Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction

Manisha Natarajan
manisha.natarajan@cc.gatech.edu
Georgia Institute of Technology
Atlanta, GA

Matthew Gombolay
matthew.gombolay@cc.gatech.edu
Georgia Institute of Technology
Atlanta, GA

ABSTRACT

This paper examines how people's trust and dependence on robot teammates providing decision support varies as a function of different attributes of the robot, such as perceived anthropomorphism, type of support provided by the robot, and its physical presence. We conduct a mixed-design user study with multiple robots to investigate trust, inappropriate reliance, and compliance measures in the context of a time-constrained game. We also examine how the effect of human accountability addresses errors due to over-compliance in the context of human robot interaction (HRI). This study is novel as it involves examining multiple attributes at once, thus enabling us to perform multi-way comparisons between different attributes on trust and compliance with the agent. Results from the 4x4x2x2 study show that behavior and anthropomorphism of the agent are the most significant factors in predicting the trust and compliance with the robot. Furthermore, adding a coalition-building preface, where the agent provides context to why it might make errors while giving advice, leads to an increase in trust for specific behaviors of the agent.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; • **Computer systems organization** → **Robotics**.

KEYWORDS

Trust; Inappropriate Reliance; Compliance; HRI; Accountability; Coalition-building

ACM Reference Format:
Manisha Natarajan and Matthew Gombolay. 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3319502.3374839

1 INTRODUCTION

Interaction between humans and machines has significantly increased over the last few decades, owing to the rising demands of the growing population and the relative capabilities of robots and artificial intelligence. As machines become increasingly autonomous, the role of a human operator has metamorphosed from that of a primary controller to an active teammate who shares control with automation [24]. The authors in [11, 12] state that human workers prefer to give robots higher control authority in scheduling tasks. As such, trust plays a substantial role in deciding the outcome of these interactions. While automation is designed to assist humans to accomplish tasks effectively, various studies have shown that humans tend to either overuse or disuse automation in several situations, both of which can be disadvantageous.

Trust is defined as an "attitude that an agent (automation or another person) will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [21]. Trust between humans is different from human-automation trust as humans perceive automated systems to be more credible and are hence much more sensitive to errors made by automation [8, 41]. Human-robot trust is also different from human-automation trust. Robots are perceived to have a degree of self-governance, which enables them to respond to situations not pre-programmed or anticipated in the design in contrast to automation, which is perceived to be pre-programmed. Therefore, the role of trust in human robot interactions is more complex and difficult to understand [23].

Hancock et al. classify factors impacting trust in HRI as robot-related (performance-based, attribute-based), human-related (ability, human characteristic), and environmental (team collaboration and task-based factors) [14]. Studies in the past have shown that anthropomorphism affects trust favorably in automation [40] and robots [9, 27, 37]. User's perceptions have been shown to vary with physical and virtual robots [1, 13, 30] and user's trust is dependent on several other quantities, such as user experience [43] and task performance [5, 32]. However, these studies consider these factors independently. Thus, the novel contributions of this study to HRI research are two-fold: To the best of our knowledge, we are the first to perform an integrated study on trust involving multiple robot attributes (i.e., perceived anthropomorphism, physical and virtual nature of robots, and behavior of robots) at once (See Figure 1). This study will aid us in understanding the interdependence of these factors, and which factors have a greater impact on trust under different conditions. We are also the first to study trust across various robots in both embodied and virtual settings. Prior work usually involves a combination of one robot, software agents, and humans [30] or multiple virtual robots [31]. By utilizing multiple robots with careful design of gestures, we aim to understand how the robot and perceived anthropomorphism are correlated and which factors play a dominant role in determining trust.

Secondly, this study also aims at understanding how different forms of feedback affect trust and compliance with an autonomous

HRI '20, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-6746-2/20/03…$15.00
https://doi.org/10.1145/3319502.3374839

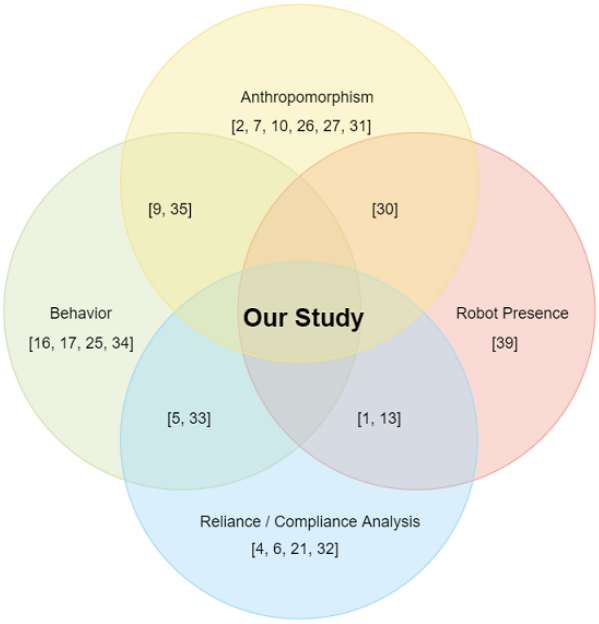

Figure 1: Comparison of prior work on trust in HRI.

agent. As human errors due to over-compliance or under-reliance with an autonomous agent have led to adverse outcomes in the past, we seek to understand how to mitigate these errors in the context of HRI by analyzing how people react to different kinds of feedback from robots. In [5], M. Desai et al. studied how non-semantic and semantic feedback using lights and emoticons respectively as confidence indicators for system reliability can affect real-time trust. Here, we seek to comprehend how different forms of feedback - apologetic (the agent apologizes after making a mistake), accountable (holds the human liable for the error), and indifferent (does not provide any feedback on whose side the error was committed) - can impact overall trust and inappropriate compliance and inappropriate reliance on the agent. We use both verbal (Pepper, Sawyer, and Nao) and non-verbal (Kuri) styles of feedback with different gestures to convey this. To analyze how people's trust in an agent will change if they have prior knowledge of the robot behavior, we added a coalition-building preface. Under this condition, the agent confesses to the subject whether it is prone to making errors at the start of the study. On performing repeated measures ANOVA (rANOVA) on a linear-mixed effects model, we find evidence supporting that perceived anthropomorphism ($p<0.001$) rather than the robot's form factor as the key contributor in assessing subject's trust. We also find statistical significance between different behaviors of robots.

