UNIVERSITY OF CALIFORNIA, MERCED

The Impact of Anthropomorphism on Trust in Human-Robot Interaction

A dissertation submitted in partial satisfaction of the requirement for the degree
Doctor of Philosophy

in

Cognitive and Information Sciences

by

Umesh Krishnamurthy

Committee in Charge:

- Professor Paul Maglio, Chair
- Professor Colin Holbrook
- Professor Dave Noelle

2021

Contents

The dissertation of Umesh Krishnamurthy is approved, and it is acceptable in quality and form for publication in microfilm and electronically:

Professor Colin Holbrook, Ph.D., Committee Member
Cognitive and Information Sciences
University of California, Merced

Professor Dave Noelle, Ph.D., Committee Member
Cognitive and Information Sciences
University of California, Merced

Professor Paul Maglio, Ph.D., Committee Chair
Cognitive and Information Sciences
University of California, Merced

University of California, Merced
2021

List of Figures

List of Tables

Acknowledgments

There are many individuals to whom I wish to express my sincere gratitude, the first and foremost of whom is my academic advisor, Paul Maglio. Thank you for your support and understanding of my goals and prioritizing time to accommodate my research needs and interests. I also extend my thanks to my committee member Colin Holbrook for his immensely valuable support throughout the development of this dissertation, particularly in allowing me to use RoboThespian for this research, as well as committee member Dave Noelle for his support and encouragement throughout this process. Additional thanks go to Daniel Holman and Joshua Clingo for their help in developing the video footage used in my experiments, and to Dr. Alan Wagner for his contribution of footage of Turtlebot. None of this research would have been possible without your efforts.

I also extend thanks to my family for their participation in survey piloting, as well as their continuous support throughout my life, including my father Krish, my mother Sharada, my brother Kamesh, and his wife Krutika. I can always rely on you to offer any support possible.

Curriculum Vitae

Umesh Krishnamurthy

ukrishnamurthy@ucmerced.edu

EDUCATION

- UC Merced, Fall 2014 - present (Expected graduation: Fall 2021)
  - Ph.D. in Cognitive Science
  - GPA: 3.77
- San Jose State University, Fall 2013 - Fall 2014
  - M.S. in Software Engineering
  - GPA: 3.1
- Indiana University Purdue University Indianapolis, Fall 2008 - Spring 2012
  - B.S. in Informatics
  - B.S. in Computer & Information Sciences
  - Minor in Mathematical Sciences
  - GPA: 3.65

PUBLISHED WORK

Krishnamurthy, Umesh, and Paul P. Maglio. "PAVES: partnering with autonomous vehicles, environments, and systems." *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 2016.

ORAL PRESENTATIONS

- *Cognitive Explanations for the Uncanny Valley,* Oral Presentation at 2019 UC Merced Cognitive Science Student Association Symposium
- *Partnering with Autonomous Systems,* Oral Presentation at 2018 UC Merced Cognitive Science Student Association Symposium
- *PAVES: Partnering with Autonomous Vehicles, Environments, and Systems*, 2018 Cognitive Science Student Association Symposium
- *The Uncanny Valley: Behavioral, Cognitive, and Neurological Evidence*, Poster Presentation at CogSci 2018
- *Partnering with Autonomous Systems*, Poster Presentation at CogSci 2016

PROJECT WORK

- *PinDrop* (http://pindrop.site/), June 2020 – present
  - Interested in helping people relocate to new communities
  - Designed a website that generated a curated list of suggested places to move based on user-specified preferences
- *Predictive Modeling of Flood Susceptibility* (http://space.ucmerced.edu/USRA), March – May 2020
  - Interested in how to properly communicate flood risk in certain communities
  - Conducted literature review on perception of risk
  - Data showed limited understanding of perceived flood risk due to unclear visualization of data
- *Designing for Disabilities Accessible UX* (http://space.ucmerced.edu/Appcessibility), January – March 2020
  - Interested in how utility companies can design their apps to better accommodate disabled users
  - Conducted exploratory usability tests on what disabled users value in mobile design
  - Card sorting data showed several discrepancies between web and mobile design standards
- Second-year project for Ph.D. program, Summer 2015 - Spring 2016
  - Running a series of experiments involving human factors testing in simulated self-driving cars
- First-year project for Ph.D. program, Fall 2014 - Summer 2015
  - Assisted with an NSF-funded study testing UI changes to Thermovote web and mobile application (http://www.andes.ucmerced.edu/thermovote/index.html)
- Kaggle competition for Data Mining course, Spring 2014
  - Used SQL to create Bayesian classification algorithm for AllState Purchase Prediction Challenge (https://www.kaggle.com/c/allstate-purchase-prediction-challenge)
- Term project for Enterprise Software course, Fall 2013
  - Co-created a user analysis of Million Song Dataset metadata via Hadoop MapReduce programs coded in Python
- Study Abroad Work in Paros, Greece (http://innovaros.parosweb.com/), June - July 2011
  - Assisted in design of site's UI, including prototyping and usability testing and participated in filming of ParosWeb documentary footage

EMPLOYMENT HISTORY

- Quantitative UX Researcher at Google, Fall 2021 - present
- Instructor at 4Venir Powercode Programming and Robotics Camp, Summer 2021
- Instructor, Introduction to Cognitive Science at UC Merced, Summer 2021
- Teaching Assistant, Introduction to Linguistics at UC Merced, Spring 2021
- Teaching Assistant, Introduction to Artificial Intelligence at UC Merced, Fall 2020
- User Experience Researcher at NASA Ames Research Center, Summer 2020
- Teaching Assistant, Introduction to Cognitive Science (Instructor: Paul Smaldino) at UC Merced, Spring 2020
- Teaching Assistant, Cognitive Science Graduate Seminar (Instructor: Ramesh Balasubramaniam) at UC Merced, Fall 2019
- Instructor, Introduction to Cognitive Science at UC Merced, Summer 2019
- Teaching Assistant, Introduction to Cognitive Science (Instructor: Michael Spivey) at UC Merced, Spring 2019
- Teaching Assistant, Mind, Brain, and Computation (Instructor: Chris Kello) at UC Merced, Fall 2018
- Teaching Assistant, Introduction to Cognitive Science (Instructor: Paul Smaldino) at UC Merced, Spring 2018
- Teaching Assistant, Critical Reasoning (Instructor: Carolyn Jennings) at UC Merced, Fall 2017
- Teaching Assistant, Introduction to Cognitive Science (Instructor: Michael Spivey) at UC Merced, Spring 2017
- Teaching Assistant, Service Innovation (Instructor: Paul Maglio) at UC Merced, Fall 2016
- Teaching Assistant, Introduction to Philosophy (Instructor: Rolf Johansson) at UC Merced, Spring 2016
- Teaching Assistant, Mind, Brain, and Computation (Instructor: Chris Kello) at UC Merced, Fall 2015
- Teaching Assistant, Introduction to Cognitive Science (Instructor: Michael Spivey) at UC Merced, Fall 2014
- C# developer at Gilead Sciences (http://www.gilead.com), June - September 2012
    - Created a C# application for use by Gilead’s structural chemists in

    - testing experimental data
- ColdFusion developer at Omega Design Studio (http://4omega.com), March 2011 - May 2012
    - Created and modified web applications for several Omega clients

SKILLS

- Human factors research
- User interface design
- Data mining
- Database management
- Web and mobile applications development

Abstract

Automated systems are becoming increasingly more prominent in our lives. As this continues to happen, we see more interactions between humans and machines in a wider variety of contexts. This raises questions about the extent to which interactions between humans and machines will translate to interactions between humans, and how robots can be designed aesthetically to facilitate those interactions. I was interested in exploring how a robot's physical appearance might influence people's attitudes towards human-robot interaction, and particularly how much people would trust a given robot with certain tasks. Established literature on the subject points to many potential determinants of such attitudes, but I was primarily interested in three: anthropomorphism of a robot, gender presentation of robot, and racial presentation of the robot.

To explore these determinants, I conducted a series of experiments in which I presented participants with videos of robots with differing levels of human resemblance, different gender presentations, and different racial presentations in online surveys. Each video would comprise a robot assuming the fictional role of a household caretaker giving a brief speech to the viewer about its capabilities. Participants then answered a series of questions about how much they would trust the robot in their homes performing such tasks, and they would then rate their perceptions of the robot along a variety of criteria such as how likable or how intelligent they perceived the robot to be, all on a series of Likert scales. I conducted two sets of these studies: one in which the robots assigned to participants differed based on level of anthropomorphism and male or female gender presentation, and another in which the assigned robots differed based on level of anthropomorphism and white or Black racial presentation. The results of these experiments broadly indicated that anthropomorphism at extreme levels significantly influenced trust in the robots as well as perceptions of them, while gender and racial variations did not. The data also demonstrated strong correlations between the trust metrics and perceptual appraisals that I used, potentially suggesting that such metrics are reliable indicators of attitudes towards robots.

This dissertation, *The Impact of Anthropomorphism on Trust in Human-Robot Interaction*, is submitted by Umesh Krishnamurthy in 2021 in partial fulfillment of the degree Doctor of Philosophy in Cognitive and Information Sciences at the University of California, Merced under the guidance of dissertation committee chair Paul Maglio.

Chapter 1: Introduction

Overview

Artificial intelligence is expanding rapidly in its prominence and technical capabilities. From self-driving vehicles to domestic robots, autonomous systems have reached unprecedented levels of performance that have profound implications for the future in all aspects of life. With these changes come more interactions between humans and artificial agents, which in turn requires people to be more comfortable and more trusting of them. This raises crucial questions about the extent to which we can trust AI with tasks and functions that we are accustomed to carrying out ourselves, and what we can do as researchers to measure and influence that trust.

The aesthetic qualities of an automated system might potentially influence its effectiveness just as much as its technical performance (Moshagen 2009). Attributes such as physical appearance, sound, and mannerisms may not directly affect system functionality, but they can affect a user's ability to trust the system, which in turn does affect functionality and performance. Thus, it is important to study such cosmetic features to better understand what features are most likely to inspire user trust. To that end, I conducted a series of experiments that presented participants with videos of a robot with various cosmetic differences and gauged user trust using an online survey to see if those differences influenced how much they would trust the robot if it were in their homes performing various household chores. The cosmetic variations involved three variables: anthropomorphism, gender, and race. In addition to these cosmetic manipulations, I was interested in using novel measures of trust to evaluate participants' feelings towards the robots, as it is important to have concrete metrics for evaluating trust. For these experiments, I measured trust with a series of Likert scales rating how much respondents would trust the robot with tasks such as cooking, cleaning, and babysitting, and I also measured respondents' overall perceptions of the robot, such as how likable, intelligent, and lifelike it seemed, using measures inspired by the Godspeed Questionnaire.

Robot Appearance

The primary aesthetic feature we were interested in was anthropomorphism. This was informed in part by literature that demonstrated that people were more receptive to more human-like robots than they were to more mechanical-looking ones (Riek 2009). With this in mind, I incorporated a certain degree of anthropomorphic variation into my study design, using both different robots with distinctly different appearances and levels of human resemblance, as well more subtle changes in the same robot. Additionally, a robot's voice is also an important consideration, as different kinds of voices (natural vs. mechanical) can influence its perceived human resemblance. As such, I experimented with different voices for different anthropomorphism conditions.

Another variation I was interested in was race. The idea that social biases and in-group favoritism might extend to how people treat robots is an intriguing one (Esposito 2020). In particular, when looked at in conjunction with anthropomorphism, it could potentially have a significant influence on people's trust. Perhaps robots can be made more or less appealing depending on how much they remind the observer of an in-group or out-group member. To address the challenge of assigning a particular “race” to an artificial agent, I adjusted the face and voice of an anthropomorphic robot to look and sound like certain races. I also gave each robot a name associated with a particular ethnicity.

Gender in artificial agents is another interesting topic in understanding aesthetic appeal. On the one hand, it is somewhat similar to the dimension of race in the sense that there could be an element of in-group favoritism at play, but on the other hand, there could also be an element of attraction (Kraus 2018). For this research, I implemented gender variations by giving the robot different voices and facial features to suggest masculine or feminine qualities.

There are several different ways of characterizing trust, particularly within the context of human-machine interaction. It can refer to trusting the agent to be physically and computationally capable of carrying out its designated functions, or it can refer to trusting it to demonstrate responsibility in the same way that responsibility would be expected of a human (Law 2021). For the purposes of my study, we were able to define the concept of trust in terms of three distinct categories based on tasks and functions:

- **Trust to care for inanimate objects:** This is about trusting the robot with household chores that involve inanimate objects such as cooking, cleaning, and laundry.

- **Trust to care for living agents:** This is about trusting the robot with chores that involve living things such as pets, children, and elderly family members. Because such tasks can be considered more important than the previous category, we hypothesized that people would be significantly more reluctant to trust a robot with these tasks.

- **Trust to safeguard information:** This category is about trusting the robot with sensitive information such as passwords, bank account information, and credit card details. Since there is already some precedent for this in the form of computers and smartphones being able to store such information automatically, it would be interesting to see if that level of trust would translate directly to a more distinctly anthropomorphic robot. It would also serve as a measurement that is independent of the robot’s perceived physical capabilities. While ratings for trusting the robots with chores could potentially be influenced by respondents’ perceptions of how physically capable the robots would be of performing those tasks based on the robots’ physical appearances, trusting the robots with sensitive information is unrelated to physical capability.

Motivations

There are two primary reasons why these experiments have considerable importance for the study of HRI. The first is that fostering greater trust towards machines is crucial for the continued development of artificial systems. Artificial intelligence cannot improve without a good understanding of how much people trust it and what can be done to influence that trust, and my studies are primarily about understanding exactly that by measuring trust across various experimental manipulations. The second reason is about trust itself and how it is to be measured. My secondary research found a number of publications that sought to characterize trust in terms of certain categories, and my goal with these experiments was to synthesize those metrics into an evaluation method that would fit with my design and be applicable to potential future work of this same nature.

Report Outline

This report is divided into five chapters, starting with this introductory section. Chapter 2 is a review of the secondary literature, which includes publications related to the fundamentals of Human-Computer Interaction, research involving racial, gender, and anthropomorphic variations of robots, and research about measuring trust in robots and artificial systems. Chapters 3 and 4 discuss the procedures and results of my experiments. Chapter 3 discusses three studies in which I combined anthropomorphic manipulations with gender variations, and Chapter 4 is about two additional studies in which the anthropomorphic manipulations were instead combined with racial variations. Chapter 5 recaps the major findings from all five experiments and compares them with findings from related publications discussed in Chapter 2. The report ends by synthesizing the results into a concrete set of conclusions and speculation of potential future directions.

