of other sites), the autonomy's performance is considered correct for true positive and true negative detection, and wrong for false positive and false negative detection.

However, such models cannot be directly applied in a sequential decision-making scenario wherein the robotic agent needs to perform complex trade-off decisions to maximize the cumulative reward, and hence, the autonomy correctness is more difficult to quantify.

In addition, previous studies have revealed the existence of different types of trust dynamics [38, 70]. In [61], the authors found two types of clusters called followers and preservers depending on the trust-dependent behavior of the participants. In [38], the authors found

three types of trust dynamics in an *episodic* task setting. We use similar clustering features and find the same clusters in a *sequential* task setting. Moreover, we associate personal characteristics with the type of trust dynamics, which has not been done in prior research.

In this study, we propose a POMDP framework with trust of the human on its autonomous partner as a state variable. Incorporating a trust update model allows our autonomous agent to explicitly consider trust and in turn, human behavior in its decision-making. We introduce a reward-based performance metric to drive the trust estimation algorithm.

Finally, from data collected through human-subject experiments, we analyze how human trust evolves with their earned reward over repeated interactions with autonomy. We find three distinct types of trust dynamics through $k-$means clustering analysis and examine associations between personal characteristics and the type of trust dynamics.

3.2 Problem Formulation

In this section, we describe the framework for modeling and incorporating trust in the decision-making system of a robotic agent. The agent provides recommendations to their human partner about the action that they should take, but the final decision of action selection lies with the human. We use the trust dynamics model described in [38] with some modifications to suit our problem. The recommendation system of the agent is modeled as a finite horizon POMDP with trust as the state variable. The agent solves an optimization problem to maximize the expected cumulative future reward. The specific scenario we target is an ISR mission in which a human soldier teams up with an *intelligent* drone to search through a town for the presence of threats. The recommendation system guides the soldier on whether s/he should breach a site directly or deploy a RARV. Using the RARV prevents any health loss to the soldier in the presence of a threat, but it takes additional time to deploy the RARV each time. On the other hand, breaching a site directly is faster, but the soldier will be harmed if a threat is present inside the site. Here, two natural (but conflicting) goals that arise are to minimize any damage to the soldier while also minimizing the time to search through all the sites.

3.2.1 Trust Aware Decision-Making

We model the ISR task as a trust-aware POMDP (Fig. 3.1) with the objective of the team being to minimize a weighted sum of the health loss to the soldier and the time taken to complete the mission. A trust-aware POMDP is a tuple of the form $(S, A, \Omega, O, H, T, R)$, where S is a set of states, A is a set of actions for the autonomous agent, Ω is the set of

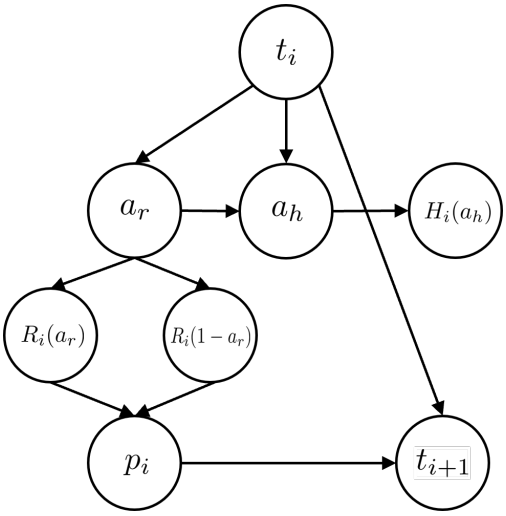


Figure 3.1: Graphical Representation of our POMDP

observations, $O(s)$ is the observation function, H is the embedded human behavior model, $T(s, a)$ is the transition function, and $R(s, a)$ is a reward function that the agent tries to maximize. Details of the definition of our POMDP are given below.

3.2.1.1 States

We use the estimated trust of the human on the robot as the state variable. More specifically, a state is specified by a tuple (α, β). Details of how this translates to the trust of the human are given in the description of transition function below.

3.2.1.2 Actions

At each search site, the two actions available to the autonomous agent are to either recommend to use the RARV or to recommend to not use the RARV.

3.2.1.3 Observations

The two possible observations are failures or successes for the recommended actions.

3.2.1.4 Reward function

We discuss the reward function first since the observation function utilizes the rewards to generate its output. The reward function uses a weighted average of health loss and time. Let h and c be the (constant) health and time losses, respectively. Let w_h and w_c be the weights for the health and time losses, respectively. We define $H_i(a_h)$ to be the immediate

reward given that the human chooses action a_h at site i. Then

$$E[H_i(a_h)] = -(1 - a_h)\hat{d}_i w_h h - a_h w_c c, \tag{3.1}$$

where $\hat{d}_i$ is the probability of threat presence at site i.

Then, we define $R_i(a_r)$ to be the immediate *task* reward at site i, given that the robot's recommendation is a_r. Now, by *fixing* the robot's recommendation to $a_r = a$, via our human behavior model defined in (3.6), we have

$$\begin{aligned} &E[R_i(a_r = a)] \\ =&E[E[R_i(a_r = a)|t_i]] \\ =&E[t_i E[H_i(a_h = a)] + (1 - t_i)E[H_i(a_h = 1 - a)]] \\ =&\hat{t}_i E[H_i(a_h = a)] + (1 - \hat{t}_i)E[H_i(a_h = 1 - a)], \end{aligned} \tag{3.2}$$

where the second equality is due to (3.6) and the third equality holds because $E[H_i(a_h = a)]$ is a constant given a fixed a.

It was noted in simulations that if we only use this *task* based reward, the autonomous agent learns that for most people, trust decreases after failures more easily than increasing after successes [38], and thus it exploits this human behavior by always recommending the opposite action. Trust being low, the human chooses the action opposite to the recommendation, thus increasing performance. We call this behavior of the autonomy as *reverse-psychology*. Since such deceptive behavior is undesirable, we add a trust gain reward to the reward function to incentivize the autonomous agent to make righteous recommendations. Therefore, we define the expected immediate reward at site i with robot's recommendation a as

$$E[\mathbf{IR}_i^a] = E[R_i(a_r = a)] + E[\underbrace{\lambda_i \cdot \mathbb{1}(A)}_{G_i(a_r=a)}], \tag{3.3}$$

where the weight for "trust-gain" is given by

$$\lambda_i = w_t \sqrt{N - i}. \tag{3.4}$$

The term $G_i(a_r = a)$ represents the trust gain reward at site i. We define A as the event when trust increases, i.e., $\mathbb{1}(A) = 1$ if the performance of the autonomous agent is a success and $\mathbb{1}(A) = 0$ otherwise (Note: we define our notion of performance via (3.5)). The parameter λ_i is a weight given to the trust gain reward that decreases with the stage number. The idea behind this is to support trust-gaining behavior near the current stage, and performance

optimizing behavior towards the later stages of planning.

In our experiments, we used the following values for the constants, $h = 100, c = 150, w_h = 0.85, w_c = 0.15, w_t = 10$. We chose these values as they resulted in about 80% threat detection accuracy for the drone in simulations.

3.2.1.5 Observation function

The performance of the recommendations based on the outcome observed defines the observation function here. At each decision stage, the intelligent agent has two, potentially conflicting, goals: to minimize any damage to the soldier and to minimize the time to search through all the sites. Since the autonomous agent's optimal recommendation comes from a reward maximization factoring in both goals, its performance cannot directly be judged by whether its recommendation matches with the presence of threat inside a building. To overcome this, we define the performance of the drone based on immediate rewards earned by the participant.

$$P_i = \begin{cases} 1 & \text{if } \ R_i(a_r) \geq R_i(1 - a_r), \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

At each site i, we observe the presence of threat and compute the realizations of the reward for following the recommendation $R_i(a_r)$, and that for doing the opposite $R_i(1 - a_r)$. We compute the performance of the drone based on these values.

3.2.1.6 Human behavior model

The embedded human behavior model encodes how a human agent responds to actions (or recommendations) made by an autonomous agent. We assume that the probability of the human to accept the recommendation given by the robotic agent is directly proportional to their level of trust. If the human does not accept the recommendation, s/he selects the opposite action of the one that was recommended. More precisely, let a_r and a_h denote the action taken by the autonomous agent and the human agent, respectively.

$$\begin{aligned} \mathbb{P}(a_h = a_r) &= t_i, \\ \mathbb{P}(a_h = 1 - a_r) &= 1 - t_i. \end{aligned} \tag{3.6}$$

Here t_i is the trust level at the i^{th} search site.

3.2.1.7 Transition function

We modify the model described in [38] to use a binary *reward-based performance* for the trust update step. We fit personalized trust parameters $(\alpha_0, \beta_0, w^s, w^f)$ for each participant to model their trust.

$$t_i \sim Beta(\alpha_i, \beta_i), \tag{3.7}$$

$$\hat{t}_i = E[t_i] = \frac{\alpha_i}{\alpha_i + \beta_i}. \tag{3.8}$$

Here, t_i is the trust level at the i^{th} stage. The parameters α_i and β_i are updated as follows, based on the performance value outputted by the observation function (Eq. 3.5).

$$\alpha_i = \begin{cases} \alpha_{i-1} + w^s, & \text{if } P_i = 1, \\ \alpha_{i-1}, & \text{if } P_i = 0. \end{cases} \tag{3.9}$$

$$\beta_i = \begin{cases} \beta_{i-1}, & \text{if } P_i = 1, \\ \beta_{i-1} + w^f, & \text{if } P_i = 0. \end{cases} \tag{3.10}$$

In our experiment, the participants report their trust level on the robot after each interaction (see section 3.3 for details). We use gradient descent on the Bayesian posterior given the performance history and the prior (from [38]) over the model parameters to update our estimates of $(\alpha_0, \beta_0, w^s, w^f)$ in real time after receiving this feedback from the participant. We use the digamma function approximation presented in [74] to approximate the gradients of the beta distribution function.

3.2.1.8 Value Iteration

We solve the POMDP using value iteration on the Bellman Equations. The action at the current state is selected by maximizing the expected reward at the current step summed together with a discounted value of the next state at the next stage.

$$V_i^a = E\left[\mathbf{IR}_i^a\right] + \sum_{(\alpha_{i+1}, \beta_{i+1}) \in S} \gamma \mathbb{P}(\alpha_{i+1}, \beta_{i+1} | \alpha_i, \beta_i, a) V_{i+1}. \tag{3.11}$$

$$V_i = \max_a V_i^a. \tag{3.12}$$

At the final site, the action that gives the maximum immediate expected reward is chosen.

$$V_N = \max_a E\left[\mathbf{IR}_i^a\right]. \tag{3.13}$$

3.3 Experiment

This section describes details of the human-subject experiment. The experiment complied with the American Psychological Association code of ethics and was approved by the Institutional Review Board at the University of Michigan.

3.3.1 Participants

A total of 46 adults participated in the study. One participant's data was discarded as the participant marked all survey questions in the middle and used significantly less time compared to other participants. The remaining 45 participants consisted of 21 females and 24 males (Age: Mean = 22.8 years, $SD = 3.6$ years). The participants were recruited over two phases. In the first phase, we had 31 participants. We modeled their trust dynamics using the model described in Section III. Clustering of the trust dynamics revealed three types with 20, 3, 8 participants respectively. As the numbers of participants in types 2 and 3 were small for statistical analysis, we recruited another 14 participants. For the second phase of participant recruitment, all potential participants filled a pre-experimental survey (see Section 3.3.3 for details) and we selected the ones whose personal characteristics profile were similar to participants identified as type 2 and type 3 in the first phase. Each participant was reimbursed with a base pay of $20 with a bonus of up to $10 based on their performance on the task. The performance was measured by the time taken by the participants to complete the task and the final health level of the soldier.

3.3.2 Testbed

We developed a 3D testbed using the Unreal Engine game development platform. A screenshot of the testbed is shown in Fig. 3.2a. It shows the *intelligent* drone, the soldier, and the RARV. Fig. 3.2b shows a screenshot of the recommendation dialog box where the participant was recommended to not use the RARV. It also shows the two bars showing the health level of the soldier and the time remaining to complete the mission, emphasizing the two objectives the participants need to optimize. Once the participant makes a choice of action, the four possibilities depending on the presence of threat and participant's selected actions are shown in Fig. 3.3. The participants are told that each time they encounter a threat without using the RARV, they will lose 5 points of health while deploying the

RARV takes approximately 10 seconds. They are told to choose an action based on their interaction history and the recommendation from the drone. After exiting each house, the participants are asked to adjust a slider to give feedback on their level of trust on the drone's recommendations.



(a) A view of the 3D environment of the testbed (b) An example of the recommendation GUI

Figure 3.2: The testbed developed in the Unreal Engine 4 game engine

3.3.3 Measures

3.3.3.1 Personality

The big 5 factors of personality (Extraversion, Conscientiousness, Agreeableness, Neuroticism, and Imagination) were measured using the 20-item mini-IPIP scale [27]. This 5-point Likert scale has been widely used in human-robot trust research [65]. This survey was administered pre-experiment.

3.3.3.2 Perfect Automation Schema

Perfect Automation Schema (PAS) was measured using the 7-item scale developed by [72]. Of these, 4 items measure high expectations from the autonomy and the other 3 items measure All-or-none thinking. This was a 7-point Likert scale. This survey was administered pre-experiment.

3.3.3.3 Propensity to Trust

Propensity to trust autonomous systems was assessed using a 6-item scale developed in [71]. This was also a 5-point Likert scale. This survey was administered pre-experiment.



(a) No Threat, RARV Not Used

(b) No Threat, RARV Used

(c) Threat, RARV Not Used

(d) Threat, RARV Used

Figure 3.3: The four outcomes based on the presence of threat inside a site and the choice of action by the participant

3.3.3.4 Moment-to-Moment Trust

During the task, the participants were asked to rate their moment-to-moment trust on the drone by adjusting a slider on a 100-point scale.

3.3.3.5 Post-experiment trust

After the experiment, we used two scales [75, 65] for assessing trust. The first one was a 8-item questionnaire with sliders while we used 6-items from the second scale all of which were 7-point Likert type questions.

3.3.3.6 Workload

Workload was measured using the NASA Task Load Index [43] administered after the experiment. We only used 5 of the 6 items as there was no physical demand from the participants in our experiment. All the items had the participants rate their feelings on a slider with values ranging from Very low to Very high.

3.3.4 Experimental procedure

Prior to the experiment, participants provided informed consent and completed several surveys assessing their demographics, personality traits, perfect automation schema, and propensity to trust automation. They were oriented to the steps of the experiment and walked through each of the screens they would see during the experiment. The two-fold objective of minimizing time and maximizing health was emphasized. Participants were told that the robotic agent was imperfect, but they were not informed of the exact reliability level. They were also told that the robotic agent's recommendations would help them achieve the two-fold objective (i.e., to minimize any damage to the soldier and to minimize the time to search through all the sites). They were informed about the performance-based bonus pay. Participants then proceeded to the experimental trials, wherein they had to search through 100 houses sequentially. After searching each house, the participants were asked to report their level of trust on the autonomous agent's recommendations. The participants took on average 51.5 minutes to complete the task. The average threat detection accuracy of the robotic agent was around 85% across the participants. At the end of the experiment, participants reported their post-experiment trust toward the automated aid using two scales. They also reported their level of workload.

3.4 Results & Discussion

3.4.1 Immediate Actual Reward as a Performance Metric

Since the participants are explicitly told to consider both the soldier's health and the time to complete the search as their objectives, we expect their trust to be correlated with the immediate reward that they receive upon choosing an action. We expect that a participant's trust would be likely to increase if following the recommendation by the drone gets them a higher reward that not following the recommendation and vice versa. Fig. 3.4 shows a representative example. In the figure, a green triangle represents the site at which following the recommendation would result in a better immediate task reward gain ($P_i = 1$) and the red triangles represent the opposite. It is quite clear that a red triangle is often followed by a decrease in trust and a green triangle is followed by an increase in trust. Thus, this reward-based performance metric is able to capture moment-to-moment trust changes of the participant. Using this performance metric in our trust update model, we get a prediction root mean squared error of 0.1266 ($SD = 0.078$) across the participants (Note: The reported trust values were between 0 and 1).

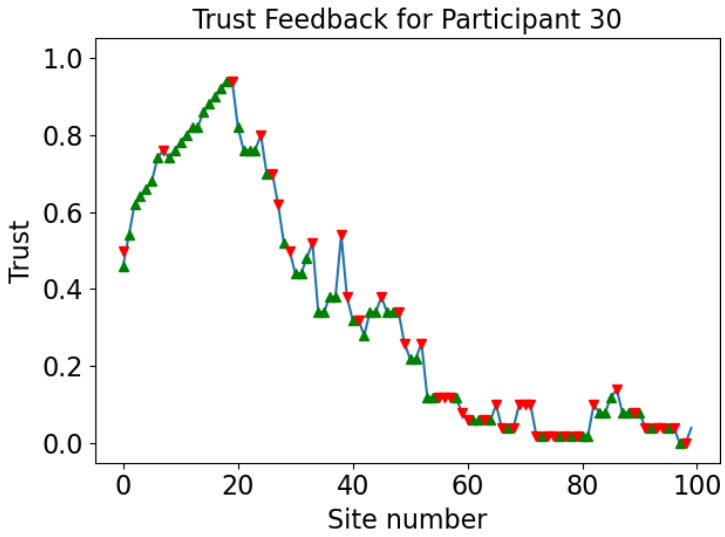
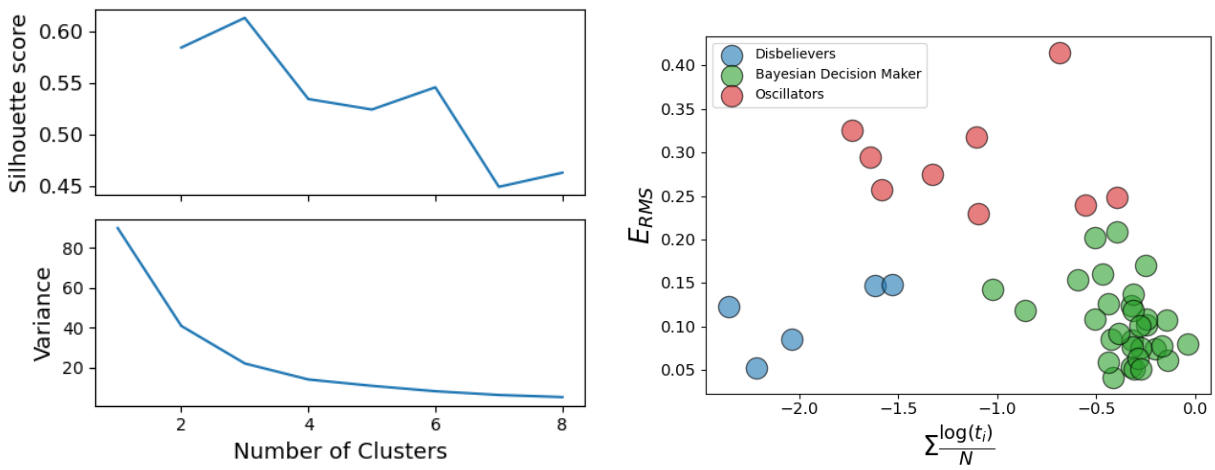


Figure 3.4: Trust feedback with overlay of red and green triangles representing the rewards. Green triangles show that the participant would receive a higher reward by following the recommendation and red triangles show that the participant would receive a higher reward by not following the recommendation

3.4.2 Clustering of Trust Dynamics

We employ k-means clustering to group together participants with similar trust dynamics. Benchmarking a prior study [3], we use two features, namely the average logarithm of trust, and the E_{RMS}, for the clustering analysis. The average logarithm of trust represents the overall trust a human agent has on the robotic agent, which is computed as the logarithm of the trust feedback values over the 100 trials. The RMS error (E_{RMS}) represents the extent to which a human agent's trust updating process is Bayesian. For computing E_{RMS}, we consider the feedback at the first 20 sites as a training set. Thereafter, we use the feedback after every 5 sites to update the parameters of our model. Thus, we compute the E_{RMS} over the last 80 sites.

Fig. 3.5 shows the results of the k-means algorithm. We note elbows in the variance plot at $k = 2$ and $k = 3$ clusters. Employing the silhouette score metric, we see that there is a peak at $k = 3$. We thus choose $k = 3$ as the optimum number of clusters. One of the clusters includes participants with small E_{RMS} and generally higher values of trust. We call this cluster *Bayesian Decision Makers* as their trust is well predicted by our model. The second significant cluster consists of participants whose E_{RMS} values are small but whose trust is generally low. We call this cluster *Disbelievers* because of their low trust value irrespective of the performance of the agent. The third significant group is one with high E_{RMS} values, whose trust cannot be predicted very well with our model. Their trust changes very rapidly



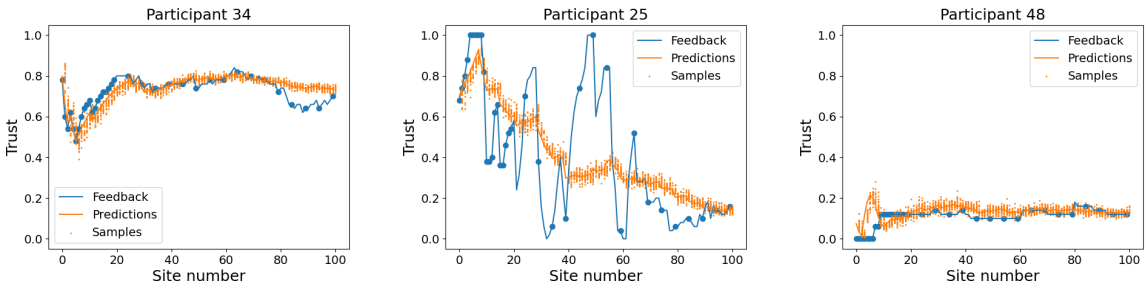
(a) Variances and Silhouette Scores

(b) Clusters

Figure 3.5: Clustering of participants according to their trust dynamics. The features used are the root mean squared error between our predictions and their feedback and their average log trust. As evidenced by the elbow in the variance plot and the maximum in the silhouette scores plot, we chose $k = 3$ as the optimum number of clusters. There are 31 Bayesian Decision Makers, 5 Disbelievers and 9 Oscillators.

from moment-to-moment, thus making it harder to predict. Representative plots of each of these clusters can be seen in Fig. 3.6. Even though we lack the "true" cluster labels, prior empirical studies examining trust in automation and human decision making provide face and external validity for the three types of trust dynamics. Disbelievers have been reported in studies investigating public's trust and acceptance of automated driving - around 40% of a JD power survey correspondents said they "would not ride in an (automated vehicle) *regardless of* what progress is made [46]." Bayesian thinkers, who form accurate prior probabilities and update their belief based on new information, have been observed in everyday cognitive judgments [35]. In addition, the oscillators could be considered as Bayesian thinkers with bounded rationality, who only update their belief based on immediate past history, resulting in significant fluctuation in trust assessment.

We find smaller E_{RMS} for the Bayesian Decision Makers (Mean = 0.093 and SD = 0.04) and Disbelievers (Mean = 0.1 and SD = 0.04), compared to the Oscillators (Mean = 0.26 and SD = 0.05), suggesting our trust dynamics model is able to accurately represent the dynamics of Bayesian Decision Makers and Disbelievers, but is unable to represent the dynamics of Oscillators very well.



(a) Bayesian Decision Maker (b) Oscillator (c) Disbeliever

Figure 3.6: Three types of trust dynamics over an interaction period of 100 sites. The blue curve shows the reported trust feedback and the orange curve shows the model predicted trust. The blue points represent the sites at which reported feedback was used to train the model. Finally, the orange dots represent points sampled from the model's beta distribution.

3.4.3 Association between Personal Characteristics and Type of Trust Dynamics

Table 3.1: Mean and Standard Deviation (SD) of personal characteristics between the three different trust dynamics
(BDM = Bayesian Decision Maker)

Personal Characteristic	BDM	Disbeliever	Oscillator
Extraversion (/20) *	9.5 (3.3)	5.8 (2.8)	11.3 (2.9)
Agreeableness (/20) *	13.5 (2.5)	10.4 (5.0)	14.1 (1.8)
Conscientiousness (/20)	13.1 (2.7)	12.4 (3.0)	12.1 (4.5)
Neuroticism (/20)	7.9 (2.7)	6.8 (3.6)	10.2 (4.7)
Intellect/Imagination (/20) †	11.7 (2.0)	9.8 (1.8)	12.2 (1.8)
High Expectations (/28) **	12.7 (3.9)	6.4 (2.8)	12.4 (4.2)
All or None Thinking (/21)	6.6 (2.9)	6.4 (3.4)	7.1 (3.1)
Trust Propensity (/30) †	20.2 (4.4)	17.2 (4.1)	22.8 (3.2)

$** - p < 0.01$, $* - p < 0.05$, $\dagger - p < 0.1$

In this section, we present significant individual differences between the three identified clusters of trust dynamics. $p < .05$ is considered significant.

We performed one-way Analysis of Variance (ANOVA) between the data of the clustered participants. Results showed significant difference between the three types of trust dynamics in Extraversion ($F(2, 42) = 4.991, p = 0.011$), Agreeableness ($F(2, 42) = 3.276, p = 0.048$),

and the High Expectations facet of the Perfect Automation Schema ($F(2, 42) = 5.752, p = 0.006$). Further, propensity to trust automation ($F(2, 42) = 3.002, p = 0.06$) and intellect/imagination ($F(2, 42) = 2.687, p = 0.08$) seemed to be different. However, these differences did not reach significance.

