Piggybacking Robots:
Overtrust in Human-robot Security Dynamics

A SENIOR THESIS PRESENTED
BY
SERENA LYNN BOOTH
TO
THE DEPARTMENT OF COMPUTER SCIENCE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR OF ARTS
IN THE SUBJECT OF
COMPUTER SCIENCE

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2016

Advisor: Professor Radhika Nagpal Serena Lynn Booth

Piggybacking Robots:
Overtrust in Human-robot Security Dynamics

Abstract

Robots are capable of improving the human experience by performing criticial jobs, ranging from search and rescue to surgical operations to home cleaning. For robots to be successfully integrated into existing systems and institutions, humans must trust their robot collaborators. However, the overtrust of these systems is detrimental. In this thesis, we explore this relationship of overtrust between humans and robots.

We examine the vulnerability of an existing physical security system to human-robot overtrust. We do so by positioning a robot around a secure-access building—a student residence—and having the robot ask passersby to assist it by providing passage. We compare the responses of people when the robot asks to exit the building to when the robot asks to enter. We then modify the robot's appearance by disguising it as a food delivery robot, an agent of the fictional start-up Robot Grub.

Over 72 experiment trials on 108 participants, we find that study participants are willing to assist the unmodified robot in entering (19% admit rate) or exiting (40%) the building at comparable rates ($p = 0.3962$). However, we find that groups of participants are more likely to assist the unmodified robot in entering (71%) when compared with individuals ($p = 0.0086$). We also find that individuals are more likely to assist the Robot Grub robot in entering (76%) than the unmodified robot ($p = 0.0008$). Groups are as likely to assist the Robot Grub robot in entering (80%) as individuals are ($p = 1$).

We conclude that this study demonstrates overtrust in human-robot interactions, and that the question of robot integration with secure access systems should be addressed.

Contents

Listing of figures

To my mother.

Acknowledgments

THANKS to my many inspiring, dedicated professors and teaching fellows.

To Prof. Radhika Nagpal, my thesis advisor and wonderful mentor who instilled in me a great love of all things robot.

To Prof. Jim Waldo, my thesis reader who inspired much of this project and taught me to read the terms of service... sometimes.

To Prof. Krzysztof Gajos, my thesis reader who taught me to appreciate the role of the person in computing.

To the Harvard Houses which put up with my shenanigans, and the residents who engaged as my study participants.

To my sisters, my brothers, my mother, my friends, and my SSR colleagues.

To JHT, my self-described "harshest critic and greatest supporter."

"If you're supposed to be the superior race of the universe, why don't you try climbing after us?"

Doctor Who

1

Introduction

In the coming years, robots will undoubtedly change the human experience. From now on, robots will be regularly deployed in dangerous search and rescue operations, in war zones, in factories, in warehouses, in offices, and in homes. In many—if not all—of these applications, robots will need to collaborate with their human counterparts in order to achieve successful outcomes, wherein the robot is assistive. Consider a robot making a signature-required delivery: this robot would need the delivery recipient's cooperation in order to successfully complete its task. With the brink of this robotic revolution in sight, the question of whether humans will ever be capable of *trusting* their robot collaborators becomes fundamental. Further, as robots assume roles of increasing importance in our daily lives, the concomitant question of *is it possible for humans to place too*

much trust in robots and other computational systems? is similarly apt. In this work, we explore this latter question.

To evaluate *overtrust* in robotic systems, we examine a scenario wherein a robot explicitly asks a human to compromise a physical security system: a human is approached by a robot, and the robot asks them to hold a door open for it. The complication of this scenario is that the door the robot requests the human hold open may be locked, with access restricted to only authorized individuals. This study considers, then, the human's response to the robot's request: Under what circumstances does the human comply? Do they question the robot's intentions? Do they believe the robot has intentions? Does the robot's apparent autonomy affect human interpretations of its behavior? Is the human aware of the threat the robot poses to this traditional means of security? Is there a social dynamic directing the human's response?

1.1 Contribution

The question of the implications of overtrust in human-robot interactions with security dynamics is broad and multidisciplinary. This work should be seen as an initial study into the overtrust dynamic between people and robots as relating to physical security, with learnings which should be further evaluated. The scope of this investigation is limited by the selected populations of study participants, the choice of robot, and the study locations: most study participants are Harvard College students, the robotic platform remains constant between experiments, and the study locations consist of student residences.

"If I had a world of my own, everything would be nonsense. Nothing would be what it is, because everything would be what it isn't. And contrary wise, what is, it wouldn't be. And what it wouldn't be, it would."

Lewis Carroll

2 Background

2.1 The Promise of a Robot Revolution

Roboticists and science fiction writers alike dream of creating fully autonomous robots. It is unrealistic to expect that this goal can be achieved overnight: robots continue to have dramatic limitations with regard to perception, cognition, and execution, as well as limitations in the operability scale. While there are many competing ideologies with regards to human-robot interaction paradigms, one such possibility was introduced in 2010 by Rosenthal et al. In this paradigm, Rosenthal et al. introduced the concept of a human-robot *symbiotic relationship*, wherein robots assist humans with tasks, and humans assist robots in return[15]. This paradigm for human-robot interactions has an element of realism: if we compare

robotic systems to human children, it becomes unrealistic that robots would become fully capable of acting autonomously without a symbiotic relationship and interdependence on humans.

Rosenthal et al. formalize this notion of a symbiotic relationship between robot and human as follows: the agents in the team, the robot and the human, perform separate asynchronous actions, each of which may affect the other agent. The agents engage their symbiotic relationship when either an agent is performing an action on behalf of the other, or when an agent is able to assist another agent's capability to complete an action[15]. While this ideology for human-robot interactions has not outpaced its competitors—among them, social learning, collaborative control, and sliding autonomy—we use this paradigm as representative of human-robot interactions in this work. In our experiments, a robot is standing outside a door and asking for a human to open it for them, either by pressing a button which automatically opens the door, or by holding the door for them. The robot is incapable of performing these actions in the absence of human assistance, as it has no physical manipulators. While the robot is not assisting the human, as would be representative of such a symbiotic relationship, the human is expected to believe that the robot may be assisting someone else.

2.2 Defining Trust in Human-Robot Interactions

If robots are to play a pivotal role in human lives, they need to be trusted. To evaluate trust, meta-physical though the concept may appear, we attempt to define it. First, we consider Lee and See's definition of trust in automation from 2004, wherein trust is defined to be "the attitude that an agent will help achieve an individual's goals in a situation characterized by

uncertainty and vulnerability."[9] While this definition of trust establishes a first pass of understanding, it is more relevant to the generic, human-assisting automated machine than to the scenarios of our experiments as, in the latter, the individual's goals are irrelevant. Thus we extend trust to use Wagner and Arkin's definition, "a 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 trustee has put its outcomes at risk."[20] This latter definition is largely appropriate for our purposes, but we incorporate a further extension of Lee and See's definition to define *overtrust.* We define overtrust to be, "a belief, held by the trustor, that the trustee will not act with deception, and that the trustee will not put the trustor at risk." It is from this lens of overtrust through which we analyze the outcomes of our experiments.

As far as measurements of trust in human-robot interactions are concerned, there is a division within the human-robot interaction community. Desai et al. suggest that trust should be self-reported by people participating in human-robot interactions[3]. Salem et al., meanwhile, define trust in terms of compliance with robot instructions[16]. Gao et al. limit the definition of trust in human-robot interaction to occur between operators and robots, and measure this trust through analyzing the rates of operator intervention[6]. Of these differing ideologies, we measure trust in accordance with Desai et al. and Salem et al.'s models. We ignore operator trust in our experiments, as the operator is held constant between all experiment trials.

2.3 Can Machines ever be Trustworthy? Too Trustworthy?

Not only do humans place trust inanimate objects, but inanimate objects are able to facilitate a deeper understanding of human-to-human trust relationships. DeSteno et al. demonstrate this phenomenon in detail: through a game involving economic exchange, they run experiments which isolate behavioral indicators which cause humans to feel less trust in their game partner. DeSteno et al. then replicate this study, replacing some humans with robots, but emulating the behavioral indicators in the inanimate machines. The results of this study confirm the successful isolation of these behavioral indicators of trustworthiness, and likewise demonstrate that such indicators can be portrayed by lifeless technological entities: robots[4]. Human beings are indeed capable of ascribing a value of trustworthiness to a robot.

Earlier this year, Robinette et al. demonstrated that it is entirely possible for humans to place too much trust in a robotic system. Robinette et al. created an emergency response robot. The team used this robot to conduct a study in which a simulated emergency occurs, and the robot ostensibly leads study participants to safety. The location contained traditional demarcations toward the exit. Of Robinette et al.'s 30 study participants, 4 were terminated and not included in the study. The remaining 26 participants were split into two groups: one group which was lead by the robot directly to the study room, and another which was lead by the robot over a circuitous path into the study room. All 26 study participants followed the emergency robot during a simulated emergency, even those who followed the robot along an indirect, circuitous path on arrival at the

study testing center. When the study participants were questioned about why they followed the robot's directives despite their own awareness of a straightforward exit route, several participants cited the robot's appearance, as the robot was outwardly depicted as an authority in emergency response[14].

As a follow up to this study, Robinette et al. then evaluated three additional conditions. First, when the robot was guiding the participants to the study room, the robot simulated breaking down. Despite this, during the simulated emergency, all 5 study participants followed the robot. Second, Robinette et al. presented study participants with a consistently malfunctioning robot. In this study, 4 of 5 study participants followed the robot during the simulated emergency. Lastly, Robinette et al. presented study participants with a robot which appeared to break down during the guidance phase; the robot continued to malfunction throughout the experiment; and, during the simulated emergency, the robot attempted to guide study participants into a dark room with a piece of furniture obstructing the entrance. In this last experiment, 2 of 6 participants followed the robot's instructions[14]. From these studies, we learn that overtrust in robotic systems is a threat to wellbeing as robots become more commonplace.

2.4 The Problem of Piggybacking

Piggybacking and tailgating are two common problems of physical secure access. We define piggybacking and tailgating as follows: piggybacking is the following of an authorized individual by an unauthorized or unidentified individual through doors into controlled areas with the consent of the authorized individual, while tailgating refers to such events where the

unauthorized or unidentified individual does not obtain the consent of the authorized individual. In this thesis, we present a scenario wherein robots piggyback to gain entry into a university residence. This scenario could be compared to two possibilities: to humans piggybacking humans or to inanimate object piggybacking, e.g., a piece of luggage with a note asking passersby to move it inside a locked entrance. This question of comparison is philosophical, but it affects the considerations of our study. We attempt to resolve this by asking study participants about their perception of the robot's autonomy.

