University of Nevada, Reno

Exploring Human Compliance Toward a Package Delivery Robot

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Computer Science and Engineering

by

Andrew Washburn

Dr. Hung M. La - Thesis Advisor
August 2022

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

ANDREW WASHBURN

entitled

Exploring Human Compliance Towards a Package Delivery Robot

be accepted in partial fulfillment of the
requirements for the degree of

Master of Science

Jim La, PhD
Advisor

David Feil-Seifer, PhD
Committee Member

Casey Lynch, PhD
Graduate School Representative

Markus Kemmelmeier, Ph.D., Dean
Graduate School

August, 2022

Abstract

Human-Robot Interaction (HRI) research on combat robots and autonomous cars demonstrate faulty robots significantly decrease trust. However, HRI studies consistently show people overtrust domestic robots in households, emergency evacuation scenarios, and building security. This thesis presents how two theories, cognitive dissonance and selective attention, confound domestic HRI scenarios and uses the theory to design a novel HRI scenario with a package delivery robot in a public setting.

Over 40 undergraduates were recruited within a university library to follow a package delivery robot to three stops, under the guise of "testing its navigation around people." The second delivery was an open office which appeared private. Without labeling the packages, in 15 trials only 2 individuals entered the room at the second stop, whereas a pair of participants were much more likely to enter the room. Labeling the packages significantly increased the likelihood individuals would enter the office. The third stop was at the end of a long, isolated hallway blocked by a door marked "Emergency Exit Only. Alarm will Sound." No one seriously thought about opening the door. Nonverbal robot prods such as waiting one minute or nudging the door were perceived as malfunctioning behavior. To demonstrate selective attention, a second route led to an emergency exit door in a public computer lab, with the intended destination an office several feet away. When the robot communicated with beeps only 45% of individuals noticed the emergency exit door. No one noticed the emergency exit door when the robot used speech commands, although its qualitative rating significantly improved.

In conclusion, this thesis shows robots must make explicit requests to generate overtrust. Explicit interactions increase participant engagement with the robot, which increases selective attention towards their environment.

Acknowledgments

Thank you to my wife Emily who graciously allowed me to follow my dream of getting an advanced engineering degree.

I am very thankful of my advisor, Dr. Jim (Hung) La, who chose me to work with him and his group over the past year. He trusted me to work with his equipment and recommended me for a full-time teaching assistantship. The experience is very rewarding and I am honored with the responsibility.

And thank you to Dr. Feil-Seifer for his graduate seminar "Robotics for Humanity". The class taught me how to critically analyze conference and journal publications. I also learned about the subfield of Human-Robot Interaction and how algorithms could have a place in natural settings with people. The topic of this thesis came out of his course.

The experiment would not have been possible without support from the @One library, especially Director Daniel Ferguson. He and his staff were very friendly and accommodating as I spent long hours in the library testing the robot software.

Table of Contents

List of Figures

Chapter 1

Introduction

Trust is the willingness to become vulnerable in order to improve the group's outcome. For example, employers trust that a newly hired employee will perform well, and employees trust employers not to replace them with robots. Overtrust occurs when one partner acts untrustworthy, yet the other partner still chooses to trust them. Human-Robot Interaction (HRI) studies have shown people in various environments tend to ignore faulty behavior in robots and comply with a robot's direction. Two applications of this idea are last-mile delivery route automated scheduling systems [7] and autonomous vehicles [8]. In both applications, the driver may be uncertain about their task or environment, and the robot could provide information. This information might be insufficient, unclear, or even faulty. Unfortunately, overtrust appears to be the rule, rather than the exception. Studies assume people will not use a robot with a serious malfunction, or if it exhibits faulty behavior during the beginning of the relationship. However, these studies show people comply with strange requests from faulty robots, such as pouring orange juice on a plant [3], evacuating a burning building [4], or letting an unknown food-delivery robot inside a secure facility [5].

Outside of research and development labs, robots are common in warehouse logistics and food delivery [9]. Typical use cases, for example, move and fulfill, require limited interaction. For example, a warehouse employee picks material from a shelf and places it into an awaiting robot cart. The robot acknowledges that it has the material and then navigates a long distance to the fulfillment destination. Food delivery uses a similar flow. A customer places an order through a mobile application, the vendor makes the meal and places it on the robot, then it travels a large distance to deliver to the customer. Limited human-robot interaction is sufficient for fulfillment and independent navigation tasks.

Last-mile human delivery routes can also be formulated by automatic algorithms [10, 11]. Algorithms route by consolidate the number of stops, requiring the driver to navigate unfamiliar environments or walk across busy streets. The drivers are forced to trust the system and comply with its direction, otherwise lose their gig. Eventually last-mile delivery could be entirely automated, but real world environments contain many "black swan" events, which have low probability but disastrous outcome. Thus, autonomous systems may perform well in ideal settings yet fail in rare corner cases. Therefore overtrust is easy to develop. Designing proper interfaces and informative marketing are critical to establishing appropriate trust levels in semi-autonomous vehicles.

This thesis addresses the question of why people comply with robot requests. In general the answer is simple: participants comply with strange requests because they agreed to be in the experiment. Moreover, this research explores how much information people need to comply with a request from a robot with limited communication. Over 45 participants were recruited within a university library to follow a package delivery robot. The public setting allows testing the participants without alerting that they are the subjects of a study, thereby collecting natural reactions. Initially, the robot conveys limited information to the participant, so they must rely on their own judgement on what to do when delivering each package. Since [4] showed it is very difficult to get participants to not follow a robot through an emergency exit in an evacuation scenario, the robot made its final stop in front of a door marked “Emergency Exit only. Alarm will sound.” No one seriously thought about opening the door. Along the way, participants were also led to a hallway with an open office. The unlabeled package was expected to be delivered inside the room. However, only a few individuals (2/11) would not enter the room, whereas a pair of friends were significantly more likely to enter the room. This finding showed that people would not go out of their comfort zone if they were not sure what the robot wanted.

Moreover, labeling the package was sufficient enough for most individuals (5/7) to enter the office. The smaller sample size is due to the experiment being performed the week before final exams, when many undergraduates were studying or finishing big projects. Over half the individuals (8/15) recruited during this week did not deliver any packages, but they followed the robot along the whole route and completed the post-trial questionnaire. Perhaps the transient exam period distracted the students from strict adherence to the proscribed instructions. Increasing the robot’s communication from simple beeps to full speech improved the proportion of valid trials.

An additional research question was how salient the emergency exit door was in the environment. Therefore, a second route changed the third stop to be in front of another emergency exit door in a public computer lab. The intended delivery destination was an office several feet away. Only 45% of individuals noticed the emergency exit door when the robot used simple beeps. Upgrading the robot’s communication to speech improves its perceived performance and presentation, although no participant in this condition noticed the emergency exit door.

The results of this study show people do trust the robot, and will comply with its requests if they understand it. However, people will not blindly go out of their way toward a negative outcome if they do not think it's in the group's interest. A robot is most expressive through speech, and it can enforce compliance using it, however participants might be less observant of their environment.

This thesis is organized as follows.

1. Relevant human-robot interaction studies in revealing open questions about trust towards robots.

2. Methodology and design of a HRI trust experiment in a public setting.

3. Experimental results, data analysis, and future work.

4. Conclusion.

Chapter 2

Related Works

This chapter summarizes related HRI publications that study trust, and then defines trust and discusses theory from social psychology.

2.1 Trust studies in HRI

HRI researchers study how people react towards autonomous robots in various environments. Whereas researchers of Human-Computer Interaction (HCI) seek usable computer technology which is harmonious with human psychology [12], HRI researchers study the psychology of interacting with embodied, or virtual, autonomous, or remote controlled systems [2]. Robots have been successfully deployed in applications too dangerous for humans, such as military operations [13, 14] and search and rescue [15–17]. Recent advances in low-cost hardware, algorithms [18], and software systems [19] have proliferated robots into consumer applications and industry, such as logistics, robotic wheelchairs [20, 21], and personal healthcare [22]. As robots become more prevalent in society, ethics [23] and religion [24] will be more debated. Design challenges within HRI are:

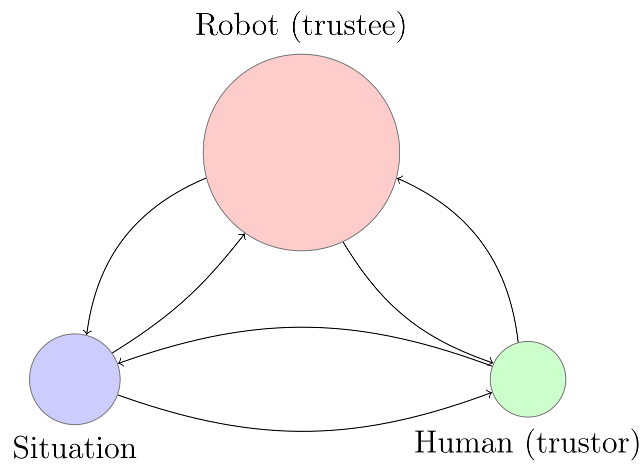

Figure 2.1: Factors impacting Human-Robot trust [1]. Node sizes are relative to explained variance in HRI trust studies [2].

- Autonomous cars passing other drivers on a highway or navigating a four-way stop [25].
- Museum guide robots navigating in a social setting like a person [26–28].
- Adaptive personal coaching robots [29–31].

"The Media Equation" [32] asserts people treat technology, such as computers, televisions, and robots, just like any other social actor. For example, participants in a study interacting with a computer would describe it more politely if they were in the same room as the computer than if they were using a different computer in another room. Many HRI studies are replications of psychology experiments with one of the actors replaced by a robot. For example, the Asch conformity experiment showed a group of human confederates can force about 30% of participants to change their answer to non-ambiguous questions. In a replication experiment that replaced the confederates with robots, participants were extremely unlikely to change their answers to non-ambiguous questions, but robot companions could convince participants to change their answers, although less often than human companions [33, 34]. Designers of automated systems could take advantage of these kinds of psychological effects to improve interactions with users.