Table 3.2: Mean and Standard Deviation (SD) of post experiment metrics between the three different trust dynamics

Personal Characteristic	**BDM**	**Disbeliever**	**Oscillator**
Trust (Muir) (/100) ***	65.4 (13.5)	15.8 (9.9)	44.7 (26.1)
Trust (Lyons) (/7) ***	4.5 (0.54)	3.1 (0.6)	3.6 (0.9)
Mental Demand (/100)	39.6 (25.2)	42.0 (36.6)	50.3 (28.6)
Temporal Demand (/100)	50.8 (27.4)	62.0 (24.5)	42.9 (21.3)
Performance (/100)	58.6 (19.7)	50.8 (30.0)	46.2 (31.7)
Effort (/100)	34.3 (23.0)	34.4 (17.4)	49.8 (32.2)
Frustration (/100) *	45.8 (22.2)	58.4 (25.4)	68.1 (14.3)

$***-p < 0.001, *-p < 0.05$

Post-hoc analysis with Bonferroni adjustment test shows that there is a significant difference in Extraversion between the Oscillators and the Disbelievers ($p = 0.009$), with the disbelievers being significantly less extroverted than the oscillators. There seemed to be a trend that the Bayesian Decision Makers were more extroverted than the Disbelievers ($p = 0.061$). However, the trend did not reach significance. Disbelievers had significantly lower expectations from automation compared to both Bayesian decision makers ($p = 0.005$) and oscillators ($p = 0.023$). In the case of Agreeableness, there seemed to be trends that the disbelievers were less agreeable than the other two groups ($p = 0.068$ between disbelievers and Bayesian decision makers and $p = 0.06$ between disbelievers and oscillators). The trends, unfortunately, did not reach significance.

Performing one way ANOVA on the post-experiment measures show significant difference between the three types of trust dynamics in their post-experiment trust reports (Trust questionnaire by Muir and Moray $F(2, 42) = 22.167, p < 0.001$, Trust questionnaire by Lyons and Guznov $F(2, 42) = 15.183, p < 0.001$) and their frustration levels ($F(2, 42) = 4.136, p = 0.023$)).

Post-hoc analysis with Bonferroni adjustment shows that there are significant differences between each of the three groups' trust reports according to the trust questionnaire by Muir and Moray ($p < 0.001$ between Bayesian decision makers and disbelievers, $p = 0.006$ between Bayesian decision makers and oscillators, and $p = 0.009$ between osciallators and

disbelievers with the highest trust for Bayesian decision makers and lowest for disbelievers). Trust reported with the trust questionnaire by Lyons and Guznov only showed significant difference between disbelievers and Bayesian decision makers ($p < 0.001$), and between oscillators and Bayesian decision makers ($p = 0.002$). We believe that a couple of reasons might be behind this difference. One, the oscillators may be inherently unsure about their level of trust on the robot, causing the high variance in their moment-to-moment trust reports. Ultimately, the oscillators may have experienced a state of suspicion in relation to the drone aid [67]. Suspicion represents the combination of high cognitive activity, high uncertainty, and perceived malicious intent [13]. A perusal of the workload data does appear to suggest that the oscillators experienced higher workload relative to the other clusters. This may have been indicative of suspicion for the oscillators whereas the Bayesian decision makers and the disbelievers were more certain of their trust and distrust, respectively. Secondly, the trust measures had some conceptual differences. The Muir and Moray measure focused on elements of trustworthiness (competence, reliability, faith, dependability, etc.) while the Lyons and Guznov measure focused on one's willingness to be vulnerable to the drone aid. It is clear that the three clusters were sensitive to variations in trustworthiness, yet it appears that only Bayesian decision makers were willing to be vulnerable to the drone aid. This has interesting implications for trust modeling given that one's intentions to be vulnerable are a precursor to risk taking.

Results also showed that oscillators were significantly more frustrated than Bayesian decision makers ($p = 0.025$)

3.5 Conclusion

In this study, we developed a framework to explicitly incorporate trust in the decision-making system of an autonomous recommendation system. We formulated the problem of finding the optimal recommendation as a POMDP with an objective function with performance-maximizing and trust-gaining components. We demonstrated the presence of three distinct types of trust dynamics - 1) Bayesian Decision Makers, who vary their trust according to the performance of the autonomy and whose trust stabilizes after repeated interactions, 2) Oscillators, whose trust varies wildly after each interaction, with little to no stabilization even after many interactions, and 3) Disbelievers, whose trust in autonomy is low irrespective of the autonomy's performance. Our trust estimation model is able to predict the trust states of Bayesian Decision Makers and Disbelievers with good accuracy. However, it is not able to capture the moment-to-moment variation of trust of an Oscillator accurately, thus pointing to a requirement to use a different model for people belonging to this category.

Given the value of establishing a robust and dynamic model of trust among individuals interacting with a machine partner, the current study suggests that there may be merit in using individual differences and state measures as a basis to evaluate the feasibility of using such methods. In particular, it would be advantageous to identify who would fall into the three taxonomies of trust dynamics. The results suggest that those individuals classified as Bayesian Decision makers evidenced high expectations of automation. When combined with other state measures such as trust and frustration, one might be able to identify a profile of the Bayesian Decision makers as the combination of individual differences and state measures could be used to parse the sample into the three categories of trust dynamics. Knowing that an individual might fall into one of the categories could influence whether or not a machine partner that is equipped with a dynamic trust model is a feasible solution for that individual.

Results of the study should be viewed in light of the following limitations. First, we assume that the human behaves according to a reverse psychology model. Incorporating a different human trust-behavior model in our POMDP formulation could be a direction for future work. Second, we currently assume that the weights in the reward function are fixed. However, these weights could vary from person to person. For example, one person can be very concerned with protecting the soldier's health, thus having a large w_h, and another person could be more concerned with taking as little time as possible to complete the search, thus having a larger w_c. It can also be possible that the weights are a function of the current health and time. Future research can use our framework, in conjunction with inverse reinforcement learning [78, 40] to define personalized weights in the reward function, thus having personalized metrics for performance of the autonomous agent. Third, we used two features in the clustering analysis. Although these two features were based on prior literature, additional features, such as human agents' decision-making biases [101], could be considered to build high-dimensional clusters in future studies. Fourth, as the number of participants in each cluster was determined post-hoc instead of *a priori*, the sample size was unbalanced. As far as we know, this is the first study to explore the association between personal characteristics and trust dynamics. Utilizing the identified associations, future research could create a more balanced dataset with higher statistical power.

CHAPTER 4

Effect of Personalized Value Alignment

In the previous chapter we introduced the trust-aware decision-making framework. We formulated the human-robot interaction scenario as a trust-aware POMDP with trust as the underlying state, the trust dynamics model as the transition function, a reward function that encourages increasing trust, and a human trust-behavior model to predict the behavior of the human based on their trust.

However, we note a few drawbacks of the reward function and the human trust-behavior model. Firstly, adding the trust-gaining reward term increases the complexity of the reward function, making it harder to study analytically. However, we noted that without the trust gaining reward term, if the robot uses the reverse psychology model for human behavior, it will eventually spiral toward decreasing human trust to reduce the uncertainty of human decisions, since the human will most likely do the opposite of what it was told. Finally, the reverse psychology model is simple in binary action scenarios where it is easy to define what the "opposite" of a recommendation is. In most cases, however, we will have multiple actions available to be recommended by the intelligent agent.

In this chapter we try to address these limitations by introducing the Bounded Rationality Disuse model of human behavior. Under this model, the recommendation system inherently prefers higher trust states and thus we can get rid of the trust-gaining reward terms making our reward function much simpler. This model also allows us to use Bayesian IRL to gain insight into the underlying reward weights of the human that the intelligent agent is interacting with. Thus, we enable a robot to be "adaptive" to the preferences of each human it interacts with.

We first provide a simulation study that shows the effects of aligning the robot's reward function with the human, leading to different "regions" of trust. We provide a mathematical analysis for the emergence of such regions.

We also test this framework in two human-subjects experiments: one in which the robot starts with an unknown prior on human reward weights and the other in which it has a good

data-driven prior on the reward weights. Our results indicate that the adaptive strategy results in higher trust and team performance in the first case. On the other hand, we find that if the robot already has a good prior on the human's reward weights, further personalization may not be necessary.

This chapter is based on work published in [8] and [9] and some unpublished work. The rest of the chapter is organized as follows: Section 4.1 introduces the work and provides background. Section 4.2 details the human-robot team task and formulates our problem as a trust-aware POMDP. Section 4.3.3 presents some initial simulation results and mathematical analysis. Section 4.4 details the human-subjects experiment. Section 4.5 discusses major results and their implications. Finally, section 4.6 concludes our study and discusses limitations and future work.

4.1 Introduction

Robots are increasingly becoming an integral part of our daily lives, marking their presence across varied domains, including healthcare, manufacturing, education, and home assistance, to name a few. As this integration deepens, robots are no longer perceived as tools performing isolated tasks; they are evolving as collaborative partners working with humans. In this human-robot partnership, research into trust between humans and robots becomes increasingly important [91, 11, 19, 103]. Without proper trust, the potential of human-robot teams remains unrealized.

A considerable amount of research has been devoted to developing robots exhibiting trustworthy behaviors, as well as investigating methods for predicting and managing the human's trust in the robot. For instance, one specific area of research focuses on providing explanations of the robots' behaviors [95, 96, 69, 64, 28] typically leading to higher perceived trustworthiness and, subsequently, trust in the robot. Other research directions include the developing real-time trust prediction algorithms [98, 38, 37, 94, 102], modeling trust dynamics [23, 102, 100], developing trust repair strategies [50, 32], and developing trust-aware planning [7, 18, 3, 17, 105, 83].

More recently, the idea of value/goal alignment – aligning the values/goals of robots with those of humans, has garnered significant attention, with the assumption that such alignment would benefit human-robot interaction [88, 104]. Recent literature in value/goal alignment is primarily focused on enabling the autonomous or robotic agent to learn the human's values/goals through preferences [40, 12, 20] or demonstrations [33, 40, 4].

However, there is a lack of research empirically examining and quantifying the effects of alignment on human-robot interaction processes and outcomes. Yet, there are at least

three reasons to suggest that such alignment could be beneficial. First, prior research has illustrated that agent adaptation to humans can enhance performance [5, 62]. Second, agent adaptation could be viewed as the agent being responsive to the human and may, in turn, increase human trust in the agent and enhance team performance [56]. Third, value alignment not only could facilitate trust establishment and enhance team performance, but it is also important for ensuring that machine partners are morally acceptable [53].

This study investigates the effect of real-time, personalized alignment of a robot's reward function to the human's values on trust and human-robot team performance through two human-subject studies. We model the interaction between the human and the robot as a trust-aware POMDP and use Bayesian IRL to estimate the reward weights of the human as they interact with the robot. We compare three types of robot interaction strategies: (1) the non-learner strategy, where the robot presumes the human's reward function mirrors its own; (2) a non-adaptive-learner strategy, in which the robot learns the human's reward function for trust estimation and human behavior modeling, but still optimizes its own reward function; and (3) an adaptive-learner strategy where the robot learns the human's reward function and aligns to it. In addition, we employ different initial conditions for the IRL learning algorithm in the two experiments, one with an informed prior and the other with an uninformed prior.

We test the following hypotheses through data collected from human-subjects experiments:

- **H1:** The adaptive-learner strategy will foster the highest amount of trust followed by the non-adaptive learner strategy and then the non-learner strategy.
- **H2:** Participants will perform better when paired with the adaptive learner strategy compared to the non-adaptive learner strategy, which in turn will perform better than the non-learner strategy.

Results indicate that when starting with an informed prior, personalized alignment to values does not seem to benefit trust or team performance, showing no support for either **H1** or **H2**. On the other hand, when an informed prior is unavailable, aligning to the human's values leads to higher trust, agreement, and reliance intentions while maintaining the same objective team performance but improvements to perceived performance by the human teammate, showing support for **H1** and partial support to **H2**. However, in either of the cases, we do not see any significant difference between the non-adaptive learner and non-learner strategies, showing that the reward weights used in solving the trust-aware POMDP is the most important when it comes to aligning with humans. To the best of our knowledge,

this is one of the few studies, if not the only, that provides empirical evidence for the benefits of value alignment.

4.2 Problem Formulation

This section describes the task for the human-robot team and the mathematical formulation of the interaction.

4.2.1 Human-Robot Teaming Task

We use the same human-robot task as in Chapter 3. The goal for the team is again to minimize damage to the soldier while also minimizing mission time. We also tried to separate the drone from the recommendation system, since the participants in our previous study were sometimes getting confused about the threat detection task and the recommended actions. Instead, we told the participants that an intelligent agent will give you recommendations based on the threat levels reported by the drone. Thus, the drone's task was only to report threat levels and the intelligent agent's only task was to recommend actions. We then ask the participants about their trust on the recommender agent.

4.2.2 Partially Observable Markov Decision Process

Similar to Chapter 3, we model the interaction between the human and the robot as a trust-aware POMDP. We update the reward function and the human behavior model in this study, as detailed below:

4.2.2.1 Reward Function

The rewards for both agents (the human and the robot) are a weighted sum of the negative cost of losing health and losing time. The weights for these costs can be different for the robot and the human. In general, for agent $o \in \{h, r\}$, the reward function can be written as,

$$R^o(D, a) = -w_h^o h(D, a) - w_c^o c(a). \tag{4.1}$$

Here, D is a random variable representing the presence of threat inside a search site, a is the action chosen by the human to implement, $o \in \{h, r\}$ represents the agent, either the human h or the robot r. Note that $h(D, a)$ is a function giving the health loss cost and $c(a)$ is a function giving the time loss cost.

4.2.2.2 Human Trust-Behavior Model

A human trust-behavior model gives the probabilities of a human choosing an action, given the robot's action, their trust level, and other factors such as the human's goals/values. In our study, we make use of the Bounded Rationality Disuse Model of human trust-behavior. This model states that the human chooses the recommended action with a probability equal to the human's current level of trust. If the human chooses to ignore the recommendation, s/he will choose an action according to the bounded rationality model of human behavior. That is, the human will choose an action with a probability that is proportional to the exponential of the expected reward the human receives with that action. Mathematically,

$$P(a_i^h = a | a_i^r = a) = t_i + (1 - t_i)q_a, \tag{4.2}$$

$$P(a_i^h = 1 - a | a_i^r = a) = (1 - t_i)(1 - q_a). \tag{4.3}$$

where t_i is the human's level of trust at the i^{th} interaction and q_a is the probability of choosing action $a \in \{0, 1\}$ under the bounded rationality model [73, 107, 33]. It is given by,

$$q_a = \frac{\exp(\kappa E[R_i^h(a)])}{\sum_{a' \in \{0,1\}} \exp(\kappa E[R_i^h(a')])}. \tag{4.4}$$

Here, κ is called the rationality coefficient of the human, with a higher value indicating a more rational human, and a value of 0 indicating a human that chooses an action at random. Note that this model can be easily extended to the case where multiple actions are possible for the human-robot team. We will just need to sum over all actions in the denominator to get the probabilities.

4.2.3 Bayesian Inverse Reinforcement Learning

We use Bayesian IRL to estimate the reward weights of the human as they interact with the recommender robot. This is done by maintaining a distribution on the possible reward weights and updating it using Bayes' rule after observing the human's selected action. More precisely, if $b_i(w)$ is the belief distribution on the reward weights before the i^{th} interaction, the distribution after the i^{th} interaction, $b_{i+1}(w)$ is given by,

$$b_{i+1}(w) \propto \begin{cases} P(a_i^h = a_i^r | a_i^r) b_i(w), & \text{if } a_i^h = a_i^r, \\ P(a_i^h = 1 - a_i^r | a_i^r) b_i(w), & \text{otherwise.} \end{cases} \tag{4.5}$$

In our formulation, we only learn a distribution over the health reward weight of the

human, w_h^h, and assume that the time reward weight is defined by $w_c^h := 1 - w_h^h$. We use the mean of the learned distribution as an estimate of the human's health reward weight. The Bayesian IRL algorithm requires a prior distribution $b_0(w)$ to get started. We ran the algorithm starting with a uniform prior on previously collected data [7] and generated an "average" distribution that can represent the weights for the general population. In Experiment 1, we use this informed prior for $b_0(w)$ and in Experiment 2, we use the uniform distribution for $b_0(w)$ to simulate a case where we lack data to set an informed prior.

4.3 Simulation Study

As a first step in this research, we designed a study in simulation that looked at the effect of the degree of (mis)alignment of reward weights between the human and the robot on the human's trust. This section describes the details of this simulation study and presents our major results.

4.3.1 Threats and Threat Levels

We follow the following strategy to set the threat levels at the search sites. First, a prior probability of threat presence d is fixed across all sites. Then, at each site independently, threat presence is determined by a Bernoulli distribution with d as the parameter. The drone gets an updated threat level after scanning a site. This updated threat level is generated by a beta distribution with a peak at 0.1 for when a threat is not present and a peak at 0.9 for when a threat is present inside the site.

4.3.2 Simulating Human Decision and Feedback

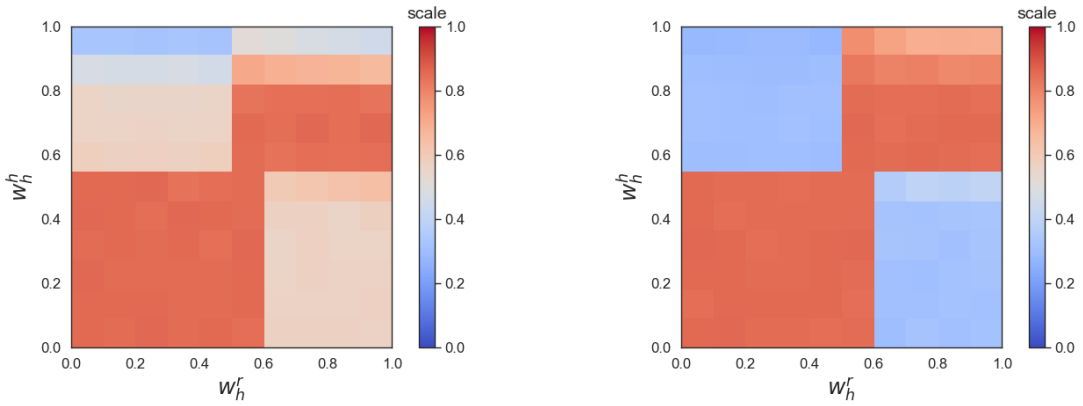
We sample from the set of trust parameters Θ that was generated from a previous study [7] to get $\theta = (\alpha_0, \beta_0, v^s, v^f)$ for the simulated human. These parameters are unknown to the robot and it estimates them via maximum likelihood estimation (for details, refer to [7]). The human also has reward weights $w_h^h, w_c^h (:= 1 - w_h^h)$ which are unknown to the robot. The human gives trust feedback by sampling from the beta distribution trust update model, using the observed rewards associated with the recommendation to judge the performance of the robot. The human chooses his/her action using the bounded rationality disuse model with $\kappa = 1$.

4.3.3 Results - Simulation Study

This section provides details about the major results from this simulation study.

4.3.3.1 Regions in End-of-mission Trust

We observed four distinct regions in the end-of-mission trust as a function of the reward weights of the human and the robot (Fig. 4.1). If a risk-averse human is paired with a risk-taking robot, the end-of-mission trust can get very low (the region in the top-left of the plots in Fig. 4.1) and vice versa. This is especially true when the risk level is higher, as can be seen in the higher difference in trust levels when the values align and when they don't in Fig. 4.1b compared to the same in Fig. 4.1a. In the figure on the left, the probability of threat being present in any site was 0.3 while the same was 0.7 in the figure on the right.



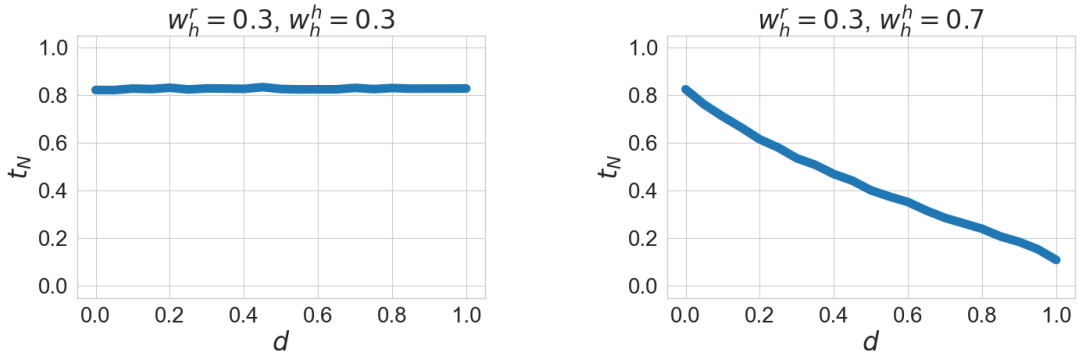
(a) $d = 0.3$ (b) $d = 0.7$

Figure 4.1: The observed regions in end-of-mission trust as a function of the health weights of the human w_h^h and the robot w_h^r. The figure on the left is the simulation result when there is a relatively low chance of threat presence at any search site ($d = 0.3$). The figure on the right is when there is a higher chance of threat presence at any search site ($d = 0.7$).

4.3.3.2 Effect of Threat Level

Since the results from the last section show some potential effects of the level of threat associated with the mission on the end-of-mission trust, we ran another set of simulations, this time, fixing the reward weights of the human and the robot (i.e. the trust region) and varying the level of threat. Fig. 4.2 shows plots of the end-of-mission trust as a function of the prior threat level d for the 4 regions of trust. Further, it is seen that if the values are roughly aligned, trust increases. If the values are not aligned, trust is still high when the

overall risk is low (at low threat levels d). However, the human starts to lose trust in the robot if the values are misaligned and the risk level is high.



(a) Region 1

(b) Region 2

(c) Region 3

(d) Region 4

Figure 4.2: The effect of the prior probability of threat presence in any house d on the end-of-mission trust, after fixing a trust region.

Fig. 4.3 shows an example when both the human and the robot are extremely risk-averse. In this case, although there is alignment in the reward weights of the human and the robot, the level of trust is generally very low. This is because the robot tries to maximize expected reward, but the human is judging performance through observed rewards. At lower risk levels, it would be better to be cautious in the long term, but it is more likely to get better rewards in the short term by being risky. Thus, it seems to the human that the robot is overly cautious, resulting in a decrease in trust.

4.3.3.3 Effect of Adapting to the Human's Rewards

We can update the reward weights of the robot at each time step by using the distribution on the reward weights of the human maintained by the robot. This results in an "adaptive"

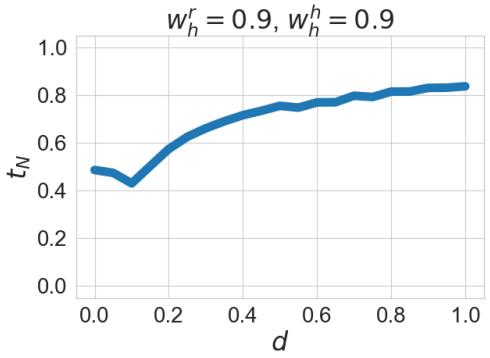
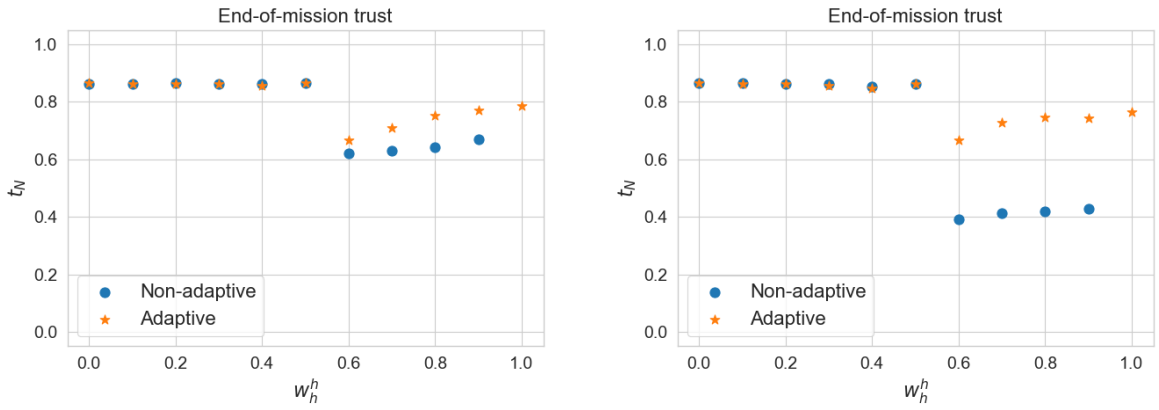


Figure 4.3: The effect of prior probability of threat presence in any house d on the end-of-mission trust when the human and the robot are both extremely risk-averse

interaction strategy in which the robot adapts to human values throughout their interaction. We ran simulations when the robot followed such a strategy, starting with a uniform distribution on the reward weights. We can then compare this adaptive strategy with a non-adaptive strategy of the robot with $w_h^r = 0.5$ (the mean of the initial uniform distribution) to see the effect of adapting to human preferences.