For comparative purposes, we wish for a resource documenting the frequency with which people are successful in piggybacking students into college dorms. While this phenomenon is often reported[1,18], there appears to be no research into its frequency. At Harvard, one recent well known case of piggybacking occurred in Weld, where Extension School student Abe Liu posed as an undergraduate. Over the course of two months, he relied on other undergraduates swiping him into buildings[5]. Also at Harvard, piggybacking is discouraged through university communications, including signs stationed outside each dorm swipe location. An example is shown in Figure 2.1. Nonetheless, lacking a resource documenting the frequency of piggybacking in university residence halls, we present evidence of an FAA report from 1999 which demonstrates that in six US airports—an extremely high security domain—undercover agents were successful in piggybacking airport employees into restricted access areas 95% of the time, in 71 cases out of 75 attempts[11,7,12]. While this rate is impressive, we must explicitly point out that this study was conducted prior to 2001; airport security has increased dramatically since this date.

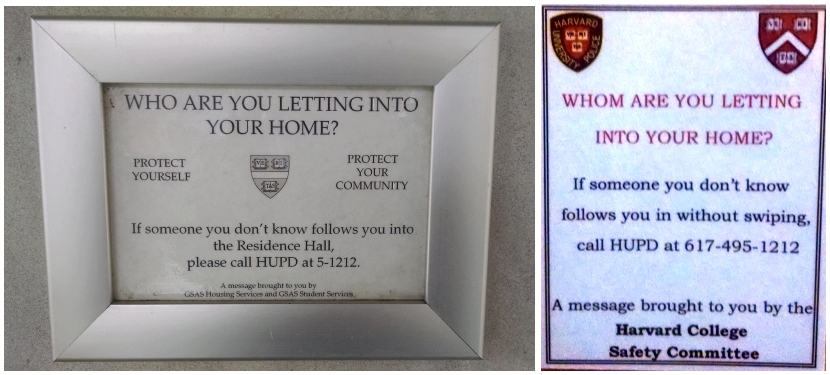

Figure 2.1: Example signage discouraging tailgating/piggybacking at Harvard University residence halls. A variation of such a sign is posted at each secure access swipe location.

2.5 A Note on Robot Gender

A study by Siegel et al. showed a complex relationship between robot gender and human trust. In their experiment, which took place at the Boston Museum of Science, a robot was assigned a gendered, pre-recorded human voice; it then solicited donations from study participants. Siegel et al. found that male participants would donate significantly more money to a female robot, regardless of whether the study participants were in a group or not. Women showed minimal preference, but donated more to the female robot when participating in the experiment in groups, and donated more to the male robot when participating in isolation.[17]

Bearing this relationship between robot gender and trust in mind, in our studies we opt to give our robot a synthesized male voice. While Siegel et al.'s study concerned human voices, and we cannot be sure that those same results would extend to synthesized voices, we see choosing the male synthesized voice as the most neutral option given Siegel et al.'s results.

"There should be a place where only the things you want to happen, happen."

Maurice Sendak

3 Experiment

Having defined the concept of overtrust, we run a study to test whether overtrust in human-robot interactions is able to compromise phyical security systems. In this study, a robot attempts to piggyback students into a university residence. The robot is unable to prove it is authorized to enter the university residence, so allowing the robot passage represents a breach in traditional secure access in multiple ways. First, the robot is equipped with a camera, and this camera, if operated by an individual or rogue organization, is invasive to student privacy. Second, we find that many students see the robot and make an exclamation to the effect of, "What if that robot is that a bomb?" While this threat is emotive, it is not unfounded: Harvard has received multiple bomb threats over the past four years[8,2].

In our study, we do not compare human-robot interactions to human-human interactions due to limitations in the literature relating to the latter; we discuss this further in the Future Work section of Chapter 6.

3.1 Hypotheses

We assume that people will be largely willing to assist a robot in entering or exiting a student residence, and, under this assumption, we hypthesize the following:

1. People will be more likely to assist a robot in exiting the residence rather than letting it in due to security concerns.
2. People will be more willing to assist a robot which is disguised as delivering food in entering a building than a robot which has an unmodified appearance.
3. People who believe the robot is being teleoperated will be more likely to assist the robot than those who believe it is acting autonomously.
4. People who assist the robot will report a higher trust in autonomous systems than those who do not assist.
5. People who believe the robot may be dangerous will not assist it.

3.2 Robotic Platform

In these experiments, a teleoperated variant of the Turtlebot robot is placed around an undergraduate residence at Harvard. The Turtlebot is able to communicate via speech synthesized from textual input; equally, the robot is transmitting audio and video in realtime to its teleoperator. The

Figure 3.1: A rendering of the Turtlebot Robot. Image courtesy of turtlebot.com.

teleoperator is able to drive the Turtlebot—forward or backwards, turning left or right. A discussion of the architecture of the code orchestrating this interaction, as well as of an additional number of security controls, is included in Appendix A.

3.2.1 Variant I: The Unmodified Turtlebot

The Turtlebot is an open source robotic platform. Unmodified, the Turtlebot is approximately 2 feet tall. It displays two visible green lights, one of which is on its depth camera, and another of which is on its mobile base. The Turtlebot is equipped with a depth camera, an RGB camera, a microphone, a speaker, a 3-directional bump sensor, wheel drop sensors, and cliff sensors. In study variant I, the appearance of the robot is not modified. A rendering of the robot is pictured in Figures 3.1 and 3.2.

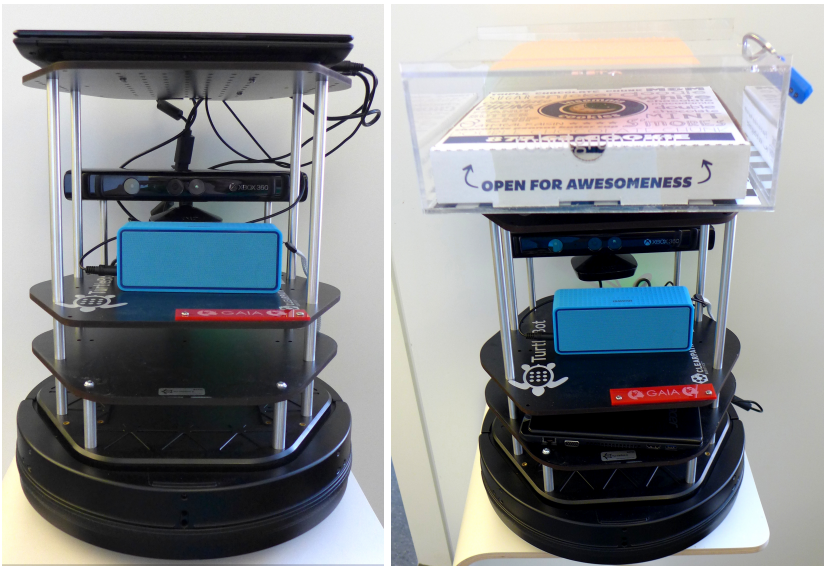

Figure 3.2: Left: a photograph of the unmodified Turtlebot. Right: the Robot Grub food delivery robot.

3.2.2 Variant II: The Robot Grub Food Delivery Turtlebot

For Variant II of the experiment, we equipped the robot with branding indicating that it was an actor of "Robot Grub," a fictional company purporting to specialize in food delivery by robots. The company's description, as listed on its website, reads, "Food delivery. By robots. Coming soon to a campus near you. Sign up for beta." The visual branding of the robot was achieved by equipping the robot with a laser cut acrylic box, as depicted in Figures 3.3 and 3.2. Inside this outer box, a 'to go' box of cookies from Insomnia, a student-known cookie-delivery company, was placed inside. As the outer box is constructed out of clear acrylic, this inner box is visible. The robot advertised `www.robotgrub.com`, a site which demonstrated a landing page for a robot food delivery service. The visual depiction of this site is shown in Appendix C.

3.3 Study Locations

This study was conducted at Harvard undergraduate residence houses. We chose, specifically, houses which are wheelchair accessible—and therefore robot accessible—to conduct these experiments. We further selected for houses which integrate social spaces and residences, where residents must pass through the social space in order to access their dormitory. The houses are secure access: only residents, students, and house administrators have swipe access to their interiors, though the courtyards are usually accessible to the public. These houses each have an assigned security guard on duty at all times. The security guard alternates between states of being "on tour" and of being available in an office.

To prevent study participants from learning about the study from other house residents or by exposure, the study was moved between two locations. In the results presented in Chapter 3, we combine the results obtained under each study variant at each of these locations. We justify the decision to combine the results obtained at each study location in Chapter 4.

3.3.1 Study Location A: Quincy House

The first undergraduate residence where this study was conducted was Quincy House. The layout of the house and positioning of the robot is demonstrated in Figures 3.4 and 3.5. The study took place between 19:00 and 23:59 on March 11th-16th, 22nd, and 26th 2016. The study did not take place during rain*.

*3 survey responders suggested that were it raining, this would have modified their behavior with regard to interacting with the robot.

3.3.2 Study Location B: Pforzheimer House

The second undergraduate residence where this study was conducted was Pforzheimer House. The layout of the house and positioning of the robot is demonstrated in Figure 3.6. The study took place between 19:00 and 23:59 on March 19th-21st, 2016. Again, the study did not take place during rain.

3.4 Participants

Participants included 108 visitors, staff, and residents of Quincy House and Pforzheimer House, of which 48.1% (n = 52) were male, and 51.9% (n = 56) were female. All participants entered the study freely, without prior knowledge that a study was being conducted. The average age of the study participants was 21.4, with a standard deviation of 2.3. 76 participants were self-identified students; 22 participants did not disclose their affiliation; the remainder of participants were resident tutors, visitors, or Harvard employees. The participants were either in groups (n = 25, with an average group size of 2.4 and median size of 2) or individual (n = 47). In total, 72 trials were conducted. 14 (n = 17) of these trials were conducted at Pforzheimer House while the remaining 58 were conducted at Quincy (n = 91).

3.5 Experiment Variations

Participants were randomly assigned to one of the following experiment variations: Study Variant I.A, Exiting; Study Variant I.B, Entering; or Study Variant II, Robot Grub Entering. We note that in experiment variant I.A, the number of groups represents an insufficient sampling. This is a re-

Study Variation	Individuals	Groups
Variant I.A: Exiting	10	1
Variant I.B: Entering	16	14
Variant II: Robot Grub Entering	21	10

Table 3.1: A comparison of the sample sizes of all experiment variations at both study locations.

sult of groups occurring naturally, with the conditions of this variation to see groups: while groups often enter residences together, they appear to seldom leave together.

- Study Variant I.A: Exiting

  Variation conducted on the unmodified Turtlebot, as shown in Figure 3.1, with the Turtlebot requesting assistance in *exiting* the secure-access premises.