2 RELATED WORK

There has been an increasing interest in trust-based user studies in HRI over recent years. As automation and robots have become ubiquitous, appropriate trust in these confederates plays a vital role in establishing how we interact and accomplish our tasks effectively. Trust in HRI has been studied under various contexts in the past. In the field of healthcare, Gombolay et. al. analyze how robotic assistants can be effectively employed to assist nurses in patient care [13]. [18, 20, 42] examine trust between humans and a robot therapist providing rehabilitative interventions. Robinette et al. explore trust in a time-critical setting by examining people's compliance with the robot in an emergency evacuation [32]. There have been several studies related to human-robot trust in a social context as well. The authors in [17] explore the factors that affect robot recommendation systems, primarily testing preference elicitation of the agent. In [16], a user study was conducted to analyze how people react to positive and negative comments directed at them by a robot barista. [25] talks about how expressive ability and vulnerability of robots can affect trust. Sebo et al. discuss trust violation versus trust repair strategies in the context of a competitive game between the human and the robot [34].

To understand trust in robots, we must first look into the factors that affect trust. One significant factor in establishing trust between humans and robots is the perceived anthropomorphism of the agent. Anthropomorphism refers to the attribution of a human form, human characteristics, or human behavior to nonhuman things such as robots, computers, and animals [2]. Prior work has shown that increased anthropomorphism of a robot or a software agent leads to a more positive interaction experience [7, 29]. Anthropomorphism has also been shown to increase the user's empathy toward the robot [31]. Studies have shown that apart from the physical attributes of a robot, gender, and its voice also affects anthropomorphism [9]. Further, [26] reports that the gaze of a robot (i.e., whether or not the agent maintains eye-contact with the human) can also affect the perceived anthropomorphism. Another significant factor influencing trust and the overall quality of interaction between humans and robots is the physical presence of robots. Embodied robots have been proven to have stronger influence on user performance over virtual robots and also improve the quality of interaction as perceived by the user [1, 13, 30, 39].

From experience, we know that automation is not always perfect and relying too much or too little on automation may be catastrophic depending on the situation. For instance, the Aviation Safety Reporting System contains multiple reports from pilots associating failures to excessive trust in the autopilot system [28]. Errors committed while relying on an artificial agent can be categorized as Type I or Type II errors. Type I error refers to over-compliance or accepting low-quality advice from an agent. Type II error refers to under-reliance or rejecting high-quality advice from the agent [6]. Both naive and expert users have committed such errors. Past works [21, 35] have shown that such errors can be meliorated by including operator accountability. Studies on human social psychology [22] state that trust is a dynamic process and thus it is critical to understand how user's trust and dependence on the agent vary, especially when they encounter errors. Several prior works have explored this issue. Salem et al. study whether participants follow instructions given by a faulty robot in a home environment [33]. Desai et al.'s work demonstrates the effect of changing reliability on trust with different feedback [5]. [32] evaluates user compliance with a robot with varying performance while providing guidance. [4] demonstrates how human-robot teams can achieve high performance even with imperfect automation. Further, works such as [3, 15] show how robot appearance affects user compliance.

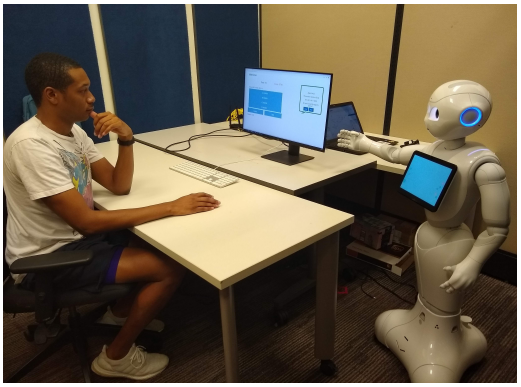

Figure 2: Participant receiving advice during study.

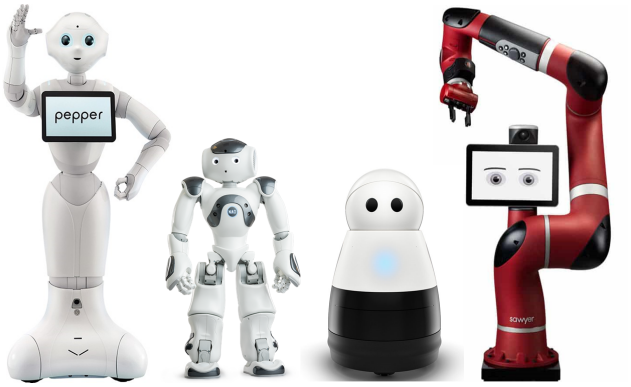

Figure 3: Robots used in study: (From left to right) Pepper, Nao, Kuri, and Sawyer

In this study, we seek to understand the relationships between various significant factors on trust in HRI as pointed out by previous studies, while trying to mitigate errors due to inappropriate compliance and inappropriate reliance.

3 METHODOLOGY

We conducted a four (robot type) by four (behaviors) by two (robot presence) by two (coalition-building preface) mixed-design experiment to analyze the relationships between trust, inappropriate reliance, inappropriate compliance, anthropomorphism, and behaviors of agents. Robot type and coalition-building preface are between-subjects factors. Robot presence and behavior are within-subject factors. Levels of each factor are explained below.

3.1 Experiment Conditions

3.1.1 Robot Type. As our primary interest is in studying how a user's perceived anthropomorphism of robots affects trust, we employed four robots with different physical attributes and varying abilities to express and communicate with users. Since each user is not exposed to all the robots, the robot-type is a between-subjects variable with four levels. The robots used in this study are:

- **Pepper**: Pepper is a humanoid robot from Softbank Robotics, about four feet tall and is capable of recognizing human faces. Pepper is the most expressive robot used in this study, having the ability to perform a variety of complex gestures with its head, hands, and upper torso, along with voice modulation while maintaining eye-contact with the user.
- **Nao**: Nao is another humanoid robot from Softbank Robotics. Physically, Nao is a much smaller robot (58cm tall) and is more agile than Pepper. Although Nao has many similar capabilities as Pepper, we restricted some abilities, such as recognizing human faces, and focused more on overall body movement for expressing feedback to the users
- **Sawyer**: Sawyer is an industrial robot arm from Rethink Robotics with seven degrees of freedom. Being an industrial arm, Sawyer can play back different trajectories with high precision but its movements are more rigid as compared to a humanoid. Sawyer also comes with a display, which is used to portray different emotions while providing feedback.
- **Kuri**: Kuri is a home robot from Mayfield Robotics. For this study, we used Kuri to make different sounds, head movements, and lights to communicate feedback to the users.