Fig. 4.4 shows the effect of the human's health reward weight on the end-of-mission trust reported by the human at two threat levels: low (0.3) and high (0.7). As can be seen, below $w_h^h < 0.5$, there is no clear difference between trust on the non-adaptive strategy and that on the adaptive strategy. This is because at $w_h^r = 0.5$, the robot has no preference between losing health and losing time. However, in expectation, not using the RARV is always better than using the RARV. Thus, the recommendation matches the human's preferred action (since the human also does not care about losing health when $w_h^h < 0.5$). As a result, trust is high for both non-adaptive and adaptive strategies. On the other hand, however, when $w_h^h > 0.5$, the human prefers to save health over saving time. Thus, the risky recommendations of the non-adaptive strategy reduce trust while the adaptive strategy is able to learn this preference and is able to maintain the human's trust.

Fig. 4.5 compares the two strategies for two different human preferences (preferring to lose health $w_h^h = 0.3$ and preferring to lose time $w_h^h = 0.7$) as the level of threat is changed. Again, we see similar behavior that when the human prefers to lose health, both strategies result in the same level of trust of the human on the robot. On the other hand, when the human prefers to lose time, the non-adaptive strategy's risky recommendations lead to a loss in trust as the level of threat increases. Again, the adaptive strategy is able to learn these preferences resulting in a much higher level of the human's trust, irrespective of the level of threat.



(a) $d = 0.3$ (b) $d = 0.7$

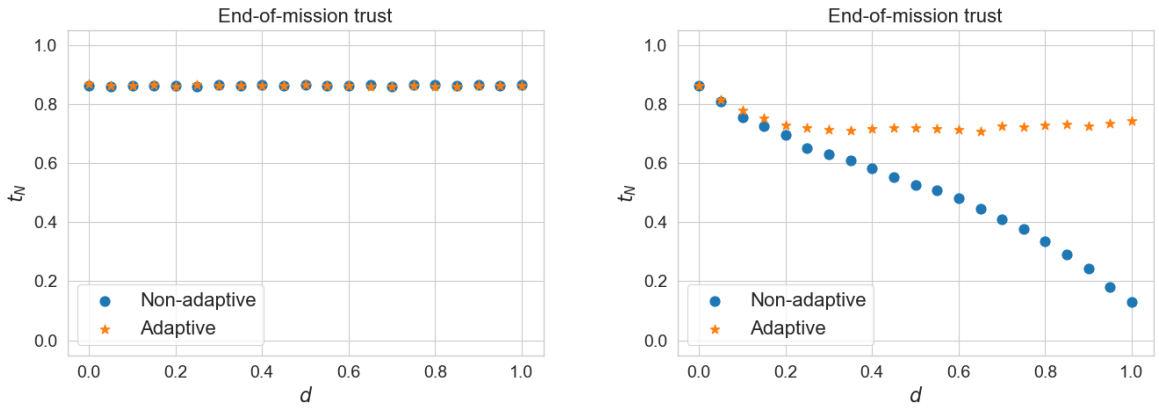
Figure 4.4: Comparing the adaptive strategy with the non-adaptive strategy in the end-of-mission trust feedback given by the simulated human for two different levels of threat. The non-adaptive strategy sets the robot's health reward weight to 0.5

A key takeaway from our results is that, under the bounded rationality disuse model value alignment is, in general, good for trust. However, at low risk settings, value alignment may not be as important for gaining trust. This conclusion is somewhat in line with the finding of [66], wherein the authors found increased reliance on automation in high risk situations compared to low risk situations. Parkhe and Stewart [81] claim that trust is only necessary when the losses can exceed potential gains (i.e. when the risk is high). We take it a step further by saying that value alignment is only necessary for trust when the risks are high. Further, since humans are generally more concerned with observed outcomes rather than expected rewards, the boundaries of the regions can result in lower end-of-mission trust despite having values aligned between the human and the robot.

In light of our results, we claim that completely aligning the robot's values to the human's may not always be the best, especially when the human is extremely risk-averse. This may result in loss of performance *and* loss of trust. On the other hand, not aligning the values at all will certainly lead to a loss in trust, especially in high risk situations. Thus, we postulate that there may be merit in looking for a middle ground that trades off between fully aligning the values and not aligning them at all. This could be a very interesting direction for future research.

4.3.4 Analyzing the Regions

We first look at the behavior of the observed and expected one-step rewards as the weights are varied.



(a) $w_h^h = 0.3$ (b) $w_h^h = 0.7$

Figure 4.5: Comparing the adaptive strategy with the non-adaptive strategy in the end-of-mission trust feedback given by the simulated human for two different levels of threat. The non-adaptive strategy sets the robot's health reward weight to 0.5

4.3.4.1 Human's Observed Immediate Rewards

The immediate rewards observed by the human after searching a site govern the performance of the recommendation. So, analyzing these should give us an insight into the transition function for the trust of the human. The observed rewards by the human after searching site i are,

$$r_i^h = \begin{cases} -(1 - w_h^h)c, & \text{if } a_i^h = 1, \\ -w_h^h h, & \text{if } a_i^h = 0, D_i = 1, \\ 0, & \text{if } a_i^h = 0, D_i = 0. \end{cases} \tag{4.6}$$

Here, a_i^h is the action chosen by the human, D_i represents the presence of threat at site i. Clearly, if $D_i = 0$, the best choice for the human is to choose to not use the armored robot $a_i^h = 0$. When a threat is present however, the best choice depends on the value of w_h^h. If $w_h^h < \frac{c}{c+h}$, the human prefers to lose health over losing time and vice versa. Thus, this fraction $w_h^* := \frac{c}{c+h}$ controls the performance of the recommendation, and in turn, the trust of the human. In other words, if $w_h^h < w_h^*$, the human always prefers to not use the armored robot, even if a threat is present inside the site. On the other hand, if $w_h^h > w_h^*$, the human prefers to not use the armored robot when there is no threat and prefers to use the armored robot when there is a threat present inside the search site. Thus, based on this, we get two distinct transition functions for the human's trust, represented in table 4.1.

Table 4.1: Transition probabilities for the trust of the human depending on the health weight of the human w_h^h. Here, $\hat{d}_i$ is the probability of threat presence after scanning site i

P_i	a_i^r	$\mathbb{P}(P_i \mid a_i^r)$	
		$w_h^h < w_h^*$	$w_h^h > w_h^*$
1	1	0	$\hat{d}_i$
1	0	1	$1 - \hat{d}_i$

4.3.4.2 One-Step Expected Reward of the Robot

We can compute the one-step expected rewards for the robot to recommend to use or not use the armored robot. They are as follows,

$$E[R_i^r(0)] = -d_i \left(t_i + (1 - t_i)p_0\right) w_h^r h - (1 - t_i)p_1 w_c^r c. \tag{4.7}$$

$$E[R_i^r(1)] = -d_i(1 - t_i)p_0 w_h^r h - (t_i + (1 - t_i)p_1)w_c^r c. \tag{4.8}$$

Here, d_i is the updated threat level the team gets after scanning site i, t_i is the trust level of the human before entering site i, p_0 and p_1 are the probabilities as computed in Eq. 4.9

$$p_i^a = \frac{\exp(\kappa E[R_i^h(a)])}{\exp(\kappa E[R_i^h(a)]) + \exp(\kappa E[R_i^h(1-a)])}. \tag{4.9}$$

Claim 1. *If the robot is less concerned about losing health, the one-step expected reward to recommend to not use the armored robot is always higher than that to recommend to use the armored robot, regardless of the risk level.*

In other words, $w_h^r < w_h^* \implies E[R_i^r(0)] > E[R_i^r(1)]$.

Claim 2. *If the robot cares about losing health, then at low risk levels, the one-step expected reward to recommend to not use the armored robot is higher than that to recommend to use the armored robot and the opposite holds when the risk level is high.*

Mathematically, this translates to: if $w_h^r > w_h^*$, then $d_i < \frac{w_c^r c}{w_h^r h} \implies E[R_i^r(0)] > E[R_i^r(1)]$ and vice versa. These claims can easily be proved by subtracting the terms in Eqs. 4.7 and 4.8 and comparing with 0.

Claim 3. *At any stage, the bigger of the two one-step expected rewards increases linearly with trust while the smaller of the two decreases with trust. Further, if the one-step expected reward for action 1 is greater than that for action 0 at some trust level $t > 0$, then it is greater for all $t > 0$ and vice versa.*

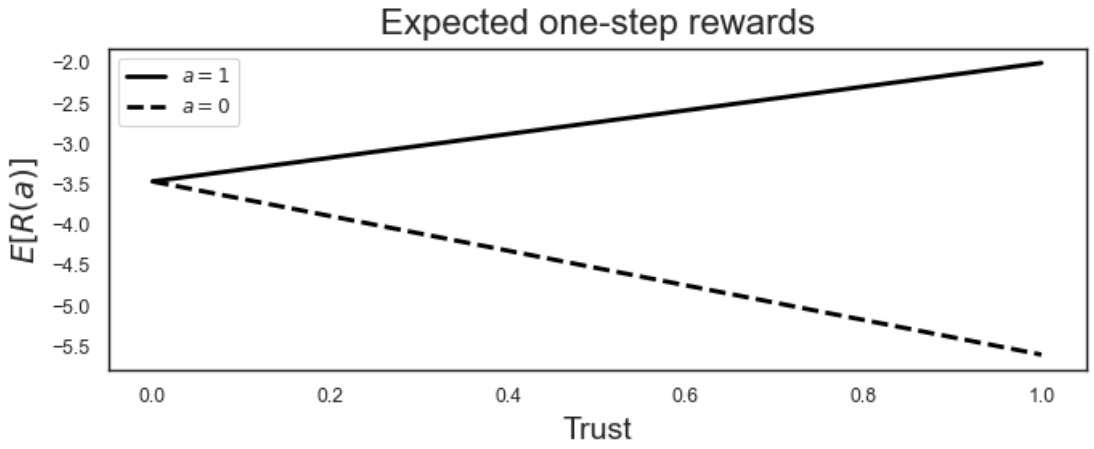


Figure 4.6: Plot of the one-step expected rewards for the robot

Proof. First, we note that both eqs. 4.7 and 4.8 are linear in trust t_i. Thus, we only need to check if the coefficient of t_i is positive or negative. Here, we note that,

$$E[R_i^r(1)] > E[R_i^r(0)] \iff d_i w_h^r h > w_c^r c.$$

This comes from claim 2. So, assuming $E[R_i^r(1)] > E[R_i^r(0)]$ we show that the claim holds true.

$$\begin{aligned}
\text{coeff of } t_i \text{ in } E[R_i^r(1)] &= -w_c^r c + d_i w_h^r h p_0 + w_c^r c(1-p_0) \\
&> -w_c^r c + w_c^r c p_0 + w_c^r c(1-p_0) \\
&= 0, \\
\text{coeff of } t_i \text{ in } E[R_i^r(0)] &= -d_i w_h^r h + d_i w_h^r h p_0 + w_c^r c(1-p_0) \\
&< -d_i w_h^r h + d_i w_h^r h p_0 + d_i w_h^r h(1-p_0) \\
&= 0.
\end{aligned}$$

A very similar argument shows that the same holds true for the other case ($E[R_i^r(1)] < E[R_i^r(0)]$) as well.

Further, by substituting $t_i = 0$ in eqs. 4.7 and 4.8, we can easily see that their values are equal at zero trust. Proving the second part of this claim is then trivial by using the fact they are linear with oppositely signed slopes (see Fig. 4.6). □

An implication of this claim is that the value function at the last decision-making stage increases with trust In other words, at the last stage, the robot prefers states with higher trust levels.

4.3.4.3 Value Function of the Robot

The value function of the robot can similarly be analyzed. It is important to note which transition function is applicable according to the health weight of the human w_h^h. We can make and prove the following claims on the value function of the robot.

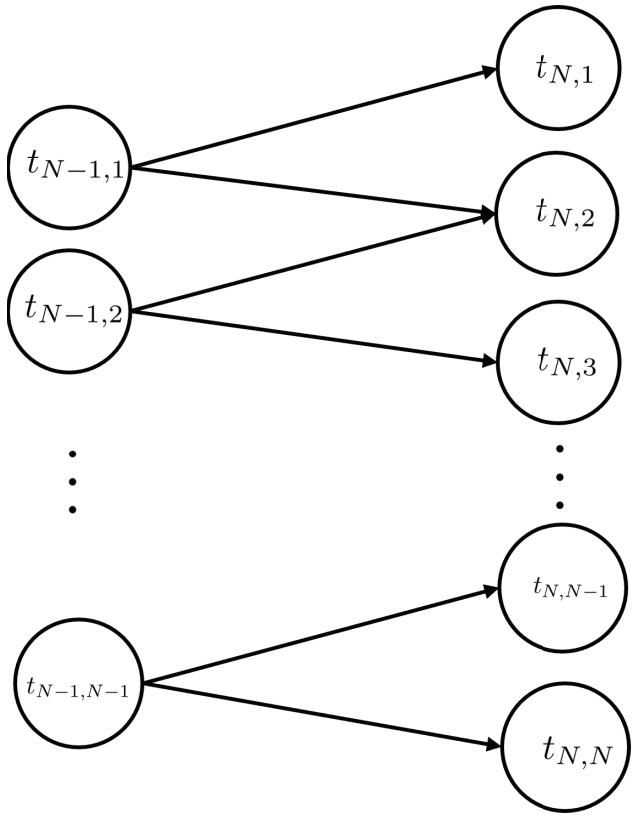


Figure 4.7: A graph representing the penultimate stage of decision making. Here, the trust decreases as we go from top to bottom in either of the two stages. The transition probabilities are a function of the health weight of the human w_h^h and the action a.

Claim 4. *The value function at any stage increases with trust.*

Proof. This can be proved via backward induction. According to the performance-based trust update model we use, at the i^{th} decision stage, there are i possible levels of trust. Let these be denoted by $t_{i,j}, j \in \{1, 2, \cdots i\}$. Without loss of generality, we assume that the following relation holds at every stage i, $k < j \implies t_{i,k} > t_{i,j}$. Fig 4.7 shows how the state transitions work at the $(N-1)^{th}$ stage of decision-making. Since the transition functions are different for regions 1 and 2 and regions 3 and 4 (see Table 4.2 for the definition of region numbers), we complete the proof separately for the two cases.

Case 1: Proof for regions 1 and 2. At the $(N-1)^{th}$ stage, the q-value function can be

written as,

$$Q_{N-1}(t_{N-1,i},0) = E_{t_{N-1,i}}[R^r_{N-1}(0)] + \gamma V_N(t_{N,i}),$$
$$Q_{N-1}(t_{N-1,i},1) = E_{t_{N-1,i}}[R^r_{N-1}(1)] + \gamma V_N(t_{N,i+1}).$$

Case 1a: Suppose $d \leq \frac{w^r_c c}{w^r_h h}$. Then, $E_{t_{N-1,i}}[R^r_{N-1}(0)] \geq E_{t_{N-1,i}}[R^r_{N-1}(1)]$. Here, clearly $Q_{N-1}(t_{N-1,i},0) > Q_{N-1}(t_{N-1,i},1)$. Further, both the terms in the expression for $Q_{N-1}(t_{N-1,i},0)$ increase with trust. Thus, the value function at stage $N-1$ also increases with trust.

Case 1b: Suppose $d > \frac{w^r_c c}{w^r_h h}$. Then, $E_{t_{N-1,i}}[R^r_{N-1}(0)] < E_{t_{N-1,i}}[R^r_{N-1}(1)]$. Further, these are equal at zero trust (claim 3) and $E_{t_{N-1,i}}[R^r_{N-1}(0)]$ decreases with trust while $E_{t_{N-1,i}}[R^r_{N-1}(1)]$ increases with trust. We have

$$\begin{aligned}
V_{N-1}(t_{N-1,i}) &\geq Q_{N-1}(t_{N-1,i},1) \\
&= E_{t_{N-1,i}}[R^r_{N-1}(1)] + \gamma V_N(t_{N,i+1}) \\
&> E_{t_{N-1,i+1}}[R^r_{N-1}(1)] + \gamma V_N(t_{N,i+1}) \\
&> E_{t_{N-1,i+1}}[R^r_{N-1}(1)] + \gamma V_N(t_{N,i+2}) \\
&= Q_{N-1}(t_{N-1,i+1},1),
\end{aligned}$$

and

$$\begin{aligned}
V_{N-1}(t_{N-1,i}) &\geq Q_{N-1}(t_{N-1,i},1) \\
&= E_{t_{N-1,i}}[R^r_{N-1}(1)] + \gamma V_N(t_{N,i+1}) \\
&> E_{t_{N-1,i+1}}[R^r_{N-1}(1)] + \gamma V_N(t_{N,i+1}) \\
&> E_{t_{N-1,i+1}}[R^r_{N-1}(0)] + \gamma V_N(t_{N,i+1}) \\
&= Q_{N-1}(t_{N-1,i+1},0).
\end{aligned}$$

The third inequality uses the latter part of claim 2.

Case 2: Proof for regions 3 and 4. At the $(N-1)^{th}$ stage, the q-value function can be written as,

$$Q(t_{N-1,i},0) = E_{t_{N-1,i}}[R^r_{N-1}(0)] + \gamma((1-d)V_N(t_{N,i}) + dV_N(t_{N,i+1})),$$
$$Q(t_{N-1,i},1) = E_{t_{N-1,i}}[R^r_{N-1}(1)] + \gamma(dV_N(t_{N,i}) + (1-d)V_N(t_{N,i+1})).$$

Case 2a: Suppose $E_{t_{N-1,i}}[R^r_{N-1}(0)] < E_{t_{N-1,i}}[R^r_{N-1}(1)]$. This implies that $E_t[R^r_{N-1}(1)]$ increases with trust and the other one decreases with trust. Again, their values are equal at

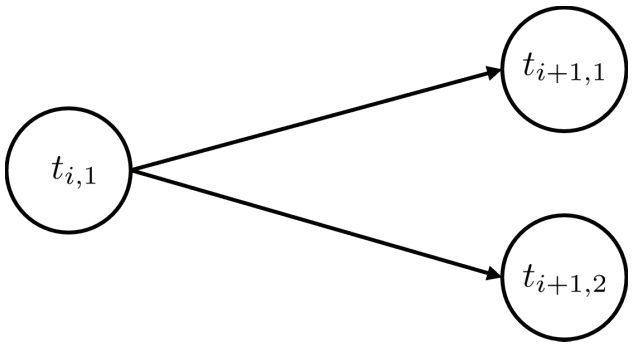


Figure 4.8: The 2-stage decision-making problem at the i^{th} search site. Here, again, the trust decreases from top to bottom

zero trust.

$$\begin{aligned} V_{N-1}(t_{N-1,i}) &\geq Q_{N-1}(t_{N-1,i}, 1) \\ &= E_{t_{N-1,i}}[R^r_{N-1}(1)] + \gamma(dV_N(t_{N,i}) + (1-d)V_N(t_{N,i+1})) \\ &> E_{t_{N-1,i+1}}[R^r_{N-1}(1)] + \gamma(dV_N(t_{N,i+1}) + (1-d)V_N(t_{N,i+2})) \\ &= Q_{N-1}(t_{N-1,i+1}, 1), \end{aligned}$$

$$\begin{aligned} V_{N-1}(t_{N-1,i}) &\geq Q_{N-1}(t_{N-1,i}, 1) \\ &= E_{t_{N-1,i}}[R^r_{N-1}(1)] + \gamma(dV_N(t_{N,i}) + (1-d)V_N(t_{N,i+1})) \\ &> E_{t_{N-1,i+1}}[R^r_{N-1}(1)] + \gamma((1-d)V_N(t_{N,i+1}) + dV_N(t_{N,i+2})) \\ &> E_{t_{N-1,i+1}}[R^r_{N-1}(0)] + \gamma((1-d)V_N(t_{N,i+1}) + dV_N(t_{N,i+2})) \\ &= Q_{N-1}(t_{N-1,i+1}, 0). \end{aligned}$$

Therefore, we have $V_{N-1}(t_{N-1,i}) > \max\{Q_{N-1}(t_{N-1,i+1}, 0) \text{ and } Q_{N-1}(t_{N-1,i+1}, 1)\} = V_{N-1}(t_{N-1,i+i})$.

Case 2b (when $E_{t_{N-1,i}}[R^r_{N-1}(0)] > E_{t_{N-1,i}}[R^r_{N-1}(1)]$) can be proved with a similar argument.

Thus, we have shown that the value function at stage $N-1$ increases with trust, given that the value function at stage N increases with trust. This result can be repeatedly applied to show that the value function increases with trust at each stage. □

An implication of this result is that the N-stage decision-making model can be effectively condensed into a 2-stage decision-making model for easier analysis. Fig. 4.8 shows the 2-stage

decision making problem. This, however, comes at the cost that we only get information about the relative orders of the value function at the two trust states. We lose information about the exact values of the value function. Nevertheless, we can infer some results about the behavior of trust in these regions. Table 4.2 associates a number with a region and defines the boundary of the region.

Table 4.2: Region numbers and their boundaries

Region No.	w_h^h	w_h^r
1	$< w_h^*$	$< w_h^*$
2	$< w_h^*$	$> w_h^*$
3	$> w_h^*$	$< w_h^*$
4	$> w_h^*$	$> w_h^*$

From our mathematical analysis, we can make the following claims about the regions,

Claim 5. *In region 1, the robot prefers to recommend to not use the armored robot irrespective of the threat level. The human also prefers losing health over losing time. Thus, trust increases after every interaction with certainty.*

Claim 6. *In region 2, if $d_i \leq \frac{w_c^r c}{w_h^r h}$, the robot prefers to recommend to not use the RARV. This increases trust with probability 1. However, if $d_i > \frac{w_c^r c}{w_h^r h}$, the robot may choose to either recommend to use or not use the armored robot depending on the discount factor γ, the prior threat level d, and other simulation parameters. Whenever the robot recommends to use the armored robot, trust decreases with probability 1.*

We will now prove the two claims above.

Proof. At search site i, the robot considers the next $N - i$ sites for computing its expected future reward. The q-value function at this stage of decision-making can be written as,

$$Q_i(t_{i,1}, 0) = E_{t_{i,1}}[R_i^r(0)] + \gamma V_{i+1}(t_{i+1,1}),$$
$$Q_i(t_{i,1}, 1) = E_{t_{i,1}}[R_i^r(1)] + \gamma V_{i+1}(t_{i+1,2}).$$

In region 1, since $w_h^r < w_h^*$, $E_{t_{i,1}}[R_i^r(0)] > E_{t_{i,1}}[R_i^r(1)]$. Also, we have shown earlier that $V_{i+1}(t_{i+1,1}) > V_{i+1}(t_{i+1,2})$. Thus, irrespective of the threat level d, $Q_i(t_{i,1}, 0) > Q_i(t_{i,1}, 0)$. Hence, in region 1, the robot always chooses action 0. In this region, since $w_h^h < w_h^*$, the human prefers losing health to losing time. Thus, the recommendation of action 0 always increases trust.

In region 2, since $w_h^r > w_h^*$, $d_i \leq \frac{w_c^r c}{w_h^r h} \iff E_{t_{i,1}}[R_i^r(0)] > E_{t_{i,1}}[R_i^r(1)]$. Thus, if this condition on d_i is satisfied, the robot will choose action 0. Since the human does not care about losing health in region 2, this action leads to a trust increase. If this condition is not satisfied, however, the robot may choose action 0 or action 1 depending on the relative values of the value function and the expected one-step rewards. □

Claim 7. *In region 3, the robot chooses to recommend to not use the armored robot when $d_i < 0.5$ which results in an increase in trust with probability $1 - \hat{d}_i$. When $d_i > 0.5$, the robot may choose either action depending on the simulation parameters.*

Claim 8. *In region 4, the robot chooses the action that gives a higher probability of trust increase when either $d_i < \min\left\{\frac{w_c^r c}{w_h^r h}, 0.5\right\}$ or $d_i > \max\left\{\frac{w_c^r c}{w_h^r h}, 0.5\right\}$. The robot may choose either action depending on the simulation parameters if d_i is in between these limits. This again enables us to give a lower bound on the probability of trust increase.*

We will now prove the two claims above.

Proof. Again, at search site i, we can write the q-value function as,

$$\begin{aligned} Q_i(t_{i,1}, 0) &= E_{t_{i,1}}[R_i^r(0)] \\ &\quad + \gamma((1 - d_i)V_{i+1}(t_{i+1,1}) + d_i V_{i+1}(t_{i+1,2})), \\ Q_i(t_{i,1}, 1) &= E_{t_{i,1}}[R_i^r(1)] \\ &\quad + \gamma(d_i V_{i+1}(t_{i+1,1}) + (1 - d_i)V_{i+1}(t_{i+1,2})). \end{aligned}$$

Here, we have the following relations,

$$\begin{aligned} d_i < \frac{w_c^r c}{w_h^r h} &\iff E_{t_{i,1}}[R_i^r(0)] > E_{t_{i,1}}[R_i^r(1)], \\ d_i < 0.5 &\iff ((1 - d_i)V_{i+1}(t_{i+1,1}) + d_i V_{i+1}(t_{i+1,2})) > \\ &\qquad (d_i V_{i+1}(t_{i+1,1}) + (1 - d_i)V_{i+1}(t_{i+1,2})). \end{aligned}$$

Thus, action 0 (1) is chosen with certainty if $d_i < \min\{0.5, \frac{w_c^r c}{w_h^r h}\}$ ($d_i > \max\{0.5, \frac{w_c^r c}{w_h^r h}\}$). This increases trust with probability $1 - \hat{d}_i$ ($\hat{d}_i$). Here, $\hat{d}_i$ is the posterior probability of threat presence after scanning a site. The action chosen will depend on the relative values of the one-step rewards and the discounted future reward if d_i is between these extremes.
In region 3, since $w_h^r < w_h^*$, $\frac{w_c^r c}{w_h^r h} > 1$. And hence, we only need to compare d_i with 0.5. □

4.4 Experiment

This section provides details about the testbed and the human-subject experiments. The experiments complied with the American Psychological Association code of ethics and were approved by the Institutional Review Board at the University of Michigan. We use the same testbed from Chapter 3, with the modified GUI elements discussed earlier.