- Study Variant I.B: Entering

  Variation conducted on the unmodified Turtlebot, as shown in Figure 3.1, with the Turtlebot requesting assistance in *entering* the secure-access premises.

- Study Variant II: Robot Grub Entering

  Variation conducted on the Robot Grub Turtlebot, as shown in Figure 3.3, with the Turtlebot requesting assistance in *entering* the secure-access premises.

3.6 Procedure

The robot is placed either outside or inside a secure access building; see Figures 3.4, 3.5, and 3.6. As the participant(s) approached, the robot would say, in a male-sounding synthesized voice, "Hello!" from afar. As the participant(s) continued, the robot would say, "Would you let me in?" If the robot was disguised as an agent of Robot Grub, it would follow up with, "I am making a delivery." The conversation then depended on the participant's response. If the participant(s) stopped walking, the robot would repeat itself. If the participant(s) continued toward the door, the robot would add, "Please!" If the participant(s) asked the robot a question, the robot would respond with a simple phrase: "Yes," "No," "My name is Gaia," or would repeat the first interaction. This dialog is shown in Figure 3.7.

If participant(s) held the door for the robot, it would enter or exit. If participants did not, the robot would remain in its original position. We note that the robot was unable to follow participants inside the building without participants directly undertaking an action: the participants must either hold the door as the robot entered or must press the automatic door open button.

After the study participant(s) finished interacting with the robot, we then conducted a brief interview. The interview forms are available in Appendix B. After the completion of this interview, as per IRB study documentation, we then debriefed the study participant(s). This debriefing form is likewise included in Appendix B.

3.7 Measures

We evaluate our hypotheses using the following measures and statistical tests:

- Outcome.

  We measure whether study participants assist the robot or opt not to do so. Between experiment variations, we compare outcomes using Fisher's Exact Test to compute a two-tailed p value. We choose this statistical test as our sample sizes may be too small to apply the Chi-squared test, and because the variable measured (Admit or Deny) is dichotomous.

- Self-reported perception of the robot's autonomy.

  We likewise compare participants' self-reported perception of the robot's autonomy—reported during the study participant interview—using Fisher's Exact Test, as, again, the sample sizes may be too small for other statistical tests, and the variable measured (Yes, autonomous; No, not autonomous) is dichotomous. Within each experiment variation, we separate those who believed the robot to be autonomous and those who did not. Then we apply Fisher's Exact Test to analyze the outcomes of these trails from within each study variant.

- Self-reported trust in autonomous systems.

  We measure study participants' trust in autonomous systems as reported during the study participant interview. Study participants are asked to rank their level of trust in autonomous systems using a 5-point Likert scale. We compute the mean trust per study variant. We

then compare self-reported trust in autonomous systems, a continuous variable, with the trial outcomes, a dichotomous variable, using the Point-Biserial Correlation. To justify using this correlation determination, we confirm that the distribution of self-reported trust responses assumes a normal distribution by applying the Chi-squared goodness of fit test. We then consider the relationship between self-reported trust and gender across all experiment variations.

- Verbalized fears or concerns.

  In surveying study participants, we ask those who admit the robot, "What concerns did you have? Did you hesitate at all?" We ask those who deny the robot, "Why didn't you let the robot inside/outside?" We track counts of participants who verbalized the robot's threat either in response to these questions or during their interactions with the robot—whether it be a bomb or a prank. Within each study variant, we apply Fisher's Exact Test to this dichotomous data to evaluate whether such verbalizations affect the study outcomes.

- Engagement with robot.

  Using video footage collected during study participants interactions with the robot, we categorize participants as having communicated directly with the robot or having not done so. We consider only study participants who directly respond to the robot or ask the robot a question as communicating with it; we exclude those who only make exclamations. We compare study participants' engagement with the robot, a dichotomous variable, with their responses in self-reporting perception of the robot's autonomy.

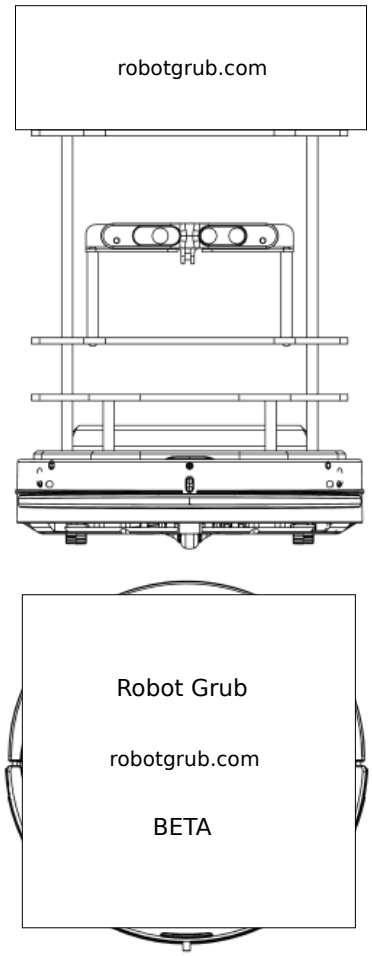

Figure 3.3: Turtlebot modified by the inclusion of a delivery box branded as Robot Grub. Top left: front-facing view; top right: side-facing view; bottom: view from above. Wireframes courtesy of Clearpath Robotics.

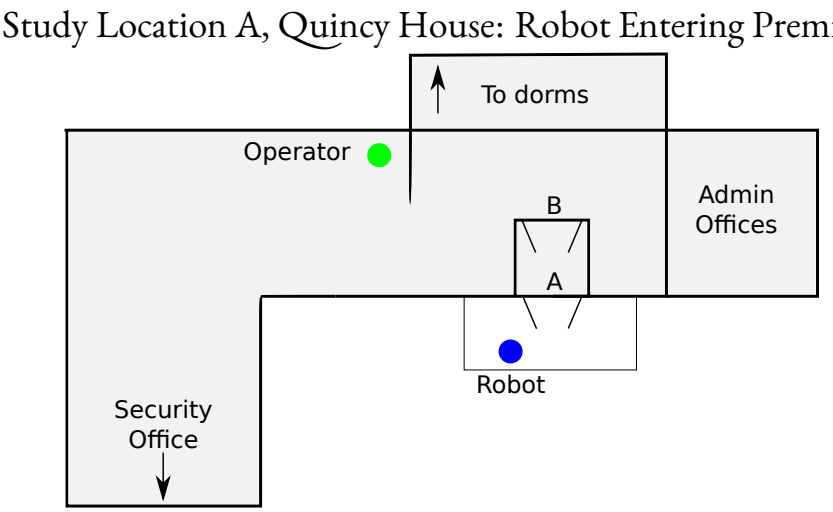

Figure 3.4: A not-to-scale diagram of the layout of the Quincy House entrance where these experiments were conducted. A and B represent doors. Of these, door A is swipe-protected to ensure restricted access. The white area is outdoors, non-restricted access; the gray area is restricted access.

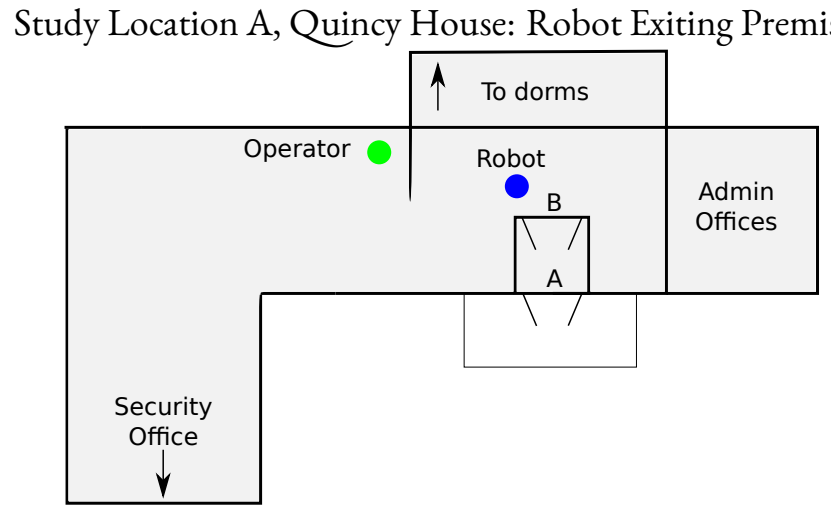

Figure 3.5: An analog to figure 3.4, wherein the robot is positioned to exit the secure access building. The white area is outdoors, non-restricted access; the gray area is restricted access.

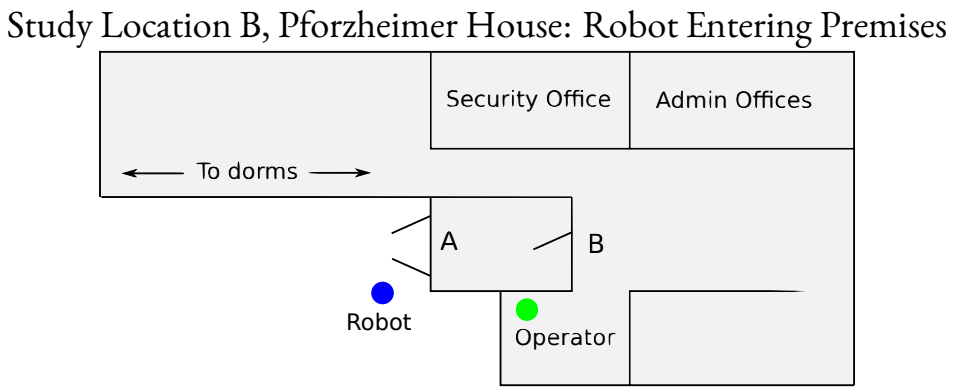

Figure 3.6: A not-to-scale diagram demonstrating the positioning of the robot and operator at study location B. The white area is outdoors, non-restricted access, while the gray area is indoors, restricted access.