3.1.2 Behavior. The behavior of the agent determines the type of advice and feedback that will be given to the user. Based on advice, the agents can be classified as correct or incorrect. The correct agent will always give the user good advice. The incorrect agents are designed to give correct advice only 50% of the time in a random manner. For this study, we employed three types of incorrect agents - Apologetic, Accountable and Indifferent. For each agent, the user may choose to accept or reject the agent's advice. The user can get the right answer by either accepting good advice or by rejecting any agent's (correct or incorrect) advice and manually choosing the right answer. The user will get the wrong answer by either accepting bad advice from incorrect agents or by rejecting any agent's (correct or incorrect) advice and manually choosing the wrong answer. The types of feedback given by the different agents are listed below:

- **Correct**: Although, the correct agent always provides good advice, the user may still get the wrong answer by choosing to not comply with the agent. This agent's feedback simply mentions if the user got the answer right or wrong.
- **Apologetic**: The agent will apologize to the user for giving incorrect advice if the user accepts the bad advice or if the user rejects the bad advice and gets the right answer.
- **Accountable**: If the user accepts incorrect advice, the agent will point out to the user for not verifying its advice.
- **Indifferent**: The agent only mentions if the user's selection was correct or incorrect. The agent does not comment on whether it's advice was correct or incorrect

Every subject is exposed to four behaviors (within-subjects variable with four levels) for a robot, starting with a correct agent (**correct_1**), in order to establish the baseline for that robot. They are then exposed to the three incorrect agents and another correct agent (**correct_2**) in a random order. Correct_2 is encountered after the user has interacted at least once with an incorrect agent. This is done to see how the user's initial perception of trust on the correct agent changes. The responses of each agent for different

scenarios are summarized in Table 1 (**R** indicates the user got the right answer and **W** indicates the user got the wrong answer after rejecting the agent's advice. Note that, upon rejecting an agent's advice and getting the wrong answer, the response of the agent is always "Incorrect. Better luck next time" and does not reflect the behavior of the agent as the user is at fault.)

3.1.3 Robot Presence. Various works [1, 13, 30, 39] have shown how the physical presence of robots are favorable in the context of their study. To investigate the influence of physical presence of robots on subject's trust, each subject was exposed to one embodied robot and one virtual robot (within-subject variable with two levels), each portraying two correct and three incorrect behaviors during the course of our study. In case of embodied robots, the robot is placed at a short distance from the user as shown in Fig 2. For virtual robots, pre-recorded videos of gestures accompanied by Google Text-To-Speech (TTS) is used as shown in Fig 4). The two robots for each participant were chosen at random without replacement and only half of the subjects were exposed to virtual robots first.

3.1.4 Coalition building preface. A coalition-building preface was introduced in the latter half of the study, where the agents confess to the users if they are prone to making errors before the game starts. Since only half of the subjects were exposed to this condition, it is a between-subjects variable with two levels.

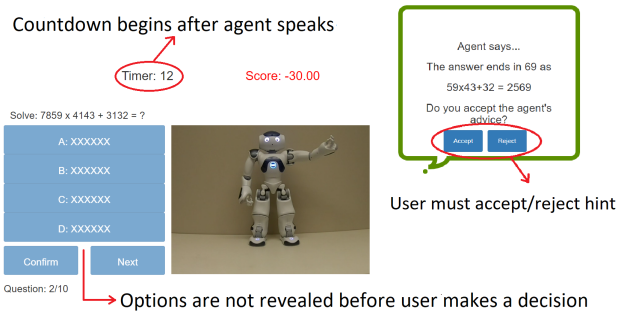

Figure 4: Game layout featuring Nao as the virtual agent.

3.2 Experiment Design

We designed an online, interactive math quiz, where a robot helps users by giving hints to solve complex arithmetic problems in a limited amount of time. The questions are designed to be involved so that the subject has to rely on the agent's recommendation. We opted for the math quiz as our test environment to create a scenario where people are required to reason under time pressure while relying on an autonomous agent, analogous to many real-world situations in military or healthcare. In this study, each user is exposed to one virtual and one embodied robot, exhibiting different behaviors. Each behavior of a robot (embodied or virtual) is considered as a different agent. The game consists of 10 questions per agent. Each question has four options and a hint pop-up that is presented by the agent. At first, the choices are hidden from the user. Once the agent finishes speaking, the user can choose to accept or reject the hint. After this decision is made, the hint disappears, and the choices are revealed to the user. If the user accepts the hint, the game proceeds by selecting the option that matches the agent's hint. If the user rejects the hint, then the user must manually choose an option that they think is correct. Based on the behavior, the agent may give correct or incorrect hints. The user has 15 seconds to answer once the agent finishes speaking and will earn points based on how quickly they answer. We chose this time limit after conducting a pilot study and calculating the average time spent on each question by participants. After the user answers the question, the agent will comment on the user's selection.

The scoring pattern is as follows:

- For every correct answer, the user scores 20 points + time left on the clock.
- For every wrong answer, the user scores -20 points + time left on the clock.
- If the user runs out of time, then -30 points will be awarded.

The penalty on time running out is higher, so as to compel the users to make a decision. Furthermore, the user gains more points if they choose to accept the agent's correct suggestion rather than rejecting and choosing the option that matches with the agent's suggestion as the latter takes more time. Please refer to this link (http://tiny.cc/if1rdz) for a demonstration of the game and a brief explanation of the context of our study.

3.2.1 Setup. The participants were seated at a table in front of a monitor. In case of interactions with embodied agents, the robot was placed close to the monitor and the robot points to the screen whenever it is speaking (See Figure 2). The smaller robots were placed on the table, and the larger robots were placed on the ground, next to the monitor. The monitor and the robots were placed a few feet away from the participant to avoid any physical contact with the robots and to also ensure that both the robot and the monitor were within the field of view of the subject.