4.4.1 Testbed

We slightly modify the testbed used in Chapter 3 for the human-subjects study in this work. We update the recommendation interface to give more information to the participants to enable them to make a more informed decision. As can be seen in Fig. 4.9, now, along with the recommendation, the interface reports the chance of threat presence reported by the drone, and the time to search with and without the RARV. This helps the participants (a) make their own decision, and (b) utilize the recommendation if they need it. This fits directly into the idea of our new human behavior model, wherein humans will choose their own action based on bounded rationality if they do not trust the recommendations enough.

We also update the trust feedback interface (Fig. 4.10) to include more information to help participants make informed decisions about their trust level. This was done to satisfy concerns from a few participants in our first study, who reported that it was slightly difficult to remember the recommendation, the outcome, and their decision after a few repeated interactions. Therefore, we added the threat level reported, recommendation, participant's choice of action, whether threat was encountered, and the time to search to the feedback slider interface. This helped participants with recall of the current interaction.

We also reduce the number of search sites from 100 to 40, since we found that if the participants' trust settled, it settled within a few interactions. Otherwise, they showed the properties of trust oscillators. We also wanted to design a within-subjects study with three conditions, and reducing the number of search sites meant that the participants could finish the experiment within a reasonable amount of time (about 75 minutes).

4.4.2 Participants

We collected data from a total of 54 participants, 30 of which participated in experiment 1 (Age: Mean 22.6 years, *SD* 3.6, 14 Female) and 24 participated in experiment 2 (Age: Mean 21.4 years, *SD* 2.3, 12 Female). All participants were students from the University of Michigan.

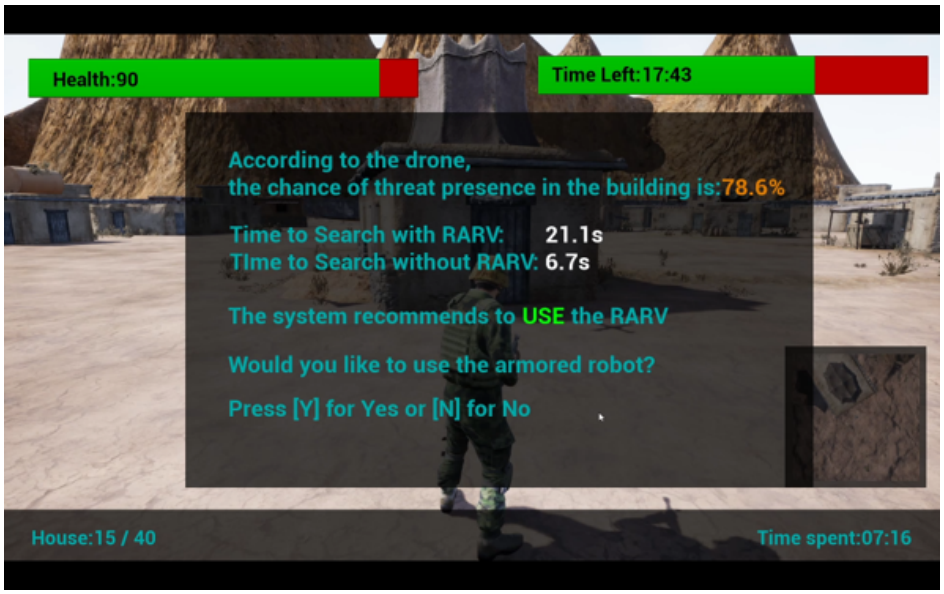


Figure 4.9: The recommendation interface

4.4.3 Experiment Design

We designed three interaction strategies for the intelligent agent:

- **Non-learner:** The intelligent agent does not learn the reward weights of the human. It assumes that the human and the intelligent agent share the same reward weights and uses these for recommendation success assessment, trust updating, human behavior modeling, and POMDP optimization.
- **Non-adaptive learner:** The intelligent agent learns personalized reward weights for each human it interacts with. It only uses these learned weights for recommendation success assessment, trust updating, and human behavior modeling. It still optimizes the POMDP with its own fixed reward weights.
- **Adaptive learner:** The intelligent agent learns personalized reward weights for each human. It uses them for recommendation success assessment, trust updating, and human behavior modeling, and also optimizes the POMDP based on these reward weights.

We employed a within-subjects design. Each participant completed three missions. In each mission, they interacted with an intelligent agent using one of the interaction strategies. To minimize potential order effects, a 3×3 Latin Square design was used.



Figure 4.10: The trust feedback slider used to get feedback from the participants after every search site. The mission timer is paused when the slider is shown to let the participants take their time in adjusting their trust.

4.4.4 Measures

4.4.4.1 Pre-experiment Measures

Prior to the experiment, participants filled in a demographic survey indicating their age, gender, academic department, nationality, frequency and skill of playing video games, and familiarity with AI/ML algorithms. Participants also filled in questionnaires about their personality, propensity to trust autonomy, and decision-making style.

4.4.4.2 Pre-mission Measures

Before each of the three missions, participants rated their preferences.

- **Task Preference:** Before the beginning of each mission, we ask the participants to rate their preference between saving health and saving time by moving a slider between these two objectives, showing their relative importance.

4.4.4.3 In Experiment Measures

After each site's search was completed (i.e., every trial), the participants were asked to report their level of trust in the intelligent agent, t_i (see fig. 4.10 for the exact question asked during the interaction). The slider values were between 0 and 100 with a step of 2 points. Additionally, for every trial, we measured whether the participants agreed with the intelligent agent. With the trial-based data, we measured the following:

- **Average Trust:** This was calculated as the empirical mean trust $\frac{1}{M}\sum_{i=1}^{M} t_i$.
- **End-of-mission Trust:** This was the participant's self-reported trust after the last trial, t_M.
- **Number of Agreement:** This was computed as the number of times the participant chose the recommended action.

Note that $M = 40$ is the number of sites in a mission.

4.4.4.4 Post-mission Measures

After every mission, participants filled out a post-mission survey gauging the following items.

- **Post-mission trust questionnaire:** This was measured using Muir's trust questionnaire [75]. It has 9 questions, each with a slider range between 0 and 100.
- **Post-mission Reliance Intentions:** This was measured using the scale developed in Lyons et al. [63]. We used 6 of the 10 items that were relevant for this task. Each item was rated on a 7-point Likert scale.
- **Workload:** Workload was measured using the NASA TLX scale [43]. We used 5 of the 6 dimensions as our experiment involved minimal physical demand. Each item was measured using a slider ranging from very low to very high.
- **Performance:** We computed the team performance by a weighted sum of the percentage health remaining of the soldier and the percentage time remaining in the mission.

$$\text{Performance} = \hat{w}_h^h \cdot (\%h_M) + \hat{w}_c^h \cdot (100 - \%c_M). \tag{4.10}$$

where $\hat{w}_h^h$ and $\hat{w}_c^h := 1 - \hat{w}_h^h$ are the reported preferences by the participant before beginning the mission, $\%h_M$ is the percent health remaining and $\%c_M$ is the percent time spent at the end of the mission. This metric allows us to convert the two conflicting objectives with different units of measurement into one unified scale. The higher the value, the better the team performance.

4.5 Results

This section summarizes our results and discusses the implications. Table 4.3 tabulates the results from the two experiments. Repeated measures ANOVA were conducted to compare

the three interaction strategies. Greenhouse-Geisser corrections to the degrees of freedom were made whenever a measure failed Mauchly's test of sphericity. In Experiment 1, we initiated the IRL learning algorithm with an informed prior. In Experiment 2, the learning algorithm started with an uninformed prior.

Table 4.3: Mean and standard deviation (SD) of dependent measures for the three interaction strategies in Experiments 1 and 2. NL - Non-learner, NAL - Non-Adaptive Learner, AL - Adaptive-learner. All values are Mean ± SD

Experiment 1: with informed prior			
	NL	**NAL**	**AL**
Average trust $\frac{1}{M}\sum_{i=1}^{M} t_i$	0.73 ± 0.16	0.71 ± 0.20	0.72 ± 0.14
End-of-mission trust t_M	0.81 ± 0.18	0.76 ± 0.21	0.76 ± 0.16
Muir's trust questionnaire	74.20 ± 14.80	71.09 ± 23.17	68.28 ± 17.4
Number of agreements	36.03 ± 3.72	35.27 ± 4.68	36.00 ± 2.60
Reliance intentions scale *	4.34 ± 1.01	4.10 ± 1.35	3.70 ± 1.27
Workload	36.87 ± 18.04	34.45 ± 17.02	39.03 ± 16.85
Performance	61.47 ± 18.12	60.25 ± 17.03	55.83 ± 20.39
Experiment 2: with uninformed prior			
	NL	**NAL**	**AL**
Average trust $\frac{1}{M}\sum_{i=1}^{M} t_i$ *	0.42 ± 0.20	0.45 ± 0.22	0.65 ± 0.20
End-of-mission trust t_M *	0.33 ± 0.30	0.35 ± 0.29	0.64 ± 0.30
Muir's trust questionnaire *	34.51 ± 19.79	34.97 ± 17.41	60.57 ± 23.71
Number of agreements *	28.17 ± 5.49	28.67 ± 4.65	35.62 ± 2.93
Reliance intentions scale *	2.16 ± 1.22	2.38 ± 1.16	3.58 ± 1.32
Workload *	38.18 ± 13.79	39.82 ± 14.94	31.13 ± 10.64
Performance	45.86 ± 16.20	49.11 ± 14.25	51.60 ± 18.68

$* - p < 0.05$

4.5.1 Trust, Agreement, and Reliance Intention

4.5.1.1 Experiment 1: with informed prior

We observed no significant difference between the three strategies in average trust ($F(2, 58) = 0.308,\ p = 0.736$), end-of-mission trust ($F(2, 58) = 1.192,\ p = 0.311$), and the Muir's trust scale ($F(2, 58) = 1.550, p = 0.221$).

Additionally, there was no significant difference in the number of agreements ($F(2, 58) = 0.755,\ p = 0.475$) across the three strategies. However, there was a significant difference in reliance intentions ($F(1.543, 44.737),\ p = 0.031$) (Fig. 4.11). Pairwise comparisons with Bonferroni adjustments revealed a lower intent to rely on the adaptive learner strategy than the non-learner strategy ($p = 0.012$).

Thus, in terms of subjective trust, we were not able to support hypothesis **H1**. In fact, for reliance intentions, the data showed a behavior opposite to our hypothesis where the non-learner was rated higher in terms of future reliance intentions than the adaptive-learner strategy. This may be due to the fact that the real-time adaptation by the adaptive-learner strategy may lead to lower predictability of the robot's recommendations compared to the non-learner strategy, whose behavior will be more consistent throughout the interaction. This unpredictability may have lowered the intention to rely on the adaptive-learner strategy.

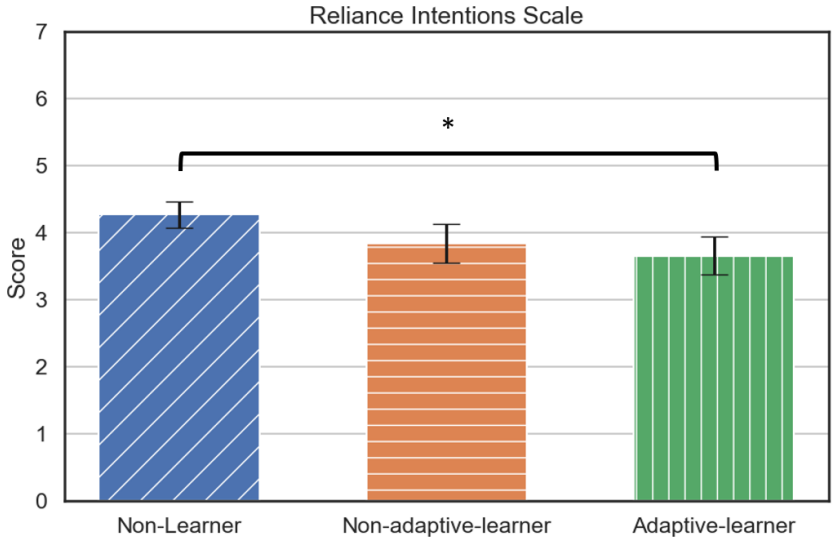


Figure 4.11: Exp 1 – Post-mission reliance intentions

4.5.1.2 Experiment 2: with uninformed prior

Figs. 4.12, 4.13, 4.14 show the comparisons of the three strategies in trust. Repeated measures ANOVA revealed significant differences between the three strategies in average trust ($F(2, 46) = 14.161,\ p < 0.001$), end-of-mission trust ($F(2, 46)) = 12.736,\ p < 0.001$, and Muir's trust scale ($F(1.586, 36.473) = 16.3,\ p < 0.001$). Pairwise comparisons with Bonferroni adjustments revealed that the adaptive-learner strategy led to higher average trust, end-of-mission trust, and post-mission trust compared to the non-learner strategy ($p < 0.001,\ p = 0.001$ and $p < 0.001$, respectively) and compared to the non-adaptive learner strategy ($p = 0.003$, $p < 0.001$, $p < 0.001$, respectively).

Regarding the number of agreements (Fig. 4.15), there was a significant difference among the three strategies ($F(1.584, 36.435) = 25.829,\ p < 0.001$). Post-hoc analysis showed that there was a significant difference between the non-learner and the adaptive-learner

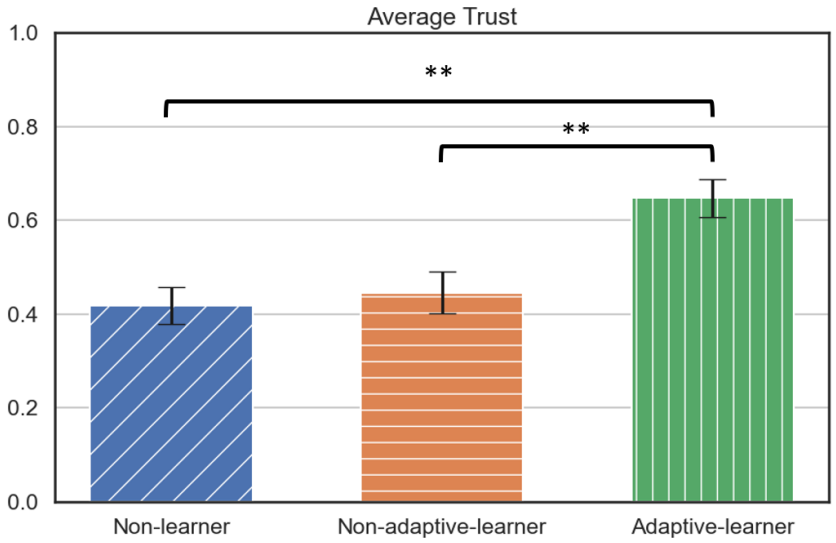


Figure 4.12: Exp 2– Average trust $\frac{1}{M}\sum_{i=1}^{M} t_i$

strategies ($p < 0.001$) and between the non-adaptive-learner and adaptive-learner strategies ($p < 0.001$).

Comparing reliance intentions (Fig. 4.16), there was a significant difference between the three strategies ($F(2, 46) = 13.691, p < 0.001$), with the adaptive-learner strategy rated higher than the non-learner strategy ($p < 0.001$) and the non-adaptive-learner strategy ($p = 0.004$).

Here we see a clear support for hypothesis **H1**, with the adaptive-learner strategy realizing the highest amount of subjective trust and reliance intentions. We still did not see any significant difference in terms of these metrics between the non-learner and non-adaptive learner strategies, showing that the reward function being used in the robot's optimization routine is most important when it comes to aligning with human values.

4.5.2 Performance

4.5.2.1 Experiment 1: with informed prior

Even though there seemed to be a decreasing performance trend from non-learner to non-adaptive learner and to adaptive learner (i.e., 61.47 ± 18.12, 60.25 ± 17.03, and 55.83 ± 20.39) the trend did not reach statistical significance ($F(2, 58) = 2.067,\ p = 0.136$).

We were not able to show support for hypothesis **H2** in this case, as the participants performed very similarly when interacting with any of the interaction strategies.

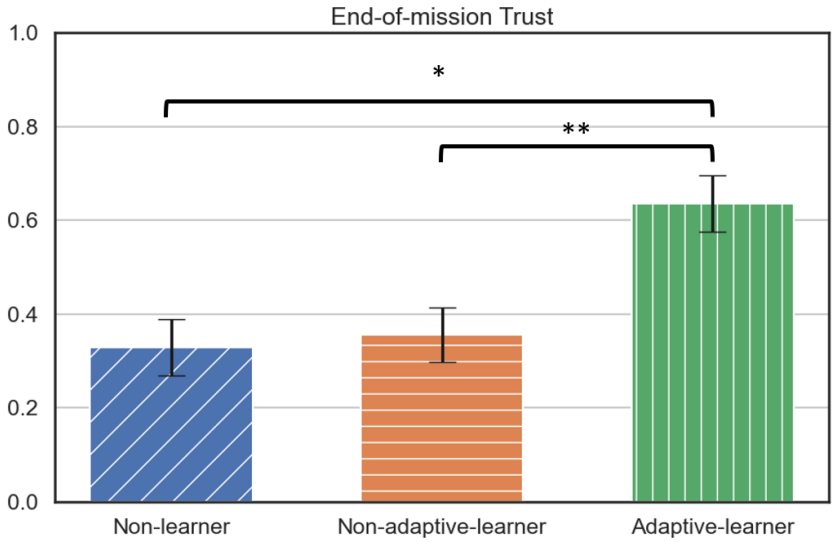


Figure 4.13: Exp 2– End-of-mission trust t_M

4.5.2.2 Experiment 2: with uninformed prior

There seemed to be an upward trend from non-learner to non-adaptive learner and to adaptive-learner. Unfortunately, it did not reach significance.

Even in the uniform prior case, we were not able to show support for hypothesis **H2** in this case, as the participants performed very similarly when interacting with any of the interaction strategies. It is interesting to see that even though the participants trusted the adaptive-learner strategy more, they performed similarly in all three conditions. This is most likely because in our scenario, since we present the participants with a chance of threat presence, it is possible for them to make their own action choice. This action choice matches the recommendation given by the adaptive learner strategy leading to higher trust and lower workload, but their action choices are what determine the performance, and hence, it remains the same between the three strategies.

4.5.3 Workload

4.5.3.1 Experiment 1: with informed prior

Comparing the average workload across the three interaction strategies showed no significant difference ($F(2, 58) = 2.634,\ p = 0.089$) (Table 4.3). Additionally, repeated measures ANOVA did not reveal any significant differences between the three strategies in any of the dimensions.

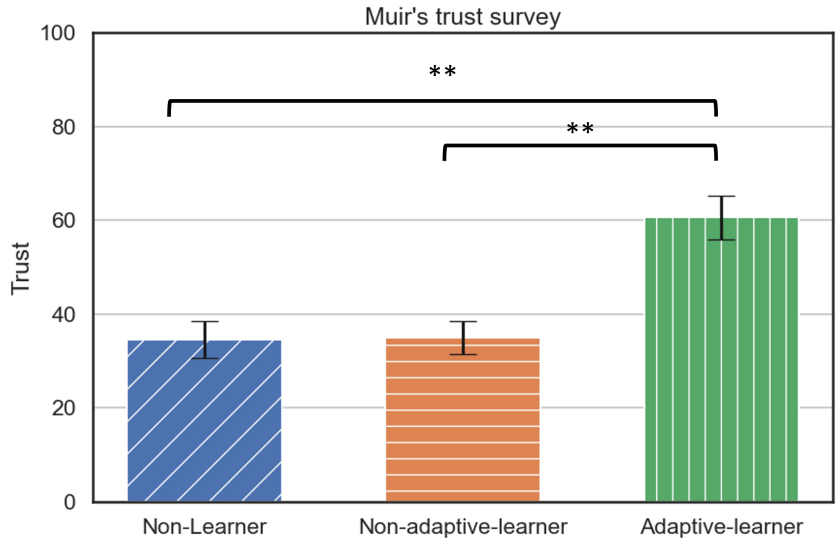


Figure 4.14: Exp 2– Post-mission trust questionnaire

4.5.3.2 Experiment 2: with uninformed prior.

Comparing the average workload (4.3) across the three interaction strategies showed a significant difference ($F(2, 46) = 10872,\ p < 0.001$). Figure 4.17 shows the participants' responses on each dimension. There were significant differences between the three strategies in performance ($F(2, 46) = 5.443,\ p = 0.008$), effort ($F(2, 46) = 4.252,\ p = 0.02$), and frustration ($F(2, 46) = 5.454,\ p = 0.007$).

Pairwise comparisons with Bonferroni adjustments showed that the adaptive-learner strategy led to higher perceived performance compared to the non-adaptive learner ($p = 0.037$) and to the non-learner ($p = 0.044$) strategies and led to lower frustration compared to the non-adaptive learner ($p = 0.032$) and to the non-learner ($p = 0.017$) strategies.

If we include the perceived performance as a measure of performance, we see some support for hypothesis **H2** in this case. The participants felt that they performed better when paired with the adaptive-learner strategies. This points to a need to look at performance not only through objective, task-based measures, but to develop more holistic measures of performance that consider trust, workload, and objective performance all in one metric.

4.6 Discussion and Conclusion

In this study, we developed and compared three robot interaction strategies: the non-learner strategy, the non-adaptive-learner strategy, and the adaptive-learner strategy, with and with-

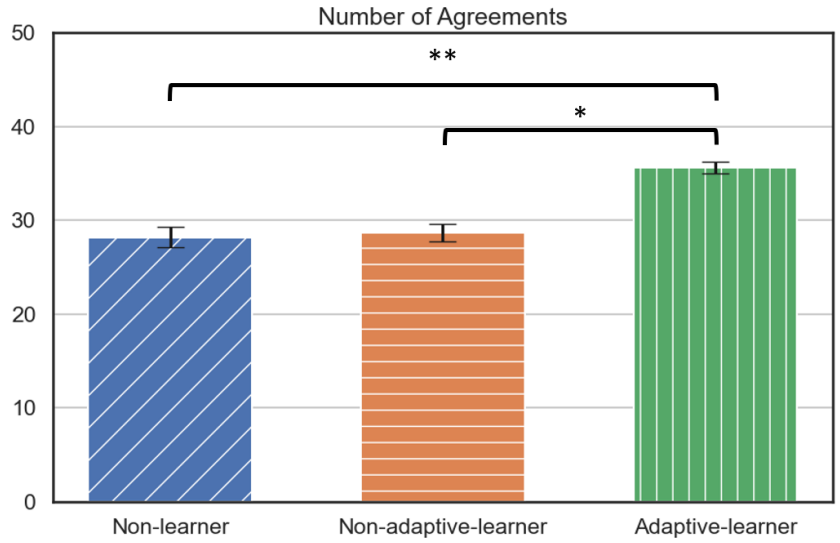


Figure 4.15: Exp 2– Number of agreements

out an informed prior for the IRL learning algorithm. We focused on evaluating their influence on various human trust in the robot, agreement, reliance, workload, and team performance.

4.6.1 Value Alignment with an Informed Prior

One critical insight from the research is the observed "uniformity" across all three interaction strategies when the IRL algorithm is initiated with an informed prior. In Experiment 1, the informed prior was calculated by training on a previously collected dataset using the same testbed and with participants from the same population as this study. Therefore, the calculated prior was considered an accurate estimation of the participants' true values/goals. The resulting prior was realistically skewed toward saving health rather than saving time. With this accurate informed prior, the three interaction strategies were more-or-less indistinguishable. Among all the dependent measures, including trust, agreement, reliance intentions, performance, and workload, we observed only one significant difference in reliance intentions. This large uniformity across the three strategies could be because, with an accurate informed prior on human values, there is no room for significant alignment to be done by the adaptive-learner strategy. This highlights that when a robot's reward function is already closely aligned with that of the human *a-priori*, adaptive strategies may exhibit negligible benefits.