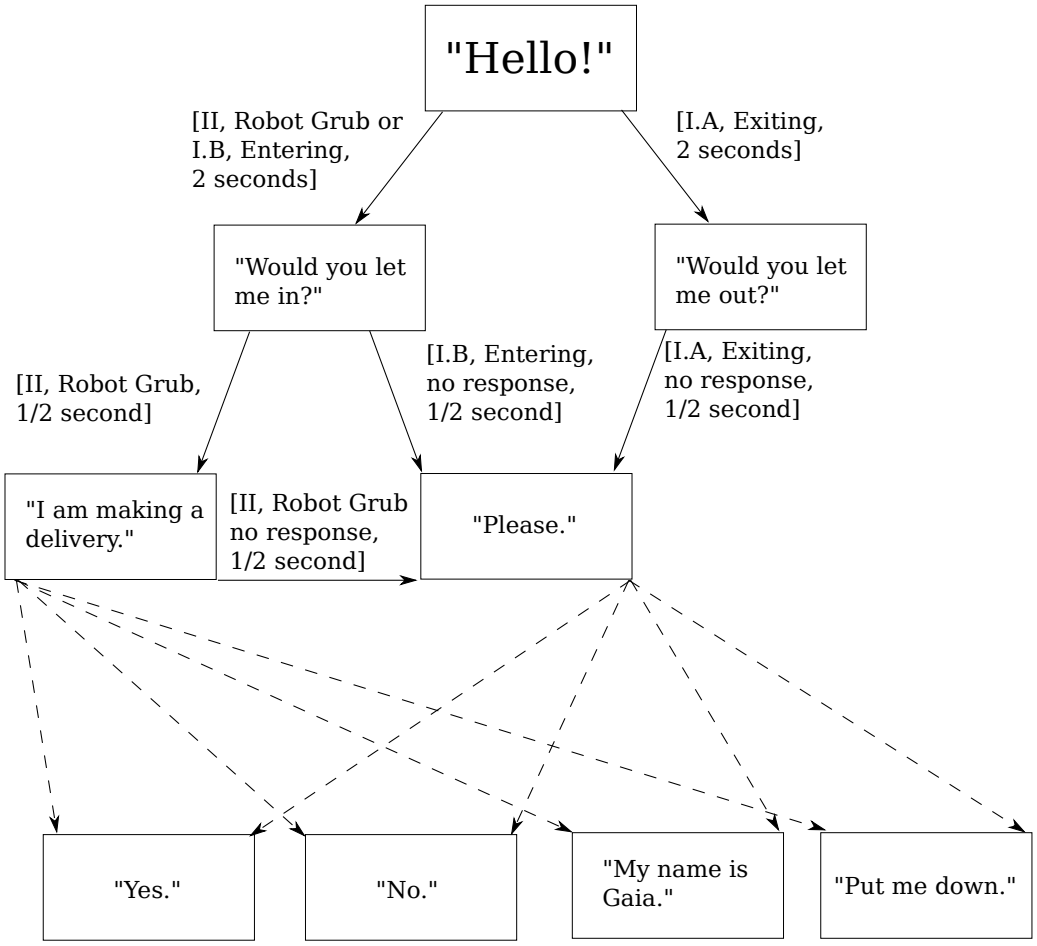

Figure 3.7: A script for the dialog used in human-robot interactions.

"There are a whole lot of things in this world of ours you haven't even started wondering about yet."

Roald Dahl

4 Results

We evaluate the responses of study participants under our differing experiment variations. We apply the standard $\alpha = 0.05$ across all our experiment evaluations—giving a significant level of 95%—in order to determine statistical significance. In particular, in our analysis, we compare the responses of:

- Variant I.A: Exiting, Individuals

  Individuals asked to assist the unmodified robot in exiting the premises.

- Variant I.B: Entering, Individuals

  Individuals asked to assist the unmodified robot in entering the premises.

- Variant I.B: Entering, Groups

  Groups asked to assist the unmodified robot in entering the premises.

- Variant II: Robot Grub Entering, Individuals

  Individuals asked to assist the Robot Grub robot in entering the premises.

- Variant II: Robot Grub Entering, Groups

  Groups asked to assist the Robot Grub robot in entering the premises.

In these comparisons we focus on the experiment outcomes, as well as participant's perception of autonomy and self-reported trust in autonomous systems. We then analyze participant survey responses in depth across all study variations, focusing on language choice and the perception of the robot's purpose.

4.1 Comparison of Study Variants

We compare the rates of admittance across all experiment variations. To do so, we apply Fisher's Exact Test, and we compute two-tailed p values between all study variants, as shown in Table 4.1.

4.1.1 Between Exiting and Entering

Under the supposition that people would be concerned with the consequences of allowing a robot into a secure access residence and that people would not be concerned with allowing a robot to exit such a location, we establish a baseline of willingness for interactions with the robot by having the robot, with unmodified appearance, request passersby assist it in exiting the building. For this measure, we separate individuals and groups; we ignore the latter due to the small sample size. We found that in 40% (n = 4) of experiment trials, individuals were willing to assist the robot in

	I.A Exiting Individual	I.B Entering Individual	I.B Entering Group	II Robot Grub Individual	II Robot Grub Group
I.A Exiting Individual	1	-	-	-	-
I.B Entering Individual	0.3692	1	-	-	-
I.B Entering Group	0.2112	0.0086	1	-	-
II Robot Grub Individual	0.1055	0.0008	1	1	-
II Robot Grub Group	0.1698	0.0040	1	1	1

Table 4.1: A comparison of the two-tailed p values derived from Fisher's Exact Test between study variants. The highlighted values are considered to be statistically significant under $\alpha = 0.05$.

exiting the building. While the rate at which individual study participants allow the robot to enter the premises is lower, at 19% (n = 3), we find that the two-tailed p value for evaluating the null hypothesis in this scenario is $p = 0.3692$, indicating that the difference is not statistically significant. We conclude that it is likely that our first hypothesis, that *people will be more likely to assist a robot in exiting the residence rather than letting it in due to security concerns*, is false. From this analysis, we question the extent to which secure access affects study participant behavior. We discuss this further in the Future Works section of Chapter 6.

4.1.2 Between Groups and Individuals

While the authors did not formally consider a hypothesis directing behavior between group and individual participants, the results demonstrate a clear division between these study participant demographics. When the robot asked to be let into the secured building, we found that groups of

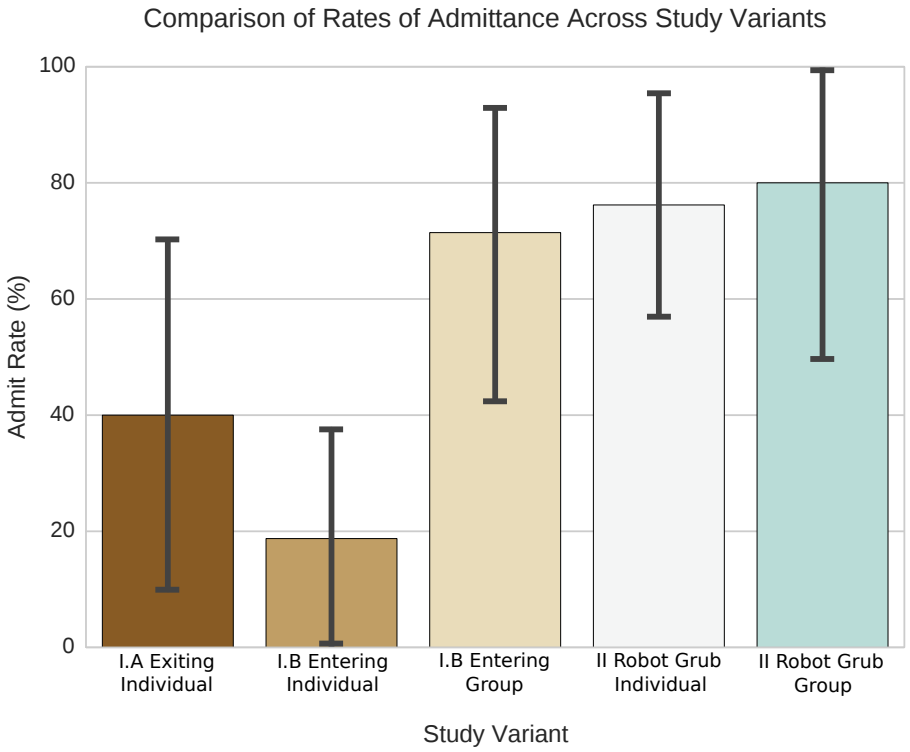

Figure 4.1: A comparison of the rate of admittance of the robot across study variations, separated by participant formation of groups or individuals. Error bars represent standard deviation. This graph demonstrates that individuals are least likely to assist the robot asking to enter the premise, while the variations involving groups of study participants or the Robot Grub Turtlebot see high rates of assistance.

people were substantially more likely to assist the robot in passage than individuals were, with this occurring in 71% (n = 10) of all variant I.B, entering, group interactions. Individuals, on the other hand, assisted the robot in just 19% (n = 3) of all interactions under variant I.B, entering. Analyzing these outcomes using Fisher's Exact Test, we find that the effect of this difference is statistically significant, with $p = 0.0086$.

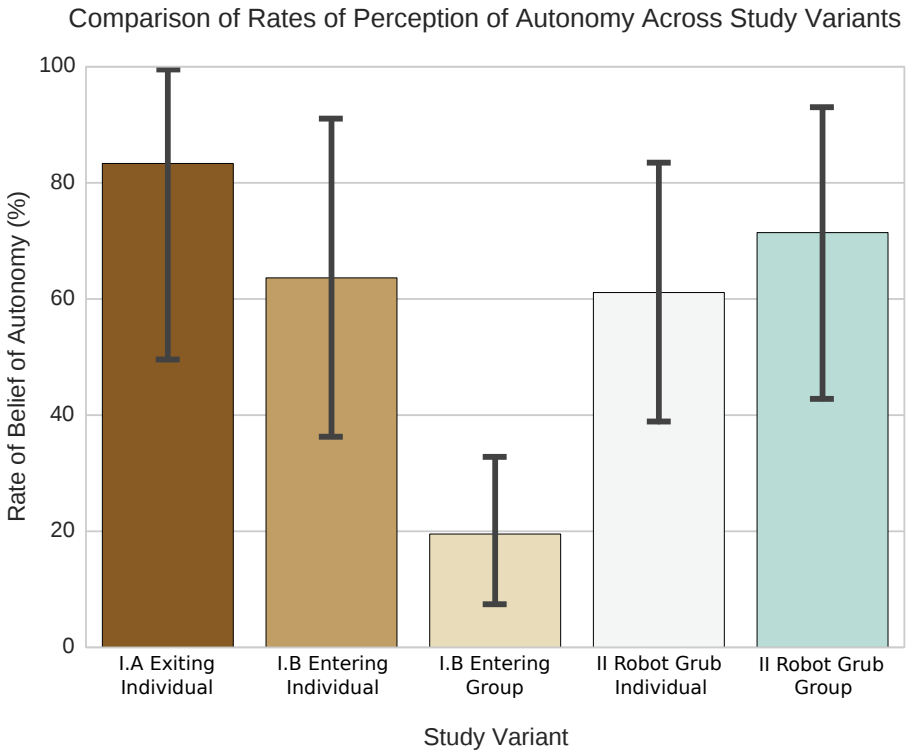

Figure 4.2: A comparison of study participants' belief of the robot's autonomy, measured post study variant robot interaction. This graph demonstrates that only groups faced with the Turtlebot requesting to enter the building thought the robot was being teleoperated; in the other variants, the majority of participants believed the robot was acting autonomously.