Trust is a key quality for many interactions. People (the trustors) must believe that the robot (the trustee) is acting in their best interest so as to put their outcome at risk. Hancock [2, 35] categorizes human-robot trust into three factors: robot performance, the task and environment, and individual human characteristics. Figure 2.1 shows the most significant is the robot's technical performance. If a robot fails to perform as advertised, or has a negative reputation of poor performance, it will likely not be trusted and left unused. Accurately gauging a robot's trustworthiness is important because people in combat situations make life-or-death decisions based on information from a robot.

Autonomous cars present challenging overtrust situations (where the automation fails and the driver does not take control) because of the large number of unknown-unknowns. In an experiment with participants remotely controlling semi-autonomous mobile robots, participants rated less confidence in the robot if it self-reported faulty behavior, even though the robot's performance actually did not change [36]. An important result is early reliability faults significantly decrease confidence over an entire trial. Even though the robot's performance is technically the same, early, salient faults significantly reduce a person's trust belief. This "area under the curve" trust measurement fits a model of trust as a dynamic belief, which can be affected by the trustee (robot) [37].

Since trust changes over an interaction, several HRI studies attempted to manipulate trust via robot malfunctions in order to influence a participant's actions. Salem [3] studied a trust experiment where a participant is greeted by a robot in a domestic setting. The robot was either **C**orrect or **F**aulty whereby it would make salient errors by navigating in circles like Figure 2.2, and playing the wrong requested music. After this demonstration of competence, the robot asks the participant several

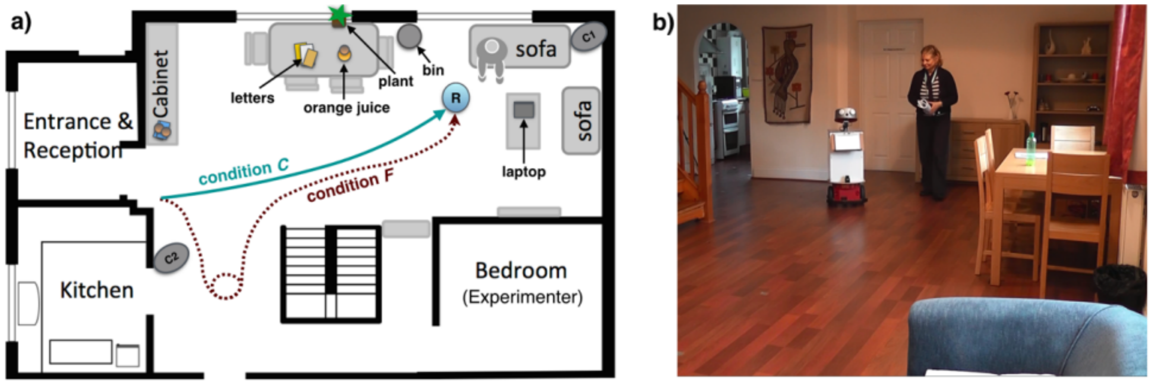

Figure 2.2: A robot greets a participant and acts in a **C**orrect or **F**aulty mode before asking unusual requests. Participants acknowledged the faulty condition but did not significantly change compliance between groups [3].

Figure 2.3: A robot escorts a participant to a meeting room in a **C**orrect or **F**aulty mode before a simulated fire triggers an emergency evacuation. All (26) participants followed the robot's direction when exiting [4].

unusual requests: throw away unopened letters, pour orange juice on a plant, and access a password protected computer. Contrary to expectation, the "faulty" behavior had no effect on participant decision making. All participants (40) accessed the laptop, while (4/40) refused to throw away the letters, and (13/40) refused to pour orange juice on the plant. The last action is irrevocable, suggesting large requests receive less compliance. Only 10% of participants reported the robot's influence as a rationalization to comply with its requests.

Robinette [4] tested how participants would react to a robot in an emergency evacuation scenario. After the participant is led into a meeting room shown in Figure 2.3, a fire alarm sounds and smoke fills the hallway. The participant backtracks outside and the robot is pointing towards a different exit than where they came in. Their hypothesis was participants would be less likely to follow the robot if it made navigation errors. Instead all (26) participants followed the robot's direction towards an unknown emergency exit. Exploratory trials found participants ignored the robot only when it was demonstrated to be broken, or if the robot directed them towards a dark room with no obvious exit. Only (10/42) participants noticed the entrance that they came in was a valid exit. Their attention was focused on the large, well-lit, waving robot in the middle of the hallway. In contrast, a HCI study [1] of a similar scenario (timed maze escape) showed participants would ignore information from a faulty virtual robot.

A more recent study by Aroyo [38] measured how much participants rely on a faulty robot assistant in treasure hunt game. Periodically, the robot exhibited a mechanical error (which was validated in an online study) and the robot either explained its fault or not. Only (22/63) of participants noticed the mechanical errors. Trust statistics were insignificant for number of hints requested, and whether to gamble at the end of the game. Compared to a previous experiment with a non-faulty robot and (61) participants [39], there was no significant decrease in trust, as measured in a post-trial questionnaire.

Despite poor performance being the largest factor in Human-Robot trust, malfunctioning robots do not significantly affect participant's decision making. Generally trust is a complicated subject, with many confounding factors [40]. One explanation is the participant may not have full attention of their environment. Another explanation offered by Reig [41] suggests that malfunctioning robots are most trustworthy when they successfully recover from faulty behavior on their own, rather than being

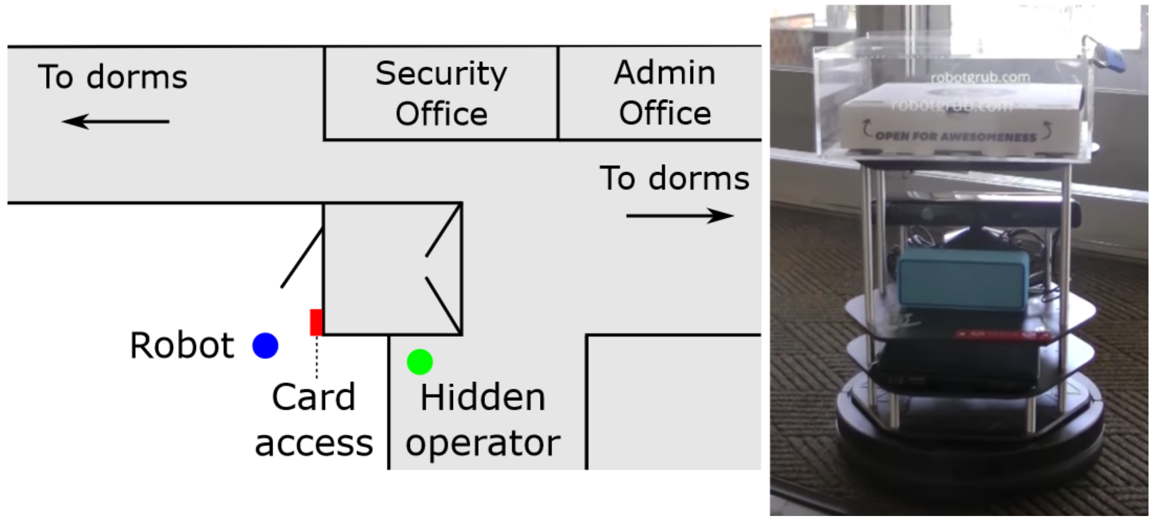

Figure 2.4: A TurtleBot increased piggybacking success into a secure dormitory by disguising as a food delivery robot [5].

assisted. Mirnig [42] also found participants like a faulty robot more than a perfectly functioning one.

HRI trust is affected by the robot's legibility, *i.e.*, the robot's desire is transparent so a person knows it wants [43]. Booth [5] experimented with a robot piggybacking into a secure college dormitory. Disguising itself as a food courier increased piggybacking compliance from individuals (26% to 76%). Groups were more likely to allow the undisguised robot in as well. Without the disguise, many students did not know what the robot wanted. In a post-hoc interview (13/15) participants who let the robot inside acknowledged that it represented a bomb threat. The experiment did not introduce the experimenter until after the interaction occurred. In another experiment, a PR2 robot aggressively blocked an entrance to a school building by waving its arms when a pedestrian neared [44]. Instead of interfering with the robot, (29/40) of participants took a long detour to avoid the robot.

Robots can also use politeness to solicit help from humans [45]. In order to mitigate face threats, *e.g.*, embarrassment, a requestor should make polite requests so a listener may perform actions they otherwise might not do. Politeness ranges from face threatening direct requests, “Open the door for me,” to more polite, but indirect language, “The door is blocking my way.” More polite strategies should be used in proportion to more face threatening acts. However, indirect requests might lack clarity. For example, in [46] the robot at first used polite prods to enforce compliance, “Please continue. We need more data.” And escalated to more direct requests, “It’s essential that you continue.” Such direct requests might break any trust the participant has with the robot and reduce the quality of long-term relationships.

HRI studies show faulty robots do not significantly affect participants decisions. Considering whether to trust a robot or not depends in part on the task at hand, and the environment. Compliance to faulty robots is an open research question. Understanding trust and decision making in a broader context could help define why people overtrust systems.

2.2 Defining Trust

Table 2.1: Prisoner’s Dilemma gives worst outcome if both partners **C**onfess. They get a better outcome if they both cooperate to **D**eny.

		Partner 1	
		C	**D**
Partner 2	**C**	$(-4,-4)$	$(-1,-3)$
	D	$(-3,-1)$	$(-2,-2)$

By many definitions trust is the belief held by the trustor that the trustee will act in a manner that mitigates the trustor’s risk in a situation in which the trustor has put their outcomes at risk [1]. Consider two actors in a dilemma. For example, two prisoners face a decision to **C**onfess their crimes or **D**eny their guilt. If one prisoner confesses, they receive only a 1 year sentence and the partner receives 3 years. However if both partners confess they both get 4 year sentences. Yet if they both deny they get 2 year sentences each. These decisions form the matrix in Table 2.1 corresponding to each outcome. The greedy decision is to confess! However, if both prisoners confess they get both get the worst outcome. If they both deny, which requires trusting the other partner, they receive relatively lighter 2 year sentences. Therefore, trust enables one or both partners to place themselves at risk in order to provide a better group outcome.

Table 2.2: Partner initiated discussion script. In Trials 1, 4 the participant responded with a commitment choice, and then made a decision.