Out of expectation, we observe a significant difference in post-mission reliance intentions,

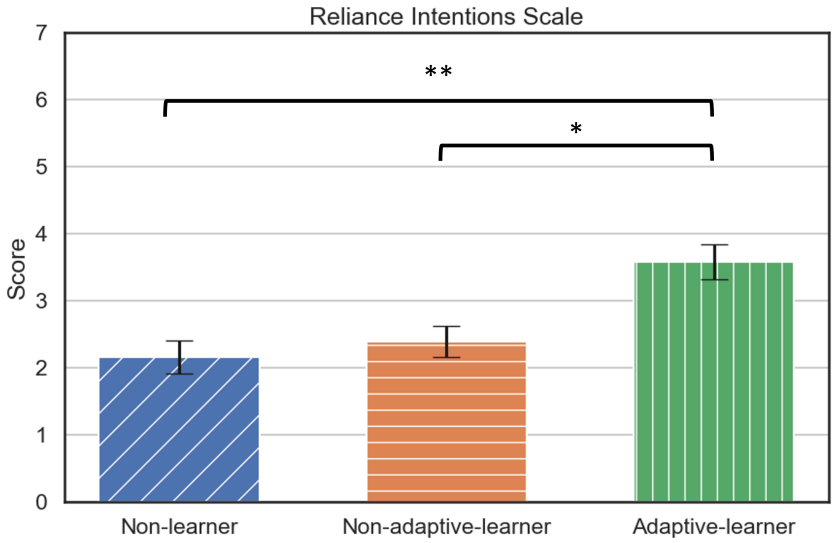


Figure 4.16: Exp 2– Post-mission reliance intentions

where participants were willing to rely on the non-learner more than the adaptive-learner strategies. This result could have been because any alignment from the adaptive-learner strategy, although very limited, could be regarded as a lack of predictability. As pointed out in research on teamwork, the ability to predict the actions of one's teammate is vital [24, 51].

4.6.2 Value Alignment with an Uninformed Prior

In Experiment 2, when the IRL learning algorithm was initiated with an uninformed uniform prior, we observed significant benefits of value alignment. The adaptive-learner strategy led to significantly higher trust, high agreement, and reliance intentions. In addition, participants had higher perceived performance and lower frustration interacting with the adaptive learner, while the team performance was maintained. Thus, in the absence of a good initial prior, our adaptive framework can be used to build trust in the robot. This scenario underlines the importance of adaptive strategies in real life where a good *a-priori* estimation of the human's values/goals is oftentimes unavailable, offering an approach to building trust while maintaining performance.

The results from Milli et al. [73] showed that robots should not be completely obedient to humans who are not acting rationally. Instead, when interacting with such humans, the robot should use its estimation of the human's underlying reward function. We extend this by providing a human-subjects study showing that personalized value alignment is only beneficial when a good prior on the human's reward weights is unavailable

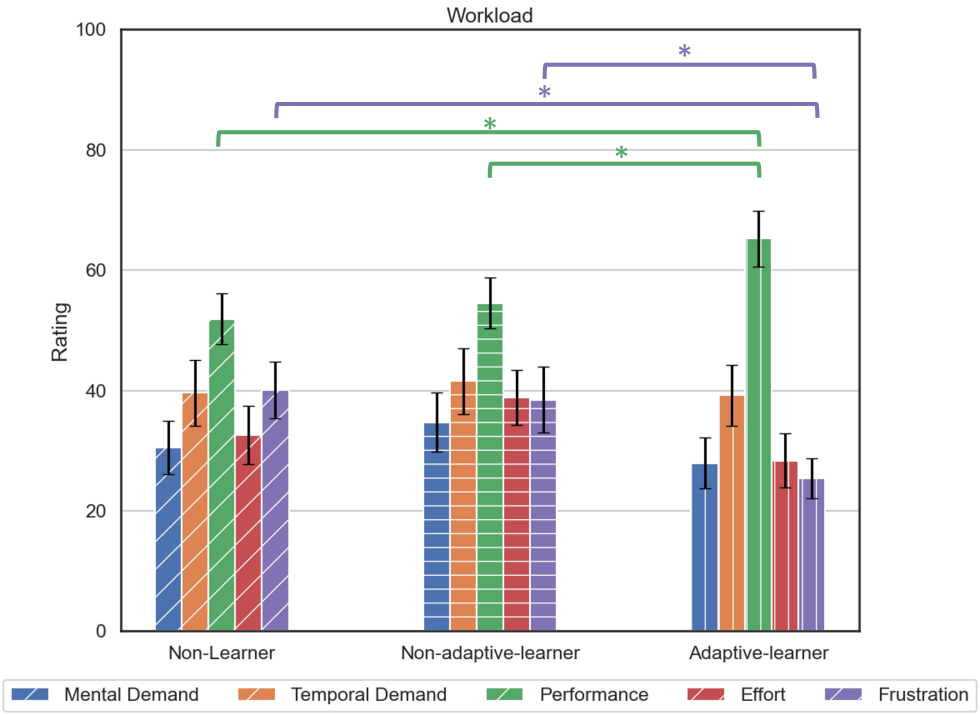


Figure 4.17: Exp 2: Responses on the NASA TLX scale

4.6.3 Implications for a Broader HRI Context

The algorithmic focus of existing work on value alignment [40, 33, 104] has one implicit assumption: aligning a robot's values with a human's is beneficial. Our study is the first attempt to examine whether and to what extent such alignment can benefit trust, workload, and team performance. We show that personalized value alignment is beneficial only when an informed prior is unavailable. The implications of our study should extend to other real-life HRI scenarios involving conflicting objectives. For instance, a rehabilitation robot must balance a patient's pain tolerance with long-term health goals when assigning the appropriate level of exercise.

Our study involved a relatively homogeneous participant group, leading to the calculation of a fairly accurate informed prior. However, achieving such accuracy in real-world settings with demographically diverse individuals is more challenging. In such cases, aligning a robot's values with those of individual human users becomes essential. The benefits obtained from value alignment are key for the acceptance and adoption of robots in homes and workplaces, highlighting the need for adaptable strategies in HRI design.

Further, the idea of incorporating a layer of trust in the decision-making system of an intelligent agent trying to align its values to that of the human user is an interesting area to

explore in other HRI domains like shared control, social robotics, etc.

4.6.4 Limitations and Future Research

The results of our study should be seen in light of the following limitations. First, we provide a demonstration in the case when there are only two components in the team's reward function. Therefore, we only need to learn the human's preference for one of the two components and can ascertain their relative preference between the two objectives. Our formulation, however, can readily be extended to the case where there are more than two objectives in the team's reward function, with additional computations required to learn and maintain a distribution over each reward weight.

Second, our simulated scenario consists of binary actions. Judging the performance of the recommendations is fairly easy in this case, since we only need to compare the rewards earned for these two actions. In case more than two actions are available, this assessment becomes more difficult. Thus, although the human trust-behavior model can be readily extended to such a case of multiple actions, extending the trust dynamics model is challenging and is an interesting avenue for future research.

Third, in our scenario, there is an expected skewness among the general population to be more concerned about saving health. It would be interesting to study cases where the two objectives are more balanced, resulting in a more balanced informed prior. In such cases, personalized adaptation may still be beneficial.

Finally, our scenario, which entails a trade-off between "saving health" and "saving time", and the decision to use or not use an armored robot, is informed by the complex decision-making scenarios in real-life HRI contexts, such as DARPA SQUAD-X program in which individuals receive recommendations from air and ground robots for various tasks. While our scenario offers insights into these types of decisions, we recognize it as a simplified representation of situations where decisions involve numerous objectives, a variety of recommendations, and possible actions. Therefore, further research is essential to determine the applicability of our findings in more complex, real-world environments and to validate the robustness of our conclusions in diverse and dynamic HRI settings.

CHAPTER 5

Refining the Reward Function: Adding State Dependence

In the previous chapter, we introduced the idea of personalization of reward weights through Bayesian IRL. However, our human-subject studies showed minimal improvement in trust and team performance in the case when the intelligent agent starts with a good prior on the reward weights of the human. We did, however, make a major assumption in that chapter - human reward weights do not change with the context of interaction. In our case, the context is the health remaining of the soldier and the time remaining to complete the mission. However, intuitively, this assumption does not seem to hold true. We think that people would be willing to take more risks (aka low reward weight for losing health) when the soldier's health is high and the mission time remaining is low. On the other hand, people would be very risk-averse (high reward weight for losing health) when the soldier's health is already low and there is plenty of mission time left. In this chapter, we try to address this limitation by explicitly including this context into the state of the trust-aware POMDP.

We provide a framework for learning the context-dependent reward weights through queries asked to human participants. We collect data from Amazon Mechanical Turk (MTurk) and learned a "state-dependent" reward function. We then compare two robot strategies - one using this state dependent reward function and the other using the constant reward function from the last chapter under differing levels of vulnerability for the soldier. Our results indicate that the state-dependent reward function results in higher trust and better team performance, especially in the high vulnerability case.

This chapter is based on work published in [10]. The remainder of the chapter is organized as follows: Section 5.1 introduces the topic and provides some background. Section 5.2 presents details of our mathematical model of human-robot interaction in a context where the robot acts as an action recommender to the human. Section 5.3 details the first study within this work, dealing with learning a data-driven state dependent reward function through Amazon Mechanical Turk. Section 5.4 presents an initial simulation study. Section 5.5

presents details about the human-subject study study where we compare a robot using this state dependent reward function with a robot using a constant reward function through an in-person human subjects experiment. Finally, Section 5.6 discusses the implications of our results and some limitations and directions of future work.

5.1 Introduction

With recent technological advances in Artificial Intelligence and Robotics, researchers are moving from a human-automation interaction perspective to a human-autonomy teaming perspective [68, 79]. As these autonomous systems become more and more capable to work in uncertain, never-before-seen environments, they continue to be seen more as teammates rather than tools to be used by a human operator.

To facilitate smooth and effective team collaboration, trust [55] in the autonomous agent has proven to be a very important factor. The area of research concerning trust-driven human-robot interaction has seen an enormous rise in popularity in recent years [57, 7, 36, 94]. Further, responsivity in autonomous agent behavior [19] has been postulated to be almost as important, if not more, than reliability in the context of human-autonomy teaming. In this work, we try to frame responsivity of autonomous systems through a state dependent reward function perspective and compare it with another system using a constant (non state dependent) reward function. We study these differences in the context of a "robot recommender - human decision-maker" scenario wherein the robot tries to maximize the reward function over the planning horizon. Our results indicate that a robot using such a state dependent reward function results in higher trust and better team performance compared to the robot using a constant reward function, especially in a context where the vulnerability of the human is high.

Our contributions in this chapter include:

- We present a framework for learning a state dependent reward function in a two objective human-robot collaboration task where the robot acts as an action recommender to the human. We demonstrate this framework through a human-subjects study done over Amazon Mechanical Turk.

- We compare two interaction strategies - one using the learned state dependent reward function and the other using a constant reward function - on trust and team performance through an in-person human-subjects study. Specifically, we test the following hypotheses:

- **H1:** The state dependent strategy will lead to higher trust compared to the constant strategy.
- **H2:** Participants will perform better when paired with the state dependent strategy compared to when they are paired with the constant strategy.
- **H3:** Vulnerability will have a moderating effect on the relationship between the robot strategy and trust/performance.

5.2 Mathematical Modeling

We use the same trust-aware POMDP framework introduced in Chapter 3, with a modified state-space, transition function, and reward function to incorporate the state-dependence of reward weights. We also use the Bounded Rationality Disuse model of human behavior introduced in Chapter 4 as the human behavior model in the framework. We detail the modifications below.

5.2.1 Trust-Aware Markov Decision Process

5.2.1.1 States

The trust of the human on the intelligent agent t, the remaining health points H, and the remaining time points C.

5.2.1.2 Reward function

Since the objective is to minimize mission time *and* minimize damage to the soldier, the reward function is modeled as a negative convex combination of the cost for losing health and the cost for losing time. Specifically,

$$R(H, C, a, D) = -w(H, C)\mathcal{H}(a, D) - (1 - w(H, C))\mathcal{C}(a). \tag{5.1}$$

Here, $w(H, C)$ is the reward weight for losing health, $\mathcal{H}(a, D)$ is the cost for losing health, $\mathcal{C}(a)$ is the cost for losing time, a is the action selected by the human, and D is a binary random variable indicating the presence of threat.

$$\mathcal{H}(a, D) = hD(1 - a), \tag{5.2}$$

$$\mathcal{C}(a) = ca, \tag{5.3}$$

where h is the cost for losing health and c is the cost for losing time. For our study, we set $h = c = 10$.

5.2.1.3 Observation function

We use the same observation function from Chapter 3, with the new reward function.

$$p_i = \begin{cases} 1 & \text{if } R_i(H, C, a_i^r, D) \geq R_i(H, C, 1 - a_i^r, D), \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

5.2.1.4 Transition function

We use the beta distribution trust model [38] with the reward-based performance metric [7] as the transition function for trust. Specifically, trust at the i^{th} interaction, t_i is modeled as a beta distributed random variable with parameters α_i and β_i, which are updated based on positive and negative experiences with the recommendations.

$$t_i \sim Beta(\alpha_i, \beta_i) \text{ where} \tag{5.5}$$

$$\alpha_i = \alpha_{i-1} + p_i v^s, \tag{5.6}$$

$$\beta_i = \beta_{i-1} + (1 - p_i) v^f. \tag{5.7}$$

We note that $(\alpha_0, \beta_0, v^s, v^f)$ are the personalized trust parameters associated with every individual that interacts with the intelligent agent, p_i is the reward-based binary performance of the recommendation at the i^{th} search site, which is given by the observation function (Eq. 5.4).

We update the health and time according to the rules mentioned above. The soldier loses health when encountering a threat without protection from the RARV. The soldier loses time for using the RARV.

$$H_{i+1} = \begin{cases} H_i - h, & \text{if } a_i^h = 0 \text{ and } D_i = 1, \\ H_i, & \text{otherwise.} \end{cases} \tag{5.8}$$

$$C_{i+1} = \begin{cases} C_i - c, & \text{if } a_i^h = 0, \\ C_i, & \text{otherwise.} \end{cases} \tag{5.9}$$

Here, i is the site index, H_i and C_i are the health and time points remaining before the i^{th} interaction, a_i^h is the action chosen by the human at the i^{th} site, and D_i is a binary variable

indicating the presence of threat at the i^{th} site.

5.3 Study 1 - Learning the Reward Function

This section describes the first phase of our overall study - data collection for learning the state dependence of reward weights. We used MTurk for collecting data. Our main insight was that assuming the bounded-rationality model of human behavior and the absence of recommendations, we can find the tipping point of the threat level at which a majority of the population will change their choice of action from risk-taking to risk-averse. This threat level (d^*) corresponds to the threat level at which the one-step expected reward for the human for both the actions is the same. Mathematically,

$$\begin{aligned} & E_D\left[R(H,C,0,D)\right] = E_D\left[R(H,C,1,D)\right] \\ \implies & w(H,C)hE[D] = (1-w(H,C))c \\ \implies & d^* = \frac{(1-w(H,C))c}{w(H,C)h}. \end{aligned} \tag{5.10}$$

So, the idea is to collect action-choice data for a set of states of (H, C) for a series of threat levels. We can then train a logistic regression model to find d^* and thus $w(H, C)$ from Eq. 5.1. Fig. 5.1 shows an example of such a query.



Figure 5.1: A sample query asked to an MTurk worker. The worker was shown this information and asked whether s/he would choose to use the RARV or not.

5.3.1 Data Collection

We defined 6 health states, 6 time states, and 11 threat levels, resulting in a total of 396 distinct queries for MTurk workers. These queries were randomly assigned into 12 separate

surveys, each containing 33 queries, to ensure the surveys could be completed within a reasonable time. At the start of the survey process, participants were asked to provide informed consent. They then watched an instructional video explaining the task, the two available actions, and their respective consequences. After viewing the video, participants had to correctly identify the full form of RARV. Those who answered incorrectly were not permitted to continue. Participants who answered correctly proceeded to the main part of the survey, which included three embedded attention check questions to maintain data quality.

The study was approved by the Institutional Review Board at the University of Michigan (ID HUM00249731). The workers completed the surveys using Qualtrics. Each of the 12 surveys was completed by at least 10 MTurk workers. In total, we obtained 4092 query responses from 124 workers.

5.3.2 Data Cleaning

We noticed that the data collected from MTurk workers contained many sub-optimal action choices, likely due to participants either misunderstanding the task or not paying sufficient attention, even though all workers passed the basic attention check questions. To improve data quality, we opted to exclude data from workers who made specific "obviously" sub-optimal choices. These involved two scenarios: (1) Using the RARV when the threat level was 0%, and (2) Not using the RARV when the threat level was 100%, when health was 10, and time was above 10. In the first scenario, choosing not to use the RARV is optimal, as it ensures no loss of health or time, while using the RARV results in unnecessary time loss. In the second scenario, failure to use the RARV is sub-optimal, as the soldier's health would drop to 0, prematurely ending the mission.

We removed the data from the corresponding MTurk workers, resulting in the removal of data from $83 \cup 10 = 85$ workers. Thus, our final dataset consisted of responses from 39 workers, corresponding to 1287 query responses.

5.3.3 Data Analysis

A logistic regression was trained using threat level as the predictor and the action choice as the target for each state (h, c) in the query set of the cleaned data. A sample logistic curve is shown in Fig. 5.2. The point at which the logistic curve reaches a value of 0.5 corresponds to d^* from Eq. 5.10. Thus, we get the raw reward weights for losing health $w(H, C)$ as a function of health remaining and time remaining at each of the queried states. This is shown as a heatmap in Fig. 5.3a.

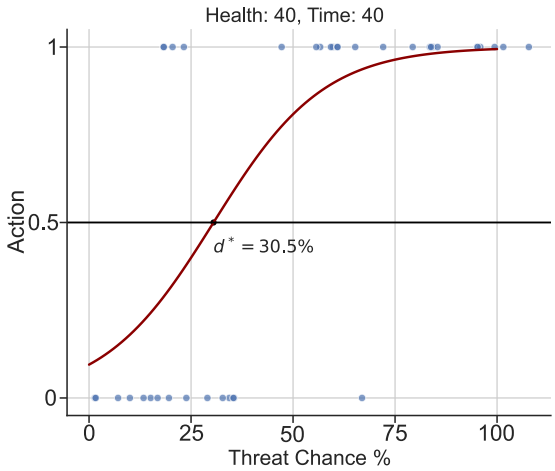


Figure 5.2: Sample logistic curve learned from the collected data

In order to get a smooth function for the health reward weight, we used the forward selection method of model selection using the Akaike Information Criterion (AIC). To do this, we incrementally added features from the set $\{H, C, H^2, C^2, HC\}$ to logistic regression models, computed the AIC, and stopped when the AIC increased. This gave us a logistic regression model with $\{H, C\}$ as the features. The final resulting model is given by,

$$w(H, C) = \frac{1}{1 + \exp(0.26H - 0.17C - 0.79)}. \tag{5.11}$$

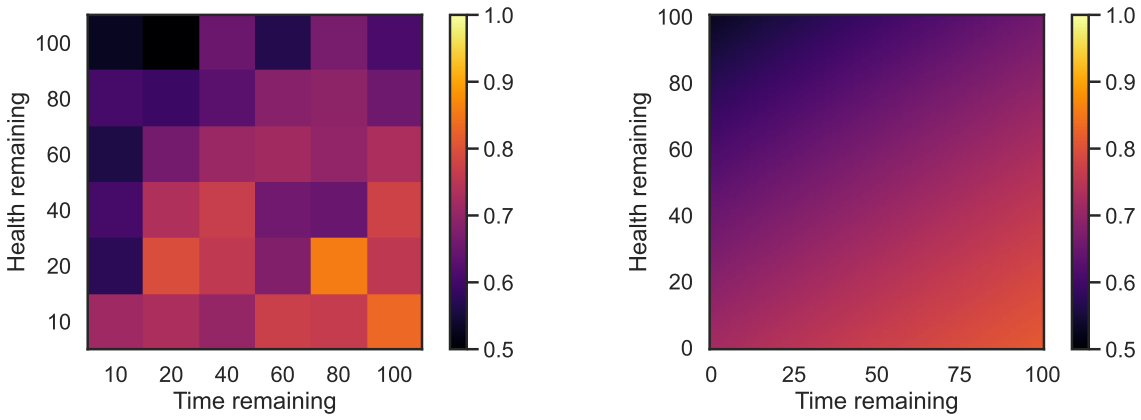
A heatmap generated using this model is shown in Fig. 5.3b. Here, the reward weight increases with increasing value of time remaining and decreasing value of health remaining.

5.4 Study 2 - Simulation Study

We simulate interactions with two robot strategies: one using a constant reward weight of 0.81 and the other using the state-dependent reward weights 5.11. The value of 0.81 was chosen from an earlier study [8] that found that using this reward weight would result in similar trust to the case when the robot is actively learning reward weights. In this study we want to see if that is the case when we explore more of the state space.

5.4.1 Simulating the Human

Our simulated human sampled trust parameters taken from a dataset collected in an earlier study [7]. The simulated human thus maintained and updated her trust using the trust dynamics model (Eq. 5.5). After receiving a recommendation from the robot, the



(a) Raw data

(b) Smoothed model

Figure 5.3: Heatmaps showing (a) raw data of learned health reward weights at each queried state and (b) the smoothed function for the state dependence of reward weights

simulated human chooses an action using the Bounded Rationality Disuse Model. We use $\kappa = 0.2$ throughout the simulation. The simulated human is assumed to be using the state-dependent reward function (Eq. 5.11). After observing the outcome of the action selection, the simulated human updates her trust parameters and reports the level of trust by sampling from the corresponding beta distribution.

The robot learns the trust parameters of each human using Maximum Likelihood Estimation (MLE) [8].

5.4.2 Setting Threats and Threat Levels

We used the following strategy for setting the threats and threat levels. In general, the prior probability of threat presence was set to 0.7 meaning that there would be threats in 7/10 search sites on average. This was to ensure that aligning the values of the robot with the human will be beneficial for trust, which is only important under high-risk scenarios [9]. Then, with 50% chance at any site independently, the threat was set randomly using a Bernoulli sample with this prior probability. To set the threat level after scanning the site by the drone, we sampled a beta distribution with a peak at 0.9 in case a threat is present and with a peak at 0.1 if a threat is not present.

With the remaining 50% chance, the threat and threat level were chosen *intelligently* to induce a difference in recommendations from the two robot strategies. As can be seen from Fig. 5.2, the value d^* is like a threshold threat level below which the action 0 is more likely and above which the action 1 is more likely. A similar behavior is also seen for recommendations

[9]. We compute d^* for the constant reward weight and the state-dependent reward weight and set the threat level to be uniformly sampled between the two values. The threat is then sampled with a Bernoulli distribution with this threat level as the parameter. In this way, we make it more likely for the two recommendations to differ.

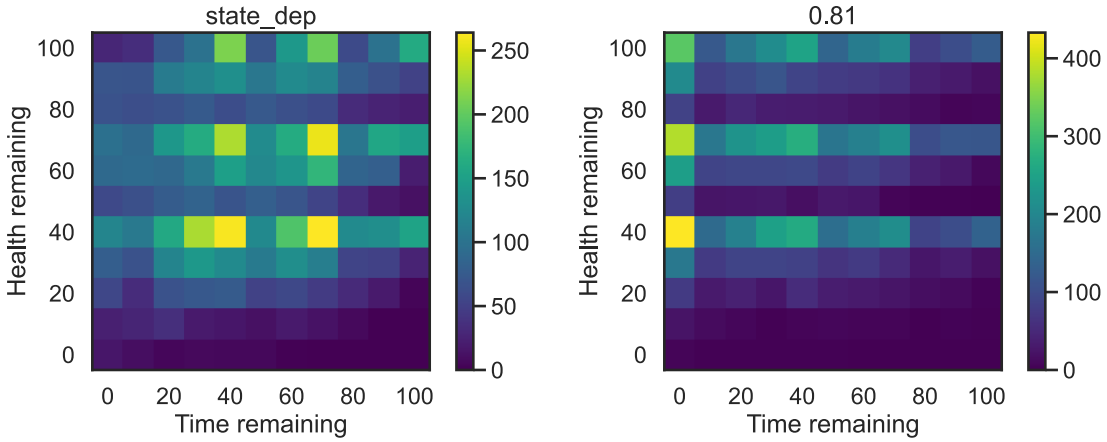
5.4.3 Results - Simulation Study

We ran 100 independent simulations for 9 starting conditions of health and time selected from the set $\{100, 70, 40\}$. The results presented below are an average of all simulation runs. For each simulation run, the human-robot team sequentially searched through 10 search sites. In all line plots in this section, the 95% confidence interval is also plotted.

5.4.3.1 States Visited

Fig. 5.4 illustrates the frequency of visits to each state during the simulation runs. Brighter colors represent states visited more frequently, while darker colors represent those visited less often. As shown, the robot strategy with constant reward weights produces more horizontally oriented patterns of brighter colors (Fig. 5.4b). This pattern suggests that health remains relatively stable throughout the interaction, reflecting a more conservative approach by the team.

In contrast, the robot strategy with state-dependent rewards leads to a broader exploration of the state space, with a greater number of states shown in brighter colors 5.4a). This pattern indicates that the team is more willing to take risks, potentially sacrificing some health to save time.



(a) State dependent rewards

(b) Constant rewards

Figure 5.4: Counts of states visited for the two robot strategies

5.4.3.2 Trust Dynamics

Fig. 5.5 compares the trust reported by the simulated human for the two interaction strategies. It is clear that the state-dependent strategy of the robot results in higher trust.