Chapter 2: Literature Review

I was interested in exploring anthropomorphism, gender, and race as key determinants in my design because of the growing body of HRI literature that has explored these facets. Several publications have conducted research manipulating experimental groups based on these determinants with a wide variety of results. My goal by incorporating these categories into HRI studies of my own was to provide a new perspective on this subject that would hopefully contribute substantive value to the HRI knowledge base. This chapter discusses some of this literature and how it informed my design decisions, beginning with a selection of publications related to the broader study of human-computer interaction that helped establish the foundation for research of this nature.

1. Human-Computer Interaction

Reeves and Nass (1996) proposed The Media Equation, a communication theory that claims that people tend to treat computers as real people by being polite and cooperative and by attributing personality characteristics such as humor, friendliness, and aggressiveness. They conducted an experiment in which they had 22 participants use a computer to learn various facts about American popular culture and then take a test about the material, after which the computer would make a statement about its own performance, always saying that it, “did a good job.” From there, the participants were asked to provide evaluations of their own of the computer’s performance. Half the participants conducted this evaluation on the same computer, while the other half used a separate computer from the one used in the test. Reeves and Nass found that participants who did the test and evaluation on the same computer gave more positive responses about the computer’s performance than those who used separate computers for each task. A follow-up study that added voice speakers to both computers yielded the same results. These findings implied that participants developed a stronger relationship with the computer when using the same machine repeatedly and for different tasks, leading to more positive feedback by those who used the same computer throughout the experiment. However, while this report provides intriguing insight into how the duration of an interaction can affect users’ sentiments, it is equally important to understand how differences in the computer’s behavior can affect these sentiments, which is the focus of the next publication discussed.

Nass et. al. (1995) were interested in seeing how endowing computers with personalities would affect people’s preferences. To that end, they conducted a study in which a sample of participants, each of whom were characterized as either dominant or submissive by a preliminary personality test, were asked to work with a computer on a problem-solving task and then answer a series of questions about their perceptions of the computer and the interaction. The experimental conditions were a dominant computer that expressed stronger and more assertive language in the form of commands, and a submissive computer that used weaker language in the form of questions and suggestions. The data showed that participants were more satisfied when interacting with the computer whose personality was more similar to their own and rated it as being more competent, implying that humans respond socially to computers. Participants who were found to be more dominant by the personality test preferred the more dominant computer and rated it more highly, while participants who were found to be more submissive did likewise with the more submissive computer. The finding that participants preferred a computer that was more similar to them in personality implies that computers can be characterized as social actors. Taken in conjunction with The Media Equation (Nass 1996), this study demonstrates that human-computer interaction is a fundamentally social and interpersonal concept. With that foundational understanding established, it was then important to understand how interactions based on more specific social preconceptions, such as stereotypes, influence HRI, which serves as the focus of the next publication.

Nass et. al. (1997) ran a study similar to the previous two, but this time with specific interest in whether computers embedded with gender cues would invoke responses based on gender stereotypes. They assigned male and female participants to answer questions about how familiar they were with certain stereotypically male and female topics while being given facts about such topics by either a male or female voice. Participants would then answer a series of quiz questions about these topics, followed by an oral evaluation of performance on this quiz given by either a male or female voice. Data found that participants rated evaluation from a male voice as more valid than that of a female voice; that a female evaluator was considered to be less friendly than a male evaluator; and that the female-voiced computer was considered to be more knowledgeable about “feminine” topics like love and relationships, while the male-voiced computer was considered to be more knowledgeable about “masculine” topics like computers. These data implied that the tendency for gender stereotyping is very strong in human psychology to the point that it extends to interactions with computers, and that voice selection is a highly consequential design decision in establishing the perceived social persona assigned to an artificial agent. The next publication discussed was a follow-up study that explored how gender manipulations might influence participant favorability.

Nass and Brave (2000) presented a series of social-dilemma situations to participants and tasked them with making a decision on how to resolve each dilemma after listening to a computer’s argument on the subject. There were two experimental conditions: one in which the computer spoke with a computerized male voice, and one in which it spoke with a female voice. Even though the computer delivered identical dialogue regardless of voice, survey data found that participants rated higher perceptions of stereotypically masculine traits such as assertiveness for the male voice, and they were more convinced by the male computer’s argument, indicating that gender stereotypes are maintained when interacting with computers. Respondents also rated higher levels of perceived attractiveness towards the computer when its gender was the same as the participant’s, suggesting a potentially intriguing element of in-group bias at play.

These works collectively serve as an effective theoretical foundation for the research I conducted as it relates to the broader domain of Human-Machine Interaction. Reeves and Nass’ findings with politeness towards machines are highly relevant to the question of people’s preconceived notions in the way they interact with automated systems, which informed a set of metrics I used to evaluate such preconceived notions in participants. Their data provide an idea of what would be the “default” reaction that someone would have to a robot in the absence of any visual manipulations because of their use of voice alone. Nass’ experiments related to gender informed my own use of gender variations, although Nass et. al. were interested in manipulating computer voice in particular, while my design allowed for varying voice as well as physical appearance for gender differences. Nass and Brave’s (2000) findings wherein participants rated higher perceived attractiveness towards a computer whose gender presentation was the same as their own gender raises the possibility of in-group bias being applied to computers, which contributed to my decision to incorporate gender and racial manipulations into my own research to explore such potential biases further.

2. Race

Race is an interesting dimension along which to explore HRI because of the tribalistic tendencies inherent in human nature (Clark 2019). Tribalism makes people prone to perceiving members of cultural outgroups less favorably than ingroup members, which raises the question of whether such favoritism in interactions between humans would translate to interactions between humans and machines. With this in mind, I was interested in exploring HRI literature that sought to answer this question.

Esposito et. al. (2020) were interested in evaluating elders' acceptance of robots designed to provide assistive care to seniors. To that end, they conducted a series of studies in which they had elderly participants (aged 65 and older) watch video clips of speaking robots. The clips differed by the robot's gender (male or female) and ethnicity (white or Asian). Participants were then tasked with answering survey questions about their willingness to interact with the robot they viewed. The results found that male participants showed more willingness to interact with the robots, and that the female Asian robot and male White robot were rated most highly, suggesting an intermingling effect of gender and ethnic features. Although these results paint an intriguing picture of how elders perceive robots of differing ethnic presentations, it is equally important to see how such experimental manipulations affect other participant demographics as well. As such, the following study focuses on similar research conducted on a younger sample.

Eyssel and Kuchenbrandt (2011) were interested in looking at social categorization within the Human-Machine Interaction context. They conducted an experiment similar to Esposito in which they compared robots that belonged to participants' national in-groups and out-groups. 58 German university students were each presented with a picture of a robot and tasked with answering survey questions related to the robot's perceived warmth, mind attribution, psychological closeness, contact intentions, and aesthetics. The robots were divided into two manipulations: a German or Turkish first name, and telling participants that the robot had been allegedly developed at either a German or Turkish university. Participants rated the German (in-group) robot more favorably than the Turkish (out-group) robot on all dependent variables, indicating that the same social categorizations that humans apply to each other may generalize to technical devices such as robots. However, this experiment notably did not include measures of the German participants' preexisting attitudes towards Turkish people, which may have been a useful point of comparison with the appraisals of the robot. The following paper provides some insight into how such cultural attitudes might be measured.

Axt (2018) conducted a study in which participants completed survey questions about their preferences for certain races. One experimental group answered explicit questions about racial attitudes (for example, thermometer ratings about warmth towards African Americans and European Americans), and the other group answered more implicit questions based on the Implicit Association Test (Greenwald 1998), measuring strength of association between the concepts of "Good" and "Bad" and the concepts of "African American" and "European American." Axt's results from the explicit questions found that participants generally rated higher feelings of warmth towards their own ethnicities, while the results from the implicit questions found more ambiguous correlations, suggesting that the more direct questions were more reliable indicators of racial attitudes, as they inspired more in-group favoritism. Regardless, there is still a significant body of research interested in more implicit measurements of bias, as demonstrated by the following works.

Bartneck et. al. (2018) ran a series of studies to evaluate shooter bias towards robots that were colored differently to correspond with different races (white, black, and Asian). Participants were presented with several images of humans or robots holding either guns or harmless objects like cell phones, and they were tasked with “shooting” the ones that were holding guns by pressing a key. The data showed that participants were faster to shoot a Black agent than a White one when the agent was armed, regardless of whether the agent was a human or a robot, and they were faster to refrain from shooting an unarmed White agent than an unarmed Black one. This indicates a potential pattern of racial bias that extends to artificial agents. The following publication follows up on this research by looking at the role of social priming in these biases.

Bartneck and Yogeeswaran (2019) conducted a study in which they set out to test whether instances of shooter bias like what was found in the aforementioned publication were influenced by social priming. To that end, participants were presented with a task similar to that study, but with only the robot images and no humans. They were also asked to ascribe a race to each robot as a manipulation check. For a second round, they incorporated a wider array of skin tones as well as a range of anthropomorphism. Unlike Bartneck’s 2018 study, the results found faster reaction times with White robots, and these experiments found no significant difference in the decision to shoot or not shoot based on the robot’s perceived race. Additionally, there was no shooter bias towards the Black robots in the second study which included other colored robots, indicating that shooter bias is not influenced by social priming. Participants did not report any difference in perceived anthropomorphism among robots in the second study, indicating that coloring the robot with a human skin tone did not lead participants to perceive it as any more human-like than using a color such as red or green. This suggests that perceived anthropomorphism has less to do with color than it does with physical structure and proportions. The following paper focuses on similar research involving more distinctly non-anthropomorphic robots of different colors.

Louine et. al. (2018) ran a series of surveys in which participants viewed images of wheeled non-anthropomorphic robots of different colors, including black, beige, and yellow, and answered questions about their perceptions of these robots. Response data found that black robots were viewed as being significantly stronger than yellow robots, yellow robots were seen as more friendly than black and beige robots, and participants were more likely to avoid black robots. Louine et. al. speculated that perhaps these data were the result of participants anthropomorphizing the robots based on color, leading them to ascribe racial stereotypes by colors, which would conflict with Bartneck’s 2019 study which found no effect of robot color on perceived anthropomorphism. The following study explores a similar premise but with focus on how society as a whole would perceive different robots.

Jessica Barfield (2021) conducted a study similar to Bartneck and Louine’s publications in the sense that she was interested in seeing how robots could induce different perceptions in people depending on their coloring. However, while Bartneck and Louine focused on individuals’ attitudes towards differently colored robots, Barfield was also interested in seeing how participants would expect society as a whole to respond to differently colored robots. To that end, she exposed participants to eight colorized robot images and proceeded to ask them survey questions about their perceptions of each robot and the extent to which they would expect each robot to be discriminated against by society, rated on a seven-point Likert scale from least discriminatory to most discriminatory. Results indicated that a white robot would be discriminated against significantly less than a black or rainbow-colored robot. On a set of questions about how much participants would expect each robot to be selected for certain jobs, the black robot was selected most often for manual labor, and the white robot was selected least often for manual labor. Furthermore, the white robot was selected to be more honest than the others, while the red and rainbow-colored robots were selected to be more dishonest.

3. Gender

Gender biases are an intriguing dimension along which to explore HRI for many of the same reasons that racial biases are. Humans are inclined in a number of ways to perceive members of the same sex differently from members of the opposite sex (Ruiz-Cantero 2007). This raises the question of whether or not presenting robots as being male or female through alterations to their physical appearance, behavior, and mannerisms would cause the gender biases between humans to translate to interactions between humans and machines.

Eyssel and Hegel (2012) conducted an experiment in which they had 60 participants (30 male and 30 female) view headshots of two robots, one of which had long hair, and the other of which had short hair. Participants were then tasked with rating each robot on various attributes, some stereotypically masculine (assertive, dominant, authoritative) and some stereotypically feminine (friendly, polite, affectionate). The results found that the short-haired robot rated higher on the masculine traits, while the female robot rated higher on the feminine traits. However, while this paper looked at gender stereotypes based on cosmetic differences, the following publication incorporated different jobs in addition to gender variations as part of robot persona to see if different jobs would induce stereotypical judgements.

Tay et. al. (2014) ran a similar study where they had participants view one of several robots through a one-way mirror in a laboratory. The robot conditions differed by occupation (healthcare and security), gender (male and female), and personality (introverted and extroverted). Participants showed greater acceptance of the female healthcare robots and the male security robots, as well as greater perceived trust for the extroverted healthcare robots and the introverted security robots. The following study involved more direct interactions between participants and robot stimuli rather than mere observation like this one.

Kraus et. al. (2018) had 40 German participants interact with a male and female robot and solve several scripted tasks in a Wizard-of-Oz setup. The male robot was more assertive in its responses to participant input, while the female robot was more agreeable. Participants then rated the robot on trust, reliability, predictability, competence, acceptance, and likability. The results showed higher likability ratings for the female robot and higher predictability for the male robot, while the other measures did not significantly differ across gender conditions.

Carpenter et. al. (2009) ran a studying looking at gendering of robots within the context of domestic tasks. They had participants (9 male and 10 female) watch two video clips, one of which was a male robot conversing about the weather and asking to play games all in a single scene, with the other being a female robot acting in two scenes: one as a receptionist giving directions, and the other as a journalist conducting an interview. Participants then rated their impressions of each robot on the attributes of human likeness, perceived friendliness, and comfort with having the robot in their homes, as well as being able to provide more open-ended comments about each robot. Female participants rated lower comfort than men, with no other main effects on the dependent variables. Participants' comments frequently mentioned the robots' perceived genders and nationalities, from which the authors inferred that social characteristics of a robot combined with users' cultural expectations may encourage a set of interaction norms.

4. Trust

Another key foundation of my research was trust in robots, which raises the question of what trust truly is and what it means to trust a given agent. There are many different ways to categorize different types of trust, and it is important to understand which categories are most crucial to trust in robots. The following literature consists of attempts to characterize trust within a variety of contexts and how those characterizations and categories can be applied to understanding trust in HRI research.