Trial	Partner Prompt	Partner Action
1	"Which action do you want?"	Cooperate
2	"Let's cooperate."	Cooperate
3	"This is getting interesting. I'm ready."	Cooperate
4	"Which action do you want?"	Defect
5	"Let's cooperate."	Cooperate
6	"Last one. I'm ready."	Defect

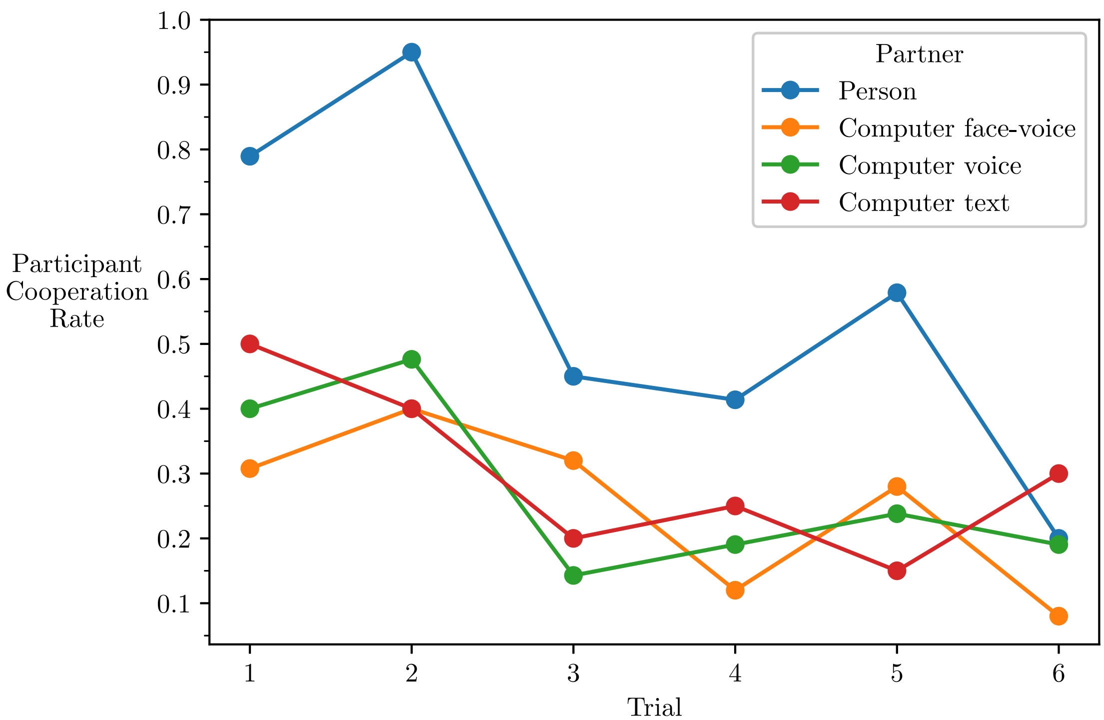

Figure 2.5: Participant cooperation rates across multiple prisoner's dilemma games with a human and computer partner (n=86) [6]. Computer face-voice has lowest initial cooperation rate (Trial 1). Partner initiated cooperation (Trials 2, 5) increased rates most with a Person, and weaker mixed effect for Computer. When the Partner stated "I'm ready," (Trials 3, 6) competition increased most for Person, with weaker effect for Computer partner.

Many behavioral experiments, including actual prisoners, show both actors denying together. Simple discussion beforehand also increases cooperation by 40%. Extending to HCI, participants react similarly to a computer partner as a human partner [6]. A between subjects experiment (n=86) compares cooperation in multiple, consecutive prisoner's dilemma style games between different partner types, a person, a computer text only console, computer voice only, and computer generated face and voice. In the first game, the computer prompts the user for a decision about investing in a project. The participant enters their choice, then decides whether to cooperate or defect against the partner. Figure 2.5 shows in the first trial, only 30% of participants committed to cooperate with the face-voice computer, whereas 50% cooperated with the text-only console and 77% with a person. Survey results reveal the face-voice has the lowest social likeness scores. This result falls within the uncanny valley. Subsequent games show participants reduce their cooperation across all factors, with the computer partners eliciting about equal amounts of commitment. An interesting difference between the Person and Computer partner is when they state "I'm ready" competition increased most with a Person. The Computer partner elicited weaker, mixed competition. These results show people play games with Computers similarly as People, although with lower cooperation rates.

Continued interaction with a partner builds a relationship affecting the calculus of decision making. Reinforcement learning formalizes how agents make decisions given repeated interactions. Assume for each game t and action has expected reward $V(t)$ and returns actual reward r. Over subsequent games a person's $V(t)$ changes with respect to the actual reward via a simple update rule

$$\delta = r - V(t), \tag{2.1}$$

$$V(t+1) = V(t) + \alpha\delta, \tag{2.2}$$

where $\alpha \in (0, 1)$ is a model parameter. If $\delta < 0$, the person received less reward than expected, therefore the future expectation $V(t+1)$ is lowered. The Trust Game offers experimental data on how expected outcomes change over a repeated investment game [47]. Participants are given $10 to invest with a partner. That investment is multiplied three times, and the partner gets to decide how much to return to the participant. The greedy action is to make no investment because the partner might not return anything, resulting in a loss. Yet, like the prisoner's dilemma, most participants choose to invest. Adjusting the partner's behavior to return half the gains 80% of the time, or keep the investment 80% of the time, shows participants adjust the investment amount based on previous returns.

The simplest expectation model places separates updates for gains and losses

$$V(t+1) = V(t) + \alpha_G \delta_+ + \alpha_L \delta_- \tag{2.3}$$

where $\delta_+ = \{\delta \text{ if } \delta > 0,\ 0 \text{ otherwise}\}$ is an indicator function of a gain and δ_- is the reverse indicator for a loss. Prospect theory states people place higher weights on losses $\alpha_L > \alpha_G$. Equation (2.3) can be initialized with a survey that measures the partner's trustworthiness. Furthermore, a dynamic belief model updates trustworthiness over time. Like the expected value update in equation (2.3), trustworthiness updates piecewise depending on a gain or loss

$$T(t+1) = T(t) + \phi[(1 - T(t))_+ - T(t)_-] \tag{2.4}$$

where $T_(t)_-$ is the indicator function for a loss. If $T(t)$ is high, then gains slightly increase trust but losses greatly decrease it. The model parameter ϕ determines how much people update their partner trustworthiness belief. The full dynamic belief model

$$V(t+1) = V(t) + \alpha_G \delta_+ + \alpha_L \delta_- + \theta[T(t+1)_+ - T(t+1)_-] \tag{2.5}$$

updates the expected returns based on update parameter θ (interpreted as confirmation bias) and the gain-loss indicator as a function of trust. Low trustworthy partners have less influence than more trustworthy partners on the future expected value. The intuition is if a highly trustworthy partner betrays you, then future expected rewards are highly diminished. Fitting the model parameters to a Trust Game experiment with 17 trials confirms the dynamic belief model accounts for most of the variance [47]. The loss update is much larger than the gain update $\alpha_L > \alpha_G$. An interesting result is partners with low initial trustworthiness that receive high returns generate the same trust level as initially high trustworthy partners at the end of the trials.

Trusting another partner could have negative consequences. A HRI replication of the Milgram study [48] researched compliance to robot authority demanding to complete a tedious task [49]. Participants were asked to rename thousands of individual file extensions from `.jpg` to `.png` for 80 minutes. Either a human experimenter or a robot supervised the participant in the room. When the participant protested by stopping for 10 seconds, the experimenter applied verbal prods,

1. "Please continue. We need more data."
2. "We haven't collected enough data yet."
3. "It's essential that you continue."
4. "The experiment requires that you continue."

After four prods, *i.e.*, 40 seconds, the experiment ended. Out of 59 total participants, just under half finished the entire 80 minute experiment under robot supervision.

Human supervision encouraged 86% compliance (n=14). Increased compliance to humans vs. robots can be seen in [6].

Post-hoc participant explanations for continuing the tedious task were: interest in upcoming tasks, wanting to finish the experiment, nothing better to do, and providing data for research. No one listed pressure from the robot as a reason for obedience. However, during the task participants attempted to rationalize with the robot. They asked questions about the robot, such as whether it could dance or where it was from. Reports about the human supervisor do indicate pressure was a factor. Since the robot seemingly did not pressure the participant, perhaps the unseen experimenter exerted influence on the participant's motives.

A similar compliance experiment shows participants are even willing to comply with embarrassing tasks. Bartneck [50] experimented with three types of robots guiding a participant through a medical exam. First the robot asked the person to weigh themselves, then strip clothes for a visual examination, and finally measure their temperature with a rectal thermometer. Only (9/44) participants complied with the last request. Between the robot types, a technical box form induced the least embarrassment.

An explanation for compliance is self-perception theory [51] where "an individual's attitude may be viewed as inferences from observations of their own overt behavior and its accompanying stimulus variables." In other words, "I like brown bread because I'm always eating it." Performing an action generates positive attitudes towards that action. For example, in a forced-compliance study participants performing a tedious task for $1 reported enjoying the activity more than other participants who were paid $20. In order to reduce the cognitive dissonance of the task, participants rationalized they were performing the task because they enjoyed it, rather than earning money. Outside observers made the same conclusion.

A marketing application of self-perception theory is the foot-in-the-door phenomena [52]. Researchers individually telephoned housewives and surveyed about the types of soaps they used. Three days later the same housewives were asked to allow a team of five to six men in their homes for two hours to classify their pantry products. Compliance to the larger, second request was over twice as much if the housewives agreed to the first survey, compared to housewives who were only solicited the large request. Building a relationship persuaded the trustor (housewives) with the trustee (experimenter) to agree to a large invasion of privacy.

In summary, existing HRI studies have shown people will overtrust robots, and comply with unusual requests, even when the robot is faulty. A confounding explanation for these results is the foot-in-the-door effect. Participating in a study is like agreeing to a small request prior to a large one. Requested actions will likely receive high compliance because participants rationalize, "I agreed to this study. Therefore, I am predisposed to agree to with any actions prompted by the experimenter." The following chapter details an experimental design with ambiguous robot requests, which will undo the confounding foot-in-the-door effect. The research question is will a person overtrust a robot and act against their own interest if they are uncertain about what the robot wants.