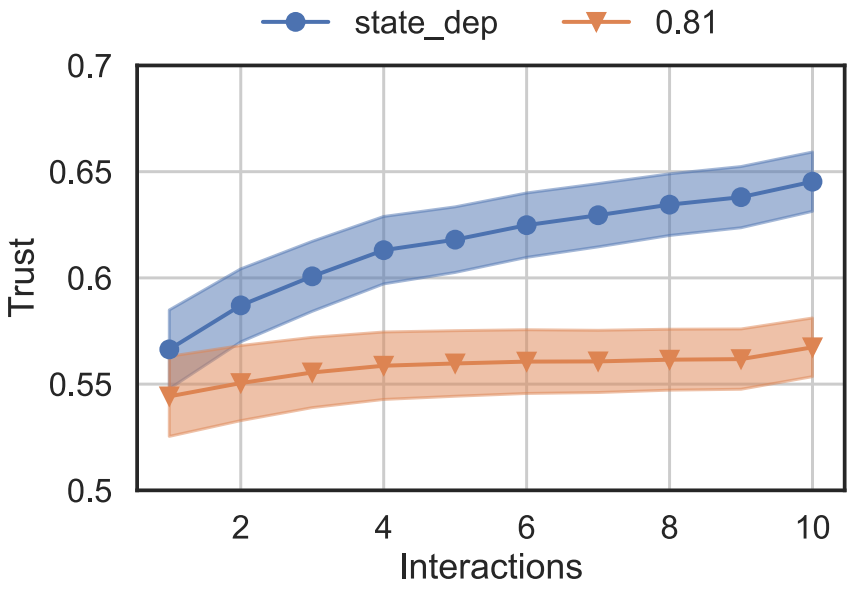


Figure 5.5: Comparison of trust reported by the simulated human for the two robot strategies. Here, trust ranges from 0 to 1.

5.4.3.3 Performance

Fig. 5.6 compares the performance of the team for the two robot strategies. Performance is measured as the sum of the health points and the time points at each stage of the mission. As is evident from the graph, the state-dependent strategy performs better than the constant strategy for the robot.

Fig 5.7 splits out the performance into its components. As can be seen, the constant strategy prefers to save health, only losing about 10 points on average throughout the mission, while losing a lot of time. The state-dependent strategy, however, loses around 20 points of health to save time.

It should be noted that the results will remain the same as long as the human's reward weights for losing health satisfies $w > 0.5$, which is the case for the state-dependent reward function and the constant reward function. This is because the performance metric will be the same as long as this condition is satisfied. Therefore, even if the simulated human was using the constant reward function, the results still hold.

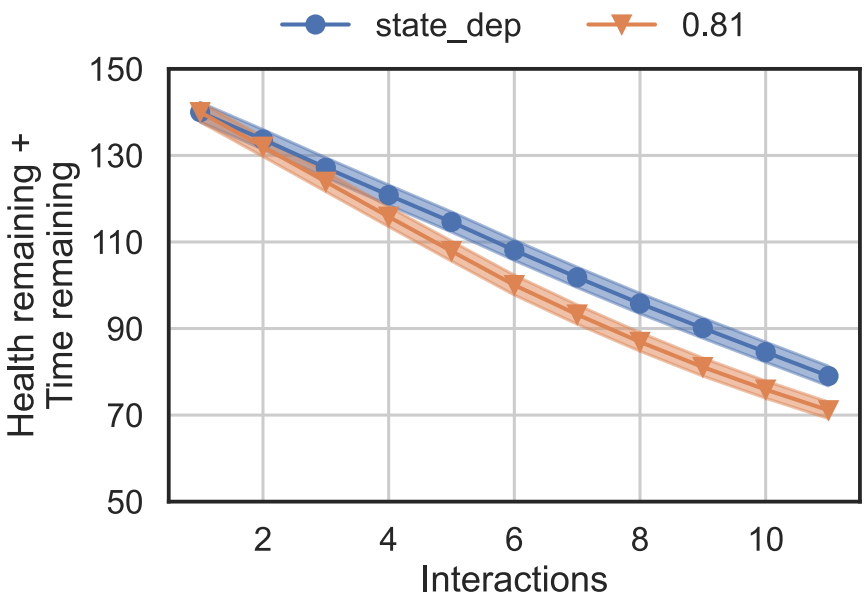


Figure 5.6: Comparison of the performance of the team, as measured by the sum of the health points and the time points

5.5 Study 3 - Studying the Effect of State Dependence of Reward Weights

This section describes the second study where we try to test the effectiveness of a robot using the state dependent reward function compared to one that uses a constant reward function.

In our previous work [8], we found that the constant reward function performed almost as well as a personalized reward function. However, we note that the constant strategy could be too conservative, especially in situations where a higher risk appetite is necessary. For example, in the context of our reconnaissance mission, if very few time points are remaining, the participants must take more risks in order to save time and finish the mission. In such situations, the constant strategy will still suggest conservative actions, disregarding the low time points remaining. However, a robot using our state dependent reward function will suggest riskier actions, helping the participants complete the mission and garnering more trust.

5.5.1 Testbed and Scenario

We modify the testbed used in Chapter 4 to incorporate the state dependent reward function. Each mission entails sequentially searching 10 sites for potential threats. The soldier starts with either 100 or 40 points of health, operationalized as high or low armor and either 100

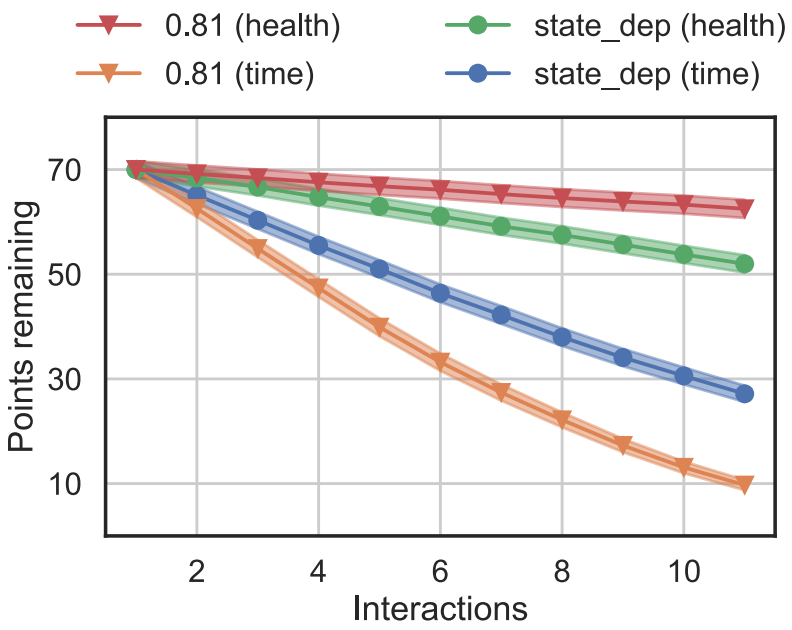


Figure 5.7: Comparison of the health and time points across the interaction for the two robot strategies

or 40 points of time, operationalized as time available to complete the mission. The mission is considered to have ended prematurely if the soldier reaches 0 health or 0 time points at any point during the mission. We ended up reducing the number of search sites from 40 in Chapter 4 to 10 here to balance the added computational cost for value iteration because of the added state dependence on time and health.

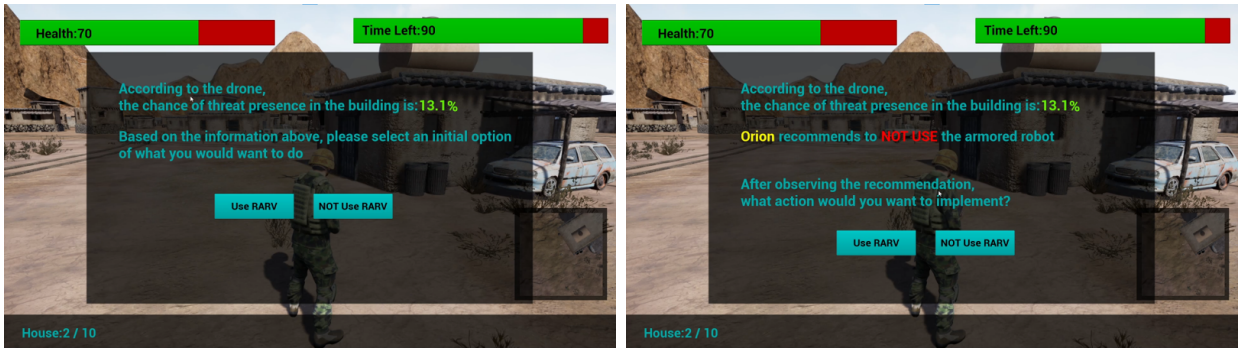
Before making their action choice, a drone enters the search site and relays information about the chance of threat present inside the site to the human-robot team. Then, the testbed asks the participant what their initial decision is regarding the choice of action. To reiterate, there are two actions available to the human - USE or NOT USE the RARV. Using the RARV results in 10 point loss of time. Encountering a threat without protection from the RARV results in 10 point loss of health. Table 5.1 shows this in a tabular format.

Table 5.1: Health and Time Loss

Used RARV?	Threat Present?	Health Loss	Time Loss
YES	YES	0	10
YES	NO	0	10
NO	YES	10	0
NO	NO	0	0

After making their initial decision, the participants are shown the recommendation from

the intelligent agent. They are then asked for their final decision about which action they want to implement. These interfaces are shown in Fig. 5.8. The participants are then shown the consequences of their action choice (shown in Fig. 5.9) and the appropriate loss of health or time is applied.



(a) Initial Action Choice (b) Final Action Choice

Figure 5.8: Interfaces to let the participants choose their (a) Initial Action Choice before receiving a recommendation from the intelligent agent and (b) Final Action Choice after receiving the recommendation.

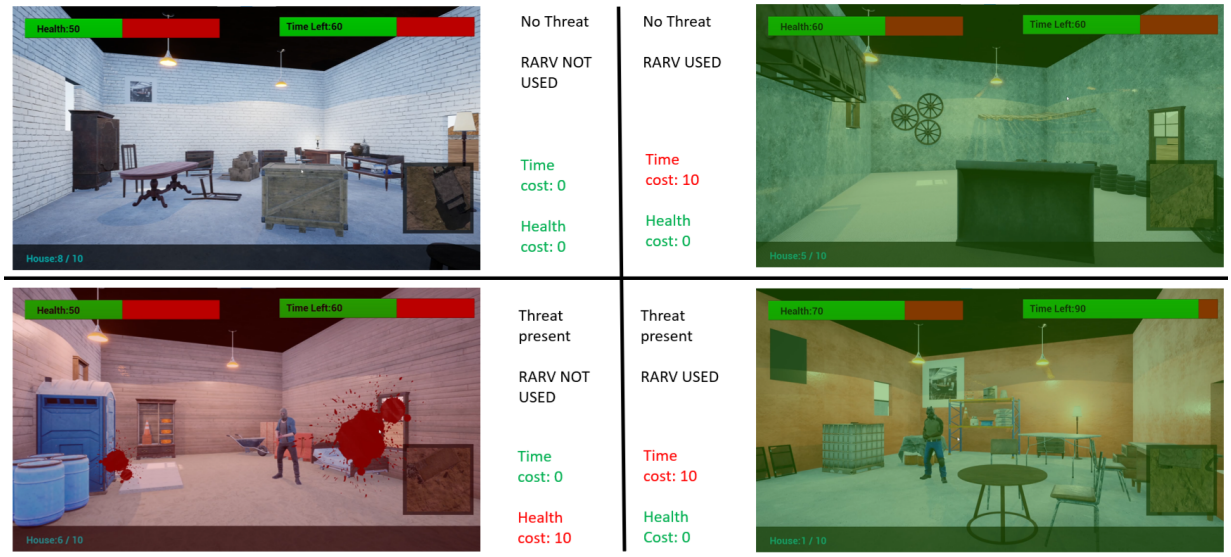
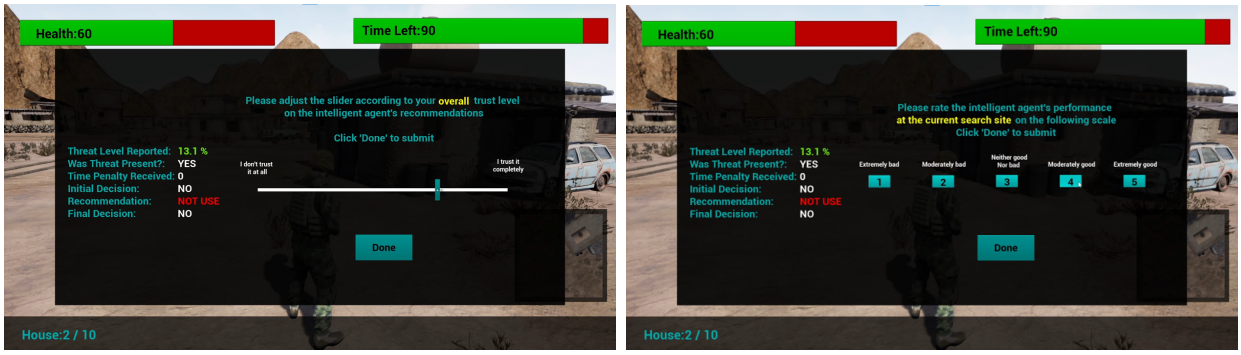


Figure 5.9: Four possible outcomes seen by the participants depending on their choice of action and whether a threat was present inside the search site.

The participants are given the objective to complete the mission while maximizing the amount of health points and the amount of time points remaining at the end of the mission. They are also told that the mission will end prematurely if either the soldier's health or the

time remaining drops to 0 at any point during the mission, and that their performance bonus will be affected negatively if this happens. This encourages the participants to perform as well as they can.

The testbed also captures the participant's trust and perceived performance of the recommendation after each search site. The interfaces are shown in Fig. 5.10. It is emphasized that they should consider their entire interaction history with the intelligent agent when rating their trust and should only consider the current interaction for rating their performance feedback.



(a) Trust feedback slider (b) Performance feedback scale

Figure 5.10: Interfaces to let the participants rate their trust and performance feedback on the intelligent agent's recommendations. The trust feedback slider converts to a value between 0 to 100 with a stepsize of 2.

5.5.2 Robot Interaction Strategies

We design two interaction strategies for the intelligent agent - a strategy using the state dependent reward weights (Eq. 5.11) and the other using a constant reward weight of $w = 0.8062$, which is based on our earlier work [8]. The participants were told that they will interact with two different robots - one called Orion and the other called Titan. Behind the scenes, these corresponded to the constant and the state dependent strategies, respectively.

5.5.3 Active Threat Selection

In order to emphasize the different risk appetites of the two robot strategies, we adopt the following strategy for setting the threats and threat levels: 50% of the time the threats are set randomly with a probability of 0.6. With the remaining 50% chance, the threat levels are set actively to induce a difference between the two interaction strategies. This is done by

leveraging the differing d^* values (Eq. 5.10) for the two strategies. Essentially, we compute d^* for both strategies and randomly generate a value between them as the threat level. This will, in general, lead to a riskier recommendation by the state dependent strategy and a more conservative recommendation by the constant strategy. The threat is then sampled from a Bernoulli distribution based on this threat level.

5.5.4 Experiment Design

We designed a 2 (robot strategy) x 2 (vulnerability) mixed design for the human-subjects study. The robot strategy was a within subjects factor while the vulnerability was a between subjects factor. Each participant completed 2 missions, one with each robot strategy. Each mission entailed searching sequentially through 10 search sites. The order for the manipulations was counterbalanced to avoid any learning effects. The details of the manipulations are presented below. The study was approved by the University of Michigan Institutional Review Board under ID HUM00262685.

5.5.4.1 Robot Strategy

The two interaction strategies are presented in Sec. 5.5.2. These were operationalized for the participants as two different intelligent agents called Orion (constant) and Titan (state dependent). The participants were not told any technical details about the differences between the intelligent agents.

5.5.4.2 Vulnerability

We had two levels of vulnerability. In the low vulnerability condition, the participants started from a health value of 100 and a time left value of 100. In the high vulnerability condition, the participants started with a health value of 40 and a time left value of 40. These were operationalized as high or low armor and high or low mission time. The participants were notified that the amount of armor only affects the total damage the soldier can take. Each time the soldier takes damage, the health loss is 10 points, irrespective of armor level.

5.5.5 Data Collection

We recruited 40 students (Age: 22.18 ± 3.05, 21 Female) from the University of Michigan. The participants came in, signed an informed consent form, and went through a detailed instructional video that briefed them about the experimental procedure. After that, they completed a pre-experimental questionnaire consisting of statistically significant items from

Chung et al. [21]. Then, they completed two missions, one with each intelligent agent. After each mission, they also completed a post-mission questionnaire about their trust in the intelligent agent, reliance intentions, and workload. Finally, the participants were compensated for their time and given a performance bonus. The base compensation was $15 with a performance bonus of up to $10 based on the amount of health and time lost during the mission. This was done to encourage the participants to lose as little health and time as possible.

We removed the data from any participants who ended the mission prematurely (reached either 0 health or time points at any point during the mission), leaving us with a dataset from 33 participants. Of these, 19 belonged to the low vulnerability condition, and 14 belonged to the high vulnerability condition.

5.5.6 Measures

We divide our measures into trust-related and performance-related measures as detailed below.

5.5.6.1 Trust-related

- **Average Trust Feedback during interaction:** Measured as the average of the trust feedback given by the participants after each search site in the testbed.
- **End of Mission Trust Feedback:** Measured as the trust feedback given by the participants after searching the last search site in the mission.
- **Post Mission Trust:** Measured using Muir's trust questionnaire Muir et al. [75], adapted to the context of autonomous systems. This 8-item slider type survey was administered after each mission the participant completed.
- **Reliance Intentions:** Measured using the questionnaire in Lyones et al. [63]. This 6-item, 7-point Likert type survey was administered after each mission the participant completed.
- **Average Initial Agreement:** Measured using the average number of times the participants' initial decision (before observing the recommendation) matched with the recommendation given by the intelligent agent. This serves as an indicator of how well the recommendation strategy aligns with what the participant would have done in that context without the recommendations.

- **Average Final Agreement:** Measured using the average number of times the participant selected the recommended action as their final choice of action. This serves as an indicator of behavioral trust.

5.5.6.2 Performance-related

- **Average Performance Rating:** Measured as the average of the performance feedback given by the participants to the intelligent agent's recommendations during the mission. This serves as an indicator of how well the participants thought the intelligent agent's recommendations were.
- **Percent Points Lost:** Measured as the percentage of total points lost (health + time) with respect to the starting amount of points. A lower amount of points lost indicates better performance.

5.5.7 Results

This section presents major results from out statistical analyses. Mixed-methods ANOVA was performed to test for statistically significant differences between the interaction strategies and their interactions with the level of vulnerability.

5.5.7.1 Average Trust During Interaction

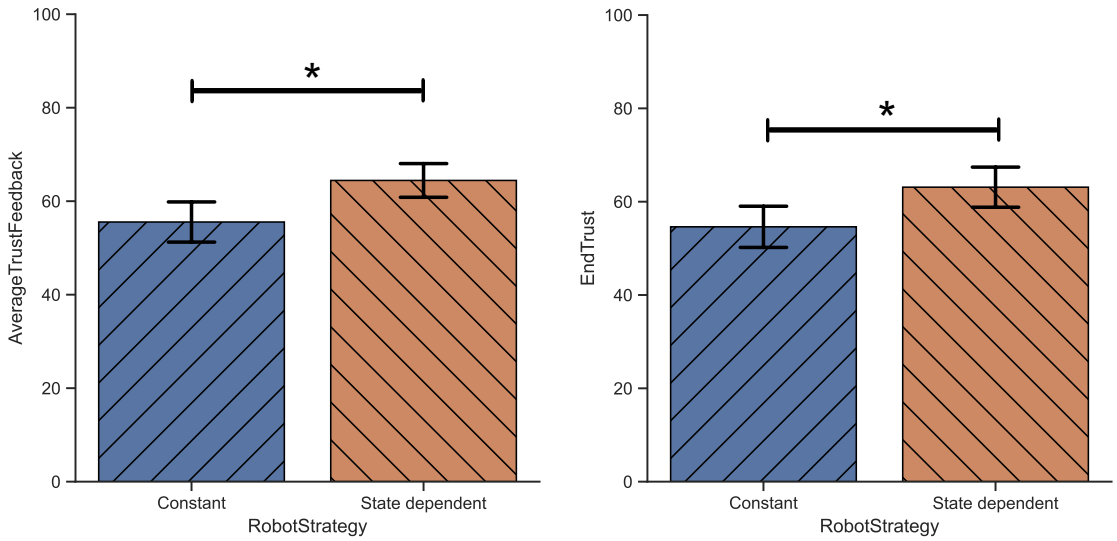
We find a significant main effect of robot strategy ($F(1, 31) = 6.270,\ p = 0.018$) (see Fig. 5.11a). The state dependent strategy results in significantly higher trust reported during the interaction.

5.5.7.2 End-of-mission Trust

We find a significant main effect of robot strategy ($F(1, 31) = 5.103,\ p = 0.032$) (see Fig. 5.11b). The state dependent strategy leads to a higher level of trust towards the end of the mission, showing that it is able to gain the trust of the participants.

5.5.7.3 Post-mission Trust Survey

We find a significant main effect of robot strategy ($F(1, 31) = 5.572,\ p = 0.025$) (see Fig. 5.12a). The participants report a higher level of trust in the state dependent strategy compared to the constant strategy, showing the alignment with the trust feedback collected in the testbed.



(a) Average Trust

(b) Trust After Last Site

Figure 5.11: Trust measures captured in the testbed: (a) Trust feedback given by the participants averaged over the 10 search sites, (b) Trust feedback given by the participants after searching the last search site.

We also see a significant interaction effect of robot strategy and vulnerability on the reported post-mission trust ($F(1, 31) = 13.210,\ p < 0.001$) (see Fig. 5.12b). This shows that the gain in trust of the state dependent strategy over the constant strategy is more pronounced in the high vulnerability condition.

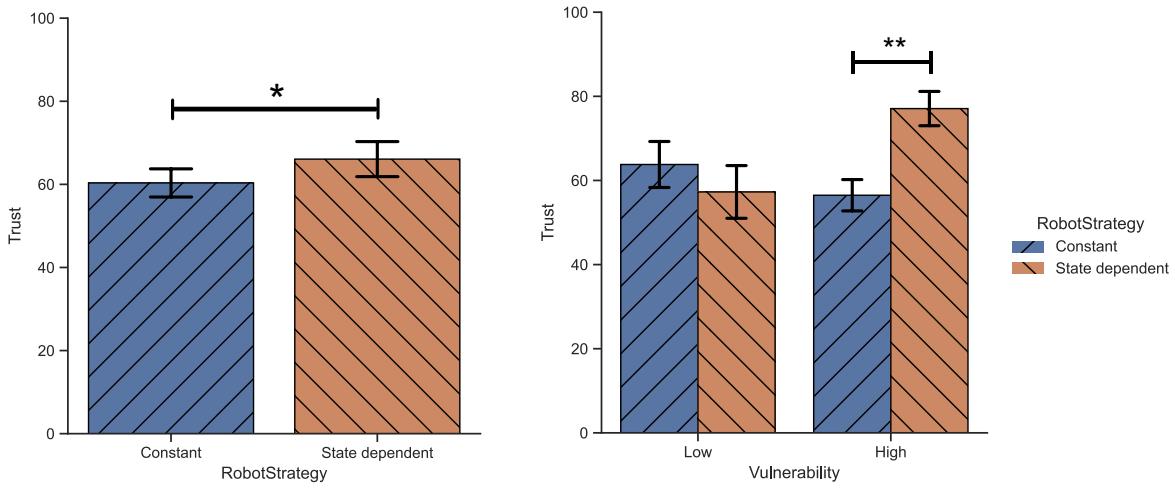
5.5.7.4 Post mission Reliance Intentions

We find a significant main effect of robot strategy on the participants' reported future reliance intentions [63] on the intelligent agent ($F(1, 31) = 9.813,\ p = 0.004$) (see Fig. 5.13a). The participants would rely on the state dependent strategy more than on the constant strategy in potential future interactions.

We also find a significant interaction effect of robot strategy and vulnerability on future reliance intentions ($F(1, 31) = 10.358,\ p = 0.003$) (see Fig. 5.13b). We see that the participants report to rely on the state dependent strategy a lot more compared to the constant strategy in the high vulnerability condition.

5.5.7.5 Average Initial Agreement

We find a significant main effect of robot strategy ($F(1, 31) = 25.530,\ p < 0.001$) on the average agreement between the initial action choice made by the participant and the recom-



(a) Main Effect

(b) Interaction Effect

Figure 5.12: Trust measured using a subjective trust questionnaire [75] after completing a mission. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability.

mendation given by the intelligent agent (see Fig. 5.14a). We see that the state dependent strategy results in a lot more agreements than the constant strategy, showing that the reward weights are aligned with the participant's preferences.