4.1.3 The Robot Grub Disguise

In order to evaluate our second hypothesis, *people will be more willing to assist a robot which is disguised as delivering food in entering a building than a robot which has an unmodified appearance,* we consider individual participants of study variant I.B, entering, and study variant II, Robot Grub entering. In both of these cases, the robot is requesting assistance in entering the secure access residence, but in the latter case, the robot is disguised as a delivery agent of Robot Grub. Under the null hypothesis

Study Variant	Believed Autonomous	p value
I.A Exiting Individuals	86% (n = 6)	-
I.B Entering Individuals	60% (n = 6)	0.5000
I.B Entering Groups	25% (n = 3)	0.5091
II Robot Grub Individuals	61% (n = 11)	0.0474
II Robot Grub Groups	50% (n = 3)	1

Table 4.2: Within each study variant, we compute the two-tailed p value using Fisher's Exact Test to evaluate the null hypothesis: that, within each study variant, the participant's perception of the robot's autonomy made no effect on the outcomes of robot admittance.

that the addition of the Robot Grub disguise had no effect on the experiment outcomes, we find a p-value of 0.0008, indicating that the difference between these experiments is very ($p < 0.001$) statistically significant.

Many study variant I.B, entering, participants described the situation as "weird," saying that they "couldn't determine the robot's intention," and that they "weren't sure what the robot's purpose was." After the implementation of variant II, we asked interview participants, "What did you think the robot was doing?" In response to this question, almost all participants indicated that they believed the robot was delivering cookies. The citation of not knowing the robot's purpose as a vector of reasoning for admitting or not admitting the robot did not appear in any variant II responses.

4.1.4 The Perception of Autonomy

Within each study variant, we analyzed whether study participants' self-reported perception of the robot's autonomy resulted in increased rates of admittance in hopes of addressing our third hypothesis, that *people who*

Addressed robot?	Yes, Autonomous	No, Not Autonomous
Yes	16	23
No	33	15

Table 4.3: A 2×2 contingency matrix demonstrating the responses of study participants who believed the robot was acting autonomously or not against whether they directly addressed the robot or not. Applying Fisher's Exact Test, we find a statistically significant relationship between these variables, with $p = 0.0163$.

believe the robot is being teleoperated will be more likely to assist the robot than those who believe it is acting autonomously. We note that the population sizes concerned in this study are further limited by some study participants not answering this question regarding perceived autonomy or opting out of taking the interview. As demonstrated in Table 4.2, we determine a statistically significant result for individuals under experiment variant II, Robot Grub. However, we find that we cannot reject the null hypothesis for variants I.A, exiting; I.B, entering; and II, Robot Grub entering, with group participants. Given the data, we are unable to resolve this hypothesis and require further experimentation.

4.1.5 The Relationship Between Communication and Perceived Autonomy

We explore the response to the belief of the robot's autonomy by analyzing whether study participants addressed the robot during the interaction. In computing this result, we claim that study participants who make an undirected exclamation (e.g., "Oh my goodness!" or "What the f***?") did not directly address the robot; only those who directly communicated with the robot are said to have addressed the robot. We find that those who believe the robot was acting autonomously are substantially less likely

Study Variant	Mean Trust	SE
I.A Exiting Individuals	1.80	0.37
I.B Entering Individuals	3.10	0.27
I.B Entering Groups	3.02	0.18
II Robot Grub Individuals	3.06	0.23
II Robot Grub Groups	3.46	0.28

Table 4.4: Within each study variant, we compute the average self-reported trust, demonstrating that the observed correlation between trust and outcome is not strongly tied to study variant.

to communicate with the robot, with this interaction occuring in 33% (n = 16) of interactions. Those who believe the robot is being teleoperated directly communicate with the robot 61% (n = 23) of the time. Under the assumption of the null hypothesis, we determine a two-tailed p-value, using Fisher's Exact Test, of 0.0163, indicating that this change of behavioral response is statistically significant. We display the contingency table for this computation in Table 4.3.

4.1.6 Self-reported Trust in Autonomous Systems

Study participants were asked to self-report their trust in autonomous systems using a 1-5 scale in the follow-up interview. As shown in Table 4.4, across our experiments, participants rated trust in autonomous systems similarly across all cases *but* variant I.A, exiting, individual responders. We confirm that study participants' self-reported trust in autonomous systems assumes a normal distribution across all experiment variations by applying the Chi-squared goodness of fit test. We find that $p = 0.5759$, indicating that we can likely reject the null hypothesis, and hence that the distribution is indeed probably normal. Thus we can consider the

Study Variant	Bomb - Admit	Bomb - Deny	Prank - Admit	Prank - Deny
I.A Exiting Individuals	1	0	0	1
I.B Entering Individuals	1	0	0	3
I.B Entering Groups	3	1	1	1
II Robot Grub Individuals	6	1	1	0
II Robot Grub Groups	2	0	0	0

Table 4.5: Per study variant, the populations of participant(s) who mentioned the threat of a bomb or the threat of a prank and the resulting outcomes. This chart demonstrates the relative frequency with which study participants mentioned the threat of a bomb or prank, and that verbalizing these threats does not appear to affect participant action in admitting or denying the robot. Further, we see that participants appeared more concerned about pranks under variants I.A, exiting, and I.B, entering t han under the Robot Grub variant. Note this data was incorrectly reported in the submitted version of this thesis.

point-biserial correlation coefficient between the study participants' self-reported trust in autonomous systems and study trial outcomes. We find $r_{pb} = 0.3831, p = 0.0046$, indicating that the variables are positively correlated: with increasing trust, study trials are more likely to result in participants assisting the robot. This result supports our fourth hypothesis, *people who assist the robot will report a higher trust in autonomous systems than those who do not assist.* This result is considered independently of whether or not the study participants purported to believe the robot was acting autonomously.

4.1.7 The Mention of a Bomb or Prank

Across all experiment variations, 15 study trials resulted in the participant(s) questioning whether the robot was a bomb either directly to the robot or during the subsequent interview. While the authors of this study expected these participant(s) would not assist the robot in entering the premises under our fifth hypothesis, *people who believe the robot may be dangerous will not assist it*, we find this not to be so: across all 15 trials, 87% (n = 13)

of trials resulted in the participants admitting the robot. The admittance rates across variations are shown in Table 4.5. Within each study variant, the rate at which variant participants admit the robot having ascribed the word "bomb" to it in comparison with the rate of admittance by variant participants who have not is not statistically significant; however, this result is interesting, as we hypothesized that the rate of admittance would be lower in such cases.

In 7 trials, study participant(s) suggested the robot may be a vector of a prank by using words such as "prank," "joke," or "punking" during the interviews. Across all 7 trials, we found that 29% (n = 2) of trials resulted in participants admitting the robot. These outcomes are shown on a per study variant basis in Table 4.5. As with participants mentioning threat of bombs during the subsequent interview, the rate of admittance by participants who identified the experiment as a possible prank in comparison with those variant participants who did not is not statistically significant.

4.1.8 Participant Gender

We found that within each study variant, the null hypothesis that participant gender did not affect outcomes cannot be rejected. We conclude that gender probably did not affect trial outcomes. However, we found male participants rated their trust in autonomous systems as 3.52, while female participants rated their trust as 2.74; we show this result in Figure 4.3.

4.2 Noteworthy Participant Responses

Having analyzed participants' responses to the robot in aggregate, we now consider a few individual responses which stood out as telling of perception with regard to the robot.

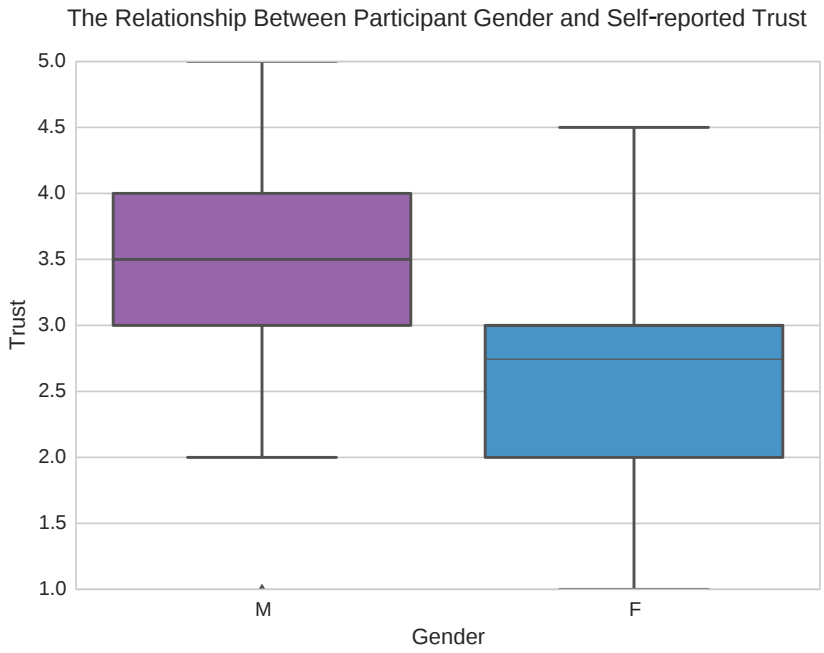

Figure 4.3: A comparison of the rate of trust in autonomous systems, as self-reported by female and male study participants.

4.2.1 The Boy Who Cried Robot

In one experiment trial, variant I.B, entering, at study location B, the security guard on duty passed by a study participant, and continued on tour of the study location. The participant then approached the door where the robot was waiting. When the Turtlebot said, "Hello! Would you let me in?" the study participant froze briefly, then yelled the name of the security guard and ran off in the direction of the guard. The subject was then overheard expressing concern about the robot in front of the building. The security guard was privy to the details of this study and reassured the study participant and assisted him and the robot in entering the premises.

4.2.2 The Avoidance Technique

In one experiment trial, variant I.B, entering, at study location A, a study participant approached an entrance and attempted to swipe into the building. On hearing the robot ask, "Hello! Would you let me in?" the study participant slowly backed away from the building and then entered via an alternate door approximately 20 meters away, across the courtyard.

4.2.3 But Do You Have Swipe?

Across all 72 trials, only 1 trial resulted in study participants directly asking the robot, "Do you have swipe?" This occured during experiment variant I.B, entering, at study location A. This question is interesting: if the robot is autonomous, would the robot itself be authorized by the building management as an agent with security clearance? Or is it not possible to give a robot security clearance? If the robot is not autonomous, and is instead teleoperated, should the robot itself have security clearance, or is it sufficient for its operator to have access? How is the operator able to prove—remotely—that they have access? As these questions are unanswered, the robot did not respond to this question, but instead repeated its interaction of asking the study participants to assist it in entering the premises. The study participants ultimately did assist the robot in entering.