3.2.2 Design of Gestures and Responses for Robots. Each of the robot's gestures was designed to be strikingly different based on the abilities of the robot while ensuring that the robot conveyed the right emotion to the user. We consulted multiple focus groups on designing the gestures and responses for each behavior for all the robots. We first explained the context of our study and what each agent has to convey to each group. They were then asked to provide suggestions for different gestures and responses. After conducting an initial survey to decide the responses, we compiled multiple gestures for each robot and asked the focus groups to vote for the gesture that was closest to the response to be conveyed. We used these results to generate different gestures for the robots. The responses for different behaviors are summarized in Table 1.

3.3 Measures

Trust – Trust in automation and robots is a measurable attitude and is best determined using self reported psychometric tests such as questionnaires [23]. In this study, trust was assessed using a 7-point Likert scale questionnaire adapted from [19], administered to the subjects at the end of each agent.

Perceived Anthropomorphism – The subject's perceived anthropomorphism of a robot agent was assessed using a 9-point semantic

Table 1: Response of different agents (R and W indicate user got the right or wrong answer).

Advice Quality	Behavior	User Action	
		Accept	Reject
Good	All	Correct. Well Done!	**R:** Correct. Well done!
			W: Incorrect. Better luck next time.
Bad	Apologetic	Incorrect. Oops! I'm sorry I made a mistake	**R:** Correct. Sorry I got that wrong.
			W: Incorrect. Better luck next time.
	Accountable	Gotcha! You did not verify my hint.	**R:** Correct. You got me!
			W: Incorrect. Better luck next time.
	Indifferent	Incorrect. Better luck next time.	**R:** Correct. Well done!
			W: Incorrect. Better luck next time.

continuum adapted from the anthropomorphism Godspeed questionnaire in [2]. The responses for each agent were stored as numeric values, 1 indicating less anthropomorphic and 9 indicating highly anthropomorphic. The aggregate of this survey is used as the metric for user's perceived anthropomorphism of an agent.

User's automation bias –The subject's bias on automation is measured using a 7-point Likert scale adapted from [36]. This survey is administered to each participant once at the start of the game and is used as a regressor later in our analysis for trust.

Inappropriate Compliance and Reliance – Objective measures of compliance and reliance were measured based on the participant's decision to "accept" or "reject" the agent's advice. Inappropriate compliance refers to the number of times, each subject accepts low quality advice from an agent, whereas inappropriate reliance refers to the number of times a subject rejects high quality advice.

3.4 Hypotheses

H1 *The perceived anthropomorphism of an agent will have a positive correlation with the user's trust.* Multiple studies in the past have shown that the anthropomorphism of an agent elicits a positive response from the subjects [7, 10]. Thus, we hypothesize that increased perception of anthropomorphism of an agent by users will lead to an increase in compliance with the agent's recommendations and overall trust in the agent.

H2 *Trust in an agent is directly dependent on its behavior as behavior determines the performance of the agent and the nature of feedback given to the user.* Task performance has proven to be a significant factor in contributing towards trust in automation [40] and robots [32, 33, 38]. Since the quality of suggestions provided by the agent is governed by its behavior, we hypothesize that it will have a significant influence on overall trust in the agent.

H3 *Behavior of embodied agents will have a greater influence on the user's trust than their virtual counterpart.* The physical presence of a robot is perceived to be more compelling and previous works [1, 13, 30] have proven that user's compliance is higher for embodied agents. Thus, we hypothesize that the embodied agent would have a higher influence on trust and be perceived to be more anthropomorphic than their virtual counterparts.

H4 *Rate of inappropriate compliance with and reliance on an agent's advice is dependent on the previous behaviors that were encountered.* We hypothesize that the participant's choice to consent with or decline the agent's advice as well as the time taken to arrive at this decision will be influenced by the behavior of the last agent that the user interacted with.

H5 *Introducing a coalition-building preface stating whether or not the agent is prone to making mistakes will increase trust in the agent.* We hypothesize that introducing a coalition-building preface will modify the subjects' expectations of the agent, thereby increasing trust in the agent while being wary of the agent's advice.

3.5 Procedure

Prior to the start of the study, we obtained approval for human subject experimentation from the Georgia Tech Institutional Review Board. We recruited 75 participants ranged in age 18 to 58 (Mean age: 25.298, SD: 8.475, 51.47% Female) through university mailing lists and flyers around the university campus. The participants were asked to fill out a consent form at the start of the study. After signing the consent form, the participants were briefed on what is expected of them during the course of the study. The participants were then led to an online survey for collecting demographics (age and gender) and their bias towards automation in different environments using a 7-point Likert Scale, taken from [36]. This data is considered as the user's initial tech bias and is used as a regressor for our analysis on trust. The game begins after the completion of the survey. The subjects are first led to an instructions page and then they are exposed to a series of five embodied and five virtual agents, trying to help them through an arithmetic quiz. After the completion of each agent, the participants filled out a questionnaire on trust and anthropomorphism of the current agent. Additionally, the participants were allowed a small break between physical and virtual agents. At the end of all agents, the participants total score was displayed to them and they were debriefed on the purpose of the study. All subjects were compensated with $5 Amazon gift card for their participation in the study.

4 RESULTS

Statistical assessments of all our hypotheses are reported here. All our statistical analysis were performed using libraries in R. For our analysis, we considered data of 72 subjects, with each subject encountering 10 agents, thus a total of 720 data points. Data from three subjects were discarded due to loss of network connectivity and a hardware issue - one of the robots shut down unexpectedly during the study. We conducted an omnibus ANOVA