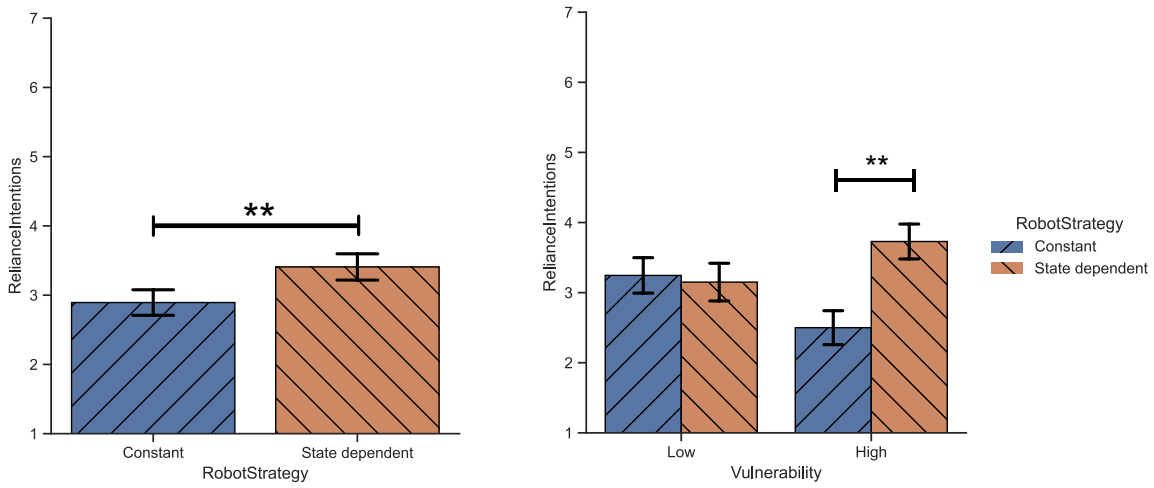
We also see a significant interaction effect of robot strategy and vulnerability on the initial agreements ($F(1, 31) = 7.144,\ p = 0.012$). We see that the difference in initial agreements between the constant and the state dependent strategies is much higher in the high vulnerability condition (see Fig. 5.14b).

5.5.7.6 Average Final Agreement

We find a significant main effect of robot strategy ($F(1, 31) = 32.278,\ p < 0.001$) on the average agreement between the final action choice of the participants and the recommendation given by the intelligent agent (see Fig. 5.15a). Again, the state dependent strategy results in more agreement with what the participants finally decide to do, indicating a higher level of behavioral trust in this strategy compared to the constant strategy.

We also find a significant interaction effect of robot strategy and vulnerability on the average final agreements ($F(1, 31) = 7.780,\ p = 0.009$). We see that the difference in agreements between the strategies is highest in the high vulnerability condition (see Fig. 5.15b).

The results above indicate support for hypothesis **H1** and **H3**, where we see a higher level



(a) Main Effect

(b) Interaction Effect

Figure 5.13: Reliance intentions measured using a 6-item 7-point Likert scale mentioned in Lyons et al. [63]. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability.

of trust reported and observed for the state-dependent strategy than the constant strategy. The interaction effects between the robot strategy and vulnerability show that vulnerability indeed moderates the effect of robot strategy on trust, where the effect of using the state-dependent strategy is higher under the conditions of high vulnerability.

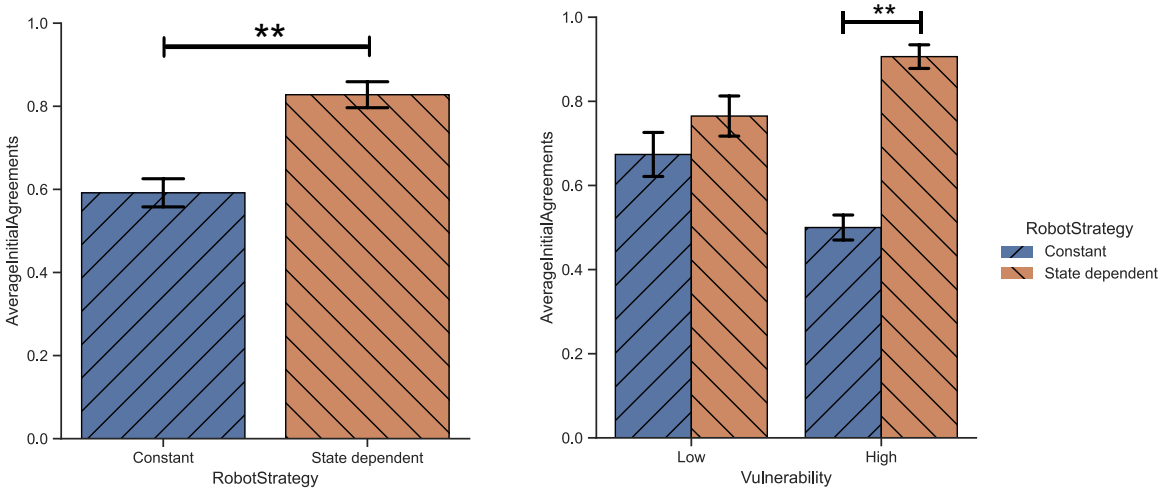
5.5.7.7 Objective Performance - Percent Points Lost

We find a significant main effect of robot strategy on the percent points lost ($F(1, 31) = 6.666,\ p = 0.015$). The percentage is computed with respect to the total points the soldier had at the beginning of the mission. A lower value indicates better performance. We see that the state dependent strategy leads to a lower loss in points (see Fig. 5.16a), indicating that the team performs better when paired with the state dependent strategy.

There is also a significant interaction effect of robot strategy and vulnerability ($F(1, 31) = 7.993,\ p = 0.008$). We see that under the high vulnerability condition, participants lose a significantly less amount of points when paired with the state dependent strategy (see Fig. 5.16b).

5.5.7.8 Subjective Performance - Average Performance Rating

We find a significant main effect of robot strategy ($F(1, 31) = 4.985, p = 0.033$) on the performance rating given to the robot's recommendations by the participants during the



(a) Main Effect

(b) Interaction Effect

Figure 5.14: Agreement between the participants' initial choice of action with the robot's recommendation, averaged over the interaction period. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability.

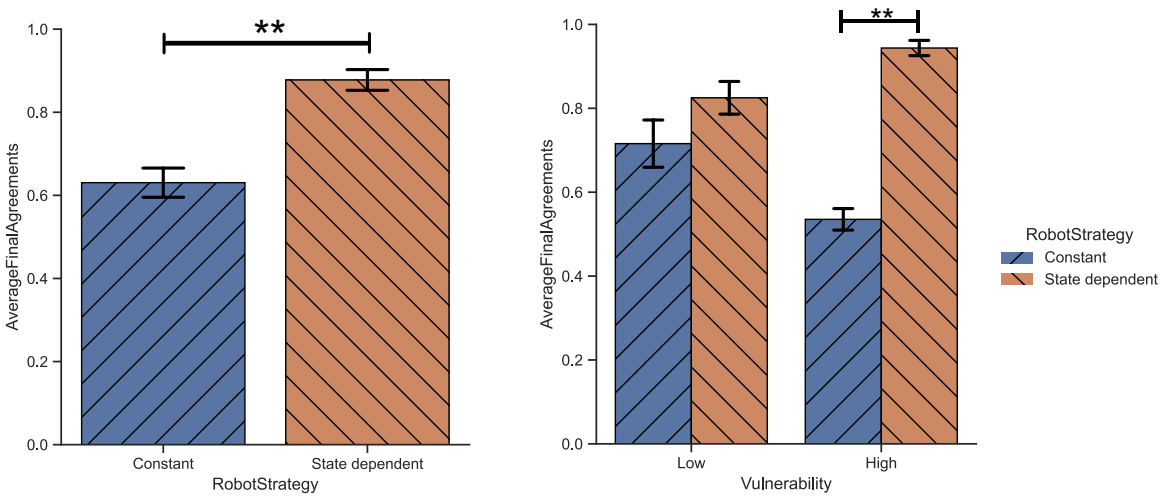
interaction, averaged over the mission. As can be seen in Fig. 5.17a, the participants rate the state dependent strategy's recommendations as better performing than the constant strategy's recommendations.

There is also a significant interaction effect of robot strategy and vulnerability ($F(1, 31) = 6.084,\ p = 0.019$) on the average performance rating of recommendations. As seen in Fig. 5.17b, the difference in performance is more apparent in the high vulnerability condition, with the state dependent strategy being rated higher in performance.

The results above indicate support for hypothesis **H2** and **H3**, where we see a higher objective team performance when interacting with the state-dependent strategy than when interacting with the constant strategy. The interaction effects between the robot strategy and vulnerability show that vulnerability indeed moderates the effect of robot strategy on performance, where the effect of using the state-dependent strategy is higher under the conditions of high vulnerability.

5.6 Conclusion

In this paper, we studied the effect of learning and using a state-dependent reward function in the context of reconnaissance missions where a robot acts as an action recommender to the human. Towards this, we conducted two studies - one through Amazon Mechanical Turk



(a) Main Effect

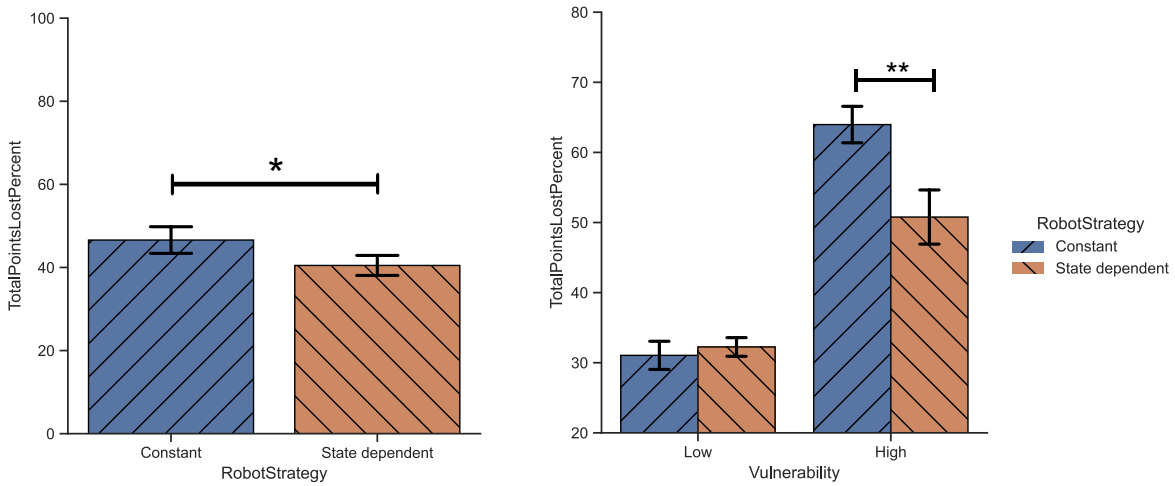
(b) Interaction Effect

Figure 5.15: Agreement between the participants' final choice of action with the robot's recommendation, averaged over the interaction period. (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability.

to collect data and learn the state dependent reward function and another in-person study which tries to see the effects of a robot using this state dependent reward function during interaction on trust and team performance. This study extends our previous work [9] by considering the context of the current state in the reward function within the trust-aware Markov Decision Process.

Our results indicate that the robot using the state dependent strategy results in higher trust, both subjective and behavioral, and better team performance, both objective and subjective, compared to the robot using a constant strategy. This is in line with prior literature like Chiou et al. [19] where the authors conclude that automation responsivity could be more relevant than reliability and reliance of automated systems. In our scenario, we incorporate responsivity through the state dependent reward function of the robot - the risk appetite of the state dependent strategy changes based on the current state of the human-robot team. Further, we see that these trust and performance benefits are more pronounced in the situation where the vulnerability is higher, which is in line with out own prior work [9] where we found that aligning the robot's reward function with that of the human is more important when risks are higher.

We conclude this chapter by discussing some limitations in this work and possible directions of future research that could try and address these limitations.



(a) Main Effect

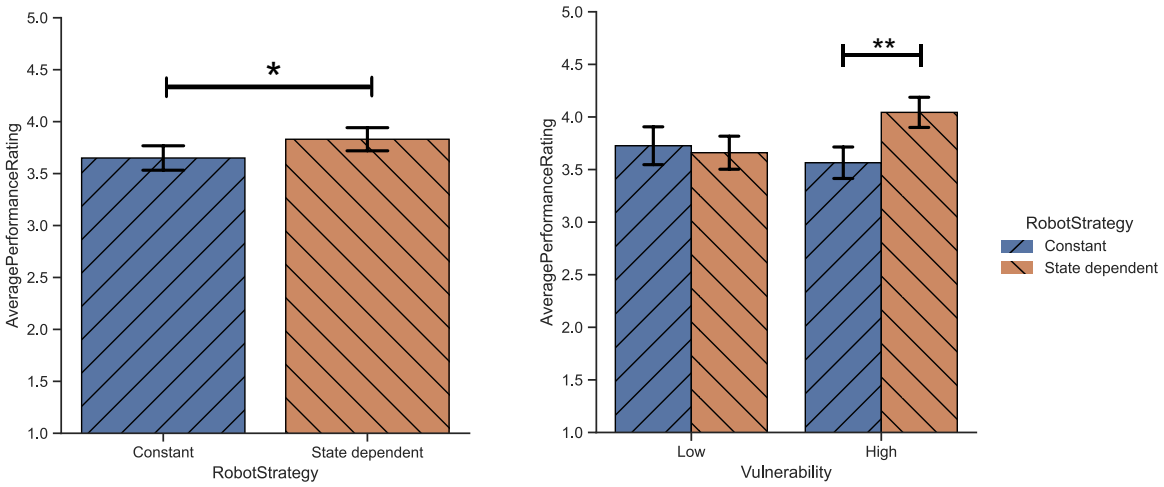
(b) Interaction Effect

Figure 5.16: Percent total points lost over the interaction period, taken with respect to the total points the soldier starts with (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability.

5.6.1 Limitations and Future Work

We have focused on a dyadic human-robot interaction where one human interacts with one robot. In real life, however, teams are usually larger and consist of multiple humans and robots. Future work should try and address this limitation by incorporating multi-agent trust models (e.g. [39]) into a decision-making framework. Some very recent work has started to tackle this problem [22, 44] but a lot more work needs to be done toward solving this problem.

We learned a reward function and demonstrated its effectiveness in a context where the conflicting objectives were about losing health and losing mission time. Most people tend to bias themselves toward saving the soldier's health, thus resulting in the reward weight for losing health being greater than 0.5 at all states (Fig. 5.11). Given that, we cannot claim that our results will directly translate to another context where there may be a larger variance in the preferences of humans about the objective. We encourage researchers to study this problem in different contexts



(a) Main Effect

(b) Interaction Effect

Figure 5.17: Subjective performance rating of the intelligent agent captured in the testbed, averaged over the interaction period (a) main effect of robot strategy, (b) interaction effect of robot strategy and vulnerability.

CHAPTER 6

Conclusion, Limitations, and Future Research Directions

6.1 Conclusion

With the rapid progression of artificial intelligence technologies, it has become increasingly important to ensure that human reliance on these systems does not inadvertently exceed appropriate boundaries. The concept of trust has been widely recognized as a critical determinant in guiding humans' willingness to depend on automated and intelligent agents. In light of this, there is a growing consensus in the literature that intelligent agents should possess mechanisms that enable them to be aware of and respond to the fluctuating trust levels of their human collaborators. Specifically, autonomous agents should be capable of dynamically adapting their behavior to foster appropriate trust, thereby safeguarding against both over-reliance and under-reliance.

Equally significant is the need to align the values underpinning intelligent agents with those held by their human counterparts. Value alignment, particularly through reward alignment, has garnered considerable attention within the research community in recent years, which has led to plenty of algorithmic advances to ensure alignment. However, several important gaps still remain in our empirical understanding of its broader implications. Most prominently, the effect of value alignment on key outcomes such as human trust on the system and team performance has yet to be systematically and empirically evaluated.

This dissertation addresses these critical gaps by investigating the interplay between trust, value alignment, and performance in human–autonomy collaborative settings. Through a series of rigorous empirical studies, we seek to elucidate the extent to which reward alignment mechanisms influence trust calibration between humans and intelligent agents, as well as the subsequent impact on collaborative team outcomes. Our findings aim to contribute to a more nuanced understanding of value-driven agent behaviors, ultimately informing the design of autonomous systems that promote trust-driven interactions with themselves.

- Chapter 3 proposes the trust-aware POMDP framework which incorporates mathematical trust models and human behavior models for enabling trust-aware decision-making by intelligent agents. By leveraging these models, the framework seeks to provide agents with the capability to dynamically adjust their decision-making strategies in accordance with observed fluctuations in human trust levels. We empirically demonstrate the effectiveness of this framework through a human-subjects experiment conducted in a custom-designed 3D testbed using Unreal Engine. By performing a clustering analysis on the collected data, we find three distinct types of trust dynamics observed among people. Further, we are able to associate personal characteristics to type of trust dynamics exhibited by an individual, making it possible for future researchers to design more personalized trust dynamics models, considering the person's individual characteristics.

- Chapter 4 augments the trust-aware framework by introducing a personalized reward learning paradigm during agent–human interaction. We propose the Bounded Rationality Disuse model, a novel human trust-behavior model that allows the robot to gain insight into the underlying preferences of the human during interaction. Using Bayesian IRL, we are able to learn these preferences during interaction. Our empirical results indicate that a robot which incorporates these learned preferences can engender higher levels of human trust and enhance overall team performance, when starting with an uninformed prior on human preferences. Conversely, our findings suggest that if the robot already possesses a well-informed prior about human preferences, the incremental benefit of real-time personalized learning may be diminished, highlighting the importance of prior knowledge in the personalization process.

- Chapter 5 delves deeper into the effects of the reward function on trust and team performance. This chapter presents a framework for learning context-dependent reward functions from human behavioral data, allowing intelligent agents to tailor their strategies to dynamically evolving human preferences. We demonstrate this framework through a study done on Amazon Mechanical Turk, allowing us to learn a context-dependent prior on the reward function of humans. Through this learned reward function, we develop a robot strategy that adapts to the changing preferences of the human depending on the context. We compare this strategy with a strategy using a constant reward function, which proved to work well in Chapter 4, in cases of low and high vulnerability. We find that our context dependent strategy garners higher trust from the human and results in better team performance, especially in the high vulnerability condition. This underscores the significance of adaptive reward mechanisms in

optimizing human–autonomy collaboration.

This dissertation makes significant advances in the development of trust-aware robotics, focusing on the design of intelligent agents that are capable of estimating trust levels, modeling and predicting human behavior as influenced by trust, and dynamically adapting their own actions to foster trust and enhance team performance. By endowing robotic agents with the ability to interpret and respond to the nuanced interplay between trust and human behavior, this research highlights the critical role of adaptability in achieving smooth and effective collaboration between humans and intelligent systems. Such trust-aware adaptability is essential in the evolution of intelligent agents, ensuring that future human–robot partnerships are characterized by increased mutual understanding, reliability, and overall effectiveness in a diverse range of task environments.

6.2 Limitations and Future Research Directions

While our research has yielded valuable insights, it is important to acknowledge several limitations inherent in the current stage of our work. These limitations not only delineate the boundaries of our present contributions but also illuminate promising avenues for future investigation within this domain. Below, we elaborate on these areas for further exploration.

6.2.1 Further Personalizing Trust Dynamics Models

Chapter 3 identified three distinct types of trust dynamics in human–autonomy teaming: Bayesian Decision Makers, Disbelievers, and Oscillators. It also established preliminary associations between these trust dynamics and certain individual characteristics, such as personality traits.

While these findings are significant, they represent an exploratory effort aimed at determining whether such distinct categories of trust dynamics exist. Consequently, the analysis was limited by the relatively small set of personal characteristics included in the questionnaires. Additional studies are therefore needed to validate these results across a wider range of human–autonomy teaming contexts and with more diverse participant profiles. Chung and Yang [21] have taken initial steps in this direction, identifying further associations between individual characteristics and trust dynamics in a sequential threat detection task. Building on this work, future research should seek to develop varied experimental scenarios and confirm the robustness of these distinct trust dynamics across contexts.

Another limitation of the current framework lies in the trust dynamics model itself, which effectively estimates trust levels for Bayesian Decision Makers and Disbelievers but performs

inadequately for Oscillators. This limitation underscores the need for more personalized modeling approaches that can capture the unique characteristics of oscillatory trust patterns. One promising avenue is the incorporation of "discounting" mechanisms that reduce the influence of outdated performance data from the autonomous agent, allowing more recent interactions to play a larger role in trust estimation. Preliminary work within our research group exploring this approach has yielded encouraging results.

Finally, while the existing model performs well for Disbelievers, it remains ineffective in increasing their trust in the autonomous agent. Future studies should therefore investigate interaction strategies such as the use of explanations, affective cues, or adaptive communication mechanisms to more effectively calibrate Disbelievers' trust to an appropriate level.

6.2.2 Tackling Diverse Human-Robot Collaborative Tasks

The research presented in this dissertation primarily focused on an ISR mission scenario, characterized by two competing objectives: minimizing damage to the soldier and minimizing overall mission duration. As expected, the findings indicate that most participants exhibit a strong bias toward ensuring the soldier's safety, often preferring to incur time losses even in relatively low-risk situations.

In Chapter 4, we observed that when the robot possessed an accurate prior estimate of the reward weights, additional personalization did not yield improvements in trust or overall team performance. This outcome likely stems from the static prior already being well-aligned with the majority of participants' preferences, such that further personalization only introduced unnecessary variability and unpredictability into the robot's behavior. However, these findings cannot be directly generalized to scenarios in which individual preferences vary more substantially. Future research should therefore investigate the effects of personalization in a broader range of human–robot collaboration contexts, where preference diversity is more pronounced. Different personalization strategies may prove more effective in such environments. Moreover, the present study is limited to a "robot–recommender–human–decision-maker" interaction paradigm. Many other forms of human–robot collaborative tasks remain unexplored from a trust-aware perspective. Continued research in diverse task domains is essential to develop a more comprehensive understanding of trust dynamics and personalization in human–robot teaming.

6.2.3 Tackling Multi-Action Scenarios

The ISR mission scenario considered in this work involved only two possible actions at each decision point: USE or NOT USE the RARV. This binary decision structure enabled the

application of the reverse psychology model of human behavior described in Chapter 3, as well as the formulation of the reward-based performance metric employed throughout this dissertation.

While the Bounded Rationality Disuse model introduced in Chapter 4 can be readily generalized to scenarios involving multiple actions, extending the trust dynamics model presents greater challenges, mainly due to the difficulty of defining an appropriate performance metric. In the binary action case, performance is easily assessed: the outcome is considered successful if the reward associated with the recommended action exceeds that of the alternative. However, in multi-action scenarios, human decision-making becomes more complex. Individuals may exhibit a satisficing behavior [92], evaluating performance as successful if the chosen action yields a reward deemed sufficiently satisfactory rather than optimal. Alternatively, humans may perceive performance on a continuum, assigning graded performance values across different actions, or they may rely on entirely different evaluative criteria. Extending the current framework to accommodate multi-action decision-making would therefore represent a valuable direction for future research, offering deeper insights into the nuances of trust, bounded rationality, and performance evaluation in more complex human–autonomy teaming contexts.

6.2.4 Shifting to the Multi-Robot Multi-Human Paradigm

The research presented in this dissertation primarily focused on a dyadic human–robot interaction paradigm, involving a single human collaborating with a single intelligent agent. While this configuration already presents substantial challenges in terms of trust and behavior modeling, value alignment, and adaptive interaction design, extending the framework to multi-human and multi-robot settings introduces an additional layer of complexity. In such environments, multiple bidirectional and cross-agent trust relationships emerge, creating a rich but difficult-to-model network of interdependencies.

This raises several fundamental research questions: How can trust dynamics be effectively modeled between each human and each robot, especially when trust may propagate indirectly through observation or communication? How should tasks and robots be optimally assigned to individual humans to maintain balanced workload distribution, foster appropriate trust calibration, and prevent over- or under-reliance? Furthermore, in single-robot multi-human contexts, how should the robot navigate potential value misalignments among its human collaborators, particularly when individual preferences or risk appetites conflict?

Recent work has begun to explore trust modeling in multi-human, multi-robot systems by incorporating both direct interaction experiences and indirect observations of others'

interactions [37, 44]. These efforts represent important early steps toward understanding collective trust dynamics and distributed decision-making in complex human–robot teams.

Nonetheless, this area of research remains in its formative stages. Substantial further work is needed to develop scalable computational models of trust that account for social influences, role differentiation, and task interdependence among multiple agents. Advancing this line of inquiry will be critical for enabling effective collaboration in future large-scale, heterogeneous human–robot teams operating in dynamic, real-world environments.

APPENDIX A

Questionnaires Used in Human-Subject Experiments

A.1 Subjective Trust Questionnaire

Survey modified from [75] to fit the context of "intelligent agents in ISR missions". All answers were given on a slider scale from 0 to 100 with anchors "None at all (0)", "Neutral (50)", and "Extremely high (100)".

Prompt: There are 8 statements below. Please move the sliders to the point that best describes your feeling or your impression on the intelligent agent that recommends if you should use the armored robot or not.

- To what extent does the agent perform its function properly?
- To what extent can the agent's behavior be predicted from moment to moment?
- To what extent can you count on the agent to do its job?
- To what extent does the agent perform the task it was designed to do in the system?
- To what extent does the agent respond similarly to similar circumstances at different points in time?
- Your degree of faith that the agent will be able to cope with other ISR missions in the future.
- Your degree of trust in the agent to respond accurately.
- Your overall degree of trust in the agent

A.2 Reliance Intentions Scale

Survey modified from [65] to fit the context of "intelligent agents in ISR missions". All answers were given on a 7-point Likert type scale with anchors "strongly disagree (1)" to "strongly agree (7)".