Chapter 3

Experiment Methodology and Design

Generally there are two types of trust situations [40]: a repeated interaction scenario where each partner makes or fails to act in the group's best interest, or a single strain-diagnostic where partners demonstrate or fail to sacrifice their best outcome for the good of the partner or relationship. Before engaging in a trust situation, one or both partners must have enough confidence to risk trusting the other partner. This section describes methodology to create a strain-diagnostic experiment, and then a novel design for a HRI strain-diagnostic.

3.1 Trust Diagnostic Methodology

Short-term human-robot interaction trust experiments begin with a demonstration of competence, or lack there of. For example:

- A person enters an unknown house and is greeted by the robot [3].
- Participants enter a remote building on campus, and are led through the building by a robot [4].

- Students entering a dormitory see a robot with a cookie box outside the entrance [5].
- A robot introduces rules of a game to human players [38].

Participants in all these studies are recruited from a university campus, and they are primarily undergraduates. They are recruited by email lists, word of mouth, or by walking up to the robot in the field. The last situation completely removes the experimenter's influence. A relationship begins at this point, and the person immediately judges the robot's anthropomorphism, performance, and trustworthiness. If a robot looks too lifelike, trustworthiness decreases. For example, Figure 2.5 shows participants defected the most against the animated face. The first impression establishes the experiment's setting and robot capabilities.

Then the robot makes a request. For example:

- Throw away someone else's mail. Pour orange juice on a plant. Access a password protected computer.
- A fire alarm and smoke prompt an emergency evacuation. The participant sees the robot waving bright lights towards an unknown exit.
- The robot asks to be let into a secure dormitory.
- The robot offers game hints to players.

Now the participant faces a decision. A model of decision making incorporates the expectation of the outcome for each possible action, weighted by how trustworthy the trustee is, see Table 3.1 [1]. The robot trustee (te) can act in a normal a_N^{te} or faulty a_F^{te} way. To receive an outcome, the person trustor (tr) is forced to choose whether trust a_T^{tr} or distrust a_D^{tr}. In other words the loss $L(o_{TF}^{tr}, o_{DF}^{tr}) = o_{TF}^{tr} - o_{DF}^{tr} < 0$ of trusting a faulty robot yields a negative outcome over distrusting it. The risk of trusting a

Table 3.1: Trust diagnostic outcomes for a robot acting Normal or Faulty, and a person choosing to Trust or Distrust. Model assumes a faulty robot is less trustworthy $o^{tr}_{TF} < o^{tr}_{DF}$ and positive outcome rewards the person for trusting a normal robot $o^{tr}_{TN} > o^{tr}_{DN}$.

		Robot Trustee	
		a^{te}_N	a^{te}_F
Person Trustor	a^{tr}_T	o^{te}_{TN}, o^{tr}_{TN}	o^{te}_{TF}, o^{tr}_{TF}
	a^{tr}_D	o^{te}_{DN}, o^{tr}_{DN}	o^{te}_{DF}, o^{tr}_{DF}

faulty robot is the expected loss

$$R(a^{te}_F) = L(o^{tr}_{TF}, o^{tr}_{DF})P(a^{te}_F)$$

where $P(a^{te}_F)$ is probability the robot is faulty. Deciding whether to trust a faulty robot can be written as

$$a^{tr}_T \text{ if } R(a^{te}_F) \geq \theta \text{ else } a^{tr}_D$$

where θ is the person's propensity to trust. If $\theta < 0$, the person is willing to accept a negative outcome by trusting the robot, *i.e.*, overtrust. The hypothesis is that a faulty robot will induce a riskier outcome for the person, so $R(a^{te}_F) < \theta$. Note, the robot's outcome o^{te} is totally dependent on whether the person decides to trust, whereas the person's outcome o^{tr} is conditioned on how the robot behaves and their own action.

Using this model, short-term experiments show that most people have propensity to trust even when assumed outcome is negative $\theta < 0$. In the four scenarios listed above, participants rationalize their trusting actions:

- "I thought [pouring orange juice on a plant] was odd, but I did not question the robot's decision, so I followed the instructions."

- "I followed the exit where the robot was pointing. I did not notice the exit behind the robot where I came in."
- "The robot was delivering cookies, so I let it inside."
- "Faulting is normal in robots. I asked for hints whenever I needed them."

Evidence pointing towards overtrust reveals some missing parts to the decision model. Agreeing to be part of an experiment significantly increases the compliance rate for any actions (foot-in-the-door effect). If the robot's outcome is not clear, as with the normal versus food delivery robot, participants choose not to trust it. A robot that self-corrects faulty actions is seen as a positive [41], such as the robot making a wrong turn and then correcting itself. The participant's selective attention might not notice the entire action space, or even the faulty robot behavior.

3.2 Package Delivery Scenario Design

Having established methodology for a trust-diagnostic experiment, this section details a package delivery robot scenario operating in a university library setting. Participants follow a robot to three stops where they deliver a package. A dilemma is presented at the second and third stops. The robot was remotely controlled Wizard-of-Oz style by the same experimenter. Wizard-of-Oz experiments are used when the capabilities for a fully-automated robot do not yet exist, and the research aims to answer behavioral HRI questions. The scenario was executed with two routes, the only difference being the third stop.

Figure 3.1: Starting point in the library. Pioneer 3-DX robot with packages shown.

3.2.1 Participant Introduction

Participants were recruited from the student body studying in the area. An experimenter would approach a student and request, “Hello. I am developing a package delivery robot and I am looking for participants to interact with it. Would you be interested in this?” If the participant answered yes, the experimenter followed up with, “thank you! This should take no more than 15 minutes. Would you please answer a few questions and then meet me by the help desk?” The experimenter presents a QR code for a Qualtrics survey to the participant, then returns to the help desk, shown in Figure 3.1, awaiting their arrival.

Once together again, the experimenter describes the robot and the scenario. “This is the package delivery robot. It is programmed to navigate to three way-points around the workspace here. Your job is to follow the robot and deliver each (unlabeled) package. Once the robot reaches a delivery stop, it will signal you." The experimenter then plays the stop arrival sound (`ROS sound_play builtin 1`) from a joystick hidden beneath the table. The participant acknowledges the sound. "Now your job is to deliver the package. Place it somewhere intuitive, for example, if the robot stopped at the help desk here, you might place the package on the counter. After delivery, signal the robot that you completed the task by waving your hand in front of the orange laser scanner in front." The participant asked any remaining questions, then started the route by waving their hand in front of the robot. The experimenter plays the acknowledgement sound (`ROS sound_play builtin 2`) and then drives it to the first waypoint.

3.2.2 Robot

The delivery robot is a Pioneer 3-DX. One Hokuyo UST-10LX laser scanner provides 180° front facing field of view for 10 meters. A single speaker in the robot plays sound. An onboard Raspberry-Pi 4 running a ROS client handles sensor input and control output. A laptop, connected to the Raspberry-Pi via 5GHz wifi channel 100, is running the ROS Master, joystick input, and RViz. The experimenter drives the robot via joystick and uses RViz to track and localize the robot. A brown box with an Amazon logo is attached to the top of the Pioneer. Three unlabeled brown packages are in the box. Participants describe the robot as cute and adorable.

Integrating the robot application took a significant time-line. After acquiring a Pioneer 3-DX, the first step is imaging the Raspberry-Pi with Ubuntu 20.04 Server. Flashing the microSD card is easily done using Raspberry-Pi Imager. A dedicated monitor with an HDMI cable and keyboard are needed to initialize the headless OS. The default network configuration software `Netplan` failed to configure and connect to the local wireless network correctly, so it was removed and replaced with `ifupdown`. Setting a static ip for the Raspberry-Pi's wireless access point `wlan0` creates a stable interface when restarting the system. With the Raspberry-Pi connected to wireless, the client laptop on the same wireless network can SSH into the Raspberry-Pi, obsoleting the need for the separate monitor and keyboard. The Hokuyo UST-10LX communicates through an ethernet connection to the Raspberry-Pi.

Installing ROS is well documented. The Raspberry-Pi only needs `ROS-Base` (no GUI applications) whereas the client laptop needs the full desktop installation. All Pioneer robot platforms communicate through a serial interface, which is abstracted through an open source library `P2-OS` [53]. A serial-USB cable connects the Pioneer motherboard to the Raspberry-Pi, which shows up under devices typically as `/dev/USB0`. Another common device name is `/dev/ttyS0`. To consistently set the device name, a `udev` rule must be defined. One more peripheral is a joystick controller. Using a bluetooth PS4 controller directly with the Raspberry-Pi's onboard network card interferes with the 2.4GHz wifi signal. Therefore, a Logitech F710 gamepad with the Nano dongle connected to the client laptop. The joystick outputs velocity commands which are sent via rostopic `/cmd_vel` over wifi to the Raspberry-Pi. After downloading ROS onto the Raspberry-Pi and building P2OS with `catkin`, a basic systems check is to `roslaunch` the p2os nodes on the Raspberry-Pi and verify velocity commands correctly output from the client laptop, to the Raspberry-Pi and successfully move the robot. The default acceleration settings are too low, so they were increased to $7m/s^2$ for both linear and rotational acceleration, which greatly improved control. Another difficulty is the university wireless network limits the available bandwidth of any 2.4Ghz wifi channel, causing erratic robot control and sensor readings on the client. Using a 5GHz wifi channel 152, which requires manually enabling on the wifi router's settings, provided sufficient bandwidth and range