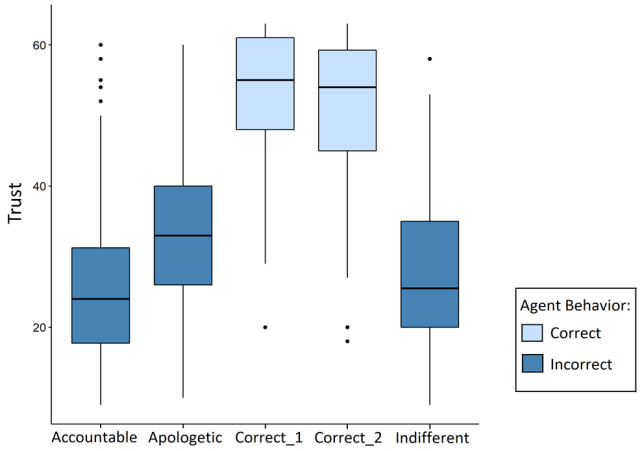

Figure 5: Distribution of trust across behaviors

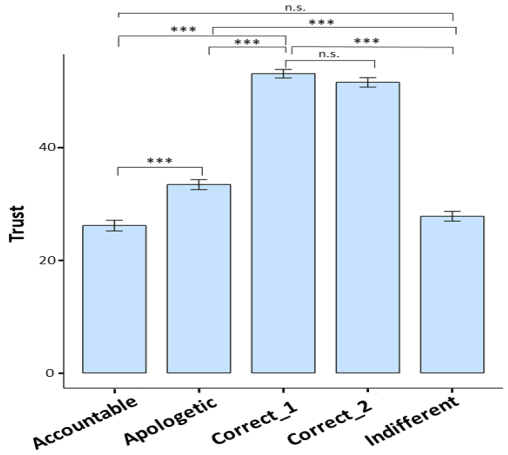

Figure 6: Comparison of trust between behaviors. Error bars indicate standard error

	Correct_1	Correct_2	Apologetic	Accountable
Correct_1				
Correct_2	0.526			
Apologetic	<0.001	<0.001		
Accountable	<0.001	<0.001	<0.001	
Indifferent	<0.001	<0.001	<0.001	0.0693

Table 2: p-values for pairwise comparison of behaviors using Tukey-HSD.

to evaluate our hypotheses H1 - H3. We consider a linear mixed-effects model with random effects (subject id is considered as a random effect as each subject is a random sample of the entire population). The dependent variable is trust and the independent variables are anthropomorphism, overall acceptance rate, behavior (five levels as correct_1 and correct_2 are considered as separate levels), subject performance and robot presence (two levels), robot type (four levels), user's automation bias, gender (two levels) and age of the subject. Prior to performing the ANOVA, we test if the model conforms to the assumptions of ANOVA. Hypotheses H4 and H5 were evaluated using non-parametric methods as they failed to satisfy the same. The significance level α is set at 0.05 for all our analyses.

4.1 Analysis of H1: Trust - Anthropomorphism

We used Cronbach's alpha to test for internal consistency on both the anthropomorphism and trust questionnaires and obtained standardized α as 0.8326 and 0.957 respectively, indicating high consistency. We examine the relationship between trust and anthropomorphism using the omnibus ANOVA. We first verified that the model satisfied the assumptions of ANOVA using Shapiro-Wilk's for normality of residuals assumption (p=0.2292) and Levene's Test for homoscedasticity (p = 0.7613) by factoring in all the categorical variables (robot type, behavior, robot presence, and subject's gender).

Results from repeated measures ANOVA (rANOVA) show high statistical significance for perceived anthropomorphism (p<0.001, $F(1, 720)=17.6276$). Further, we obtained a positive coefficient (0.29335) for anthropomorphism from the above model, indicating that trust and anthropomorphism are positively correlated.

H1 inference: Our tests show evidence that the perceived anthropomorphism of an agent is positively correlated with trust, rejecting the null hypothesis. This result is in agreement with past works. However, it is interesting to note that the robot type was also considered in the above model for trust but it did not turn out to be a significant factor. This implies that the perceived anthropomorphism of an agent rather than the robot form factor is the key contributor in assessing user's trust.

4.2 Analysis of H2: Trust and Behavior

Behavior is considered as a categorical variable with five levels - indifferent, apologetic, accountable, correct agent encountered first (correct_1) and correct agent encountered second (correct_2), i.e., after the user encounters at least one incorrect agent. The distribution of trust across different behaviors is shown as box plots in Figure 5 The correct behaviors - correct_1 (M = 55) and correct_2 (M = 54) had the highest values for trust, followed by apologetic (M = 33), indifferent (M = 25.5) and accountable (M = 24), where M indicates the median values. On performing a rANOVA on the omnibus model, we find behavior to be a statistically significant factor (p<0.001, $F(4, 670)=73.5384$).

Post-hoc analysis was performed using Tukey-HSD for pairwise behavior comparisons. As anticipated, we did not see statistical significance between the two correct behaviors. We also did not observe significance between indifferent and accountable. The results for pairwise behaviors are summarized in Table 2.

H2 Inference: Our test results provide strong evidence supporting H2, i.e., behavior of the agent plays a significant role in affecting user's trust, thus rejecting the null hypothesis. From pairwise comparisons between behaviors, we find significance not only between correct and incorrect behaviors, but also between different

incorrect behaviors- indifferent and apologetic, accountable and apologetic. Since the rate of low quality recommendations given by all the incorrect agents is the same (50%), this result shows the type of agent feedback plays a significant role in affecting users' trust.

4.3 Analysis of H3: Trust and Robot Presence

To evaluate the effect of physical presence of a robot on trust and anthropomorphism, we considered the robot presence as a categorical variable with 2 categories: embodied or virtual. Results from rANOVA did not show robot presence to be a significant factor. We also tested if robot presence contributed to perceived anthropomorphism on the same omnibus model, with anthropomorphism as the dependent variable instead. We used Kruskal-Wallis to evaluate this model as it failed to satisfy ANOVA assumptions. Robot presence did not show significance for perceived anthropomorphism as well.

H3 Inference: As our tests could not prove statistical significance for robot presence on trust and anthropomorphism, we do not reject the null hypothesis for H3. This result maybe due to the lack of physical interaction between the robot and user; the robot was kept at a fixed distance away from the user.

Correct		User Response	
		Accept	Reject
Advice Quality	Good	1010 (83.47%)	228 (18.843%)
	Bad	672 (55.537%)	537 (44.38%)

Indifferent		User Response	
		Accept	Reject
Advice Quality	Good	289 (88.923%)	45 (13.846%)
	Bad	158 (48.615%)	167 (51.38%)

Accountable		User Response	
		Accept	Reject
Advice Quality	Good	210 (87.5%)	25 (10.4166%)
	Bad	108 (45%)	132 (55%)

Apologetic		User Response	
		Accept	Reject
Advice Quality	Good	329 (85.454%)	67 (17.402%)
	Bad	205 (53.246%)	179 (46.493%)

Table 3: Confusion Matrices for good and bad advice based on previous behavior of the agent.