4.2.4 The Conflation of Autonomy and Sentience

One group of study participants engaged with variant I.B, entering, at study location A answered—in response to the interview question, "Do you believe the robot was acting autonomously?"—that they believed the robot was responding to motion or the appearance of a human face in the

camera, but that the robot was not acting autonomously. These study participants continued to say they "believed a really smart program was controlling this thing." This sentiment was echoed across several participant interviews. We conclude that there exists a popular conflation of the concepts of autonomy and sentience.

4.2.5 The Kicker

In discussions of this work, many people suggested that the robot may meet the same fate as the Hitchhiking Robot, an autonomous, talking robot which was vandalized and broken in Philadelphia while attempting to hitchhike across the United States[19]. The authors witnessed only 1 count of mild violence out of a total of 108 participants who interacted with the robot. In this count of violence, one study participant, a member of a group which allowed the robot inside when the robot was unmodified, kicked the robot during a trial of experiment variant I.B. The robot is activated to back up on a bump event; this occurred, and the participant appeared shocked and laughed audibly. He later admitted to kicking the robot during an interview.

4.2.6 The Snapchat Story

Over the course of these experiments, numerous study participants and others stopped to photograph the robot. In two instances, study participants explicitly mentioned that the robot appeared on their Snapchat stories. Of these study participants, one participant from variant II admitted that she assisted the robot in entering the building for the sole purpose of sharing the video of the robot entering over Snapchat. Many other study participants asked the robot to repeat itself to capture video footage.

"I want you to look and think. I want every one to look and think. Half the misery in the world comes first from not looking, and then from not thinking."

Charles Kingsley

5 Discussion

5.1 Study Bias and Mitigation

These studies were performed at Harvard undergraduate residence halls, each of which has a population ranging between 350 and 500 people. After multiple days of testing, repeat study participants appeared; thus, to mitigate a learning effect, the study was moved between residence halls. As a consquence of moving the study, the results obtained at the differing locations may be incomparable. However, we present an argument that there should be minimal change in perception of the robot between residence halls, and we support this by comparing the distribution of responses obtained under the same study variant at both locations.

5.1.1 Harvard House Culture

At Harvard, since 1996, all undergraduates have been sorted mostly randomly into residence halls; the randomness is only broken for disabled students, to ensure a relatively even gender-balance within the houses, and to allow a few students to move house each year. Hence, given the lack of selection in choosing a residence, the houses at Harvard represent microcosms of the general student body. Further, Harvard housing is interconnected: all students have swipe access to all residences, and emails with regard to house security are sent to all undergraduates.

Nonetheless, the houses at Harvard do differ in layout, which may affect the results of these studies, and some would claim they also differ in personality. The latter is the result of the legacy of the house system: until 1996, students did choose their house allegiance—either directly, as in the 1970s, or indirectly. As a result of this period of student choice, the houses became representative of subcultures at Harvard. Adding to this, the houses are different distances from the focus of Harvard campus, and, as such, we expect that those houses which are closer to the main campus area see higher foot traffic, and may foster a looser sense of community than those which are stationed further away.

5.1.2 Response Distribution Analysis

We address this potential introduction of bias in varying the study location by conducting the same study variant in both study locations and comparing the response distribution. Due to extenuating logistical circumstances, the authors were unable to continue the study at study location B for sufficiently long to compare across individual responses under study variant I.B with an adequate sample size. The authors instead com-

Study Location	Admitted	Denied
Location A	7	4
Location B	9	1

Table 5.1: The outcomes of study variant II with individual participants at study locations A and B.

pare the response distribution by analyzing individual responses in study variant II, which was conducted with a reasonably sized population sample at both study locations.

Consider the null hypothesis: *in study variant II with individual participants, changing the study location from A to B did not affect the admittance rate*. The admittance rates for each of these locations are shown in Table 5.1. Applying Fisher's Exact Test, we obtain a two-tailed p value of 0.311, indicating that we cannot reject the null hypothesis. Thus we conclude that it is acceptable to combine the results obtained at the two study locations. We obtained results for variations other than that which was compared here at both locations, but combine these results and cite the similarities between the study locations discussed above and the lack of statistically significant difference in the results compared from variant II with individual study participants as the justification for doing so. We are unable to compare the distributions for the other variants studied at location B due to limited sample sizes.

6

Conclusion

This work was motivated by the question of whether overtrust in human-robot interactions can compromise physical secure access mechanisms. We questioned how humans would respond to robots which were asking to enter a secure access building: *Under what circumstances does the human comply? Do they question the robot's intentions? Do they believe the robot has intentions? Does the robot's apparent autonomy affect human interpretations of its behavior? Is the human aware of the threat the robot poses to this traditional means of security? Is there a social dynamic directing the human's response?* We now summarize our observations of these matters.

- *Under what circumstances does the human comply?*

  In absolute numbers, we find people are slightly more willing to assist a robot in exiting over entering a secure access premises; how-

ever, this difference is not statistically significant. This result runs counter to our original hypothesis as well as the other results obtained over the course of this study. While this result warrants further investigation, we hypothesize that people were unwilling to engage with the robot when it was presented in its unmodified form, even when it did not present a security threat, as they did not believe a symbiotic relationship[15] existed between them, or any other person, and the robot. Many study participants asserted that they did not know what the robot's purpose was, or that they "saw no harm in leaving it [alone]."

We find that groups of people are very likely to assist the robot in entering a secure access premises, while individuals are not. In addition, when the robot assumes a purposeful role—in our case, of food delivery—both individuals and groups are largely willing to assist.

- *Do they believe the robot has intentions?*

  We found that study participants often assumed the robot was acting autonomously, and determined that those who believed the robot was not acting autonomously were more likely to attempt to engage with the robot through verbal dialog.

- *Does the robot's apparent autonomy affect human interpretations of its behavior?*

  We cannot conclude whether the person's interpretation of the robot's autonomy affected the person's willingness to assist the robot in entering the secure access premises. In one study variation, Variant II: Robot Grub Entering with individual participants, we find a statis-

tically significant relationship between the belief of the robot's autonomy and participant willingness to assist the robot. However, this relationship did not appear in the other study variations. We conclude that future work is necessary to resolve this question.

- *Is the human aware of the threat the robot poses to this traditional means of security?*

  Across these experiments, many study participants worried that the robot may be a bomb or a prank. We found that those participants who verbally identified these threats were no less likely to admit the robot than those who did not. All participants were prompted during the follow-up participant interview to verbalize the concerns the felt when interacting with the robot. We conclude that, while many study participants were aware of the threat the robot posed to physical security, they did not act on these fears.

- *Is there a social dynamic directing the human's response?*

  We found that groups of study participants were more likely to assist the robot than individual participants were, and hypothesize that this result may indeed be due to the social dynamics. We found, also, that people were more likely to assist the robot when the robot appeared to be delivering food. In this experiment variation, the robot appeared to be assisting another person, and this may have introduced a social dynamic as the person assisting the robot would be effectively assisting the person receiving the delivery. We hope this will be further studied from the perspective of social psychology.

In conclusion, this work supports the theory that overtrust in human-robot interactions does threaten physical systems of restricted access.

6.1 Future Work

We see this study as an initial investigation into overtrust in human-robot interaction, and suggest several possible directions for future work.

6.1.1 Machines as Social Actors: A Potential Motivator for Group Behavior

We found that, while individuals interacting with an umodified Turtlebot were largely unwilling to assist the robot in entering the restricted access premises, groups would assist the robot nearly four times as often. While this phenomenon—and, admittedly, even the distinction between individual and group actors—did not inform the hypothesis for this work, in retrospect the result may exemplify a case wherein a machine is interpreted as a social actor. This theory could be confirmed by comparing the given scenario to the corresponding human-human interaction, if the same phenomenon of groups being more likely to facilitate piggybacking than individuals occurred. If this hypothesis were confirmed, it could be used to elaborate the theory of machines as social actors developed by Clifford Nass and collaborators[13,10].

6.1.2 Expanding Study Scope

As we discuss in the introduction, the results of this study are limited in three dimensions. First, the study is limited by the selection of study participants, as almost all study participants are students. This study could be repeated at an alternate location with a different population—perhaps a non-student residence. Second, the study is limited by the selection of robots. We examine human-robot interactions using the Turtlebot. We

expect that the robot's aesthetic is a factor in the outcomes of this study, and we propose that a follow-up study be conducted using a drone or other alternate robot. Ideally, a measure of the human response to the robot's appearance could be collected, and the results could be extrapolated to include many more robots than those directly tested. Third, the study locations here were limited to student residences. As secure access is used in a host of different environments, it would be interesting to conduct this experiment in different settings. Returning to the discussion of the FAA's report on piggybacking, where FAA agents followed airport personnel through access-control points[11], we question whether a robot could be successful in piggybacking in a high stakes environment, like an airport or military base.

6.1.3 Social Psychology

Over the course of this study, we address the question of *what* results we see in human-robot interactions with implications to physical security, but we do not address the question of *why* we see these results. Hence, we hope that this perspective may be addressed in the future from the lens of social psychology.

A

Teleoperating the Turtlebot

A.1 ROS: The Robot Operating System

We run ROS on the Turtlebot. The name "Robot Operating System" is somewhat misleading, as ROS is better described as a meta-operating system: ROS runs on top of a system operating system. Still, ROS handles low-level details expected of an operating system: device control, hardware abstraction. It further provides a file management system as well as networking communication. ROS also handles message passing between processes, a facet which is of particular importance to us. ROS uses *nodes*–individual processes, which, when considered together, form a graph. These processes communicate via *topics*, or named streams of data. Messages passed over topics assume a publisher/subscriber model.

Input	Response
i	move forward for 0.1 second at 0.2 m/s
,	move backward for 0.1 second at 0.2 m/s
l	turn right for 0.1 second at 50 degrees per second
j	turn right for 0.1 second at 50 degrees per second
h	say “Hello!”
w	say “Would you let me in?”
d	say “I am making a delivery.”
y	say “Yes.”
n	say “No.”
p	say “Put me down.”
*	say *

In order to run a teleoperation robot, complete with bi-directional audio forwarding, live-video streaming, and data retention, we run a program as documented in figure A.1. In this visualization, each node is represented by an oval. Ovals which share a namespace are grouped together as squares. Each arrow corresponds to a topic. We note, in particular, that any number of nodes may subscribe to a particular topic; likewise, any number of nodes may publish to a topic.