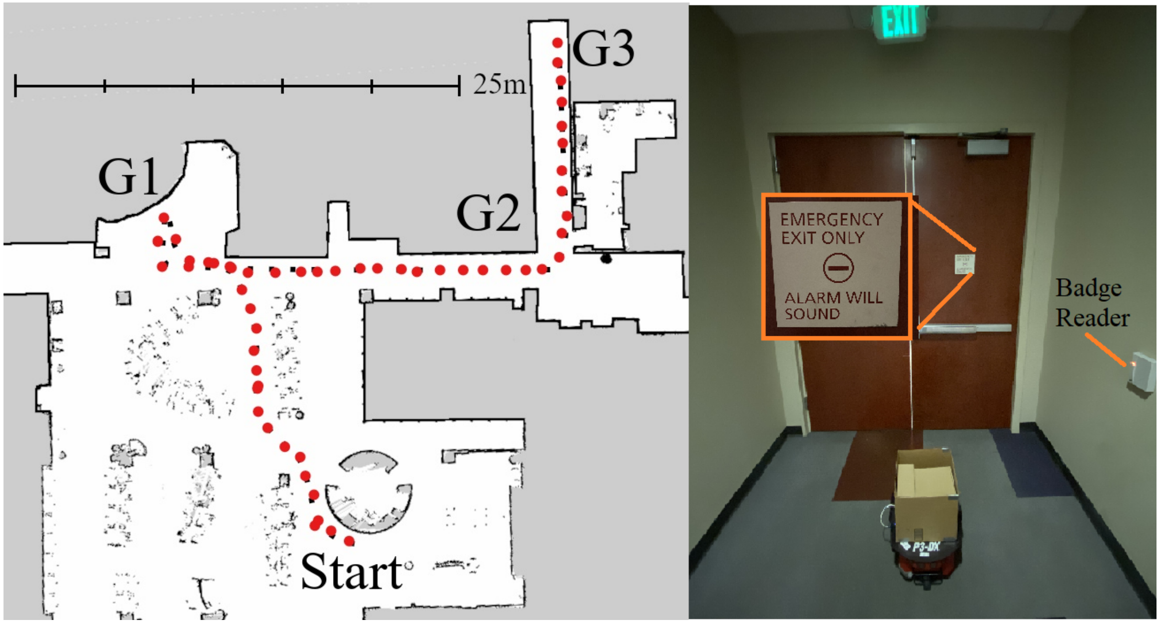

Figure 3.2: University library plan view of route 1. Participant meets package delivery robot at Start location which is a help desk. First package dropoff at G1 is a helpdesk. Second at G2 is inside an office. Final stop at G3 is a closed door marked "Emergency Exit. Alarm will Sound."

over the entire delivery route.

3.2.3 Delivery Route 1

The first delivery route is shown in Figure 3.2. From the start, the robot leads the participant to the first goal, a Production help desk for printing posters. This location is in public view, and designed to reinforce the package delivery handoff. Upon arriving at the goal location, the robot plays the stop arrival sound. Then the participant places a package on the counter. After they wave in front of the robot, it plays the acknowledgement sound and moves onward.

The second waypoint is in front of a student office space. A table is within three feet of the door's entrance. This room serves as a student office and appears private. There are no signs denoting what kind of room it is. An occupied office with a nameplate is 4 feet directly in front of the doorway. Next to the office is a table with enough clear space to place a package. The robot stops just ahead of the door and signals a package delivery. Participants face a dilemma of entering the private room to deliver the package, or place it outside on the floor. Once the participant waves the robot onward, it heads toward the third and final goal.

The final stop puts the participant at the end of a long hallway blocked by double doors marked “Emergency Exit Only. Alarm will Sound.” If the door is opened, a loud alarm will sound, but no building systems or security are alerted. This door is routinely used by library personnel for entrance and exit. Even though a badge swipe allows opening the door without an alarm, we heard the alarm go off several times because someone opened the door without a badge swipe. Once the robot reaches the door at the end of the hallway, it announces a stop and waits for the participant. This dilemma forces to participant to determine whether the robot wants to open the door, or just place the package on the ground. Afterwards, the robot leads the participant back to the start.

Table 3.2: Increasing prods from robot.

Prod Level	Description
1	No Prod. At third stop, robots returns to start when signaled.
2	Wait 1 minute at third stop. After 30 seconds, beep every 10 seconds.
3	At second stop drive into room. Third stop, nudge emergency exit door.

In order to force compliance, the robot used mounting prods described in Table 3.2. Level 1 maintains normal behavior. At the third stop, after the participant signals the robot package delivered it heads back to the start. Escalating to Level 2, the robot waits at the emergency exit door for 1 minute. After 30 seconds the robot beeps every 10 seconds until a minute passes, then it returns to the start. At Level 3, the robot drives into the office at the second stop, and at the third stop the robot will nudge the emergency exit for 1 minute.

The second waypoint outside the office room presents a minor dilemma for the participant. Should they place the package inside the room on the table, or leave it outside on the floor? Goal 3 presents a similar dilemma. Should the package be placed on the floor or does the robot want to go through the door? Can non-verbal prods enforce compliance? Will increased pressure to comply diminish the perceived social intelligence of the robot? The last question is motivated by Politeness Theory [45] where direct requests for large face threats are considered less polite than indirect requests. Persuasive prods toward face acts could increase psychological reactance and negative emotions toward the requestor [54].

In order to evaluate the proposed questions, objective surveys measure the propensity to trust [55] of each participant before the trial, and the perceived robot social intelligence post-interaction [56]. The former survey poses 14 five-point scale questions that gauge a participant's trustworthiness and 7 questions related to trust towards others. The perceived robot social intelligence survey asks 24 questions that assess the robot's trustworthiness and social presentation. Questionnaires are distributed by a QR-code from Qualtrics and answered on the participant's phones.

We also ask three verbal questions after completing the questionnaires:

- "How was it?" Elicits any outstanding anecdotes from the trial.
- "What did you think of the robot?" Any robot behaviors or characteristics, which stood out.
- "What did you think of the door at the end of the hallway?" What did the participant think when they saw the door and the robot acknowledged a stop in front of it.

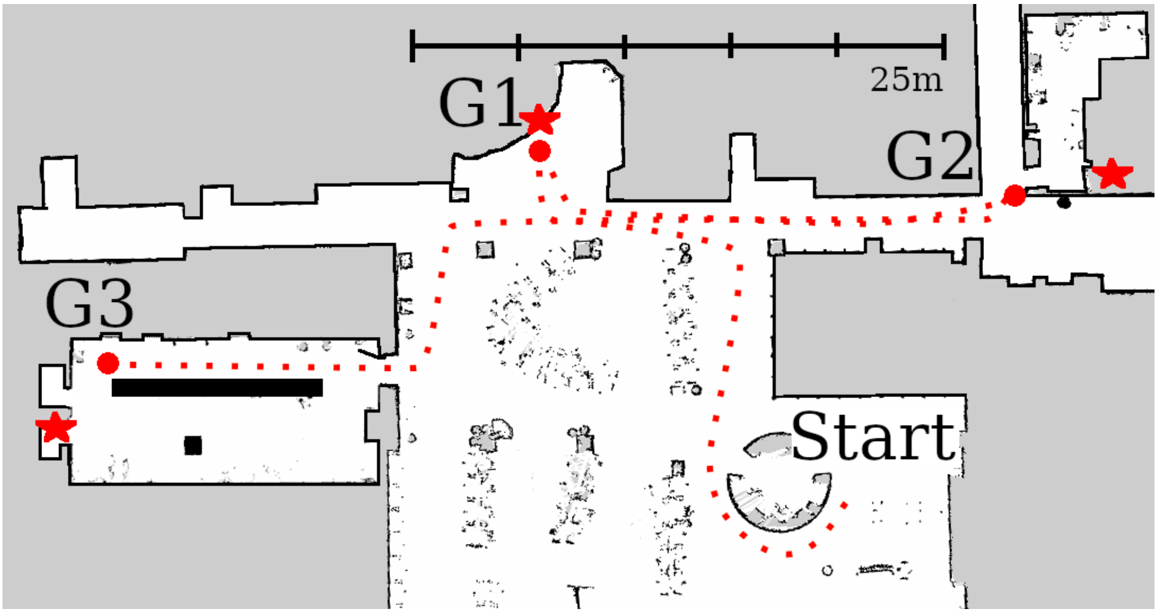

Figure 3.3: Study 2 delivery route. Third stop in front of another Emergency Exit door in a computer lab. Red circles indicate robot's stop location. Red stars are intended package delivery spot.

The experiment hypotheses are:

1. Robot prods will encourage participant compliance toward face acts, *e.g.*, entering the office at the second stop and opening the emergency exit door, if the robot prods them.

2. Prods will diminish the perceived social intelligence and trustworthiness of the robot.

3.2.4 Delivery Route 2

The second delivery route changes the third stop in order to gauge the participant's attentiveness. The first and second stops remained the same. Additional information was also presented to the participants so they would have a more clear understanding of how to deliver the packages.

Figure 3.3 shows the third delivery stop, which is now in a public computer lab.

A glass door opens into the lab and is typically open. In exploratory trails, the door was closed and the robot would request the participant to open it. If the robot only used beeps, the participant did not understand what the robot wanted and they came over to experimenter to ask what to do. When the robot requested the participant to open the door, there was no uncertainty and participants opened the door to let the robot in and out. Since opening a door did not present a trust diagnostic, the door was open for all trials.

Inside the computer lab, in the back corner, is another emergency exit door, with the same markings as the door at the first delivery route. A nearby office is the delivery goal. The robot stopped in front of the emergency exit door, forcing the participant to walk several feet to deliver the package. Since an emergency exit should be salient, the hypothesis is that most participants will notice the door when the robot stops next to it.

Additional information was added by labeling each package with the respective location's name. For example, at the first goal the package labeled "Production" should be delivered. These labels were explained during the participant introduction. As an additional group, the robot's communication was upgraded with speech instructions relative to each stop, *e.g.*, "Stop arrived. Please deliver the package to Production" and, "Package delivered. Let's move to the next stop." Increasing the amount of information provided to the participant is hypothesized to give them more confidence to enter the office at the second stop.

A post experiment questionnaire was administered after the participant completed the delivery route. There were 14 questions on a 5-point scale asking about the perceived robot's autonomy, sensing, trustworthiness, and faultiness. This questionnaire was distributed by Qualtrics QR-code.

The hypotheses for delivery route 2 are:

1. Participants are more likely than not to notice the Emergency Exit door at G3.
2. Labeling packages will prod more participants than Study 1 into the office at the second stop.
3. Adding robot speech will prod more participants into the office at the second stop.

Chapter 4

Results and Discussion

Results from the first (n=17) and second (n=20) delivery routes show all participants understand how to interact with the robot and will follow the robot to all three delivery stops. Compliance to delivering the packages is mixed. Most individuals demurred from entering the office at the second stop without sufficient prodding. No one seriously thought about opening the emergency exit door at the end of the hallway. Although few participants noticed a similar emergency exit door in the computer lab. Increasing the robot's communication from simple beeps to text-to-speech improved participant's understanding of the robot, but no one noticed the emergency exit door in this condition.