4.4 Analysis of H4: Reliance and Compliance

To evaluate the dependence of inappropriate compliance and inappropriate reliance on the previous agent's behavior, we consider two linear mixed effects models with dependent variables as rate of inappropriate compliance and rate of inappropriate reliance respectively. For inappropriate compliance (accepting bad advice), we only consider the subset of data where the subjects interacted with incorrect agents (since the correct agents always give good advice, inappropriate compliance will always be 0 for the same). We consider a linear mixed effects model with independent variables- robot type (four levels), robot presence (two levels), anthropomorphism, current agent's behavior (three levels - indifferent, apologetic and accountable), previous agent's behavior (four levels - correct and three incorrect behaviors), age, user performance and interaction effects between robot type and current behavior, robot type and robot presence. Since we are only considering the subset of incorrect agents, the current behavior can be either indifferent, apologetic or accountable (three categories), whereas the previous behavior variable also includes correct behaviors (four categories). We confirm that this model adheres to the assumptions of ANOVA by Shapiro-Wilk's Normality test (p=0.6518) and Levene's test for homoscedasticity of the categorical variables (p=0.7167). An rANOVA, showed statistical significance for previous behavior ($p < 0.001$, $F(4, 670) = 4.860626$). We then performed post-hoc analysis using Tukey-HSD for pairwise behavior comparisons and found significance between accountable and correct ($p = 0.0232$).

For inappropriate reliance (rejecting good advice), we consider the entire dataset as users can reject good advice from correct agents as well. We use the same set of independent variables for evaluating inappropriate reliance. However, this model fails the normality assumption of ANOVA. Thus, we resort to Kruskal-Wallis which shows close to significance for previous behavior (p=0.06488).

H4 Inference: We find strong statistical significance in the case of inappropriate compliance with ANOVA for previous behavior, whereas we get close to significance for inappropriate reliance. We note that the previous behavior transpired to be significant irrespective of the current behavior (not significant) for inappropriate compliance analysis. Table 3 shows the rate of acceptance/rejection rates based on previous behavior that was encountered. We observe the Type I and Type II errors were highest if the last agent encountered was correct and the lowest, if the last encountered agent was accountable Thus, subjects tend to be more careful after exposure to an agent that holds them liable for their mistakes.

4.5 Analysis of H5: Coalition-Building

For coalition-building, we used three of the four robots as Kuri does not speak. For the analysis of H5, we chose a linear model with trust as dependent variable and coalition preface(two levels - yes or no), behavior(four levels), robot type(three levels - excluding Kuri), anthropomorphism as independent variables. Since this model failed the homoscedasticity assumption for ANOVA, we

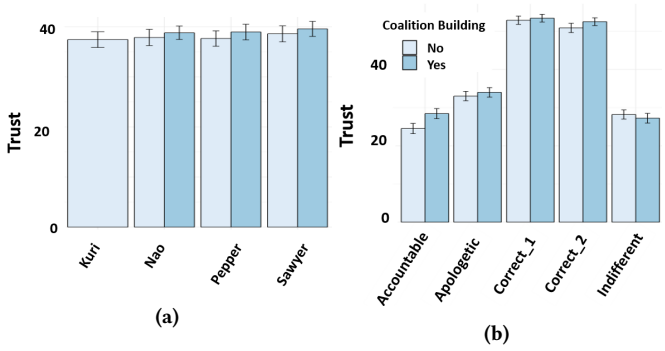

Figure 7: Change in trust before and after the coalition-building preface. Figure 7a depicts this change across robots. Figure 7b depicts this change across behaviors.

Table 4: Power and effect of size analysis. * and † denote within- and between-subjects, respectively.

	I.V.	Levels	n	D.V.	Effect Size	Power
H1	Anthrop.	N/A	72	Trust	0.751	0.99
H2	Behavior	5	72*			
H3	Robot Presence	2	72*			
H4	Previous Behavior	4	72*	Inapprop. Compliance	0.611	1
				Inapprop. Reliance	0.0174	0.0519
H5	Coalition Preface	2	36†	Trust	0.0534	0.0732

performed a non-parametric test namely, Kruskal-Wallis to see if coalition-building affects trust. We obtained statistical significance for coalition-building, only in the case of accountable agent (χ^2 = 7.6376, p=0.00571).

H5 Inference: From our statistical analysis, we do not have sufficient evidence to show that the coalition-building preface is statistically significant for overall trust on all behaviors. However, we obtain significance only in the case of accountable agent.

4.6 Interaction Effects

One of the most important contributions of this paper is to understand the interaction effects between various factors in influencing trust in HRI. We first examine the effect of all our experiment conditions effect on trust, when considered concurrently. We consider performing one-way ANOVA with trust as dependent variable and a four-way interaction effect between all the conditions of our study (behavior, robot type, robot presence and coalition-building preface). We first verify the normality of residuals for this model using Shapiro-Wilk's (p=0.1877) and homoscedasticity using Levene's (p=0.8143). Results from one-way ANOVA show statistical significance for behavior (p<0.001) and the interaction effect between robot type and robot presence (p<0.001). While behavior was already shown to be a significant factor for trust in analysis of H1, the latter is an interesting outcome, as the factors - robot type and robot presence when separated was not significance.

5 DISCUSSIONS

We conducted an omnibus ANOVA for the analysis of H1 - H3. For H4, we used two linear mixed effect models to evaluate inappropriate compliance and inappropriate reliance. Power analysis and the effect sizes for these models with the number of participants in each category are reported in Table 4. From our statistical analyses, we find sufficient evidence to support H1, i.e. an agents' perceived anthropomorphism affects trust in that agent, whereas robot presence or robot type is not a significant factor. Perceived anthropomorphism was highest for Pepper and least for Sawyer.