Prompt: In the following 6 items, consider your attitudes toward the intelligent agent responsible for the recommendations you received. Use the scale provided to select the best response to the item. (Scale: strongly disagree = 1, to strongly agree = 7)

- I would rely on the agent without hesitation.
- I think using the agent will lead to positive outcomes.
- I would feel comfortable relying on the agent in the future.
- If I had my way, I would NOT let the agent have any influence over issues that are important to the task (search and rescue). (reverse coded)
- I really wish I had a good way to monitor the decisions of the agent. (reverse coded)
- I would be comfortable allowing the agent to implement its search decision, even if I could not monitor it.

A.3 NASA Task Load Index (TLX)

Obtained from [43] to measure the workload of the human. We removed the physical workload component from the questionnaire since it was not relevant to our case. Each item was rated on a slider from 0 to 100 with anchors "Very low (0)", "Neutral (50)", and "Very high (100)".

Prompt: Please move the 5 sliders below to the points which best describe your feeling.
Descriptors:
Mental Demand – How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching etc.). Was the task easy or demanding, simple or complex, exacting or forgiving?
Temporal Demand – How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance – How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort – How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level – How insecure, discouraged, irritated stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

- Mental Demand
- Temporal Demand
- Performance (reverse coded)
- Effort
- Frustration

A.4 Perfect Automation Schema

Perfect Automation Schema (PAS) was measured using the 7-item scale developed by Merritt et al. [72]. Of these, 4 items measure high expectations from the autonomy and the other 3 items measure All-or-none thinking. Each item was rated on a 7-point Likert type scale, with anchors at "Strongly disagree (1)" to " Strongly agree (7)".
Prompt: There are 7 statements below for you to rate your beliefs regarding automated systems. Please select the value on each line that best describes your feeling. (Note: strongly disagree=1; strongly agree=7).

- Automated systems have 100% perfect performance.
- Automated systems rarely make mistakes.
- Automated systems can always be counted on to make accurate decisions.
- People have NO reason to question the decisions automated systems make.
- If an automated system makes an error, then it is broken.
- If an automated system makes a mistake, then it is completely useless.
- Only faulty automated systems provide imperfect results.

A.5 Big 5 Personality Traits

We use the 20-item mini-IPIP scale [27] to measure the big 5 personality traits of Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N). Each item is rated on a 5-point Likert scale.

Prompt: There are several scales for you to rate your personality. Please select the value on each line at the point which best describe your feeling. (Note: strongly disagree=1; strongly agree=5).

- I am the life of the party. (E)
- I sympathize with others' feelings. (A)
- I get chores done right away. (C)
- I have frequent mood swings. (N)
- I have a vivid imagination. (O)
- I don't talk a lot. (E, reverse)
- I am not interested in other people's problems. (A, reverse)
- I often forget to put things back in their proper place. (C, reverse)
- I am relaxed most of the time. (N, reverse)
- I am not interested in abstract ideas. (O, reverse)
- I talk to a lot of different people at parties. (E)
- I feel others' emotions. (A)
- I like order. (C)
- I get upset easily. (N)
- I have difficulty understanding abstract ideas. (O)
- I keep in the background. (E, reverse)
- I am not really interested in others. (A, reverse)
- I make a mess of things. (C, reverse)

- I seldom feel blue. (N, reverse)
- I do not have a good imagination. (O, reverse)

A.6 Trust Propensity

Propensity to trust autonomous systems was assessed using a 6-item scale developed by Merritt et al. [71]. Each item was rated on a 5-point Likert type scale with anchors "strongly disagree (1)" and "strongly agree (5)".
Prompt: There are 6 statements below about autonomous systems. Please select a value that best describes your feeling for each statement. (Note: strongly disagree=1; strongly agree=5)

- I usually trust autonomous systems until there is a reason not to.
- For the most part, I distrust autonomous systems. (reverse coded)
- In general, I would rely on an autonomous system to assist me.
- My tendency to trust autonomous systems is high.
- It is easy for me to trust autonomous systems to do their job.
- I am likely to trust an autonomous system even when I have little knowledge about it.

BIBLIOGRAPHY

[1] Kumar Akash, Neera Jain, and Teruhisa Misu. Toward Adaptive Trust Calibration for Level 2 Driving Automation. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 538–547, October 2020. arXiv:2009.11890 [cs].

[2] Kumar Akash, Katelyn Polson, Tahira Reid, and Neera Jain. Improving human-machine collaboration through transparency-based feedback – part i: Human trust and workload model. *IFAC-PapersOnLine*, 51(34):315–321, 2019. 2nd IFAC Conference on Cyber-Physical and Human Systems CPHS 2018.

[3] Kumar Akash, Tahira Reid, and Neera Jain. Improving human-machine collaboration through transparency-based feedback – part ii: Control design and synthesis. *IFAC-PapersOnLine*, 51(34):322–328, 2019. 2nd IFAC Conference on Cyber-Physical and Human Systems CPHS 2018.

[4] Saurabh Arora and Prashant Doshi. A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*, 297:103500, August 2021.

[5] Hebert Azevedo-Sa, Suresh Kumaar Jayaraman, X. Jessie Yang, Lionel P. Robert, and Dawn M. Tilbury. Context-adaptive management of drivers' trust in automated vehicles. *IEEE Robotics and Automation Letters*, 5(4):6908–6915, 2020.

[6] Chandrayee Basu, Erdem Bıyık, Zhixun He, Mukesh Singhal, and Dorsa Sadigh. Active Learning of Reward Dynamics from Hierarchical Queries. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 120–127, Macau, China, November 2019. IEEE.

[7] Shreyas Bhat, Joseph B. Lyons, Cong Shi, and X. Jessie Yang. Clustering trust dynamics in a human-robot sequential decision-making task. *IEEE Robotics and Automation Letters*, 7(4):8815–8822, 2022.

[8] Shreyas Bhat, Joseph B. Lyons, Cong Shi, and X. Jessie Yang. Evaluating the Impact of Personalized Value Alignment in Human-Robot Interaction: Insights into Trust and Team Performance Outcomes. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 32–41, Boulder CO USA, March 2024. ACM.

[9] Shreyas Bhat, Joseph B. Lyons, Cong Shi, and X. Jessie Yang. *Value Alignment and Trust in Human-Robot Interaction: Insights from Simulation and User Study*, pages 39–63. Springer Nature Switzerland, Cham, 2024.

[10] Shreyas Bhat, Joseph B. Lyons, Cong Shi, and X. Jessie Yang. Effects of learning state dependence of reward weights on trust and team performance in a human-robot sequential decision-making task. In *2025 IEEE 5th International Conference on Human-Machine Systems (ICHMS)*, pages 35–40, 2025.

[11] Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, and Peter A. Hancock. Human-robot interaction: Developing trust in robots. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 109–110, 2012.

[12] Erdem Biyik, Nima Anari, and Dorsa Sadigh. Batch Active Learning of Reward Functions from Human Preferences. *ACM Transactions on Human-Robot Interaction*, 13(2):1–27, June 2024.

[13] Philip Bobko, Alex J Barelka, and Leanne M Hirshfield. The Construct of State-Level Suspicion: A Model and Research Agenda for Automated and Information Technology (IT) Contexts. *Human Factors*, 56(3):489–508, 2014.

[14] Daniel S Brown, Jordan Schneider, Anca Dragan, and Scott Niekum. Value alignment verification. In *International Conference on Machine Learning*, pages 1105–1115. PMLR, 2021.

[15] Matthew Brzowski and Dan Nathan-Roberts. Trust measurement in human–automation interaction: A systematic review. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):1595–1599, 2019.

[16] Jessie Y. C. Chen, Shan G. Lakhmani, Kimberly Stowers, Anthony R. Selkowitz, Julia L. Wright, and Michael Barnes. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, 19(3):259–282, 2018.

[17] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. Planning with trust for human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, page 307–315, New York, NY, USA, 2018. Association for Computing Machinery.

[18] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. Trust-aware decision making for human-robot collaboration: Model learning and planning. *J. Hum.-Robot Interact.*, 9(2), jan 2020.

[19] Erin K. Chiou and John D. Lee. Trusting automation: Designing for responsivity and resilience. *Human Factors*, 65(1):137–165, 2023. PMID: 33906505.

[20] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[21] Hyesun Chung and X. Jessie Yang. Associations between Trust Dynamics and Personal Characteristics. In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, pages 1–6, Toronto, ON, Canada, May 2024. IEEE.

[22] Hyesun Chung and X. Jessie Yang. Communication dynamics and team performance in multiple-operator-multiple-agent (moma) team. In *2025 IEEE 5th International Conference on Human-Machine Systems (ICHMS)*, pages 81–86, 2025.

[23] Myke C. Cohen, Mustafa Demir, Erin K. Chiou, and Nancy J. Cooke. The dynamics of trust and verbal anthropomorphism in human-autonomy teaming. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*, pages 1–6, 2021.

[24] Nancy J Cooke, Jamie C Gorman, Christopher W Myers, and Jasmine L Duran. Interactive team cognition. *Cogn. Sci.*, 37(2):255–285, March 2013.

[25] Resul Dagdanov, Milan Andrejevic, Dikai Liu, and Chin-Teng Lin. Improving Trust Estimation in Human-Robot Collaboration Using Beta Reputation at Fine-grained Timescales. *IEEE Robotics and Automation Letters*, 10(8):8562–8569, August 2025. arXiv:2411.01866 [cs].

[26] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 251–258, 2013.

[27] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, 18(2):192, 2006.

[28] Na Du, Jacob Haspiel, Qiaoning Zhang, Dawn Tilbury, Anuj K. Pradhan, X. Jessie Yang, and Lionel P. Robert. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation Research Part C: Emerging Technologies*, 104:428–442, July 2019.

[29] Clémence Dubois and Jerome Le Ny. Adaptive task allocation in human-machine teams with trust and workload cognitive models. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3241–3246. IEEE, 2020.

[30] Mary T Dzindolet, Scott A Peterson, Regina A Pomranky, Linda G Pierce, and Hall P Beck. The role of trust in automation reliance. *International journal of human-computer studies*, 58(6):697–718, 2003.

[31] Brochu Eric, Nando Freitas, and Abhijeet Ghosh. Active preference learning with discrete choice data. *Advances in neural information processing systems*, 20, 2007.

[32] Connor Esterwood and Lionel P. Robert Jr. Three Strikes and you are out!: The impacts of multiple human–robot trust violations and repairs on robot trustworthiness. *Computers in Human Behavior*, 142:107658, May 2023.

[33] Jaime F. Fisac, Monica A. Gates, Jessica B. Hamrick, Chang Liu, Dylan Hadfield-Menell, Malayandi Palaniappan, Dhruv Malik, S. Shankar Sastry, Thomas L. Griffiths, and Anca D. Dragan. Pragmatic-pedagogic value alignment. In Nancy M. Amato, Greg Hager, Shawna Thomas, and Miguel Torres-Torriti, editors, *Robotics Research*, pages 49–57, Cham, 2020. Springer International Publishing.

[34] Michael W. Floyd, Michael Drinkwater, and David W. Aha. Learning Trustworthy Behaviors Using an Inverse Trust Metric. In *Robust Intelligence and Trust in Autonomous Systems*, pages 33–53. Springer US, Boston, MA, 2016.

[35] Thomas L. Griffiths and Joshua B. Tenenbaum. Optimal predictions in everyday cognition. *Psychological science*, 17(9):767–773, 2006.

[36] Yaohui Guo, Cong Shi, and Xi Jessie Yang. Reverse psychology in trust-aware human-robot interaction. *IEEE Robotics and Automation Letters*, 6(3):4851–4858, 2021.

[37] Yaohui Guo, X. Yang, and Cong Shi. Enabling Team of Teams: A Trust Inference and Propagation (TIP) Model in Multi-Human Multi-Robot Teams. In *Robotics: Science and Systems XIX*. Robotics: Science and Systems Foundation, July 2023.

[38] Yaohui Guo and X. Jessie Yang. Modeling and predicting trust dynamics in human-robot teaming: A bayesian inference approach. *International Journal of Social Robotics*, 12 2021.

[39] Yaohui Guo, X. Jessie Yang, and Cong Shi. TIP: A Trust Inference and Propagation Model in Multi-Human Multi-Robot Teams. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 639–643, Stockholm Sweden, March 2023. ACM.

[40] Dylan Hadfield-Menell, Stuart J Russell, Pieter Abbeel, and Anca Dragan. Cooperative inverse reinforcement learning. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[41] P A Hancock, Theresa T Kessler, Alexandra D Kaplan, John C Brill, and James L Szalma. Evolving Trust in Robots: Specification Through Sequential and Comparative Meta-Analyses. *Human Factors*, 63(7):1196–1229, 2020.

[42] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5):517–527, 2011. PMID: 22046724.

[43] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

[44] Drew T. Hidalgo, Yaohui Guo, Joseph B. Lyons, Cong Shi, and X. Jessie Yang. Trust propagation in multi-operator multi-autonomy (moma) teams: Friends versus strangers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 0(0):10711813251370736, 0.

[45] Kevin Anthony Hoff and Masooda Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3):407–434, 2015. PMID: 25875432.

[46] J. D. Power. Automated Vehicles and Insurancc Pulse Policy. Technical report, J. D. Power, 2018.

[47] Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1):53–71, 2000.

[48] Craig J. Johnson, Mustafa Demir, Nathan J. McNeese, Jamie C. Gorman, Alexandra T. Wolff, and Nancy J. Cooke. The impact of training on human–autonomy team communications and trust calibration. *Human Factors*, 65(7):1554–1570, 2023. PMID: 34595958.

[49] Poornima Kaniarasu, Aaron Steinfeld, Munjal Desai, and Holly Yanco. Potential measures for detecting trust changes. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, page 241–242, New York, NY, USA, 2012. Association for Computing Machinery.

[50] Ulas Berk Karli, Shiye Cao, and Chien-Ming Huang. "What If It Is Wrong": Effects of Power Dynamics and Trust Repair Strategy on Trust and Compliance in HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '23, pages 271–280, New York, NY, USA, March 2023. Association for Computing Machinery.

[51] G. Klien, D.D. Woods, J.M. Bradshaw, R.R. Hoffman, and P.J. Feltovich. Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 19(6):91–95, November 2004. Conference Name: IEEE Intelligent Systems.

[52] Spencer C. Kohn, Ewart J. De Visser, Eva Wiese, Yi-Ching Lee, and Tyler H. Shaw. Measurement of Trust in Automation: A Narrative Review and Reference Guide. *Frontiers in Psychology*, 12:604977, October 2021.

[53] Michael Laakasuo, Jussi Palomäki, Anton Kunnari, Sanna Rauhala, Marianna Drosinou, Juho Halonen, Noora Lehtonen, Mika Koverola, Marko Repo, Jukka Sundvall, Aku Visala, and Kathryn B. Francis. Moral psychology of nursing robots: Exploring the role of robots in dilemmas of patient autonomy. *European Journal of Social Psychology*, 53(1):108–128, 2023.

[54] John Lee and Neville Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992.

[55] John D. Lee and Katrina A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004. PMID: 15151155.

[56] Huao Li, Tianwei Ni, Siddharth Agrawal, Fan Jia, Suhas Raja, Yikang Gui, Dana Hughes, Michael Lewis, and Katia Sycara. Individualized mutual adaptation in human-agent teams. *IEEE Transactions on Human-Machine Systems*, 51(6):706–714, 2021.

[57] Yang Li, Jiaxin Xu, Di Guo, and Huaping Liu. Trust-Aware Human–Robot Fusion Decision-Making for Emergency Indoor Patrolling. *IEEE Transactions on Automation Science and Engineering*, 22:4596–4605, 2025.

[58] Zhanrui Liao and Yue Wang. Trust-based variable impedance control of human–robot cooperative manipulation. *Robotics and Computer-Integrated Manufacturing*, 88:102730, August 2024.

[59] Bin Liu. A Survey on Trust Modeling from a Bayesian Perspective. *Wireless Personal Communications*, 112(2):1205–1227, May 2020.

[60] Chang Liu, Jessica B Hamrick, Jaime F Fisac, Anca D Dragan, J Karl Hedrick, S Shankar Sastry, and Thomas L Griffiths. Goal inference improves objective and perceived performance in human-robot collaboration. *arXiv preprint arXiv:1802.01780*, 2018.

[61] Jundi Liu, Kumar Akash, Teruhisa Misu, and Xingwei Wu. Clustering human trust dynamics for customized real-time prediction. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 1705–1712, 2021.

[62] Ruikun Luo, Yifan Weng, Yifan Wang, Paramsothy Jayakumar, Mark J. Brudnak, Victor Paul, Vishnu R. Desaraju, Jeffrey L. Stein, Tulga Ersal, and X. Jessie Yang. A workload adaptive haptic shared control scheme for semi-autonomous driving. *Accident Analysis & Prevention*, 152:105968, 2021.

[63] Joseph Lyons and Svyatoslav Guznov. Individual differences in human–machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, 20(4):440–458, 2019.

[64] Joseph B. Lyons, Izz aldin Hamdan, and Thy Q. Vo. Explanations and trust: What happens to trust when a robot partner does something unexpected? *Computers in Human Behavior*, 138:107473, January 2023.

[65] Joseph B. Lyons, Chang S. Nam, Sarah A. Jessup, Thy Q. Vo, and Kevin T. Wynne. The role of individual differences as predictors of trust in autonomous security robots. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pages 1–5, 2020.

[66] Joseph B Lyons and Charlene K Stokes. Human-human reliance in the context of automation. *Hum. Factors*, 54(1):112–121, February 2012.

[67] Joseph B. Lyons, Charlene K. Stokes, Kevin J. Eschleman, Gene M. Alarcon, and Alex J. Barelka. Trustworthiness and IT suspicion: An evaluation of the nomological network. *Human Factors*, 53(3):219–229, 2011.

[68] Joseph B. Lyons, Katia Sycara, Michael Lewis, and August Capiola. Human–Autonomy Teaming: Definitions, Debates, and Directions. *Frontiers in Psychology*, 12:589585, May 2021.

[69] Joseph B. Lyons, Thy Vo, Kevin T. Wynne, Sean Mahoney, Chang S. Nam, and Darci Gallimore. Trusting autonomous security robots: The role of reliability and stated social intent. *Human Factors*, 63(4):603–618, 2021. PMID: 32027537.

[70] Griffon McMahon, Kumar Akash, Tahira Reid, and Neera Jain. On modeling human trust in automation: Identifying distinct dynamics through clustering of markovian models. *IFAC-PapersOnLine*, 53:356–363, 2020.

[71] Stephanie M. Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 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, 2013.

[72] Stephanie M. Merritt, Jennifer L. Unnerstall, Deborah Lee, and Kelli Huber. Measuring individual differences in the perfect automation schema. *Human Factors*, 57(5):740–753, 2015.

[73] Smitha Milli, Dylan Hadfield-Menell, Anca Dragan, and Stuart Russell. Should Robots be Obedient?, May 2017. arXiv:1705.09990 [cs].

[74] Cristinel Mortici. The proof of muqattash–yahdi conjecture. *Mathematical and Computer Modelling*, 51:1154–1159, 05 2010.

[75] Bonnie Muir and Neville Moray. Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39:429–60, 04 1996.

[76] Changjoo Nam, Phillip Walker, Michael Lewis, and Katia Sycara. Predicting trust in human control of swarms via inverse reinforcement learning. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 528–533, August 2017. ISSN: 1944-9437.

[77] Manisha Natarajan, Esmaeil Seraj, Batuhan Altundas, Rohan Paleja, Sean Ye, Letian Chen, Reed Jensen, Kimberlee Chestnut Chang, and Matthew Gombolay. Human-Robot Teaming: Grand Challenges. *Current Robotics Reports*, 4(3):81–100, August 2023.

[78] Andrew Y. Ng and Stuart Russell. Algorithms for inverse reinforcement learning. In *in Proc. 17th International Conf. on Machine Learning*, pages 663–670. Morgan Kaufmann, 2000.

[79] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 64(5):904–938, August 2022.

[80] Raja Parasuraman and Victor Riley. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2):230–253, 1997.

[81] Arvind Parkhe and Stewart Miller. The structure of optimal trust: A comment and some extensions. *The Academy of Management Review*, 25:10–11, 01 2000.

[82] Claudia Pérez-D'Arpino, Rebecca P. Khurshid, and Julie A. Shah. Experimental assessment of human–robot teaming for multi-step remote manipulation with expert operators. *J. Hum.-Robot Interact.*, 13(3), August 2024.

[83] Charles Pippin and Henrik Christensen. Trust modeling in multi-robot patrolling. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 59–66, 2014.

[84] Deepak Ramachandran and Eyal Amir. Bayesian inverse reinforcement learning. In *IJCAI*, volume 7, pages 2586–2591, 2007.

[85] Nicole Robinson, Jason Williams, David Howard, Brendan Tidd, Fletcher Talbot, Brett Wood, Alex Pitt, Navinda Kottege, and Dana Kulić. Human-robot team performance compared to full robot autonomy in 16 real-world search and rescue missions: Adaptation of the darpa subterranean challenge. *J. Hum.-Robot Interact.*, 14(1), December 2024.

[86] Lucero Rodriguez Rodriguez, Carlos E. Bustamante Orellana, Erin K. Chiou, Lixiao Huang, Nancy Cooke, and Yun Kang. A review of mathematical models of human trust in automation. *Frontiers in Neuroergonomics*, 4:1171403, June 2023.

[87] Dorsa Sadigh, Anca Dragan, Shankar Sastry, and Sanjit Seshia. Active Preference-Based Learning of Reward Functions. In *Robotics: Science and Systems XIII*. Robotics: Science and Systems Foundation, July 2017.

[88] Lindsay Sanneman and Julie A. Shah. Validating metrics for reward alignment in human-autonomy teaming. *Computers in Human Behavior*, 146:107809, September 2023.

[89] Kristin E. Schaefer. *Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI"*, pages 191–218. Springer US, Boston, MA, 2016.

[90] Kristin E. Schaefer, Jessie Y. C. Chen, James L. Szalma, and P. A. Hancock. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, 58(3):377–400, 2016. PMID: 27005902.

[91] Thomas B. Sheridan. Human–Robot Interaction: Status and Challenges. *Human Factors*, 58(4):525–532, 2016. Publisher: SAGE Publications Inc.

[92] Herbert A Simon. Satisficing. In *The new Palgrave dictionary of economics*, pages 11933–11935. Springer, 2018.

[93] Indramani L. Singh, Robert Molloy, and Raja Parasuraman. Automation- induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, 3(2):111–122, 1993.

[94] Harold Soh, Yaqi Xie, Min Chen, and David Hsu. Multi-task trust transfer for human–robot interaction. *The International Journal of Robotics Research*, 39(2-3):233–249, March 2020.

[95] Ning Wang, David V. Pynadath, and Susan G. Hill. The impact of pomdp-generated explanations on trust and performance in human-robot teams. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, AAMAS '16, page 997–1005, Richland, SC, 2016. International Foundation for Autonomous Agents and Multiagent Systems.

[96] Ning Wang, David V Pynadath, Susan G Hill, Ning Wang, and David V Pynadath. Building Trust in a Human-Robot Team with Automatically Generated Explanations. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, (15315):1–12, 2015.

[97] Anqi Xu and Gregory Dudek. Trust-driven interactive visual navigation for autonomous robots. In *2012 IEEE International Conference on Robotics and Automation*, pages 3922–3929, St Paul, MN, USA, May 2012. IEEE.

[98] Anqi Xu and Gregory Dudek. Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 221–228, 2015.

[99] Anqi Xu and Gregory Dudek. Maintaining efficient collaboration with trust-seeking robots. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3312–3319, 2016.

[100] X. Jessie Yang, Yaohui Guo, and Christoper Schemanske. *From Trust to Trust Dynamics: Combining Empirical and Computational Approaches to Model and Predict Trust Dynamics In Human-Autonomy Interaction*, pages 253–265. Springer International Publishing, Cham, 2023.

[101] X. Jessie Yang, Christopher Schemanske, and Christine Searle. Toward Quantifying Trust Dynamics: How People Adjust Their Trust After Moment-to-Moment Interaction With Automation. *Human Factors*, page 00187208211034716, 2021.

[102] X. Jessie Yang, Christopher Schemanske, and Christine Searle. Toward Quantifying Trust Dynamics: How People Adjust Their Trust After Moment-to-Moment Interaction With Automation. *Human Factors*, 65(5):862–878, 2023.

[103] X. Jessie Yang, Vaibhav V. Unhelkar, Kevin Li, and Julie A. Shah. Evaluating Effects of User Experience and System Transparency on Trust in Automation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, pages 408–416, New York, NY, USA, 2017. ACM.

[104] Luyao Yuan, Xiaofeng Gao, Zilong Zheng, Mark Edmonds, Ying Nian Wu, Federico Rossano, Hongjing Lu, Yixin Zhu, and Song-Chun Zhu. In situ bidirectional human-robot value alignment. *Science Robotics*, 7(68):eabm4183, 2022.

[105] Zahra Zahedi, Mudit Verma, Sarath Sreedharan, and Subbarao Kambhampati. Trust-Aware Planning: Modeling Trust Evolution in Iterated Human-Robot Interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 281–289, Stockholm Sweden, March 2023. ACM.

[106] Michelle Zhao, Reid Simmons, and Henny Admoni. The Role of Adaptation in Collective Human–AI Teaming. *Topics in Cognitive Science*, 17(2):291–323, April 2025.

[107] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.