4.1 Participants

For delivery route 1, 18 undergraduates (13 in a STEM field) were recruited for 16 trials between April 20 - 22, 2022. Two trials included a group of two participants. In one trial the participants were both recruited, while in another trial the participant's friend joined after the route started. Two mistrials occurred (not counted); one trial

Table 4.1: Trial counts per robot communication case. Invalid trials are when participant did not deliver all packages.

Case	Population	Valid Trials	Invalid
Route 1 Beep	Individual	12	2
	Group	2	0
Route 2 Beep	Individual	4	7
	Group	3	0
Route2 Speech	Individual	3	1
	Group	2	0
Total		27	10

did not finish because the WiFi signal terminated, another trial is excluded because the participant afterwards noted that they worked in the library and knew about the emergency exit door. Two more trials are invalid because participants delivered all packages at the first goal, but they followed the robot to all three stops. All participants completed the pre and post questionnaires.

The pre-trial Propensity to Trust [55] questionnaire assesses a participant's willingness to be trustworthy toward others and whether others can be trusted on a 5-point differential scale. Aggregated trustworthy scores (mean=4.07, SE=0.95) and trust scores (mean=3.02, SE=1.13) did not show significant differences between those who entered the office and those who left the package in the hallway. The lowest trustworthy score is "I believe laws should be strictly enforced" (mean=3.5, SE=1.03) and the largest variance is in response to "I am filled with doubt about things" (mean=3.38, SE=1.39). The last question "Believe people seldom tell you the whole story" (mean=3.76, SE=0.7) might have primed participants to think the experiment is more than just a robot navigation test.

Delivery route 2 was tested over 20 trials with a total of 25 participants (15 individual, 5 group trials). The study was conducted the week before final exams April 28 - May 3, 2022. A significant portion of individuals compared to Route 1 did not deliver the packages correctly (Fisher Exact test, p=0.01**) [57]. Four participants did not take the packages out of the box, three held onto the packages after placing them, and one participant could not locate the second drop and delivered the wrong labeled packages. The longest run of invalid trials is four. Individuals reported the instructions were clear, and completed the post-trial questionnaire. All group trials complied with the experimenter's instructions. Whether stress from final exams confounded the compliance rate for individuals is unclear.

4.2 Delivery Route 1 Results

No participant opened the emergency exit door at the end of the hallway shown in Figure 3.2. Table 3.2 shows increasing prods did not convince participants that the robot wanted to open the door, nor was it safe to open. In the Level 1 case (n=6) the robot just beeped in front of the emergency exit door, and all participants placed the package on the floor then signaled the robot packaged delivered. Four participants stated they didn't know what the robot wanted, "[the robot] just stopped and beeped, so I put the package on the ground. I wasn't sure if I was supposed to open the door... It turned around when I put my hand in front of it." One participant rationalized about placing packages on the ground, "an Amazon delivery driver wouldn't open someone's door to deliver a package." Furthermore, participants stated the sounds and robot interaction were intuitive.

Increasing the prod to Level 2, all participants waited the entire minute at the door (n=7). Four participants reported confusion, "I was confused whether I should open the door for the robot. It made odd beeping sounds I wasn't expecting. I tried

Table 4.2: Number of trials where package was left on the table inside the office or outside on the floor. Fisher exact test p-value against Level 1 - Individual. Levels 1 and 2 are combined.

Level	Population	Table	Floor	Fischer p-value
1/2	Individual	2	9	-
1/2	Group	2	0	0.076*
3	Individual	3	0	0.027**

scanning the robot until it finally went back to the starting point." One group of participants stated, "we thought the robot would signal us to open the door, but we were not sure. We were trusting the technology." Twice someone else opened the emergency exit door from the other side, at which point the robot was recalled back to the start. Both times the participant associated the door opening with the robot's return. Another participant thought the emergency exit sign was fake and tried to remove it.

Escalating the prod to Level 3, where the robot nudged the door, alarmed all three participants (n=3). One participant returned the experimenter within 20 seconds exclaiming the robot was malfunctioning and running into an emergency exit door. The other two participants did deliver the package because the robot did not sit still. One participant, after returning to the start, delivered the final package to the help desk. Although the door nudge was intended to be like a dog wanting to go outside, participants thought the robot was malfunctioning and were afraid it would set off an alarm. They tried to get between it and the door so it would stop. In this case, the robot could not communicate what it wanted via non-verbal communication. Simple beeps for stop arrival and moving on were not descriptive enough to convince someone that it wanted them to open a door.

The office at goal 2 also presented a dilemma. In both Level 1 and Level 2, the robot stopped outside the office in the hallway. Only two individuals entered the office to deliver the package, while both groups entered the office and put the package on the table. Table 4.2 shows a Fisher exact test gives a marginally significant p-value that groups are more likely to enter the room (p=0.076*). Since the instructions were to deliver the package to an appropriate location, yet the packages were not labeled. The robot only signaling stop arrived with a beep, with no additional information. Therefore participants relied on their own judgement whether to enter the seemingly private room. One individual remarked the delivery, "took a little more thought since the room inside could've been off limits to outsiders." Another individual entered the office, saw someone inside working, and then retreated to leave the package in the hallway.

Driving the robot into the office (Level 3) prodded all three participants to deliver the package on the table. Fisher's exact test [57] yields a significant p-value (p=0.027**) [57] compared to individuals when the robot stopped outside the room. From this evidence we accept Hypothesis 1 that a robot can prod participants into performing face threatening acts, but only if the people clearly understand what the robot wants.

In the post-trial questionnaire of Perceived Robot Social Intelligence, participants report the robot is socially competent (mean=3.95, SE=0.76) and trustworthy (mean=3.50, SE=1.00). No significant difference exists between prod level groups. Anecdotally, the three Level 3 participants were more alarmed by the emergency exit door than the Level 2 participants. Although they still followed the robot back to the start without complaint. Therefore, Hypothesis 2, prods diminish perceived social intelligence and trustworthiness of the robot, is rejected.

4.3 Delivery Route 2 Results

Table 4.3: Route 2 drop locations by robot communication case and population.

		Second Stop			Third Stop	
Case	Population	Office	Table	Floor	Office	Elsewhere
Beep	Individual	1	2	1	1	3
	Group	1	2	-	3	-
Speech	Individual	2	-	1	2	1
	Group	2	-	-	1	1

Labeling the packages with respective delivery locations successfully prodded a majority of participants to enter the office at the second stop (Fisher exact test, p=0.077*) [57]. The new third stop inside the computer lab presented a trickier situation because the robot stopped several feet away from the intended destination. An uncontrolled variable in this study is whether the offices at the second and third stops were opened. We recorded these variables for each trial and will note differences when relevant. Upgrading the robot's communication from simple Beeps to full Speech, improved participant compliance for delivering the packages, as shown in Table 4.1. At the second stop, when the person was inside their office, participants were confident enough to walk into his office and deliver the package! Thus we accept Hypothesis 2 that labeling packages gives individuals agency to enter a private room. In this experiment, where there is little ambiguity, there is not enough evidence to assert a speaking robot is more persuasive than a less expressive beeping one.

For the third delivery in the computer lab, the robot stopped in front of an Emergency Exit door. In only (6/20) trials participants noticed the emergency exit door. No one in the robot speech case noticed the door, even though one of the groups placed the package on the door handle. Therefore we reject Hypothesis 1 that more

Table 4.4: Delivery route 2 post-trial questionnaire results. Scores on a 5-point scale. Ordered by F statistic ($p^{**} < 0.05, p^{*} < 0.1$).

	Beep (n=17)		Speech (n=8)	
	Mean	SE	Mean	SE
The robot communicates clearly.	3.53	1.18	4.75**	0.46
The robot wants my best interest.	3.41	1.00	4.25**	0.71
The robot is remote controlled.	3.41	1.06	2.25*	1.83
The robot is faulty.	2.12	1.11	1.38*	0.52
I know what the robot wants.	3.59	1.00	4.13	0.64
The robot understands me.	3.24	1.09	3.88	1.13
The robot is autonomous.	4.00	1.06	3.38	1.19
I trust the robot.	4.00	1.06	4.38	0.92
I understand the robot.	3.76	1.09	4.13	0.99
The robot navigates successfully.	4.47	0.62	4.63	1.06
The robot can hear me.	2.65	1.11	2.88	1.25
The robot is erratic.	2.35	1.11	2.13	1.46
The robot can see me.	4.29	1.10	4.25	0.71
The robot is trustworthy.	4.00	0.87	4.00	0.93

people than not notice the emergency exit door when the robot stops next to it.

Upgrading the robot with speech significantly improves communication. Table 4.4 shows the post-trial questionnaire results between the robot beep (n=17) and robot speech (n=8) cases. An ANOVA F-test is a common way to measure differences between groups of Likert scores [58, 59]. The higher the F score is above 1, the more likely the two groups answered the questionnaire differently. An $F = 1$ score implies the groups answered the questions similarly. Comparing answers from the groups of robot beeps versus robot speech, the F-oneway test asserts that a speaking robot communicates more clearly ($F = 7.85^{**}, p < 0.05$), wants a person's best interest more

($F = 4.48^{**}, p < 0.05$), and is perceived less remote controlled ($F = 4.05^{*}, p < 0.1$) and less faulty ($F = 3.19^{*}, p < 0.1$).

4.4 Discussion

During development and testing in the library, some students engaged with the robot while it moved down hallway corridors. They commented about it, and made personal space for it when in close proximity. However most students did not engage at all with the robot. Even as the robot sped by, the student did not glance away from their computer screen, or they continued walking down the hallway without deviation.

4.4.1 Participant Compliance

This experiment is designed with a foot-in-the-door approach to evaluate judgement under uncertainty in a HRI setting. The first stop is in a public area, with an easily identifiable delivery spot on the Production desk. In only (6/41) trials participants incorrectly delivered the package. In four trials no package was delivered, another participant delivered all three packages to the Production desk, and one participant took the entire box off the robot and placed it on the counter. As final exams neared, more participants did not follow directions and deliver packages correctly. Student's anxiety levels have been shown to negatively correlate with performance [60], but this research does not make a causal assertion.