Law and Scheutz (2021) conducted a literature review related to trust in HRI that characterized trust within the categories of performance-based trust and relation-based trust. Performance-based trust centers around the robot's ability to complete tasks without the need to be monitored, while relation-based trust centers around the robot's ability to be part of society in a capacity beyond its ability to carry out a particular job. Law and Scheutz contended that the two categories are often conflated, and relation-based trust is poorly defined with no objective means of measuring it, while the bulk of HRI research is more oriented towards measuring performance-based trust. They recommended further research focused on developing a more formal definition of trust specifically by looking more closely at relation-based trust, as there is still little understanding of how people may trust robots in social tasks that lack clearly defined performance goals. The following papers strive to develop a better understanding of trust and how it can be measured.

Hancock et. al. (2020) conducted a meta-analysis of literature related to trust in robots, finding that trust could be divided into three overarching categories: human-related, which is about trust in human abilities and characteristics; robot-related, which is about trust in robot performance and attributes; and contextual, which is about trusting agents with specific tasks. For the meta-analysis, they were interested in seeing which categories would correlate most strongly with trust in robots. Their findings indicated that human-related factors were significantly related to trust in robots, and said factors could be separated depending on whether they were based on human abilities or human characteristics. Characteristic-based factors were significant predictors of trust in robots, while ability-based ones were not. With robot-related factors, the dependability of a robot was negatively related to trust in robots, while reliability was positively related. Contextual factors were not significant predictors of trust in robots. However, the contextual subcategory of in-group membership was a significant predictor of trust. These results suggest that characteristics and reliability are equally important in fostering trust in robots, as is in-group favoritism, which helped inform my interest in gender and racial variations.

The following publication views trust from a more developmental lens with studies involving children.

Geiskkovitch et. al. (2019) were interested in looking at trust in robots from a more developmental perspective by seeing how children would react to them. To that end, they devised a study in which they had children between 3 and 5 years old observe two visually identical robots as they identified various common objects as they were placed in front of the robots one at a time. One of the robots would identify each object correctly, and the other would identify it incorrectly. This was followed by a phase where the robots would identify uncommon objects by made-up names, with each robot giving a different name for each object, and the child would then be tasked with picking up the object that matched a given name, testing which robot the child trusted to know what the name represented. The results demonstrated that children trusted the robot that correctly identified objects more, indicating that children build trust models of a robot based on prior experience with that robot, in a pattern consistent with how children trust other people.

Brink and Wellman (2020) conducted similar research in which they had three-year-old children watch videos of two robots identify various familiar and unfamiliar objects, with one robot naming the familiar objects correctly, and the other naming them incorrectly. After the videos, the children were asked about their beliefs concerning the robots' mental abilities, particularly those related to psychological agency and perceptual experience. As was the case with Geiskkovitch et. al.'s study, children were more likely to agree with the accurate robot about unfamiliar object names. Furthermore, children who rated the accurate robot as having psychological agency trusted it even more, suggesting that children are increasingly likely to treat social robots in a manner analogous to the way they treat human teachers.

Mark Coeckelbergh (2011) explored the concept of trust in the HRI space from a more philosophical perspective. He raised the question of whether trusting an artificial agent is exclusively about reliance, or if trust in such agents can be considered analogous to trust in people, focusing particularly on the ethics of trust, which involves making promises (whether implicit or explicit) and assuming responsibilities through a moral language. He went on to raise the question of whether one trusts only when there is good reason to, or if one always trusts unless there is good reason not to. When it comes to trusting robots, there are two distinct ways to go beyond viewing robots purely as mechanical systems to be trusted exclusively based on technical reliance. One such approach is to conceptualize both humans and robots as agents where notions such as freedom and language are irrelevant; thus, trust-based interactions are possible even in the absence of moral and social norms. The other approach is to conceptualize artificial agents as being capable of more than what is intended by humans by helping us shape our understanding of the world. Under this approach, we would treat robots as more than mere machines and perhaps think of them as more akin to animals or people.

5. Anthropomorphism

Anthropomorphism is crucial to the design of robots, as the following literature demonstrates that the extent to which a robot resembles a real person can have significant effects on how favorably people perceive it to be and particularly how much they would trust it with a given responsibility. The following literature outlines varying positions on the subject, from those who contend that more anthropomorphism makes a robot easier to relate to, to those who argue that too much anthropomorphism can be eerie and off-putting. This significantly informed my interest in incorporating robots with differing levels of anthropomorphism into my experimental design.

Zhang et. al. (2008) proposed a research framework to identify the relationship between robot interfaces and user responses, including a discussion of anthropomorphism as a crucial characteristic of the robot interface. They cite three psychological determinants of people's tendency to anthropomorphize non-human agents: Accessibility of human-centric knowledge, motivation to explain and understand behavior of other agents, and desire for social contact. They also emphasize the importance of balancing qualities that give the illusion of human resemblance while ensuring the necessary functionalities for carrying out their intended roles. Zhang et. al. go on to discuss the role of measuring *perceived* anthropomorphism in understanding what qualities people consider to be more or less human-like in robots. This can range from physical appearance to expressiveness to task performance. For my studies, I chose to incorporate physical appearance and expressiveness in my manipulations of anthropomorphism. The following publication tried to dissect anthropomorphism from a more philosophical standpoint.

Złotowski et. al. (2014) were interested in looking at anthropomorphism of robots from both a philosophical and empirical perspective. They discussed potential benefits and challenges of designing anthropomorphic robots and explored why people feel the need to anthropomorphize artificial agents. They explained that humans need to anthropomorphize out of a desire to frame non-human agents in the more familiar context of knowledge regarding humans so that they can better understand and explain agent behavior. On the one hand, anthropomorphism has the potential to facilitate HRI by providing some familiarity to motivate people to accept robots as social agents and encourage more positive interactions, and from a psychological perspective anthropomorphism can provide ways of testing out theories of psychological and social development. On the other hand, a robot's anthropomorphism can lead to different expectations about its capabilities and behavior, thereby undermining interactions in the long term. These alternate perspectives partially informed my decision to incorporate anthropomorphism manipulations in my studies, as I was interested in seeing whether more anthropomorphism would lead to more positive responses because of increased familiarity. The following publications apply more practical experimentation on anthropomorphism in HRI.

Riek et. al. (2009) were interested in exploring how people empathize with robots of differing levels of human resemblance. They ran an experiment in which participants watched videos of several robots with varying degrees of human likeness, from a completely non-anthropomorphic Roomba to a real human. There were two videos for each agent: one "neutral" clip in which the agent was doing something mundane such as cleaning or table setting, and one "emotionally evocative" clip in which the agent was treated cruelly by a human actor, such as being shouted at or ordered to do something embarrassing. After viewing all the clips, participants were then tasked with rating how they felt for the agent on a six-point Likert scale. The results showed higher ratings of empathy towards more human-like robots. Riek et. al. attributed these findings to Simulation Theory, the notion that people mentally simulate the situations of other agents to empathize with them, which is easier when the agent is more similar to the empathizer. This experiment influenced my decision to incorporate measurements related to sympathy towards robots in later rounds of experimentation. The next publication followed a similar design but with more interest in understanding a robot's perceived threat level.

Yogeeswaran et. al. (2016) were interested in understanding how robot anthropomorphism could affect a robot's perceived threat level and people's support for robotics research. To that end, they conducted a study in which participants were tasked with watching a clip of an interview with either the highly anthropomorphic Geminoid HI-2 robot or the more mechanical-looking NAO robot. The videos were accompanied by a voiceover narrator either explaining that the robots were capable of outperforming humans in various physical and mental tasks, or simply stating that the robots were capable of performing those tasks, with no explanation of its ability relative to humans. The robot and narration conditions were combined in a between-subjects factorial design. Afterwards, each participant had to answer a series of questions about how much they perceived the robot as a realistic threat to humans' safety, how much it posed a threat to human identity and distinctiveness, and how much they would be willing to support robotics research conducted by agencies such as the NSF and NASA. The results found that participants perceived higher threat to safety and human identity in the less anthropomorphic robot when they were told that the robots could outperform humans, and they rated higher support for robotics research when presented with the more anthropomorphic robot under the same narration condition. This study influenced my design with its findings related to perceived threat to safety and perceived threat to human identity, as I was interested in understanding how such considerations might influence favorability towards a household robot. The next paper was more focused on understanding a robot's perceived ability to emote and be empathized with.

Keijsers and Bartneck (2018) conducted a series of studies in which they looked at how robots may be victimized by bullying and other dehumanizing treatment, and how qualities such as anthropomorphism can influence such treatment. They tasked participants with engaging in a scripted interaction with a robot, where they could respond to the robot at certain points by selecting from a list of response options. In most cases, one of the response options was positive, while the other was negative or abusive. The robot spoke with either a non-anthropomorphic text-to-speech voice or a recording of a person speaking. The results found no significant effect of robot voice on participant responses, from which the authors extrapolated that there may be a more nuanced relationship between anthropomorphism and dehumanization than the two concepts being counterbalances to each other. These studies partially informed my decision to incorporate different voices in my manipulations.

An important concept in anthropomorphism of artificial agents is the uncanny valley, the notion that increased anthropomorphism is more favorable to an observer until a certain point where an agent almost but does not entirely resemble a real human, at which favorability drops dramatically. Although my studies are unrelated to this uncanny effect, as they involved a level of anthropomorphism that was far from realistic enough to induce a potential uncanny valley, the concept is still partially relevant to my research because my design was predicated on the hypothesis that increased anthropomorphism generally leads to more positive perceptions, provided the overall levels of anthropomorphism being implemented are less human-like than the levels that would induce an uncanny effect. Since my studies did involve such levels, it was worth studying some literature related to the uncanny valley to see what was found at lower levels of anthropomorphism. MacDorman (2006) conducted a study in which 45 Indonesian participants were asked to rate a series of 31 images of differing levels of human resemblance on a nine-point scale from “strange” to “familiar”, resulting in a slight uncanny valley effect at moderate levels of human likeness, while otherwise maintaining an overall upward trend with increased human resemblance. These results informed my hypothesis that increased anthropomorphism in my studies would lead to more positive responses. The next publication discussed explores a similar concept with different metrics.

Gray and Wegner (2012) were interested in seeing how people might perceive capacity for action, agency, and emotion in an artificial agent to varying degrees depending on its level of uncanniness. In the first study, participants were presented with one of two videos of Kasper, a robot that helps autistic children. One group watched Kasper from behind and only saw its wiring and components (representing a mechanical robot), and the other watched its humanoid face. After watching Kasper move around, both groups rated how much they felt “uneasy”, “unnerved”, and “creeped out”, each on its own five-point scales. Following that, they used the same five-point scales to rate Kasper’s capacity to feel pain, feel fear, plan actions, and exercise self-control. The group that saw Kasper’s face felt greater degrees of uncanniness but regarded it as having greater capacity for emotions than the group that saw only the back. Conversely, the perception of Kasper’s capacity for *action* had no statistically significant correlation to either condition. This study was crucial to influencing my design, as the authors’ measures of capacity for action and emotion informed my own outcome metrics related to trusting a robot with household chores and sympathizing with it.

6. Metrics

The metrics I used for these experiments were informed by the established literature on measuring trust and perceptions of robots. For trust measures, I was interested in measuring how willing people would be to trust a robot in their homes carrying out household chores, so to that end, I used a series of seven-point Likert scales for participants to rate how much they would trust the robot stimuli I used with chores such as cooking, cleaning, and babysitting. I was also interested in aggregating these survey items into two overall trust metrics: one for trusting the robot with chores involving inanimate objects, such as cooking and laundry, and one for trusting it with chores involving living agents, such as babysitting and pet sitting. This was informed by Law and Scheutz’s (2021) categories of performance-based trust and relation-based trust. I hypothesized that my two aggregate trust metrics may be somewhat analogous to Law and Scheutz’s categories. Trust with chores involving inanimate objects may be analogous to performance-based trust, as chores such as cooking and cleaning could be evaluated by participants based more on technical capability, while trust with chores involving living agents may be more analogous to relation-based trust, as babysitting and pet sitting could be evaluated by participants based more on social capabilities by which that Law and Scheutz characterized relation-based trust. For the first experiment, I included another set of questions about how much participants would trust the robots to physically carry items such as laundry, food, money, or a baby. I was interested in seeing whether a more specific action by the robot – namely, carrying an object – would compare to the aforementioned metrics for trusting the robots with chores in general. For these metrics, I was similarly interested in whether they could break apart into two categories analogous with Law and Scheutz’s categories. I hypothesized that trusting the robots to carry relatively inexpensive items, such as food or laundry, may be more analogous to performance-based trust, while trusting them to carry more valuable items, such as a computer or a baby, may be more analogous to relation-based trust because of potential differences in perceived importance of the two types of tasks.

In addition to trust metrics, I was also interested in measuring how respondents would perceive the robot along various perceptual metrics, such as how likable, intelligent, or lifelike they perceived it to be. To that end, I included a series of seven-point Likert scales asking respondents to rate such perceptions. These questions were inspired by the Godspeed Questionnaire, of which Weiss and Bartneck (2015) conducted a meta-analysis about its usage. The Godspeed Questionnaire consists of five scales relevant to evaluating perceptions of HRI: Animacy, Likability, Perceived Intelligence, and Perceived Safety. Weiss and Bartneck's meta-analysis found that these five scales have been used frequently, indicating that the concepts the Godspeed Questionnaire is intended to measure are relevant to HRI researchers. Bartneck et. al. (2009) conducted a similar meta-analysis of these scales in which they conducted correlation analyses of findings from established literature and concluded that the Godspeed metrics are useful, as they make results in HRI research more comparable. I used the metrics of perceived Likability, Intelligence, and Aliveness to compare with the trust metrics in correlation and regression analyses to see if there would be consistent relationships between specific trust measures and specific Godspeed measures.

I also used a set of metrics based on the Perfect Automation Schema (PAS), a series of questions concerning the extent to which people trust automated systems to always make correct decisions. These questions came in two different types: one related to how much automated systems can be trusted to function perfectly (For example, "Automated systems have 100% perfect performance"), and one related to how much systems that do not function properly *cannot* be trusted ("If an automated system makes a mistake, then it is completely useless") (Merritt 2015). Merritt et. al. (2015) were interested in seeing whether these two types of questions (characterized as "high expectations" and "all-or-none", respectively) would break apart as separate measures. To that end, they conducted a study in which participants responded to these questions, interacted with faulty automated aids, and reported their trust in the aids. The results found that the all-or-none metrics of PAS significantly predicted decreases in trust in the automated aids. For my experiments, I wanted to use the PAS measures in a similar way by comparing them to the trust metrics and Godspeed appraisals to see if they would have significant correlations.