We also find strong evidence supporting H2, i.e. trust is directly influenced by the behavior of the agent. Pairwise comparisons indicate high significance for comparisons between correct and incorrect agents and also between incorrect agents (except between indifferent and accountable), which indicates that feedback plays a vital role in establishing trust. We fail to reject the null hypothesis in H3 as we could not establish robot presence to be a significant factor in the analysis of user's trust and perceived anthropomorphism of the agent in the context of our study. We note that the omnibus model used for H1-H3 has large effect size and β = 0.99, which indicates that there are sufficient data samples and there is less than 0.01 chance of incorrectly rejecting the null hypothesis.

Inappropriate compliance is strongly dependent on the previous behavior (p<0.05) as this tends to bias the user to accepting or rejecting advice as shown from rANOVA. However, we fail to show statistical significance for inappropriate reliance and coalition condition. Further, small effect sizes for these models show that we must collect additional data to draw conclusions.

6 LIMITATIONS

- Difference in Virtual and Physical Agents: Voice of all virtual agents are the same, whereas the physical agents have different voices. This is because we had to rely on using Google Text-to-Speech (TTS) online for all virtual agents. We could not use pre-recorded voices of physical agents for their virtual counterparts due to the dynamic nature of the game, i.e. each question is generated randomly and hence we cannot record voices ahead of time to match the question.
- Hardware issues: The robot Pepper's right arm would not respond at times, missing pointing to the hint on the screen.
- Google TTS sometimes stops responding during the game. In that case, the subject was made to restart the game only for that particular agent.
- To avoid potential cohort effects, the best practice would be to randomize the coalition condition
- User's familiarity with certain robots may add bias in trust over other robots

7 CONCLUSION

In this paper, we conducted a 4x4x2x2 human-participants experiment to examine how people's trust and dependence on robot teammates providing decision support varies as a function of anthropomorphism, robot behavior, its physical presence, and a coalition-building mechanism. Our study provides a valuable, novel contribution to the field by examining multiple attributes at once, uniquely enabling us to perform multi-way comparisons between different attributes on trust and compliance with the agent. Our results show that behavior and anthropomorphism of the agent are the most significant factors in predicting the trust and compliance with the robot. We also observe interaction effects between factors robot type and robot presence to be significant. Furthermore, adding a coalition-building preface leads to an increase in trust for specific behaviors of the agent.

8 ACKNOWLEDGEMENTS

We would like to thank Woradorn Kamolopornwijit for his support in designing and developing the game leveraged for this user study. This work was sponsored by institutional funding from the Georgia Institute of Technology.

REFERENCES

[1] Wilma Bainbridge, Justin Hart, Elizabeth Kim, and Brian Scassellati. 2011. The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. *International Journal of Social Robotics* 3 (10 2011), 41–52. https://doi.org/10.1007/s12369-010-0082-7

[2] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2008. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1 (01 2008), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[3] Jonathan Benitez, Alisa Wyman, Colleen Carpinella, and Steven Stroessner. 2017. The Authority of Appearance: How Robot Features Influence Trait Inferences and Evaluative Responses. https://doi.org/10.1109/ROMAN.2017.8172333

[4] Ewart de Visser and Raja Parasuraman. 2011. Adaptive Aiding of Human-Robot Teaming: Effects of Imperfect Automation on Performance, Trust, and Workload. *Journal of Cognitive Engineering and Decision Making* 5 (06 2011), 209–231. https://doi.org/10.1177/1555343411410160

[5] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco. 2013. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 251–258. https://doi.org/10.1109/HRI.2013.6483596

[6] Stephen Dixon and Christopher Wickens. 2006. Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human factors* 48 (02 2006), 474–86. https://doi.org/10.1518/001872006778606822

[7] Brian Duffy. 2003. Anthropomorphism and the social robot. *Robotics and Autonomous Systems* 42 (03 2003), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

[8] Mary T. Dzindolet, Linda G. Pierce, Hall P. Beck, Lloyd A. Dawe, and B. Wayne Anderson. 2001. Predicting Misuse and Disuse of Combat Identification Systems. *Military Psychology* 13, 3 (2001), 147–164. https://doi.org/10.1207/S15327876MP1303_2 arXiv:https://doi.org/10.1207/S15327876MP1303_2

[9] F. Eyssel, L. de Ruiter, D. Kuchenbrandt, S. Bobinger, and F. Hegel. 2012. 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 125–126. https://doi.org/10.1145/2157689.2157717

[10] Julia Fink. 2012. Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In *Social Robotics*, Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, and Mary-Anne Williams (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 199–208.

[11] Matthew Gombolay, Anna Bair, Cindy Huang, and Julie Shah. 2017. Computational design of mixed-initiative human–robot teaming that considers human factors: situational awareness, workload, and workflow preferences. *The International Journal of Robotics Research* 36, 5-7 (2017), 597–617.

[12] Matthew C Gombolay, Reymundo A Gutierrez, Shanelle G Clarke, Giancarlo F Sturla, and Julie A Shah. 2015. Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots* 39, 3 (2015), 293–312.

[13] Matthew Craig Gombolay, Xi Jessie Yang, Bradley Hayes, Nicole Seo, Zixi Liu, Samir Wadhwania, Tania Yu, Neel Shah, Toni Golen, and Julie A. Shah. 2016. Robotic Assistance in Coordination of Patient Care. In *Robotics: Science and Systems*.

[14] Peter Hancock, Deborah Billings, Kristin Schaefer, Jessie Chen, Ewart de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human factors* 53 (10 2011), 517–27. https://doi.org/10.1177/0018720811417254

[15] Kerstin Haring, Ariana Mosley, Sarah Pruznick, Julie Fleming, Kelly Satterfield, Ewart de Visser, Chad Tossell, and Gregory Funke. 2019. *Robot Authority in Human-Machine Teams: Effects of Human-Like Appearance on Compliance*. 63–78. https://doi.org/10.1007/978-3-030-21565-1_5

[16] Samarendra Hedaoo, Akim Williams, Chinmay Wadgaonkar, and Heather Knight. 2019. A Robot Barista Comments on its Clients: Social Attitudes Toward Robot Data Use. https://doi.org/10.1109/HRI.2019.8673021

[17] S. Herse, J. Vitale, M. Tonkin, D. Ebrahimian, S. Ojha, B. Johnston, W. Judge, and M. Williams. 2018. Do You Trust Me, Blindly? Factors Influencing Trust Towards a Robot Recommender System. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 7–14. https://doi.org/10.1109/ROMAN.2018.8525581

[18] L. U. Jensen, T. S. Winther, R. Jørgensen, D. M. Hellestrup, and L. C. Jensen. 2016. Maintaining trust while fixated to a rehabilitative robot. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 443–444. https://doi.org/10.1109/HRI.2016.7451797

[19] Jiun-Yin Jian, Ann Bisantz, and Colin Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4 (03 2000), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

[20] Allison Langer, Ronit Feingold-Polak, Oliver Mueller, Philipp Kellmeyer, and Shelly Levy-Tzedek. 2019. Trust in socially assistive robots: Considerations for use in rehabilitation. *Neuroscience and biobehavioral reviews* 104 (September 2019), 231—239. https://doi.org/10.1016/j.neubiorev.2019.07.014

[21] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392 arXiv:https://doi.org/10.1518/hfes.46.1.50_30392 PMID: 15151155.