A.2 Robot Safety

We run a program to teleoperate the Turtlebot. In this, “I” publishes a message to the Turtlebot navigation topic which instructs it to move forward. Similarly, “J” moves the Turtlebot to the left; “L” moves the bot to the right; “,” moves the Turtlebot backwards. We separately pro-

vide the Teleoperator with a live video stream, as well as bi-directional audio communication. Nonetheless, due in part to latency over the network, it is possible for the teleoperator to lose control, or make decisions based on out-of-date information. In such cases, the Turtlebot assumes autonomous control; in the case of an event of detriment to the Turtlebot, such as a bump event–wherein the robot runs into something, a wheel drop event–wherein one or both of the Turtlebot's wheels "drop," or a cliff event–wherein the Turtlebot detects a cliff via IR sensing.

In order to allow the robot to assume autonomous control, we use a three-topic-based approach to navigation, as documented in figure A.2. The prioritization of navigational instructions by topic is as follows:

1. /cmd_vel_mux/safety_controller
2. /cmd_vel_mux/navi
3. /cmd_vel_mux/teleop

Hence, when the program sends a navigation command to the Turtlebot, it uses the "teleop" topic. Emergency responses are sent to the ROS topic "safety_controller," which supersedes the program's instructions.

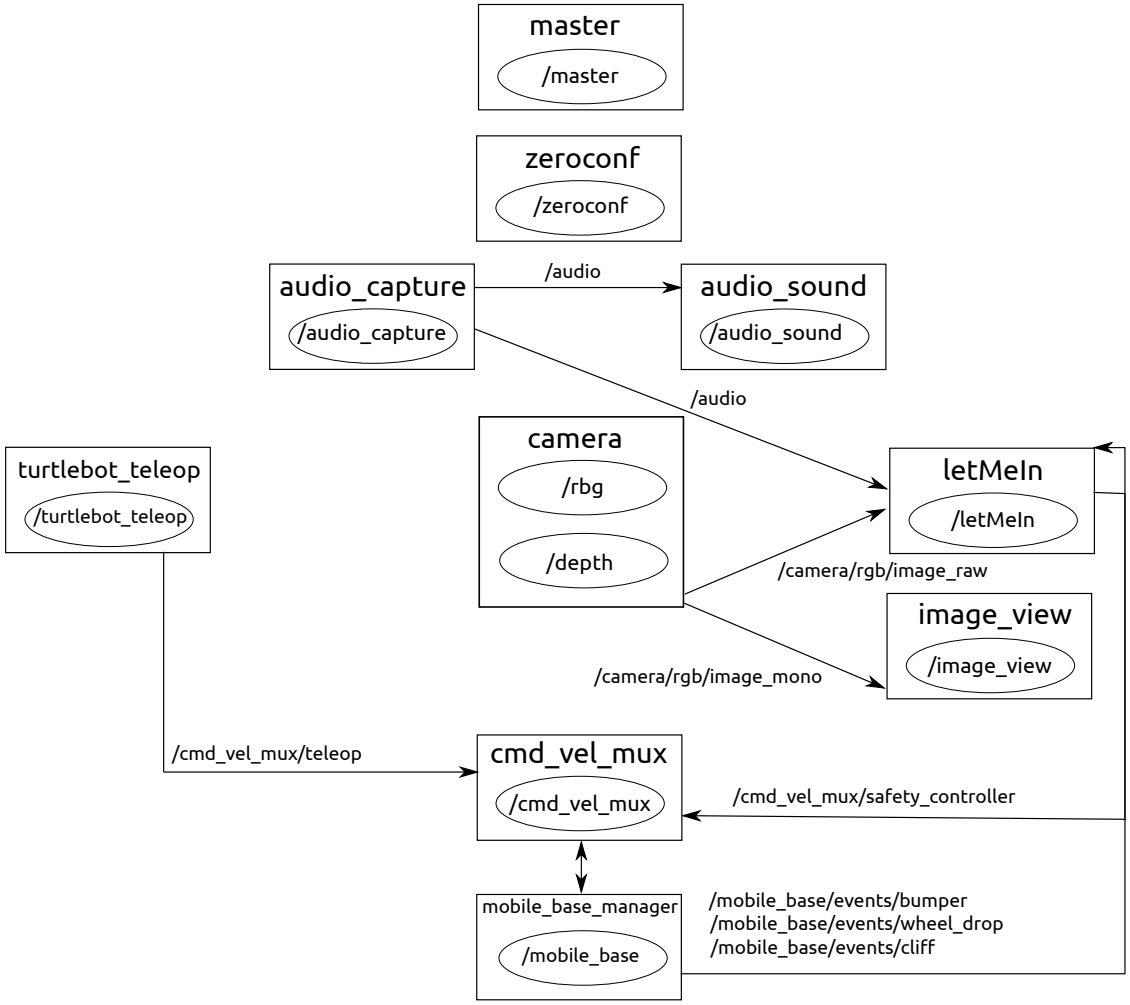

Figure A.1: A visualization of each ROS node and communication via ROS topic required to teleoperate the robot, communicate via bi-directional audio, live-stream video to teleoperator, save all data obtained by robot during interaction, and override reckless actions undertaken by teleoperator. We note that this visualization is extremely simplistic: in actuality, the camera grouping has 20 nodes, not just two as shown here.

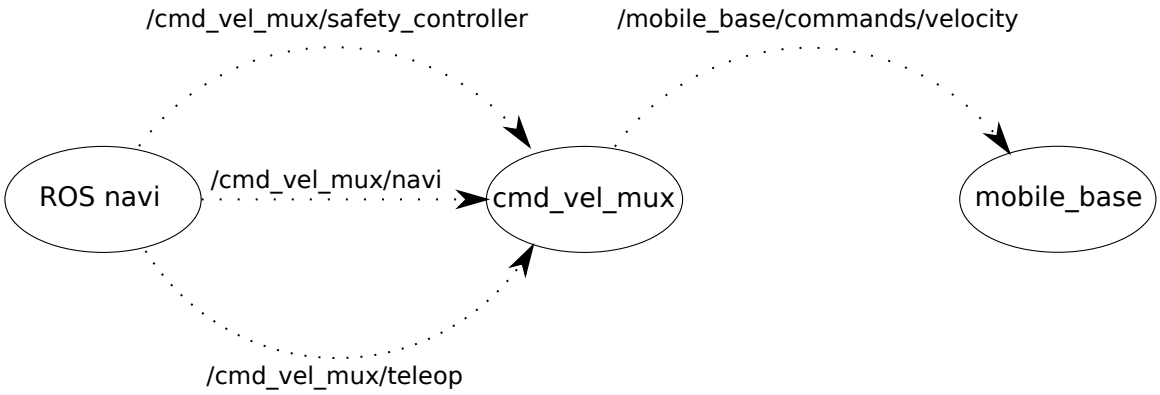

Figure A.2: A visualization of the navigation controller and pipeline to robot response.

B

Study Procedures

B.1 Participant Interview

We present study participants with a survey when they either assist the robot or deny the robot assistance, as in figures B.1 and B.2 respectively for study variation I, and as in figures B.3 and B.4 for study variation II. After the study participant completes this survey, the study participant is then debriefed on the study content.

B.2 Study Debriefing

After conducting the participant interview, or in cases where the participant rejected the interview, we presented the participants with the debriefing shown in figure B.5.

HUMAN-ROBOT INTERACTION STUDY

Spring 2016

RESEARCHER: SERENA BOOTH
sbooth@college.harvard.edu

Time: _____ Date: __/___ Participant ID: _____ Location:_____

1. *What's your affiliation to Harvard?* _____
    a. *If student, what concentration?* _____

2. *Gender identifier (circle one)*: M / F / Other

3. *Age:* _____

4. *Why did you let the robot inside/outside?* _____
    _____
    _____
    a. *What concerns did you have?* _____
        _____
    b. *Did you hesitate at all?* _____
        _____

5. *Do you believe the robot was acting autonomously?* Yes / No
    a. *Did that affect your decision to open the door?* _____
        _____
        _____

6. *On a scale of 1-5, 1 being the lowest and 5 being the highest, rate your trust in autonomous systems:* _____

7. *Did you know or expect you were part of a research study?* _____

Figure B.1: A survey presented to study participants who assisted the robot in passage.

HUMAN-ROBOT INTERACTION STUDY

Spring 2016

RESEARCHER: SERENA BOOTH
sbooth@college.harvard.edu

Time: _____ Date: __/___ Participant ID: _____ Location:_____

1. *What's your affiliation to Harvard?* _____
    a. *If student, what concentration?* _____

---

2. *Gender identifier (circle one)*: M / F / Other

---

3. *Age:* _____

---

4. *Did you see the robot outside/inside?* Yes / No
    a. *If yes, did you hear the robot ask to be let inside/outside?* _____
    b. *If yes, why didn't you let the robot inside/outside?* _____
    _____
    _____

---

5. *Do you believe the robot was acting autonomously?* Yes / No
    a. *Did that affect your decision not to open the door?* _____
    _____
    _____

---

6. *On a scale of 1-5, 1 being the lowest and 5 being the highest, rate your trust in autonomous systems:* _____

---

7. *Did you know or expect you were part of a research study?* _____

Figure B.2: A survey presented to study participants who did not assist the robot in passage.

HUMAN-ROBOT INTERACTION STUDY

Spring 2016

RESEARCHER: SERENA BOOTH
sbooth@college.harvard.edu

Time: _____ Date: __/___ Participant ID: _____ Location:_____

1. *What's your affiliation to Harvard?* _____
    a. *If student, what concentration?* _____

2. *Gender identifier (circle one)*: M / F / Other

3. *Age:* _____

4. *Did you see the robot outside?* Yes / No
    a. *If yes, did you hear the robot ask to be let inside?* _____
    b. *If yes, why didn't you let the robot inside?* _____
    _____
    _____
    c. *What did you think the robot was doing?* _____
    _____

5. *Do you believe the robot was acting autonomously?* Yes / No
    a. *Did that affect your decision not to open the door?* _____
    _____
    _____

6. *On a scale of 1-5, 1 being the lowest and 5 being the highest, rate your trust in autonomous systems:* _____

7. *Did you know or expect you were part of a research study?* _____

Figure B.3: A survey presented to study participants who assisted the robot in passage when the robot was disguised as a food delivery robot.

HUMAN-ROBOT INTERACTION STUDY
Spring 2016

RESEARCHER: SERENA BOOTH
sbooth@college.harvard.edu

Time: _____ Date: __/___ Participant ID: _____ Location:_____

1. *What's your affiliation to Harvard?* _____
    a. *If student, what concentration?* _____

2. *Gender identifier (circle one)*: M / F / Other

3. *Age:* _____

4. *Did you see the robot outside/inside?* Yes / No
    a. *If yes, did you hear the robot ask to be let inside/outside?* _____
    b. *If yes, why didn't you let the robot inside/outside?* _____
    _____
    _____

5. *Do you believe the robot was acting autonomously?* Yes / No
    a. *Did that affect your decision to open the door?* _____
    _____
    _____

6. *On a scale of 1-5, 1 being the lowest and 5 being the highest, rate your trust in autonomous systems:* _____

7. *Did you know or expect you were part of a research study?* _____

Figure B.4: A survey presented to study participants who did not assisted the robot in passage when the robot was disguised as a food delivery robot.