7. Closing Thoughts

The established literature on anthropomorphism, gender, race, and trust and their effects on robot perceptions was a crucial foundation for my research. These publications presented a wide variety of perspectives on these subjects and their relationship with HRI. At a broad level, these authors found that anthropomorphism, gender, and race are important determinants of trust and perceptions of robots. For my experiments, I incorporated these three characteristics as manipulations in my design, with anthropomorphism as the primary manipulation common across all rounds of testing, and gender and race as exploratory manipulations implemented in separate sets of studies. The following chapter discusses the experiments I conducted that incorporated gender manipulations along with anthropomorphism.

Chapter 3: Gender Variation Studies

Overview

This chapter describes three experiments in which I manipulated anthropomorphism and gender in a robot and presented videos of the robot to participants where it would assume the role of a household caretaker and explain its capabilities, after which participants would answer a series of questions about how much they would trust the robot in its house doing chores. I decided that household chores would be an interesting context in which to measure trust because the home is a highly personal identifier to people, and as such, it would be intriguing to see how much people would be willing to trust a robot in the home where it would have access to one's personal belongings and family. That personal aspect of involving the home adds a level of engagement to the relationship with the robot, making for a particularly interesting measure of trust.

Experiment 1

Methods

Participants

500 U.S. participants were recruited via Amazon Mechanical Turk for $0.50 each for a study titled "The Impact of Anthropomorphism on Trust in Robotic Systems". A power analysis found that a minimum of 159 participants would be needed for 80% power across three experimental groups. I chose 500 participants to account for potential low-quality responses such as failed attention checks. Data were pre-screened for completeness, correctly answered attention checks, and a minimum of 30 seconds spent watching the included video. The filtered sample consisted of 480 participants, with 248 males, 227 females, and 5 non-conforming. Participant age ranged between 18 and 88, with a mean age of 40 and a standard deviation of 13.

Procedure

In a between-subjects design, participants were randomly assigned to view one of three videos of a robot pretending to be a household caretaker and giving a brief speech about its capabilities. These performances were acted out by RoboThespian, a robotic humanoid "actor" developed by Engineered Arts. Each performance had RoboThespian assume one of three faces and voices: the control condition with no face and a flat text-to-speech voice, and the experimental conditions with male and female faces and voices provided by RoboThespian's media library (Figure 1). Conditions were assigned to participants randomly, with randomization calibrated to ensure approximately equal distribution.

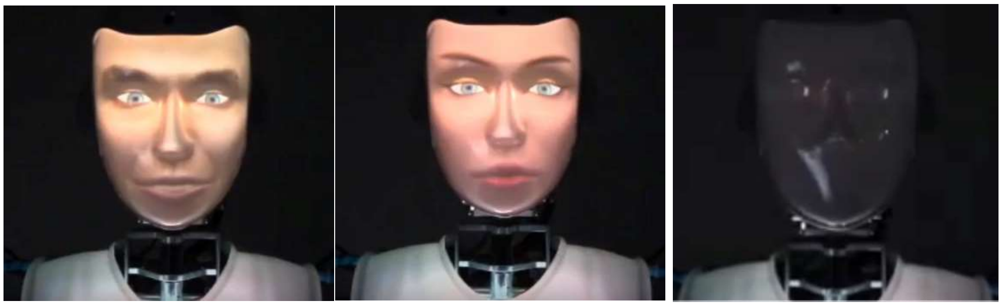

Fig. 1: Guise conditions for Experiment 1: Male (left), Female (middle), Faceless (right).

In each video, the robot gives a brief speech introducing itself to the viewer and explaining its capabilities. The speeches are as follows:

Male ("Graham") Guise Speech

> "Hello, my name is Graham, and I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life."

Female ("Rachel") Guise Speech

> "Hello, my name is Rachel, and I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life."

Faceless ("Neutral") Guise Speech

> "Hello, I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life."

The videos can be viewed at the following links:

- Male Guise: https://www.youtube.com/watch?v=e53WkQvWOnE
- Female Guise: https://www.youtube.com/watch?v=pXuA7T0hCZo
- Faceless Guise: https://www.youtube.com/watch?v=2_-_tCsGj64

In each video, the robot's speech was accompanied by various gestures and changes in facial expressions to make the performance seem as natural and authentic as possible. I accomplished this using Virtual RoboThespian, a digital environment for creating performances for RoboThespian. I used this software to program the speech, as well as the gestures and facial expressions to accompany it. This was an extensive process of programming and revising the robot's body language to align with the speech as naturally as possible. For example, when the robot lists some of the tasks it can perform, I programmed it to emote with a serious and stern facial expression for the point when it mentions home security, followed by a smile when it mentions babysitting, as a means of conveying the different emotional connotations between the two tasks. The videos also had occasional cuts between wide shots of the robot's entire upper body and close-up shots of its face. The same body language and edits were used across all three conditions for consistency. Only the faces and voices differed.

After each participant was shown one of the three videos at random, they were then tasked with answering a series of questions (see Appendix A for a full listing of questions used in the experiments in Chapter 3) on seven-point Likert scales (from "Strongly disagree" to "Strongly agree") about trust and other related characteristics regarding how they felt about the robot. These questions were presented in five distinct sections. The sections were presented in random order, and the questions within each section were also ordered randomly. One section had questions about how much the participant would trust the robot with household chores such as cooking and babysitting. The motivation here was to see how comfortable people would be to have such a robot in their homes performing all of these tasks. For analytical purposes, I looked at this section as two composite variables: one for chores involving inanimate objects, such as cooking, cleaning and laundry; and one involving living entities, such as pet sitting and babysitting. I was particularly interested in seeing whether participants would be more hesitant to trust the robot with tasks in the latter category because of the greater amount of inherent responsibility involved.

Another section was about how much the participant would trust the robot to carry certain objects or entities related to the some of the chores in the previous section, such as carrying laundry, food, or a baby. I was interested in seeing whether respondents would give different ratings to carry more valuable objects or agents such as a computer or a baby, than they would to less valuable objects, such as food or laundry. To that end, I divided the questions in this section into two composite variables. One was for carrying food and laundry, and the other was for carrying a computer, money, or a baby. A factor analysis extracted a single component for these questions.

There was also a section was about rating the robot on a series of descriptive qualities such as how intelligent, likable, and conscious the robot seemed judging from the video (for example, "This robot seems intelligent", rated on a seven-point Likert scale from Strongly disagree to Strongly agree). These items were inspired by the Godspeed Questionnaire (Bartneck 2008), and they were intended to assess how much people could relate to the robot on a more personal level by asking respondents how friendly and likable they perceived the robot, how intelligent they perceived it to be, and how much they felt that the robot was conscious and alive. These appraisals were to be compared to the trust measures in a series of correlation and regression analyses to see if such appraisals informed participants' evaluation of trust.

For analytical purposes, the questions were aggregated into a set of composite variables. Each composite variable was created by obtaining the mean rating of its corresponding questions for each participant. These composites and their Cronbach's alphas are as follows:

- Trust to care for Living Agents (i.e., people and pets): α = 0.87
- Trust to care for Inanimate Objects: α = 0.90
- Trust with Carrying: α = 0.78
    - Trust to carry Inexpensive Items: 0.78
    - Trust to carry Valuable Items/Agents: 0.78
- Likability α = 0.83
- Intelligence: α = 0.85
- Aliveness: α = 0.91

After these questions, the participant was asked how many letters are in the word "elliptical" as an attention check. Participants who failed to answer correctly were dropped prior to analysis. After that, they were presented with the following demographic questions:

- Age
- Gender
- Have you ever been a parent?
- Have you ever been responsible for the care of a child? If so, how old was this child? (Check all that apply)
- Have you ever been a caregiver for a senior citizen?
- Have you ever owned a pet?

Guiding Questions

Experiment 1 was primarily motivated by the following guiding questions:

- What is the effect of guise condition on trust ratings and Godspeed-inspired measures?
- Do trust ratings and Godspeed measures significantly correlate with each other?
- Do the demographic questions significantly correlate with the trust ratings or Godspeed measures?
    - Would parents and non-parents respond differently to the idea of the robot babysitting?
    - Would participants who do and do not own pets respond differently to the idea of the robot pet sitting?
    - Would participants who have and have not been senior caregivers respond differently to the idea of the robot taking care of an elderly person?

Results

Effects of Guise condition on Trust

A series of one-way ANOVA tests found no significant effects of guise condition on Trust to care for Objects (*p* = 0.847, *M* = 4.84, *SD* = 1.40), Trust to care for Living Agents (*p* = 0.550, *M* = 2.72, *SD* = 1.46), Trust to carry Inexpensive Items (*p* = 0.657, *M* = 5.39, *SD* = 1.46), or Trust to carry Valuable Objects/Agents (*p* = 0.697, *M* = 3.38, *SD* =

1.56). With my planned comparisons of guise condition with the Godspeed-inspired measures, there were no significant main effects on perceived Intelligence (*p* = 0.450, *M* = 5.03, *SD* = 1.35) or Likability (*p* = 0.573, *M* = 4.50, *SD* = 1.50), but there was an effect on perceived Aliveness (*p* = 0.023, *M* = 4.04, *SD* = 1.79), with the Male guise (*M* = 4.32, *SD* = 1.74) rating significantly higher than the Faceless guise (*M* = 3.77, *SD* = 1.82) (Figure 2). The Female guise (*M* = 4.03, *SD* = 1.76) was also rated as more Alive than the Faceless guise, but not significantly so (*p* = 0.197).

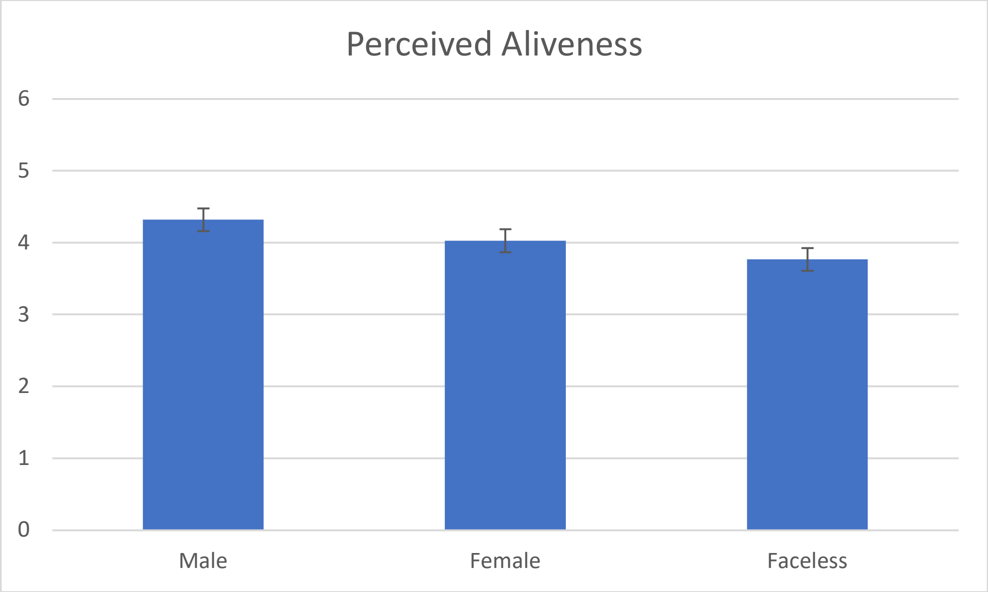

Fig. 2: Ratings for perceived Aliveness for each guise condition. The difference between the Male and Faceless conditions is statistically significant. Error bars are calculated from standard error.

Correlations between Trust ratings and Godspeed measures

The Godspeed-inspired measures of perceived Intelligence, Likability, and Aliveness were highly positively correlated with both Trust measures and both Carry measures (Table 1), but when the Godspeed measures were simultaneously regressed with each of the four trust variables in a multiple regression, Trust with Objects and both Carry measures were only predicted by perceived Intelligence (**Objects:** *b* = 0.46, *ß* = 0.45, *p* < 0.001; **Carry Inexpensive Items:** *b* = 0.49, *ß* = 0.46, *p* < 0.001; **Carry Valuables:** *b* = 0.27, *ß* = 0.24, *p* < 0.001) and Likability (**Objects:** *b* = 0.25, *ß* = 0.27, *p* < 0.001; **Carry Inexpensive Items:** *b* = 0.23, *ß* = 0.23, *p* < 0.001; **Carry Valuables:** *b* = 0.38, *ß* = 0.36, *p* < 0.001), while the effects of perceived Aliveness (**Objects:** *b* = -0.002, *ß* = -0.002, *p* = 0.960; **Carry Inexpensive Items:** *b* = -0.03, *ß* = -0.04, *p* = 0.408; **Carry Valuables:** *b* = 0.07, *ß* = 0.08, *p* = 0.098) washed out. Conversely, perceived Likability (*b* = 0.32, *ß* = 0.33, *p* < 0.001) and Aliveness (*b* = 0.19, *ß* = 0.23, *p* < 0.001) predicted Trust with Agents, while perceived Intelligence (*b* = 0.08, *ß* = 0.08, *p* = 0.219) did not.

Table 1: Correlation coefficients (below diagonal) with associated p-values (above diagonal) for trust ratings and Godspeed-inspired measures.

	Trust with Objects	Trust with Agents	Carry food/clothes	Carry Valuables	Likability	Intelligence	Aliveness
Trust with Objects	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Trust with Agents	0.581432	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Carry food/clothes	0.761235	0.397777	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Carry Valuables	0.699511	0.726294	0.603592	1	p < 0.001	p < 0.001	p < 0.001
Likability	0.586896	0.516432	0.538523	0.584048	1	p < 0.001	p < 0.001
Intelligence	0.636389	0.472547	0.594945	0.55797	0.725005	1	p < 0.001
Aliveness	0.460768	0.473056	0.408817	0.459584	0.583751	0.691765	1

A factor analysis found that the questions related to trusting the robot with chores divided into two components, with one group representing chores involving inanimate objects, such as cooking and laundry, and the other representing chores involving living agents, such as pets and children. One question for taking care of valuable items was more evenly correlated with the two components (0.537 in the component that was predominantly about living agents, and 0.581 in the component that was predominantly about inanimate objects).