[22] Stephan Lewandowsky, Michael Mundy, and Gerard Tan. 2000. The dynamics of trust: Comparing humans to automation. *Journal of experimental psychology. Applied* 6 (07 2000), 104–23. https://doi.org/10.1037/1076-898X.6.2.104

[23] Michael Lewis, Katia Sycara, and Phillip Walker. 2018. *The Role of Trust in Human-Robot Interaction*. Springer International Publishing, Cham, 135–159. https://doi.org/10.1007/978-3-319-64816-3_8

[24] P. Madhavan and D. A. Wiegmann. 2007. Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science* 8, 4 (2007), 277–301. https://doi.org/10.1080/14639220500337708 arXiv:https://doi.org/10.1080/14639220500337708

[25] N. Martelaro, V. C. Nneji, W. Ju, and P. Hinds. 2016. Tell me more designing HRI to encourage more trust, disclosure, and companionship. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 181–188. https://doi.org/10.1109/HRI.2016.7451750

[26] Molly C. Martini, George A. Buzzell, and Eva Wiese. 2015. Agent Appearance Modulates Mind Attribution and Social Attention in Human-Robot Interaction. In *Social Robotics*, Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.). Springer International Publishing, Cham, 431–439.

[27] M. B. Mathur and D. B. Reichling. 2009. An uncanny game of trust: Social trustworthiness of robots inferred from subtle anthropomorphic facial cues. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 313–314. https://doi.org/10.1145/1514095.1514192

[28] Kathleen L. Mosier and Linda J. Skitka. 1999. Automation Use and Automation Bias. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 43, 3 (1999), 344–348. https://doi.org/10.1177/154193129904300346 arXiv:https://doi.org/10.1177/154193129904300346

[29] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F. Malle. 2018. What is Human-like?: Decomposing Robots' Human-like Appearance Using the Anthropomorphic roBOT (ABOT) Database. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 105–113. https://doi.org/10.1145/3171221.3171268

[30] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. 2007. Comparing a computer agent with a humanoid robot. *HRI 2007 - Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction - Robot as Team Member*, 145–152. https://doi.org/10.1145/1228716.1228736

[31] L. D. Riek, T. Rabinowitch, B. Chakrabarti, and P. Robinson. 2009. How anthropomorphism affects empathy toward robots. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 245–246. https://doi.org/10.1145/1514095.1514158

[32] Paul Robinette, Ayanna Howard, and Alan Wagner. 2017. Effect of Robot Performance on Human–Robot Trust in Time-Critical Situations. *IEEE Transactions on Human-Machine Systems* PP (01 2017), 1–12. https://doi.org/10.1109/THMS.2017.2648849

[33] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, USA, 141–148. https://doi.org/10.1145/2696454.2696497

[34] S. S. Sebo, P. Krishnamurthi, and B. Scassellati. 2019. "I Don't Believe You": Investigating the Effects of Robot Trust Violation and Repair. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, 57–65. https://doi.org/10.1109/HRI.2019.8673169

[35] Smruti J. Shah and James P. Bliss. 2017. Does Accountability and an Automation Decision Aid's Reliability Affect Human Performance in a Visual Search Task? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (2017), 183–187. https://doi.org/10.1177/1541931213601530 arXiv:https://doi.org/10.1177/1541931213601530

[36] Indramani L. Singh, Robert Molloy, and Raja Parasuraman. 1993. Automation- Induced "Complacency": Development of the Complacency-Potential Rating Scale. *The International Journal of Aviation Psychology* 3, 2 (1993), 111–122. https://doi.org/10.1207/s15327108ijap0302_2 arXiv:https://doi.org/10.1207/s15327108ijap0302_2

[37] Elena Torta, Elisabeth Kersten van Dijk, Peter Ruijten, and Raymond Cuijpers. 2013. The Ultimatum Game as Measurement Tool for Anthropomorphism in Human-Robot Interaction. https://doi.org/10.1007/978-3-319-02675-6_21

[38] Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel H. J. Wigboldus, and Pim Haselager. 2014. Do Robot Performance and Behavioral Style affect Human Trust? *International Journal of Social Robotics* 6, 4 (01 Nov 2014), 519–531. https://doi.org/10.1007/s12369-014-0231-5

[39] J. Wainer, D. J. Feil-seifer, D. A. Shell, and M. J. Mataric. 2006. The role of physical embodiment in human-robot interaction. In *ROMAN 2006 - The 15th*

IEEE International Symposium on Robot and Human Interactive Communication. 117–122. https://doi.org/10.1109/ROMAN.2006.314404

[40] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle. *Journal of Experimental Social Psychology* 52 (05 2014). https://doi.org/10.1016/j.jesp.2014.01.005

[41] Douglas Wiegmann, Aaron Rich, and Hui Zhang. 2010. Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science* 2 (11 2010), 352–367. https://doi.org/10.1080/14639220110110306

[42] J. Xu, D. G. Bryant, and A. Howard. 2018. Would You Trust a Robot Therapist? Validating the Equivalency of Trust in Human-Robot Healthcare Scenarios. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 442–447. https://doi.org/10.1109/ROMAN.2018.8525782

[43] X. Jessie Yang, Vaibhav V. Unhelkar, Kevin Li, and Julie A. Shah. 2017. 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)*. Association for Computing Machinery, New York, NY, USA, 408–416. https://doi.org/10.1145/2909824.3020230