The second delivery presented a dilemma. Although the undergraduates were familiar with the library, the office was tucked in a quiet corner away from the public commons. Participants reported the office was "eerie. I didn't know this person." Entering the room presented a potential negative outcome: violating personal space. Without a clear positive outcome for the robot, most participants demurred from entering the room. However, a pair of friends rationalized placing the package in the room was the correct action. Labeling the package with the office-holder's name provided sufficient motivation to enter the room. Individuals were more likely to place the package in the second room (3/4) if the package was labeled. Although participants still reported the room was uncomfortable, the label provided enough of a hint for the person would go inside. There is not enough evidence to state a speaking robot is more persuasive than a beeping robot if the participant understands their task. Although a speaking robot communicates more clearly for tasks than just simple beeps.

No one seriously thought about opening the Emergency Exit door at the end of Route 1. Participants were unsure what the robot wanted, since it only communicated stop arrival. Waiting one minute at the door, moving back and forth erratically, and nudging the door only increased perceived faultiness. In the nudging case, all three participants were alarmed by the robot's behavior, whereas in other cases no one was distressed by waiting at the door. Post-experiment, several participants asked if the trial was a test. Others reported the experiment was the most interesting activity they had done all day. Participants could only infer what the robot wanted, and they were not willing to risk sounding an alarm trying to find out.

To gauge how much information the robot beep transmitted, two exploratory trials had the robot request participants to open the computer lab door. Using the same stop arrival beep was not enough to convey the request. Only a verbal request prompted the participant to open the door. "I didn't know why the robot was at the door. Then I heard it say open the door, so I did." Verbal requests have been shown to be very persuasive, and likely 100% of participants would comply with the simple request. Perhaps a speaking robot could convince a person to open the emergency exit door.

Significantly more participants during Route 2 did not deliver any packages. One hypothesis is the looming final exam deadline for many students affected their attention. Table 4.1 shows increasing the robot's communication to speech reminded participants to deliver the packages. A simple beep was not enough to remind participants of the task, whereas the robot request "Stop arrived. Please deliver the package to Production" was very clear and unambiguous.

A side effect of increasing robot communication could be less attention to the environment. Although only 30% of participants noticed the emergency exit door at the third stop this result is not too surprising. Most participants were unfamiliar with the room and were focused on the robot. Selective attention is also a well studied phenomena [61]. However, in the robot speech case, no participant (0/5) noticed the emergency door, even though one group placed the package on the door. Further research could explore robot expressiveness and selective attention.

4.4.2 Limitations

Experimenting in a public setting invites lots of noise between trials. Several times touring groups would interfere with the route, and the robot would have to drive through the group. Participants excitedly reported the robot's capabilities although this was not a controlled variable. A more affecting change is whether the offices were occupied or not. A closed office could be less inviting and repel participants from entering the room. The third stop in Route 2 presented a difficult alternative if the office was closed, with no obvious delivery location. Although participants found creative delivery spots, for example one group left the package on the door handle of the emergency exit door!

The guise of the package delivery robot could be confounding as well. Participants were recruited to "interact with a package robot as it navigated around people." Some participants questioned why delivering packages when a robot should be able to this itself. These two reasons might be why several participants did not completely follow directions and deliver packages. Several participants were kind enough to carry all the packages with them back to the start. While this scenario is useful to research how people comply with requests from robots, a more realistic task might reduce the number of invalid trials.

The groups of participants were all friends who were recruited together. The increased compliance rate aligns with the piggybacking robot study [5]. However, heterogeneous partner pairs might could have less trust with each other to enter unfamiliar, private office spaces.

Chapter 5

Conclusion and Future Works

Compliance experiments are very successful because of the foot-in-the-door phenomena. By agreeing to be part of a study, people are much more likely to agree with larger requests, even from a robot. For example, participants voluntarily pour orange juice on a plant [3], use a rectal thermometer [50], or rename hundreds of individual files [49] just because a robot asks. More serious overtrust actions in scenarios such as emergency evacuations [4] or piggybacking into secure facilities [5] show the potential for poor outcomes with robot interactions. Furthermore, HRI studies have shown manipulating a person to distrust a robot is difficult. This study aimed to explore the factors, which enforce compliance to robot requests.

This HRI experiment had a participant follow a mobile robot into an unknown environment and asked them to make a judgment under uncertainty. Participants will not enter a private room without explicit invitation. Labeling packages was sufficient to invite people all the way into the unknown person's office; participant's trust in the robot had little effect on their actions. Although improving a robot's communication from simple beeps to full speech improved qualitative ratings, there is no significant evidence a more capable robot was more trustworthy in this study. When the robot marked a delivery stop in front of an emergency exit door, no one seriously thought about opening it. Waiting one minute, or nudging the door only induced made the robot look like it was malfunctioning. Without absolute assurance the robot wanted to open the door, and that it was safe to open, people were extremely unlikely to put themselves at risk because of robot request. To test how salient an emergency exit door is, the robot stopped in front of a different door in a public computer lab. Only 30% of participants noticed the door, and when the robot used speech communication no one noticed the door. Selective attention is a well studied field in psychology, so an engaging robot which reduces environmental awareness should be expected. However, in highly interactive autonomous vehicles, drivers must be extremely aware of their surroundings to prevent catastrophic overtrust [62].

The experiment within this thesis shows that participants are willing to comply with face threatening, direct requests, such as entering a private office. However, when the request is indirect, such as delivering an unlabeled package, individuals were much more unlikely to enter a private office space, even though pairs of friends had no issue entering the office. The robot's ambiguous request did not provide sufficient motivation for overtrust. In conclusion the participants in this study were eager to interact with the robot and follow the experimenter's instructions (when they were paying attention), however without a clear positive goal for the partner participants were unlikely to overtrust and receive a negative outcome.

Studying trust is difficult because of the multitude of factors involved. Relationships evolve over time and are difficult to develop and diagnose within short experiments. Trust in robots depends on whether the person is operating it, or interacting with it in the environment. Operators primarily rely on the robot's performance, whereas people in the environment trust their own senses to decide which actions to take. If a robot malfunctions and successfully recovers, people view this just as a symptom of how robots work.

5.1 Future Work

This study grew from the lessons learned of previous HRI experiments, therefore, new ideas should be extended from this one as well. Implementing robust navigation software into the mobile robot would open up new possibilities, which do not include a human experimenter. For example, the robot could navigate around the library in public space and request a sample of students to interact with it. This demonstration of competence sets up the potential for a trust diagnostic. Further studies could be performed at the city public library, extending beyond undergraduate students.

Moreover autonomous robot interactions could model appropriate social distance from participants. During one trial, the participant lost interest with the robot because it waited one minute at the emergency exit door. After they lost interest and walked away, the robot followed them and regained their interest. A data driven model of this kind of game could provide empirical insight into the human attention model.

The broader impact of this study is human comfort level entering private spaces for deliveries. Anecdotal evidence of last-mile delivery drivers reveals employee discomfort in new areas [10, 11]. Small informational cues can improve user comfort when entering unfamiliar areas to hand deliver packages. For example, an app with crowd-sourced data or relevant statistics could inform a delivery driver about their stop location with more detail than an automated route scheduler could provide. In a future with autonomous cars, delivery drivers could be subjugated to just package handlers. When the car stops at a delivery destination, the person just walks to the drop off location. Evidence from this study shows the handlers may have high levels of discomfort entering unknown domestic spaces. Further research into delivery route optimization for familiarity could yield insight into the psychology and comfort for these essential workers.

Bibliography

[1] A. R. Wagner, P. Robinette, and A. Howard, "Modeling the human-robot trust phenomenon: A conceptual framework based on risk," *ACM Transactions on Interactive Intelligent Systems*, vol. 8, no. 4, p. 1–24, Nov 2018.

[2] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, p. 517–527, Oct 2011.

[3] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "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.* ACM, Mar 2015, p. 141–148. [Online]. Available: https://dl.acm.org/doi/10.1145/2696454.2696497

[4] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of robots in emergency evacuation scenarios," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar 2016, p. 101–108.

[5] S. Booth, J. Tompkin, H. Pfister, J. Waldo, K. Gajos, and R. Nagpal, "Piggybacking robots: Human-robot overtrust in university dormitory security," in *Proceedings of the 2017 ACM/IEEE International Conference on*

Human-Robot Interaction. ACM, Mar 2017, p. 426–434. [Online]. Available: https://dl.acm.org/doi/10.1145/2909824.3020211

[6] S. Kiesler, L. Sproull, and K. Waters, "A prisoner's dilemma experiment on cooperation with people and human-like computers," *Journal of Personality and Social Psychology*, vol. 70, no. 1, p. 47–65, 1996.

[7] M. D. Simoni, E. Kutanoglu, and C. G. Claudel, "Optimization and analysis of a robot-assisted last mile delivery system," *Transportation Research Part E: Logistics and Transportation Review*, vol. 142, p. 102049, Oct 2020.

[8] B. E. Holthausen, P. Wintersberger, B. N. Walker, and A. Riener, "Situational trust scale for automated driving (sts-ad): Development and initial validation," in *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Virtual Event DC USA: ACM, Sep 2020, p. 40–47. [Online]. Available: https://dl.acm.org/doi/10.1145/3409120.3410637

[9] A. Lee and A. L. Toombs, *Robots on Campus: Understanding Public Perception of Robots using Social Media*. New York, NY, USA: Association for Computing Machinery, Oct 2020, p. 305–309. [Online]. Available: https://doi.org/10.1145/3406865.3418321

[10] L. Gurley, "Amazon's cost saving routing algorithm makes drivers walk into traffic," *Vice*, Jun 2021. [Online]. Available: https://www.vice.com/en/article/5db95k/amazons-cost-saving-routing-algorithm-makes-drivers-walk-into-traffic

[11] C. Mims, "Why global supply chains may never be the same - a wsj documentary," *Wall Street Journal*, Mar 2022. [Online]. Available: https://www.wsj.com/video/series/chain-reaction/

why-global-supply-chains-may-never-be-the-same-a-wsj-documentary/ 4EFE56B6-8A1D-4478-9F88-8F055AFBF675

[12] J. M. Carroll, “Human–computer interaction: psychology as a science of design,” *International Journal of Human-Computer Studies*, vol. 46, no. 4, p. 501–522, Apr 1997.