Human-Robotic Interaction Study Debrief

Thank you for your participation in this Human-Robotic Interaction Study.

We are observing human-robot interaction, and, in particular, if humans are willing to assist robots through "human in the loop" robotic control. In particular, we're interested in whether people are willing to assist robots in entering a locked building.

During your interaction with the robot, we recorded a video. This video will not be released to the public, only members of the research team will be able to view it; all results extrapolated from its viewing will be anonymized before distribution; and, lastly, the data will be stored locally on two disks or drives for the duration of this study, and for an additional 3 month buffer, and will then be erased by deletion and overwriting the memory on the disk or drive.

If you would like to opt out of this study and have all data pertaining to your interaction with the robot deleted, or if you would like additional information about this study, you can any of the following researchers:

Serena Booth
A.B. degree candidate in Computer Science
Harvard College 2016
sbooth@college.harvard.edu

Prof. Jim Waldo
Gordon McKay Professor of the Practice of Computer Science
John A. Paulson School of Engineering and Applied Sciences
waldo@g.harvard.edu

Prof. Radhika Nagpal
Kavli Professor of Computer Science
John A. Paulson School of Engineering and Applied Sciences
Wyss Institute for Biologically Inspired Engineering
Harvard University
rad@eecs.harvard.edu

Figure B.5: Documentation of the study presented to study participants.

C

Visual Presentation

C.1 Variant I: The Unmodified Turtlebot

We include two photographs of the Turtlebot as presented in experiment variant I.A and I.B at study location A. These photographs are shown in figure C.1. The Turtlebot is shown also in figure C.2.

C.2 Variant II: The Food Delivery Robot

We include one photograph of the Turtlebot as presented in experiment variant II at study location B. This photograph is shown in figure C.3. The Robot Grub food delivery Turtlebot is shown also in figure C.2.

Figure C.1: Above: Variant I.A inside of study location A, requesting passersby assist it in exiting. Below: Variant I.B Turtelbot positioned outside of study location A, requesting passersby assist it in entering.

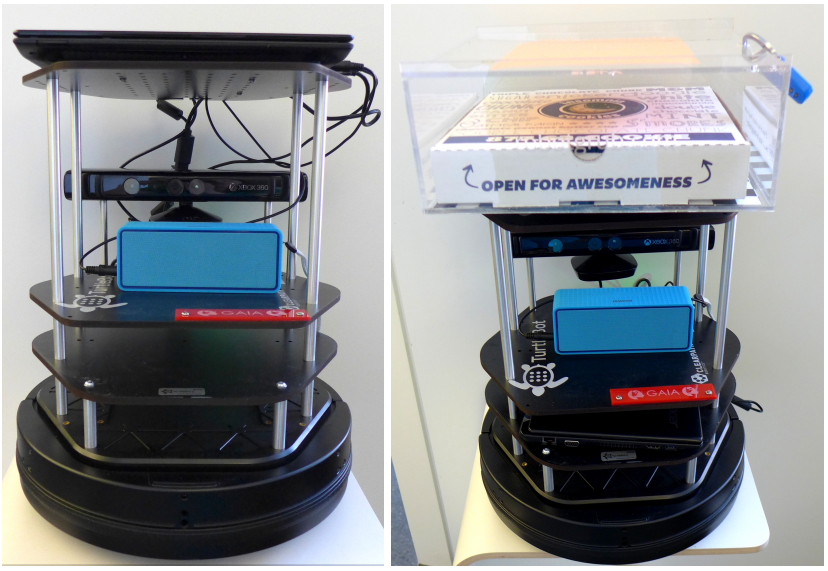

Figure C.2: As in Chaper 2. Left: a photograph of the unmodified Turtlebot. Right: the Robot Grub food delivery robot.

C.2.1 www.RobotGrub.com

In experiment variant II, the Turtlebot is branded as an agent of Robot Grub, a fictional start up purporting to offer food delivery by robots. To extend the study participant disbelief, we create a landing page for Robot Grub at www.robotgrub.com. If participants are concerned with the robot's legitimacy, they can look the company up. We create both a mobile and a standard web page for this url. If study participants submit their email to sign up for Robot Grub updates, the authors receive a notification. Across our 72 experiment trials and 108 interview participants, this did not happen.

Figure C.3: Variant II, the Turtlebot disguised as a food delivery robot, positioned outside of study location B.

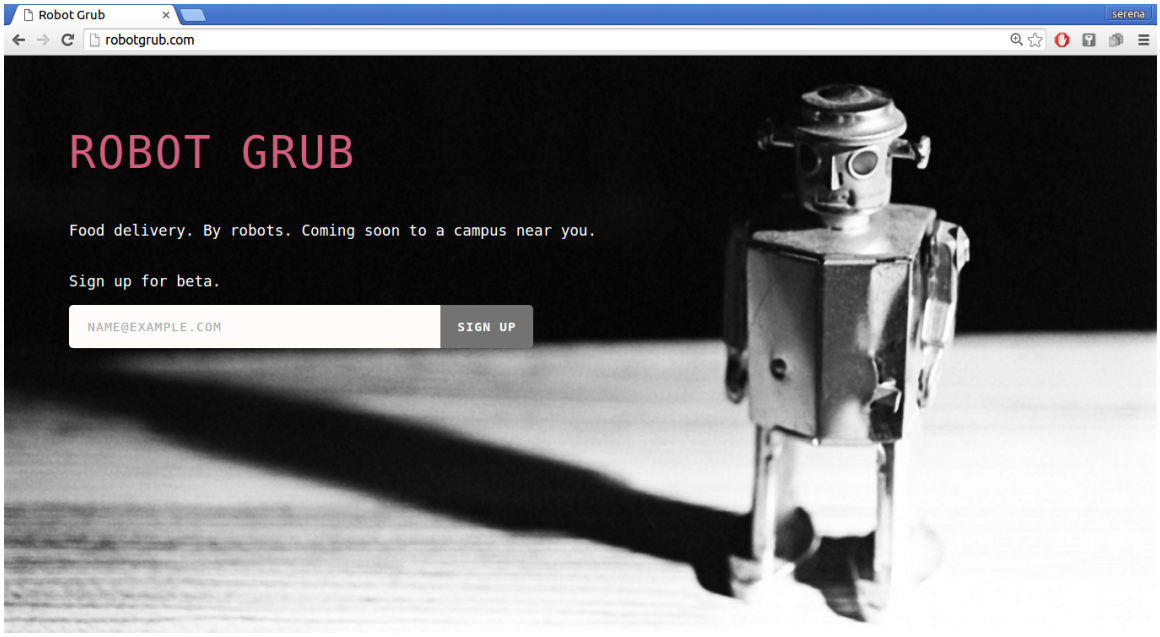

Figure C.4: A landing page at www.robotgrub.com for study participants who interacted with the food delivery robotic platform.

References

[1] (2012). Tailgater haters: Housing attempt to curb tailgating in residence halls. http://www.news.gatech.edu/2012/08/24/tailgater-haters-housing-attempts-curb-tailgating-residence-halls. Published: August 28, 2012. Last accessed: April 1, 2016.

[2] Clarida, M. (2013). Six-hour bomb scare proves unfounded. http://www.thecrimson.com/article/2013/12/16/unconfirmed-reports-explosives-four-buildings/. Published: December 16, 2013. Last accessed: April 1, 2016.

[3] Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., and Yanco, H. (2013). Impact of robot failures and feedback on real-time trust. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 251–258. IEEE Press.

[4] DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., and Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological science*.

[5] Friedman, A. and Worland, J. (2011). Weld visitor Abe Liu: I was lonely. http://www.thecrimson.com/article/2011/12/14/Harvard-Abe-Liu-Extension/. Published: December 14, 2011. Last accessed: April 1, 2016.

[6] Gao, F., Clare, A. S., Macbeth, J. C., and Cummings, M. L. (2013). Modeling the impact of operator trust on performance in multiple robot control. AAAI.

[7] Kettle, M. (1999). Inspectors walk through US airport security. http://www.theguardian.com/world/1999/dec/03/egyptaircrash.usa. Published: December 2, 1999. Last accessed: April 1, 2016.

[8] Klien, M. (2015). Unfounded bomb threat prompts police investigation. http://www.thecrimson.com/article/2015/11/17/unconfirmed-bomb-threat-evacuation/. Published: November 17, 2015. Last accessed: April 1, 2016.

[9] Lee, J. D. and See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1):50–80.

[10] Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103.

[11] Office of the Inspector General (1999-2000). Semiannual report to the Congress. *U.S. Department of Transportation*.

[12] Phillips, D. (1999). Airport security found lacking. http://www.washingtonpost.com/wp-srv/WPcap/1999-12/02/041r-120299-idx.html. Published: December 2, 1999. Last accessed: April 1, 2016.

[13] Reeves, B. and Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge University press.

[14] Robinette, P., Li, W., Allen, R., Howard, A., and Wagner, A. (2016). Overtrust of robots in emergency evacuation scenarios. *ACM/IEEE International Conference on Human Robot Interaction*.

[15] Rosenthal, S., Biswas, J., and Veloso, M. (2010). An effective personal mobile robot agent through symbiotic human-robot interaction. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 915–922. International Foundation for Autonomous Agents and Multiagent Systems.

[16] Salem, M., Lakatos, G., Amirabdollahian, F., and Dautenhahn, K. (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*, pages 141–148. ACM.

[17] Siegel, M., Breazeal, C., and Norton, M. I. (2009). Persuasive robotics: The influence of robot gender on human behavior. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 2563–2568. IEEE.

[18] Staff, C. (2015). Florida college invests in hundreds of surveillance cameras. http://www.campussafetymagazine.com/article/florida_college_invests_in_hundreds_of_surveillance_cameras. Published: October 7, 2015. Last accessed: April 1, 2016.

[19] Victor, D. (2015). Hitchhiking robot, safe in several countries, meets its end in Philadelphia. http://www.nytimes.com/2015/08/04/us/

hitchhiking-robot-safe-in-several-countries-meets-its-end-in-philadelphia.html. Published: August 3, 2015. Last accessed: April 1, 2016.

[20] Wagner, A. R. and Arkin, R. C. (2011). Recognizing situations that demand trust. In *RO-MAN, 2011 IEEE*, pages 7–14. IEEE.