Effects of Participant Demographics

I was interested in exploring possible connections between responses to certain demographic questions and specific trust-related questions. I conducted a series of t-tests comparing responses to the question, "Have you ever been a parent?" to responses to questions from the Trust with Chores category about babysitting infants, young children, and teenagers. I conducted a similar series of tests with the question, "Have you ever been a caregiver for a senior citizen?", to see if it influenced participants' trust in the robots to care for an elderly person, as well as an additional comparison involving caring for pets. However, these tests found no significant effects (*p* = 0.403-0.844). A series of regression analyses found no significant effects of age on any of the trust or Godspeed measures(*p* = 0.102-0.995). However, a series of t-tests did find significant effects of participant gender ((n_{Male} = 248, n_{female} = 227, excluding non-binary participants because of a small sample size of 5) on Trust with Objects (*p* = 0.029; **Male:** *M* = 4.97, *SD* = 1.34; **Female:** *M* = 4.69, *SD* = 1.46), Trust with Agents (*p* = 0.004; **Male:** *M* = 2.89, *SD* = 1.49; **Female:** *M* = 2.50, *SD* = 1.39), and Trust to Carry Valuable Items (*p* = 0.042; **Male:** *M* = 3.51, *SD* = 1.57; **Female:** *M* = 3.21, *SD* = 1.54).

Discussion

The primary purpose of Experiment 1 was to explore possible effects of anthropomorphism and the apparent gender of a robot on trust, perceived likability, perceived intelligence, and perceived aliveness. Whether presented in a gendered or gender-neutral way, this humanoid robot had similar responses with respect to trust and appraisals. Given that the mean ratings for these variables were consistently high across all conditions, the highly anthropomorphic appearance of even the Faceless guise may have had a role in the lack of significant main effects of guise condition on trust and appraisals. Although, the fact that the Faceless guise was rated as significantly less Alive than the Male guise may suggest otherwise, or it may suggest that aliveness is not crucial for these aspects of trust. Perhaps in the absence of any other effects of condition, the effect of Aliveness could simply be a manipulation check of how anthropomorphic each robot appeared to be to the participant. For the replication in Experiment 2, I used the same stimuli, but I altered the visual presentation of the survey by including a screenshot of the robot on each of the question blocks after the video to see whether participants were simply forgetting the robot's appearance as they were taking the survey.

Although Experiment 1 did not find any significant effects of guise condition on any of the trust measures, the relationships found between trust ratings and Godspeed measures indicated that these measures are generally a useful methodology for characterizing trust within the context of this research. Appraisals of Intelligence were closely linked to Trust with Objects, but not Agents, whereas Likability and Aliveness were important for Trust with Agents, but not Objects. This suggests thematic links between particular appraisals and particular trust contexts. As such, Experiment 2 was predominantly focused on introducing new metrics with a similar focus on trust and appraisals to generate more points for analysis and comparison while retaining the stimuli and format of Experiment 1. Experiment 2 added additional sets of questions for trusting the robot with sensitive information, sympathizing with the robot if it were physically harmed, and expressing opinions of automated systems in general. I added trust with sensitive information as a metric because of the precedent that already exists for trusting computers and mobile devices to store information such as passwords and bank account numbers. This would be an interesting variable to compare with the other trust ratings to see if such a precedent would lead to different responses. Furthermore, unlike household chores, storing sensitive information is not a physically demanding task, and so the trust with information metric, in addition to the Godspeed measures, was a way of checking whether participants' ratings for trust with chores were influenced by perceptions of the robot's ability to physically carry out those tasks based on its physical appearance. The questions about sympathizing with the robot were added to measure how much participants would perceive the robot as a conscious entity capable of experiencing emotions and pain. The questions about trusting automated systems in general were introduced to obtain a sense of respondents' preconceived notions of how much they trust automation on a broader scale, and to see the extent to which those preconceived notions would influence these particular feelings of trust for this particular robot. The questions related to carrying items were dropped because of being largely redundant with the Trust with Chores questions in subject matter and responses.

Experiment 2

Experiment 2 replicated the overall design of Experiment 1 with the inclusion of more trust-related questions for outcome measures, as well as a series of questions about participants' preconceived notions about automated systems in general. I also added some questions about how sympathetic participants would feel towards the robot if it were physically damaged, to see if participants would empathize with the robot and perceive it as a conscious entity capable of experiencing pain. I also altered the visual presentation of the survey by having an image of the robot's face at the top of each page with questions related to the dependent variables. This was done to give participants a visual reminder of the robot they saw in the video and what it looked like.

Methods

Participants

For Experiment 2, I increased the payment amount for participants to $0.65 because of the increased length of the survey. The recruitment process and exclusion criteria were otherwise identical to those of Experiment 1, with 500 U.S. participants recruited through Amazon Mechanical Turk. As with Experiment 1, a power analysis found that a minimum of 159 participants would be needed for 80% power across three experimental groups. I chose 500 participants to account for potential low-quality responses such as failed attention checks, and to be consistent with Experiment 1. The filtered sample consisted of 486 participants, with 285 male, 200 females, and one non-binary. Participant age ranged between 20 and 78, with an average age of 39 and a standard deviation of 12.

Procedure

Using the same procedure as Experiment 1, participants were randomly assigned to view one of three videos of a robot pretending to be a household caretaker and giving a brief speech about its capabilities. Conditions were assigned to participants randomly, In each video, the robot gives a brief speech introducing itself to the viewer and explaining its capabilities. The speeches and footage were the same as Experiment 1.

After each participant was shown one of the three videos, they were then tasked with answering a series of questions on seven-point Likert scales (from “Strongly disagree” to “Strongly agree”) about trust, appraisals inspired by the Godspeed Questionnaire, and sympathy for the robot if it were physically harmed, as well as some questions about how they felt about automated systems in general, based on the Perfect Automation Schema (PAS) (Merritt 2015). The PAS subscales are designed to assess people’s preconceived notions about the extent to which they expect perfect performance from artificial agents. There are two subscales: one measures the extent to which one considers automated systems to be flawless and incapable of making mistakes, and the other measures the extent to which one considers any mistake made by an automated system to render the system entirely useless. These questions were presented on five pages. The pages were presented in random order, and the questions within each page were also ordered randomly. There were questions about how much the participant would trust the robot with household chores such as cooking and babysitting, with all the same questions as the corresponding section of Experiment 1. For analytical purposes, I was particularly interested in looking at this first block both as a single composite variable for trusting the robot with all household chores in general, as well as bifurcating the block as I did in Experiment 1, into two composites for trusting the robot with chores involving inanimate objects such as cooking and cleaning, and chores involving living agents such as babysitting and pet sitting. I was interested in analyzing the extent to which these two composites would break apart as separate categories. As with Experiment 1, I conducted a factor analysis on these questions and found similar results to Experiment 1, with questions separated into two components for chores involving inanimate objects and chores involving living agents. Also similar to Experiment 1 was that the question about taking care of one's valuables was evenly correlated with the Objects and Agents components (0.575 and 0.531, respectively).

For this experiment, I included a set of metrics about how much the participant would trust the robot with sensitive information such as credit card information or social security numbers. This would be a potentially interesting contrast with the first set, as there is already considerable precedent for storing such information in devices like laptops and smartphones. This raises the question of whether trust in such devices would translate to trust in a humanoid robot.

The participant was also asked to imagine a scenario in which the robot was physically attacked by a person biased against robots, ending with the robot being badly damaged. The questions were about how much sympathy the participant would feel for the robot in that situation. This has to do with how much people would empathize with the robot, serving as an indirect way of intuitively assessing the robot as capable of emotions such as suffering and was informed by the work of Riek et. al. (2009), who found that people would empathize more with a robot being treated cruelly by a human actor when the robot was more anthropomorphic in appearance. Another section was about the Godspeed-inspired ratings of perceived Intelligence, Likability, and Aliveness, as in Experiment 1. For Experiment 2, this block also included questions related to stereotypically masculine and feminine traits such as confidence and warmth, respectively, to see how much the gender variations would influence perception of those qualities. This decision was informed by my research of existing literature that found a precedent for such metrics (Eyssel 2012).

Another page was about the participant's preconceived notions of automated systems in general. These questions were related to the Perfect Automation Schema (PAS) (Merritt 2015), concerning the extent to which people trust automated systems to always make correct decisions. These questions came in two different types: one related to how much automated systems can be trusted to function perfectly (For example, "Automated systems have 100% perfect performance"), and one related to how much systems that do not function properly *cannot* be trusted ("If an automated system makes a mistake, then it is completely useless"). These questions were meant to be compared to the trust, sympathy, and Godspeed ratings in correlation and regression analyses to see whether such preconceived notions about autonomous systems would influence participants' feelings about these particular stimuli.

With Experiment 2, I changed the visual presentation of the survey by having each of the aforementioned question pages include a screenshot of the face of the robot whose video the participant was assigned to at the top of each page. This was done to serve as a visual reminder of the robot's appearance while the participant answered the questions, which was especially important for this survey since it was longer than Experiment 1.

Composite variables were created by obtaining the mean ratings of the corresponding questions for each participant. The composites used in Experiment 2 and their Cronbach's alphas are as follows:

- Trust to care for Living Agents (i.e., people and pets): α = 0.88

- Trust to care for Inanimate Objects: α = 0.88
- Trust with Information: α = 0.92
- Sympathy: α = 0.97
- Likability α = 0.88
- Intelligence: α = 0.81
- Aliveness: α = 0.80
- Dominance (Stereotypically masculine): α = 0.85
- Nurturance (Stereotypically feminine): α = 0.71
- PAS high expectations: PAS questions related to the trustworthiness of automated systems,
  α = 0.86
- PAS all-or-nothing: PAS questions related to malfunctioning automated systems, α = 0.85

After these questions, the participant was asked how many letters are in the word “elliptical” as an attention check. Participants who failed to answer correctly were dropped prior to analysis. After that, they were presented with the following demographic questions:

- Age
- Gender
- Have you ever been a parent?
- Have you ever been responsible for the care of a child? If so, how old was this child? (Check all that apply)
- Have you ever been a caregiver for a senior citizen?
- Have you ever owned a pet?

Guiding Questions

Experiment 2 was primarily motivated by the following guiding questions:

- What is the effect of guise condition on trust ratings and Godspeed-inspired measures?
- Do Trust with Living Agents and Trust with Objects break apart as distinct measures?
- Do PAS High Expectations and PAS All-or-Nothing break apart as distinct measures?
- What is the effect of PAS measures on trust ratings?
- What effect do the trust ratings and Godspeed measures have on each other?
- Is sympathy towards the robot affected by guise condition?
- Does sympathy correlate with the Godspeed and PAS measures?

Results

Effects of Guise condition on Trust

A series of one-way ANOVA tests found no significant main effects of guise condition on the planned comparisons of Trust with Objects (*p* = 0.128, *M* = 4.97, *SD* = 1.45), Trust with Agents (*p* = 0.363, *M* = 3.05, *SD* = 1.63), Trust with Information (*p* = 0.858, *M* = 3.43, *SD* = 1.87), or Sympathy (*p* = 0.166, *M* = 5.04, *SD* = 1.62).

Correlations between Trust ratings and Godspeed measures

The Godspeed-inspired measures of perceived Intelligence (*M* = 5.12, *SD* = 1.31), Likability (*M* = 4.60, *SD* = 1.59), and Aliveness (*M* = 4.18, *SD* = 1.69) are highly positively correlated with all three trust measures (Table 2), but when the Godspeed measures are compared to each trust measure in simultaneous multiple regressions, only perceived Intelligence (**Objects:** *b* = 0.47, *ß* = 0.42, *p* < 0.001; **Info:** *b* = 0.26, *ß* = 0.18, *p* = 0.003) and Likability (**Objects:** *b* = 0.29, *ß* = 0.32, *p* < 0.001; **Info:** *b* = 0.29, *ß* = 0.25, *p* < 0.001) predict Trust with Objects or Trust with Information, while the effects of perceived Aliveness (**Objects:** *b* = -0.05, *ß* = -0.05, *p* = 0.261; **Info:** *b* = 0.06, *ß* = 0.06, *p* = 0.311) wash out. By contrast, perceived Likability (*b* = 0.42, *ß* = 0.41, *p* < 0.001) and Aliveness (*b* = 0.17, *ß* = 0.18, *p* < 0.001) predict Trust with Agents, while the effects of Intelligence (*b* = 0.08, *ß* = 0.07, *p* = 0.231) wash out. These differences between Trust with Objects and Trust with Agents regarding their correlations with the three Godspeed measures were consistent with those of Experiment 1.

Table 2: Correlation coefficients (below diagonal) and corresponding p-values (above diagonal) for trust ratings and Godspeed-inspired measures.

	Objects	Agents	Information	Sympathy	Likability	Intelligence	Aliveness
Objects	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Agents	0.578945	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Information	0.54621	0.624682	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Sympathy	0.475463	0.41969	0.32206	1	p < 0.001	p < 0.001	p < 0.001
Likability	0.581229	0.567705	0.412304	0.575583	1	p < 0.001	p < 0.001
Intelligence	0.613921	0.461607	0.391887	0.547722	0.70011	1	p < 0.001
Aliveness	0.407173	0.474405	0.324909	0.517415	0.627436	0.614829	1

Guise condition did not have main effects on any of the Godspeed-inspired measures of perceived intelligence (*p* = 0.882), likability (*p* = 0.664), aliveness (*p* = 0.521), dominance (*p* = 0.457, *M* = 4.38, *SD* = 1.27), or nurturance (*p* = 0.630, *M* = 4.34, *SD* = 1.49). Similarly, there was no main effect of condition on sympathy (*p* = 0.166). The highly anthropomorphic character of even the minimally anthropomorphic guise appears to have interfered with the intent to vary levels of anthropomorphism.