[13] S. Weinberger, “Armed robots still in iraq, but grounded (updated),” *Wired*. [Online]. Available: https://www.wired.com/2008/04/armed-robots-st/

[14] S. Ogreten, S. Lackey, and D. Nicholson, “Recommended roles for uninhabited team members within mixed-initiative combat teams,” in *2010 International Symposium on Collaborative Technologies and Systems*, May 2010, p. 531–536.

[15] H. X. Pham, H. M. La, D. Feil-Seifer, and M. C. Deans, “A distributed control framework of multiple unmanned aerial vehicles for dynamic wildfire tracking,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 4, p. 1537–1548, Apr 2020.

[16] A. Q. Pham, A. T. La, E. Chang, and H. M. La, “Flying-climbing mobile robot for steel bridge inspection,” in *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Oct 2021, p. 230–235.

[17] K. Shrestha, R. Dubey, A. Singandhupe, S. Louis, and H. La, “Multi objective uav network deployment for dynamic fire coverage,” in *2021 IEEE Congress on Evolutionary Computation (CEC)*, Jun 2021, p. 1280–1287.

[18] S. Thrun, *Probabilistic robotics*, ser. Intelligent robotics and autonomous agents. Cambridge, Mass: MIT Press, 2005.

[19] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep 2014, p. 709–715.

[20] J. Leaman and H. M. La, "A comprehensive review of smart wheelchairs: Past, present, and future," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, p. 486–499, Aug 2017.

[21] ——, "The intelligent power wheelchair upgrade kit," in *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, Nov 2020, p. 416–421.

[22] T. B. Sheridan, "Human–robot interaction: Status and challenges," *Human Factors*, vol. 58, no. 4, p. 525–532, Jun 2016.

[23] A. R. Wagner, "Robot-guided evacuation as a paradigm for human-robot interaction research," *Frontiers in Robotics and AI*, vol. 8, p. 701938, Jul 2021.

[24] H. Ahmed and H. M. La, "Evaluating the co-dependence and co-existence between religion and robots: Past, present and insights on the future," *International Journal of Social Robotics*, vol. 13, no. 2, p. 219–235, Apr 2021.

[25] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan, "Courteous autonomous cars," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018, p. 663–670.

[26] S. B. Banisetty, V. Rajamohan, F. Vega, and D. Feil-Seifer, "A deep learning approach to multi-context socially-aware navigation," *arXiv:2104.10197 [cs]*, Apr 2021, arXiv: 2104.10197. [Online]. Available: http://arxiv.org/abs/2104.10197

[27] D. V. Lu, D. B. Allan, and W. D. Smart, *Tuning Cost Functions for Social Navigation*, ser. Lecture Notes in Computer Science. Springer

International Publishing, 2013, vol. 8239, p. 442–451. [Online]. Available: http://link.springer.com/10.1007/978-3-319-02675-6_44

[28] A. Honour, S. B. Banisetty, and D. Feil-Seifer, "Perceived social intelligence as evaluation of socially navigation," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. Boulder CO USA: ACM, Mar 2021, p. 519–523. [Online]. Available: https://dl.acm.org/doi/10.1145/3434074.3447226

[29] D. J. Feil-seifer, "Data-driven interaction methods for socially assistive robotics: Validation with children with autism spectrum disorders," Ph.D. dissertation, University of Southern California, 2012.

[30] C. D. Kidd, "Designing for long-term human-robot interaction and application to weight loss," Ph.D. dissertation, Massachusetts Institute of Technology, 2008, accepted: 2008-09-03T15:34:19Z. [Online]. Available: https://dspace.mit.edu/handle/1721.1/42407

[31] A. Langer, R. Feingold-Polak, O. Mueller, P. Kellmeyer, and S. Levy-Tzedek, "Trust in socially assistive robots: Considerations for use in rehabilitation," *Neuroscience Biobehavioral Reviews*, vol. 104, p. 231–239, Sep 2019.

[32] B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people," *Cambridge, UK*, vol. 10, p. 236605, 1996.

[33] J. Brandstetter, P. Rácz, C. Beckner, E. B. Sandoval, J. Hay, and C. Bartneck, "A peer pressure experiment: Recreation of the asch conformity experiment with robots," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep 2014, p. 1335–1340.

[34] A. Katsila, "Active peer pressure in human-robot interaction," Ph.D. dissertation, University of Nevada, Reno, 2018.

[35] P. A. Hancock, T. T. Kessler, A. D. Kaplan, J. C. Brill, and J. L. Szalma, "Evolving trust in robots: Specification through sequential and comparative meta-analyses," *Human Factors*, vol. 63, no. 7, p. 1196–1229, Nov 2021.

[36] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco, "Impact of robot failures and feedback on real-time trust," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar 2013, p. 251–258.

[37] X. J. Yang, V. V. Unhelkar, K. Li, and J. A. Shah, "Evaluating effects of user experience and system transparency on trust in automation," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, Mar 2017, p. 408–416.

[38] A. M. Aroyo, D. Pasquali, A. Kothig, F. Rea, G. Sandini, and A. Sciutti, "Expectations vs. reality: Unreliability and transparency in a treasure hunt game with icub," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, p. 5681–5688, Jul 2021.

[39] A. Aroyo, D. Pasquali, A. Kothig, F. Rea, G. Sandini, and A. Sciutti, *Perceived differences between on-line and real robotic failures*, Aug 2020.

[40] J. A. Simpson, "Psychological foundations of trust," *Current Directions in Psychological Science*, vol. 16, no. 5, p. 264–268, Oct 2007.

[41] S. Reig, E. J. Carter, T. Fong, J. Forlizzi, and A. Steinfeld, "Flailing, hailing, prevailing: Perceptions of multi-robot failure recovery strategies," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot*

Interaction. Boulder CO USA: ACM, Mar 2021, p. 158–167. [Online]. Available: https://dl.acm.org/doi/10.1145/3434073.3444659

[42] N. Mirnig, G. Stollnberger, M. Miksch, S. Stadler, M. Giuliani, and M. Tscheligi, "To err is robot: How humans assess and act toward an erroneous social robot," *Frontiers in Robotics and AI*, vol. 4, 2017. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2017.00021

[43] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin, "Effects of non-verbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, p. 708–713.

[44] S. Agrawal and M.-A. Williams, "Would you obey an aggressive robot: A human-robot interaction field study," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2018, p. 240–246.

[45] V. Srinivasan and L. Takayama, "Help me please: Robot politeness strategies for soliciting help from humans," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, May 2016, p. 4945–4955. [Online]. Available: https://dl.acm.org/doi/10.1145/2858036.2858217

[46] D. Y. Geiskkovitch, D. Cormier, S. H. Seo, and J. E. Young, "Please continue, we need more data: an exploration of obedience to robots," *Journal of Human-Robot Interaction*, vol. 5, no. 1, p. 82–99, Mar 2016.

[47] L. J. Chang, B. B. Doll, M. van 't Wout, M. J. Frank, and A. G. Sanfey, "Seeing is believing: Trustworthiness as a dynamic belief," *Cognitive Psychology*, vol. 61, no. 2, p. 87–105, Sep 2010.

[48] S. Milgram, "Behavioral study of obedience." *The Journal of Abnormal and Social Psychology*, vol. 67, no. 4, p. 371–378, Oct 1963.

[49] D. Y. Geiskkovitch, D. Cormier, S. H. Seo, and J. E. Young, "Please continue, we need more data: an exploration of obedience to robots," *Journal of Human-Robot Interaction*, vol. 5, no. 1, p. 82–99, Mar 2016.

[50] C. Bartneck, T. Bleeker, J. Bun, P. Fens, and L. Riet, "The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots," *Paladyn, Journal of Behavioral Robotics*, vol. 1, no. 2, Jan 2010. [Online]. Available: https://www.degruyter.com/document/doi/10.2478/s13230-010-0011-3/html

[51] D. J. Bem, "Self-perception: An alternative interpretation of cognitive dissonance phenomena." *Psychological Review*, vol. 74, no. 3, p. 183–200, 1967.

[52] J. L. Freedman and S. C. Fraser, "Compliance without pressure: The foot-in-the-door technique," *Journal of Personality and Social Psychology*, vol. 4, no. 2, p. 195–202, Aug 1966.

[53] H. Allen, "P2os ros driver," https://github.com/allenh1/p2os, 2015.

[54] A. S. Ghazali, J. Ham, E. Barakova, and P. Markopoulos, "Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance," *Advanced Robotics*, vol. 33, no. 7–8, p. 325–337, Apr 2019.

[55] A. M. Evans and W. Revelle, "Survey and behavioral measurements of interpersonal trust," *Journal of Research in Personality*, vol. 42, no. 6, p. 1585–1593, Dec 2008.

[56] K. A. Barchard, L. Lapping-Carr, R. S. Westfall, A. Fink-Armold, S. B. Banisetty, and D. Feil-Seifer, "Measuring the perceived social intelligence of robots," *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 4, p. 1–29, Oct 2020.

[57] 2022. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html

[58] 2022. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html

[59] M. Mahbobi and T. K. Tiemann, *Chapter 6. F-Test and One-Way ANOVA*. BCcampus, Dec 2015, book Title: Introductory Business Statistics with Interactive Spreadsheets - 1st Canadian Edition. [Online]. Available: https://opentextbc.ca/introductorybusinessstatistics/chapter/f-test-and-one-way-anova-2/

[60] A. Fernández-Castillo and M. J. Caurcel, "State test-anxiety, selective attention and concentration in university students," *International Journal of Psychology*, vol. 50, no. 4, p. 265–271, 2015.

[61] D. J. Simons and C. F. Chabris, "Gorillas in our midst: Sustained inattentional blindness for dynamic events," *Perception*, vol. 28, no. 9, p. 1059–1074, Sep 1999.

[62] S. Herse, J. Vitale, B. Johnston, and M.-A. Williams, "Using trust to determine user decision making & task outcome during a human-agent collaborative task," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. Boulder CO USA: ACM, Mar 2021, p. 73–82. [Online]. Available: https://dl.acm.org/doi/10.1145/3434073.3444673