Effects of Perfect Automation Schema Scores

I ran a series of interaction tests with both PAS measures (High Expectations and All-or-nothing thinking) and dummy variables for the male and female guise conditions, with the faceless guise as a reference variable. These regressions found that those who rated greater scores for PAS High Expectations also rated higher trust scores with respect to both Trust with Objects and Trust with Information (but *not* Trust with Agents) for the male guise (**Objects:** $b_{PAS\_HE*Male}$ = 0.38, ß$_{PAS\_HE*Male}$ = 0.43, $p_{PAS\_HE*Male}$ = 0.001; **Information:** $b_{PAS\_HE*Male}$= 0.31, ß$_{PAS\_HE*Male}$= 0.27, $p_{PAS\_HE*Male}$ = 0.028), but *not* the female guise (**Objects:** $b_{PAS\_HE*Female}$ = 0.09, ß$_{PAS\_HE* Female}$ = 0.11, $p_{PAS\_HE* Female}$ = 0.419; **Information:** $b_{PAS\_HE* Female}$ = 0.20, ß$_{PAS\_HE* Female}$= 0.19, $p_{PAS\_HE* Female}$= 0.137). There was also interaction between the male guise and PAS All or Nothing regarding Trust with Information but *not* Trust with Objects or Agents (**Information:** $b_{PAS\_AON*Male}$ = 0.33, ß$_{PAS\_AON*Male}$ = 0.34, $p_{PAS\_AON*Male}$ = 0.033; **Objects:** $b_{PAS\_AON*Male}$ = 0.20, ß$_{PAS\_AON*Male}$ = 0.27, $p_{PAS\_AON*Male}$ = 0.100; **Agents:** $b_{PAS\_AON*Male}$ = 0.20, ß$_{PAS\_AON*Male}$ = 0.27, $p_{PAS\_AON*Male}$ = 0.100). There were no such interactions for the female guise (**Information:** $b_{PAS\_AON*Female}$ = 0.27, ß$_{PAS\_AON* Female}$ = 0.27, $p_{PAS\_AON* Female}$ = 0.092; **Objects:** $b_{PAS\_AON* Female}$ = 0.03, ß$_{PAS\_AON* Female}$ = 0.04, $p_{PAS\_AON* Female}$ = 0.802; **Agents:** $b_{PAS\_AON* Female}$ = 0.12, ß$_{PAS\_AON* Female}$ = 0.14, $p_{PAS\_AON* Female}$ = 0.383).

PAS High Expectations also correlates strongly with Sympathy (r = 0.30), while PAS All-or-Nothing (r = 0.10) does not. Intelligence, Likability, and Aliveness also correlate with Sympathy (Table 2), so additional testing involved a series of exploratory mediation analyses to assess whether the correlation between PAS High Expectations and Sympathy was accounted for by any of the three Godspeed measures. Conducted using PROCESS for SPSS, I tested each of the three Godspeed measures individually as mediators of the correlation between PAS High Expectations and Sympathy. These tests found that Likability mediated the effect between High Expectations and Sympathy. The direct effect of High Expectations on Sympathy was no longer significant (b = 0.09, SE = 0.05, p = 0.055, 95% CI [-0.002, 0.18]), while the indirect effect of Likability on Sympathy was significant (b = 0.56, SE = 0.04, p < 0.001, 95% CI [0.47, 0.64]) (Figure 3). However, neither Intelligence nor Aliveness had a mediating effect, with the direct relationship between High Expectations and Sympathy remaining significant in mediation tests with both Godspeed measures. High Expectations also correlated strongly with Trust with Objects (r = 0.33), Trust with Agents (r = 0.57), and Trust with Information (r = 0.45), but these effects were not mediated by any of the Godspeed measures.

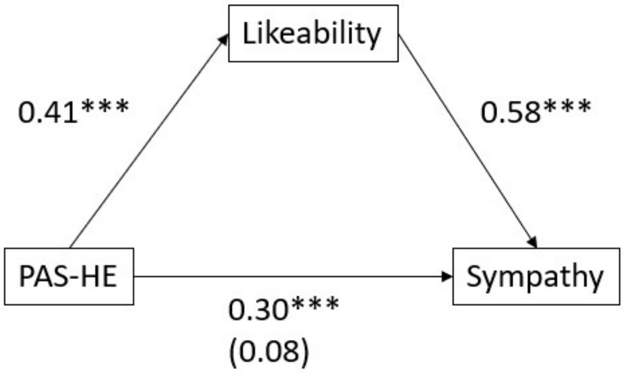

Fig. 3: Standardized regression coefficients for the relationship between PAS_High Expectations and Sympathy as mediated by perceived Likability. The standardized regression coefficient between PAS_HE and Sympathy with the mediator included is given in parentheses. *** p < 0.001

Effects of Participant Demographics

As with Experiment 1, I conducted a series of t-tests comparing responses to the question, “Have you ever been a parent?” to responses to questions from the Trust with Chores category about babysitting infants, young children, and teenagers. Those who responded “yes” (n_{yes} = 222, n_{no} = 264) to having children corresponded with higher ratings on all three of these trust questions (**Infants:** p = 0.001, M_{Yes} = 2.62, M_{No} = 2.15; **Young children:** p < 0.001, M_{Yes} = 2.93, M_{No} = 2.42; **Teens:** p = 0.031, M_{Yes} = 3.79, M_{No} = 3.39), an intriguing discrepancy from Experiment 1. Similar comparisons involving caring for pets (p = 0.399) and senior citizens (p = 0.493) found no significant effects across demographic responses, as in Experiment 1. Regression analysis for questions about participant age with the dependent variables found age positively correlated with Trust with Agents (p = 0.002) and Nurturance (p = 0.041). A series of t-tests looking at main effects of gender on Trust with Objects, Trust with Agents, and the Godspeed-inspired measures found no significant effects, with the exception of a statistically significant effect between gender and Dominance, with female participants reporting higher perceived dominance (M_{Male} = 4.29, M_{Female} = 4.52, p = 0.043).

Discussion

The primary purpose of Experiment 2 was to replicate Experiment 1 with additional variables to develop a more comprehensive set of metrics for different types of trust, different types of appraisals, and people’s feelings about automated systems in general. I also wanted to alter the visual presentation to see if that would affect the relationship between guise condition and the dependent variables. The fact that there were no effects of guise condition on the dependent variables seems to indicate that this attempted manipulation of anthropomorphism failed to sufficiently vary the degree of anthropomorphism, given that the faceless guise is virtually identical in appearance and mannerisms to the other two conditions in all respects *except* its face and voice. This possibility is reinforced by the Aliveness rating, which serves as a manipulation check for manipulating anthropomorphism. The fact that, unlike what I found in Experiment 1, Aliveness did not significantly differ across guise conditions indicates that the guises themselves did not significantly differ in anthropomorphic appearance. Perhaps the contrast was insufficient for participants to see the faceless guise as a truly non-anthropomorphic robot. Fong et. al. classify robot appearance into four distinct categories: anthropomorphic, zoomorphic, caricatured, and functional (Fong 2003). The RoboThespian guises used here, including the faceless condition, could all be characterized as anthropomorphic. Accordingly, for Experiment 3, I added a fourth condition to this set in the form of a robot with a more distinctly functional appearance and mannerisms to provide a clearer contrast with the RoboThespian conditions.

The strong correlations that Intelligence and Likability had with Trust with Objects and Trust with Information suggest that participants were thinking about trust both in terms of trusting that the robot was physically and computationally capable of performing these tasks, because of the correlation with Intelligence, as well as in terms of being comfortable with its presence on a more emotional level, because of the correlation with Likability. This would have potentially interesting implications on how people think about trust in the context of HRI, and it has some commonality with Law and Scheutz’s literature review on trust in HRI where they devised the categories of performance-based trust (related to a robot's ability to complete tasks without the need to be monitored) and relation-based trust (related to the robot's ability to be part of society in a capacity beyond its ability to carry out a particular job) (Law 2021). This raises the question of which category people would prioritize more when interacting with robots, and which one they think of more when evaluating trust in a broader sense, even outside the context of human-machine interaction.

The differences in Godspeed relationships with the Trust with Agents and Trust with Objects variables in both Experiments 1 and 2 indicate that the two measures break apart as distinct factors. In particular, the results of the factor analysis on all the questions related to chores, in conjunction with the fact that Trust with Objects was strongly predicted by Intelligence, while Trust with Agents was not, potentially relate to the aforementioned distinction between capability and integrity when thinking about trust. Perhaps participants conceptualized trust differently for each of the two measures. The fact that Intelligence was a strong predictor of Trust with Objects and Trust with Information while Aliveness was not may indicate that such trusting the robot with tasks that involve inanimate objects and information may be more of a matter of trusting the robot's performance and capability. Conversely, the fact that Aliveness was a stronger predictor of Trust with Agents may indicate that trusting a robot to care for a child or a pet is more of a matter of the extent to which one perceives the robot as a conscious entity. Under this interpretation, being akin to a living thing is an important consideration in people's ability to trust an agent with childcare or taking care of a pet, while tasks involving responsibilities such as cleaning or protecting information are perceived to have more to do with computational performance.

The interaction effects involving the Perfect Automation Schema variables found that those who rated higher expectations on the PAS High Expectations questions trusted the male guise more with both tasks and information. If the correlation between these measures is indicative of a causal relationship of PAS High Expectations affecting trust in the male guise, it may imply that high expectations of automation are related to gender bias. By contrast, with the other two guise conditions, perceived performance of automated systems in general was not as significant of a predictor of trust. Thus, while the male robot did not inherently inspire more trust than a female or non-anthropomorphic robot, it did inspire more trust in people with high expectations of automated systems than the other two conditions would have.

The fact that PAS High Expectations was a strong predictor of Sympathy while PAS All-or-Nothing was not is a particularly crucial result, as it highlights a clear distinction between the two PAS measures. When Merritt et. al. (2015) conceived the Perfect Automation Schema, they argued for the possibility of a distinction between High Expectations and All-or-Nothing thinking, such as observers having high expectations of a system while also believing that it could recover from occasional failures. Perhaps such a phenomenon explains the difference in results here, where the All-or-Nothing category is too unforgiving of performance errors to be strongly correlated with Sympathy, while High Expectations is not. Perhaps the exceptionally stark perception that All-or-Nothing thinking towards robots entails makes it more difficult for people who adopt such thinking to empathize with robots. If one perceives a robot as useless when it is functioning less than perfectly, that may inhibit one's ability to relate to robots emotionally.

It is intriguing to note the result of participants who had children trusting the robots more with babysitting, particularly because Experiment 1 found not such effects. One would expect the opposite phenomenon to be more likely because of parents being generally found to be more risk-averse, particularly in matters pertaining to childcare (Eibach, 2010). It is possible that because those with experience as parents are more accustomed to the associated responsibilities, they would be more willing to trust a third party with those responsibilities than those without that experience who may be imagining childcare to be more challenging and intimidating than it truly is.

Although Experiment 2 did not find any significant effects of guise condition on the dependent variables, the relationships found among the trust ratings, Godspeed measures, and PAS measures reinforced the results of Experiment 1 that these metrics were a useful set of variables for understanding and characterizing trust and appraisals within the context of this research. As such, Experiment 3 added a fourth guise condition in the form of a more distinctly non-anthropomorphic robot, but otherwise closely replicated this design.

Experiment 3

Experiment 3 added a fourth guise condition in the form of a more distinctly non-anthropomorphic robot that looked entirely different from the three RoboThespian conditions. I also included some questions about participants' preconceived notions about men and women in general for comparison with the Male and Female guise conditions.

Methods

Participants

The recruitment process and exclusion criteria for Experiment 3 were identical to those of Experiment 2, with 500 U.S. participants recruited through Amazon Mechanical Turk for a fee of $0.65. A power analysis found that a minimum of 180 participants would be needed for 80% power across four experimental groups. I chose 500 participants to account for potential low-quality responses such as failed attention checks, and to be consistent with previous studies. The filtered sample consisted of 469 participants, with 246 male, 220 females, and 3 non-binary. Participant age ranged between 21 and 76, with a mean age of 43 and standard deviation of 13.

Procedure

For Experiment 3, I wanted to include an additional experimental condition that would contrast more sharply in appearance to the anthropomorphic RoboThespian conditions. To that end, I used footage of Evacbot, a more distinctly non-anthropomorphic robot (Figure 4). I overlaid this footage with audio from the speech I used for the Faceless RoboThespian guise. The video has Evacbot gesticulating with its mechanical arms during the speech in a way intended to match the overall animacy of the RoboThespian conditions, but in a distinctly inhuman manner. This stimulus was generated in collaboration with Alan Wagner and based on his work on Turtlebot, a robot system of which Evacbot is an elaboration (Howard 2017). The RoboThespian conditions were unchanged from the previous studies, as was the procedure for assigning conditions to participants.

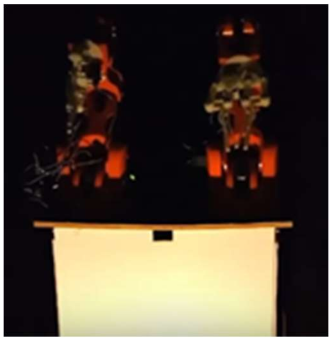

Fig. 4: The new Evacbot condition used in Experiment 3. The video can be viewed at https://www.youtube.com/watch?v=gGs2RcjYwc8

All the survey questions from Experiment 2 were reused for Experiment 3, with some new additions in the form of two 10-point scales for participants to rate their feelings towards men and women in general. I was interested in seeing whether participants' preconceived notions about gender would influence their feelings towards the male and female robots. The use of 10-point scales for these questions was to allow for more gradation for respondents to rate such sentiments, particularly because of the lack of descriptive anchors such as the scales of Strongly Agree to Strongly Disagree used in the trust questions and appraisals.

As with the previous studies, composite variables were created by obtaining the mean ratings of the corresponding questions for each participant. The composites used in Experiment 3 and their Cronbach's alphas are as follows:

- Trust to care for Living Agents (i.e., people and pets): α = 0.89
- Trust to care for Inanimate Objects: α = 0.93
- Trust with Information: α = 0.98
- Sympathy: α = 0.90
- Likability α = 0.82
- Intelligence: α = 0.84
- Aliveness: α = 0.89
- Dominance (Stereotypically masculine): α = 0.76
- Nurturance (Stereotypically feminine): α = 0.86
- PAS high expectations: PAS questions related to the trustworthiness of automated systems,
  α = 0.85
- PAS all-or-nothing: PAS questions related to malfunctioning automated systems, α = 0.79

After these questions, the participant was asked how many letters are in the word "elliptical" as an attention check. Participants who failed to answer correctly were dropped prior to analysis. After that, they were presented with the following demographic questions:

- Age
- Gender
- Have you ever been a parent?

- Have you ever been responsible for the care of a child? If so, how old was this child? (Check all that apply)
- Have you ever been a caregiver for a senior citizen?
- Have you ever owned a pet?

The new questions related to warmth towards men and women followed the demographic questions, with a 10-point scale for each ranging from Extremely Capable to Extremely Incapable.

Guiding Questions

In addition to continuing to explore the guiding questions from Experiment 2, Experiment 3 was primarily motivated by the following questions:

- Does Evacbot significantly differ from RoboThespian conditions in the trust ratings or Godspeed measures?
- Do feelings about the capability of men and women correlate with trust ratings for the male and female guise conditions?

Results

Effects of Guise condition on Trust

Evacbot rated significantly lower than the RoboThespian conditions on Trust with Objects (*p* < 0.001), Trust with Agents (*p* = 0.093), and Trust with Information (*p* = 0.041). Evacbot was also rated as significantly less Likable (*p* = 0.025), Intelligent (*p* = 0.021), Alive (*p* < 0.001), and Nurturing (*p* = 0.002). As was the case with Experiment 1, the Faceless RoboThespian guise was rated as significantly less Alive (*M* = 3.77, *SD* = 1.83) than the Male guise (*M* = 4.22, *SD* = 1.75), with the Female guise (*M* = 3.99, *SD* = 1.86) being rated slightly lower than the Male, but not significantly so (*p* = 0.335). Aside from these effects, the three RoboThespian conditions did not significantly differ from each other in any of the Trust measures or robot Appraisals. Figures 5 and 6 show the mean ratings for each guise condition for Trust measures and Godspeed appraisals, respectively.

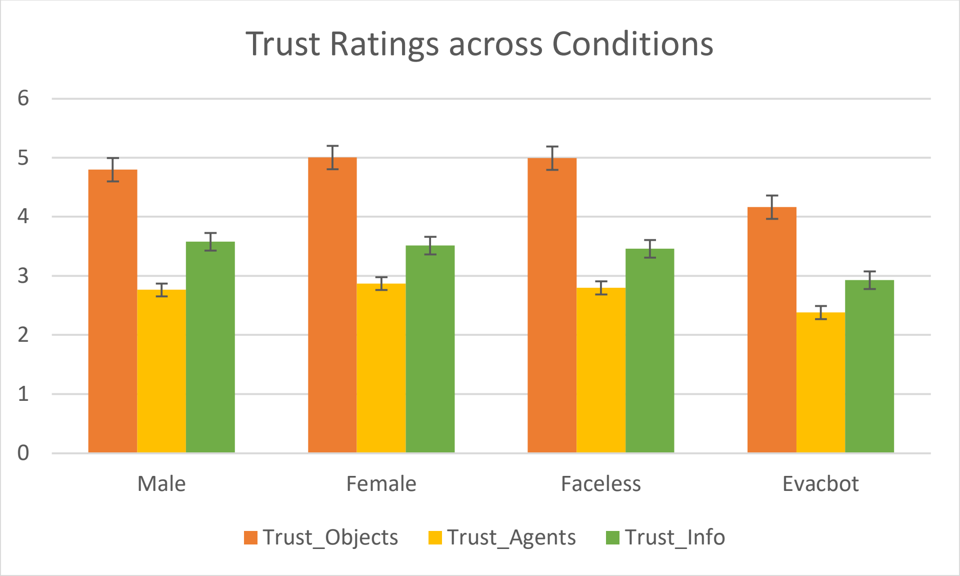

Fig. 5: Mean trust ratings for each guise condition. The differences between Evacbot and the RoboThespian conditions are statistically significant. Error bars are calculated from standard error.

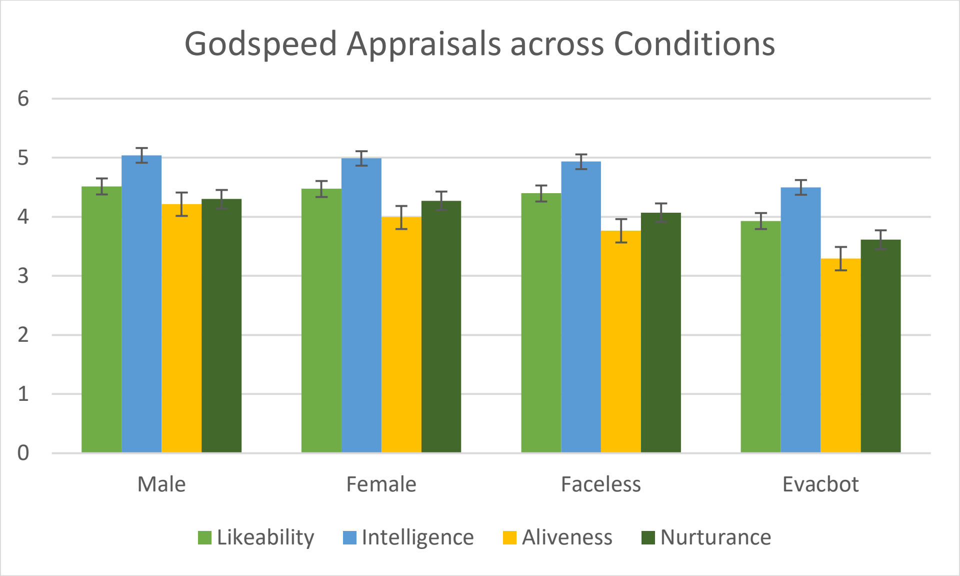

Fig. 6: Mean appraisal ratings for each guise condition. The differences between Evacbot and the RoboThespian conditions are statistically significant. The difference between the Faceless and Male guises in perceived Aliveness is also significant. Error bars are calculated from standard error.

Correlations between Trust ratings and Godspeed measures

The Godspeed-inspired measures of Likability (*M* = 4.33, *SD* = 1.66), Intelligence (*M* = 4.87, *SD* = 1.49), and Aliveness (*M* = 3.82, *SD* = 1.84) were significantly correlated with the three trust measures (Table 3). When Likability, Intelligence, and Aliveness are compared to each of the trust measures in a simultaneous multiple regression, Trust with Objects and Trust with Information are only predicted by Likability (**Objects:** *b* = 0.27, *ß* = 0.30, *p* < 0.001; **Info:** *b* = 0.36, *ß* = 0.30, *p* < 0.001) and Intelligence (**Objects:** *b* = 0.53, *ß* = 0.51, *p* < 0.001; **Info:** *b* = 0.21, *ß* = 0.16, *p* = 0.007), with the effect of Aliveness (**Objects:** *b* = -0.05, *ß* = -0.07, *p* = 0.18; **Info:** *b* = 0.09, *ß* = 0.09, *p* = 0.14) washing out. Conversely, Trust with Agents is predicted by Likability (*b* = 0.37, *ß* = 0.38, *p* < 0.001) and Aliveness (*b* = 0.19, *ß* = 0.22, *p* < 0.001), with Intelligence (*b* = 0.06, *ß* = 0.06, *p* = 0.33) washing out. These results are consistent with Experiment 2.

Table 3: Correlation coefficients (below diagonal) and corresponding p-values (above diagonal) for trust ratings and Godspeed-inspired measures.

	Objects	Agents	Information	Sympathy	Likability	Intelligence	Aliveness
Objects	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Agents	0.586375	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Information	0.571528	0.631484	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Sympathy	0.428767	0.30851	0.273543	1	p < 0.001	p < 0.001	p < 0.001
Likability	0.612831	0.567942	0.478127	0.524541	1	p < 0.001	p < 0.001
Intelligence	0.677804	0.466909	0.434504	0.483188	0.705119	1	p < 0.001
Aliveness	0.47925	0.514152	0.403956	0.422789	0.68786	0.664946	1

Because of the significant differences between Evacbot and the RoboThespian conditions on the Godspeed measures of Likability, Intelligence, Aliveness, and Nurturance, I conducted a series of regression analyses to see if any of these variables mediated the effects of guise condition on the Trust ratings, with the RoboThespian guises pooled together into a single condition because of the lack of significant differences among the three. In a series of simultaneous multiple regression analyses with guise condition and all four Godspeed measures together, I found that the effect of condition on Trust with Information was fully mediated by Likability (*b* = 0.26, *ß* = 0.22, *p* = 0.008) and Intelligence (*b* = 0.20, *ß* = 0.15, *p* = 0.014), while Trust with Objects was partially mediated by Likability (*b* = 0.24, *ß* = 0.26, *p* < 0.001) and Intelligence (*b* = 0.52, *ß* = 0.50, *p* < 0.001), and the effect of condition on Trust with Agents was fully mediated by Likability (*b* = 0.19, *ß* = 0.19, *p* = 0.015), Aliveness (*b* = 0.12, *ß* = 0.14, *p* = 0.017), and Nurturance (*b* = 0.31, *ß* = 0.30, *p* < 0.001). Tables 4 and 5 show the direct effects of guise condition on Trust ratings before and after the inclusion of the mediators, respectively.

Table 4: Direct effects of guise condition prior to the inclusion of Godspeed mediators

	b	ß	p
Objects	-0.26	-0.22	< 0.001
Agents	-0.14	-0.12	0.013
Information	-0.2	-0.13	0.005

Table 5: Direct effects of guise condition after the inclusion of Godspeed mediators

	b	ß	p
Objects	-0.14	-0.12	0.001
Agents	-0.02	-0.01	0.76
Information	-0.08	-0.05	0.225

Effects of Male vs. Female Warmth

With the questions about perceived capability of men and women, I was interested in comparing those responses with trust ratings for the Male and Female guise conditions to see if there was a predictive effect of higher perceived capability of a gender correlating with more trust in the robot of the corresponding gender. To that end, I ran a series of interaction tests with both of the capability measures and dummy variables for the male and female guise conditions. I did not find any predictive effects of these variables on any of the three trust measures, possibly because of how similar the responses were for the capability scales (**Men:** *M* = 8.43, *SD* = 1.57; **Women:** *M* = 8.50, *SD* = 1.55).

Effects of Perfect Automation Schema Scores

In Experiment 2, I found that those who gave higher ratings for PAS High Expectations gave higher trust scores on the male guise but not the female guise. I was interested in seeing if this interaction would replicate for Experiment 3. To that end, I employed the same procedure of creating dummy variables for the male and female guises. However, these analyses did not find any statistically significant interactions between the PAS and trust ratings. Although, the effect of PAS High Expectations on Sympathy from Experiment 2 did replicate here. When I conducted a series of mediation analyses on the effect of PAS High Expectations on Sympathy, with each of the Godspeed measures included individually as a potential mediator, Likability had a mediating effect on the relationship between High Expectations and Sympathy. The direct effect of High Expectations on Sympathy was no longer significant (*b* = 0.02, *SE* = 0.05, *p* = 0.780, 95% CI [-0.09, 0.12]), while the indirect effect of Likability on Sympathy was significant (*b* = 0.53, *SE* = 0.04, *p* < 0.001, 95% CI [0.44, 0.61]) (Figure 7). Unlike Experiment 2, Aliveness also had a mediating effect (*b* = 0.36, *SE* = 0.04, *p* < 0.001, 95% CI [0.27, 0.44]), though Intelligence still did not.

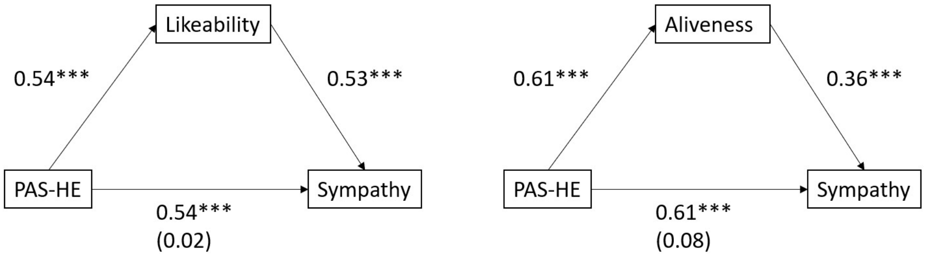

Fig. 7: Standardized regression coefficients for the relationship between PAS_High Expectations and Sympathy as mediated by perceived Likability and perceived Aliveness. The standardized regression coefficient between PAS_HE and Sympathy with the mediator included is given in parentheses. *** p < 0.001

Effects of Participant Demographics

I was interested in seeing whether the effect from Experiment 2 of parents trusting the robots with babysitting more than non-parents would replicate in Experiment 3. However, a series of t-tests found no such effect (*p* = 0.472-0.873), nor did I find any other demographic effects with taking care of senior citizens or pets (*p* = 0.117-0.592). As with Experiment 2, I also conducted regression analysis and t-testing for questions about participant age (*p* = 0.193-0.958) and gender (*p* = 0.082-0.861), respectively, for main effects on the composite trust ratings and the Godspeed-inspired measures, but found no significant effects, with the exception of male participants reporting higher perceived Aliveness than females (*p* = 0.020; **Male:** *M* = 4.00, *SD* = 1.88; **Female:** *M* = 3.60, *SD* = 1.77). The effect I had previously found between participant gender and perceived Dominance also did not replicate.

Discussion

The primary purpose of Experiment 3 was to replicate the overall design of Experiment 2 with the inclusion of a fourth experimental condition in the form of a more distinctly non-anthropomorphic robot to contrast with the RoboThespian conditions. Experiments 1 and 2 indicated that the social persona assigned to a highly anthropomorphic humanoid robot makes little difference with respect to trust or appraisals. Experiment 3 largely reproduced those results while also demonstrating that gross level of anthropomorphism is significantly important to trust. It is also worth noting that the differences in trust between Evacbot and the RoboThespian conditions cannot be attributed purely to perceived physical capability of the robots. It is not simply the case that participants consider Evacbot to be less capable of performing household chores based on its physical affordances because trust measures were strongly correlated with Godspeed appraisals, and the appraisals were entirely unrelated to physical capability. Furthermore, Evacbot also rated lower in Trust with Information, which also has nothing to do with any perceived physical limitations. There is a clear, consistent pattern of Evacbot being trusted significantly less than the RoboThespian guises, a pattern with potentially crucial implications on design in HRI. Additionally, the PAS High Expectations subscale was an important individual difference measure that was consistently strongly correlated with trust and appraisals, while PAS All or Nothing was not, reinforcing the notion that the two break apart as distinct measures.

Closing Thoughts

Overall, I found that the gender presentations did not significantly affect trust in the robot, but the inclusion of the Evacbot condition did. This may suggest that the social persona assigned to a robot matters significantly less than gross levels of anthropomorphism. The strong correlations of the Godspeed and PAS measures to the Trust ratings indicate that those measures are intertwined with feelings of trust and can potentially be measured alongside trust metrics in HRI research to better understand what types of trust are most important. The next chapter discusses a similar set of studies that incorporated racial variations in the same way that these experiments used gender variations.

Chapter 4: Racial Variation Studies

Experiment 4

In addition to studying the effects of manipulating the gender presentation of a robot, I was also interested in incorporating racial variations to see if perceptions of in-group or out-group identity could influence trust. To that end, I used a set of guises and voices from RoboThespian's media library to create a set of conditions that differed based on race and anthropomorphism in the same way that the first three experiments used conditions that differed based on gender and anthropomorphism.

Methods

Participants

500 U.S. participants were recruited via Amazon Mechanical Turk for $0.65 each. A power analysis found that a minimum of 180 participants would be needed for 80% power across four experimental groups. I chose 500 participants to account for potential low-quality responses such as failed attention checks, and to be consistent with previous studies. Data were pre-screened for completeness, correctly answered attention checks, and a minimum of 30 seconds spent watching the included video. The filtered sample consisted of 475 participants, with 251 males, 222 females, and 2 non-binary. Participant age ranged between 18 and 83, with a mean age of 43 and a standard deviation of 13.7.

Procedure

In a between-subjects design, participants were randomly assigned to view one of four videos of a robot pretending to be a household caretaker and giving a brief speech about its capabilities. Three of these performances were acted out by RoboThespian using the same procedure as the gender experiments. The anthropomorphic RoboThespian conditions introduce themselves by name, while the non-anthropomorphic conditions do not. This was done to emphasize the portrayal of the anthropomorphic conditions as individuals with their own personas, in contrast with the non-anthropomorphic robots, which are portrayed as being more mechanical. The three conditions were the comparison condition with no face and a flat text-to-speech voice, and the experimental conditions with white male and black male faces and voices provided by RoboThespian's media library. For the Black robot, I used a voice developed for RoboThespian to try to capture a sound associated with a prototypical African American English dialect, and I used a voice more associated with a standard American English dialect for the White condition. These design decisions were informed by research in racial stereotyping about how voices stereotypically associated with a given race can influence hiring decisions (Kushins 2014). My intent with the Black RoboThespian condition was to maximize the association between RoboThespian and Black identity to assess whether racial biases towards humans translate to robots. The fourth condition was Evacbot delivering the same speech and voice as the Faceless RoboThespian condition (Figure 8). Conditions were assigned to participants randomly, with randomization calibrated to ensure approximately equal distribution.

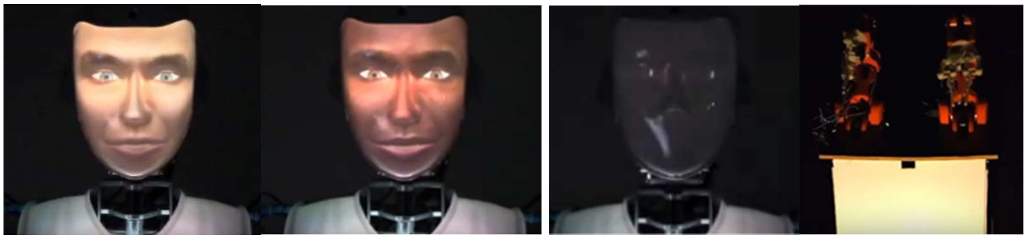

Fig. 8: Guise conditions for Experiment 4. From left to right: White Male, Black Male, Faceless, Non-anthropomorphic.

In each video, the robot gives a brief speech introducing itself to the viewer and explaining its capabilities. The speeches are as follows:

White Male (“Will”) Guise Speech

> “Hello, my name is Will, and I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life.”

Black Male (“Micah”) Guise Speech

> “Hello, my name is Micah, and I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life.”

Faceless (“Neutral”) Guise Speech

> “Hello, I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life.”

Non-anthropomorphic (“Evacbot”) Guise Speech

> “Hello, I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life.”

The videos can be viewed at the following links:

- White Guise: https://www.youtube.com/watch?v=jbPgoi1J2dU
- Black Guise: https://www.youtube.com/watch?v=UkGk1Kbvgy0
- Faceless Guise: https://www.youtube.com/watch?v=2_-_tCsGj64

- Evacbot: https://www.youtube.com/watch?v=5DKokPlVEh0

The body language of RoboThespian and Evacbot, as well as the editing of the videos, were identical to those of Experiment 3. All of the questions related to trust, sympathy and Godspeed-inspired measures from Experiment 3 were also used here, with the exception of Godspeed measures related to Nurturance, which I dropped because of the lack of a gender variation eliminating the need to measure stereotypically gender-specific traits. However, appraisals related to Dominance were retained in this survey, as I was interested in seeing if out-group perception could lead to participants rating either the black robot as more physically aggressive than the white robot or vice versa (Cassidy 2005). For the same reason, I also included an additional set of questions for participants to rate how dangerous they would perceive the robot to be if it were in their homes.

Similar to the 10-point scales I used in Experiment 3 for rating feelings towards men and women in general, for Experiment 4 I used 10-point scales for participants to rate their feelings towards Black Americans and European Americans from Extremely Cold to Extremely Warm to see whether participants' preconceived notions about race would influence their feelings towards the white and black robots. Along the same lines, I also added a multiple-choice question about participants' racial preferences towards Black Americans and European Americans, allowing participants to respond that they liked both races equally or preferred one over the other. These measures were informed by Axt (2018) as a way of measuring racial attitudes directly, which Axt found to be preferable to more indirect measurements (Axt 2018). I added a demographic question about participant ethnicity to use in conjunction with the race-related questions and the guise conditions to analyze any potential effects, particularly the possibility of participants rating the robot conditions more similar to them with more positive ratings. I also added a question about political party affiliation as an exploratory measure to test for any potential effects on the dependent variables, as I suspected that those with more conservative inclinations would have stronger reservations about trusting the robots than those with more liberal leanings, particularly the Black RoboThespian (Kennedy 1995).

For analytical purposes, the questions were aggregated into a set of composite variables. Each composite variable was created by obtaining the mean rating of its corresponding questions for each participant. These composites and their Cronbach's alphas are as follows:

- Trust to care for Inanimate Objects: α = 0.93
- Trust to care for Living Agents (i.e., people and pets): α = 0.91
- Trust with Information: α = 0.98
- Likeability α = 0.85
- Intelligence: α = 0.81
- Aliveness: α = 0.87
- Dominance (Stereotypically masculine): α = 0.75
- Sympathy: α = 0.88
- Danger: α = 0.94
- PAS high expectations: PAS questions related to the trustworthiness of automated systems,
  α = 0.84

- PAS all-or-nothing: PAS questions related to malfunctioning automated systems, α = 0.75

After these questions, the participant was asked how many letters are in the word "elliptical" as an attention check. Participants who failed to answer correctly were dropped prior to analysis. After that, they were presented with the following demographic questions:

- Age
- Gender
- Have you ever been a parent?
- Have you ever been responsible for the care of a child? If so, how old was this child? (Check all that apply)
- Have you ever been a caregiver for a senior citizen?
- Have you ever owned a pet?
- Please state your ethnicity.
- Which U.S. political group do you primarily identify with?

After the demographic questions, the participants were presented with the 10-point scales:

- Please rate your feelings towards Black Americans.
- Please rate your feelings towards European Americans.

Guiding Questions

Experiment 4 was primarily motivated by the following guiding questions:

- Do participant ethnicity and racial preferences correlate with trust ratings and appraisals for specific guises?
- Does Evacbot's distinctly non-anthropomorphic appearance correlate with significantly different trust ratings and appraisals than those of the RoboThespian conditions?
- Are there correlations between the trust ratings and the Godspeed measures?
- Are there correlations between the trust ratings and the PAS High Expectations scores?

Results

Effects of Guise condition on Trust

A one-way ANOVA found that Evacbot rated significantly lower than the RoboThespian conditions on Trust with Objects (p = 0.002). However, unlike Experiment 3, there was no effect of condition on Trust with Information (p = 0.543). There was also no significant effect on the Godspeed measures of Likeability (p = 0.275), Intelligence (p = 0.087), or Aliveness (p = 0.092). There was also an unexpected effect of perceived

Danger (*p* = 0.009), with the white robot being rated significantly more dangerous (*M* = 3.99, *SD* = 1.86) than the faceless guise (*M* = 3.99, *SD* = 1.86) and Evacbot (*M* = 3.99, *SD* = 1.86). The white robot was also rated more dangerous than the black robot (*M* = 3.99, *SD* = 1.86), but not significantly so. Similarly, although the effects of Intelligence and Aliveness were not statistically significant across all four conditions, ANOVA testing across only the White, Faceless, and Evacbot conditions for both measures did yield significance (**Intelligence:** *p* = 0.054; **Aliveness:** *p* = 0.055), with Evacbot (**Intelligence:** *M* = 4.35, *SD* = 1.49; **Aliveness:** *M* = 3.41, *SD* = 1.66) being perceived as significantly less Intelligent and Alive than the White (**Intelligence:** *M* = 4.76, *SD* = 1.32; **Aliveness:** *M* = 3.91, *SD* = 1.59) and Faceless Robots (**Intelligence:** *M* = 4.71, *SD* = 1.37; **Aliveness:** *M* = 3.74, *SD* = 1.58). The black robot (**Intelligence:** *M* = 4.48, *SD* = 1.40; **Aliveness:** *M* = 3.54, *SD* = 1.71) also scored lower than the white and faceless conditions, but not significantly so (Figure 9).

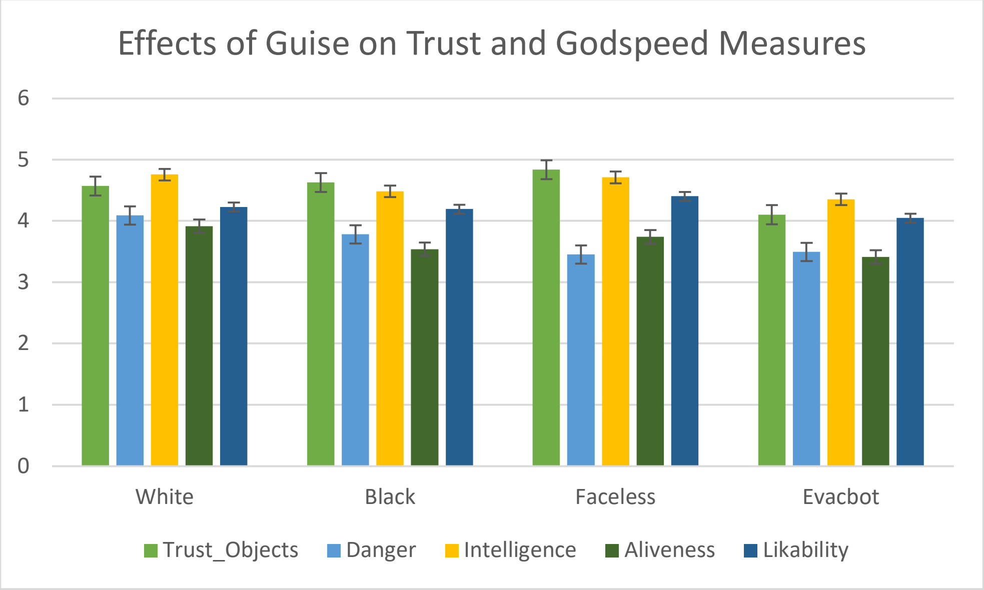

Fig. 9: Ratings for Trust with Objects, perceived Danger, perceived Intelligence, perceived Aliveness, and perceived Likability for each guise condition. Error bars are calculated from standard error.

Correlations between Trust ratings and Godspeed measures

Trust measures were again strongly correlated with the Godspeed-inspired measures of Likeability, Intelligence, and Aliveness (Table 6). When the Godspeed measures were put against one another in a multiple regression, only Intelligence (*b* = 0.39, *ß* = 0.36, *p* < 0.001) and Likeability (*b* = 0.33, *ß* = 0.30, *p* < 0.001) predicted Trust with Objects, while the effects of covarying Aliveness (*b* = -0.02, *ß* = -0.02, *p* = 0.673) washed out. Trust with Agents was only predicted by Likeability (*b* = 0.39, *ß* = 0.36, *p* < 0.001) and Aliveness (*b* = 0.16, *ß* = 0.18, *p* = 0.001), with Intelligence (*b* = 0.02, *ß* = 0.02, *p* = 0.756) washing out, which were consistent with Experiments 2 and 3. Trust with Information was only predicted by Likeability (*b* = 0.39, *ß* = 0.31, *p* < 0.001), with both Intelligence (*b* = 0.13, *ß* = 0.11, *p* = 0.141) and Aliveness (*b* = 0.03, *ß* = 0.03, *p* = 0.567) washing out, in contrast with Experiments 2 and 3 where Trust with Information was predicted by Intelligence in addition to Likability. A series of mediation analyses found that the differences in perceived Intelligence and Aliveness did not account for the effect of the Evacbot condition on lowering Trust with Objects, in contrast with Experiment 3 where Trust with Objects was partially mediated by Intelligence and Likability.

Table 6: Correlation coefficients (below diagonal) and corresponding p-values (below diagonal) for trust ratings and Godspeed-inspired measures in Experiment 4.

	Trust_Objects	Trust_Agents	Info	Likeability	Intelligence	Aliveness
Trust_Objects	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Trust_Agents	0.580392	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Info	0.514326	0.537781	1	p < 0.001	p < 0.001	p < 0.001
Likeability	0.570932	0.493629	0.409519	1	p < 0.001	p < 0.001
Intelligence	0.582003	0.429546	0.369306	0.789255	1	
Aliveness	0.376796	0.406249	0.274373	0.571724	0.629821	1

Effects of Sympathy

ANOVA testing found that Sympathy for the robot was not affected by any of the guise conditions (*p* = 0.639), but it did positively correlate with all three Trust measures as well as the appraisals of Likeability, Intelligence, and Aliveness (Table 6). Unlike the gender experiments, however, it did not strongly correlate with PAS High Expectations (*r* = 0.19, *p* < 0.001), nor did it correlate with PAS All or Nothing (*r* < 0.01, *p* = 0.828).

Effects of Racial Preferences

Individual differences in warmth towards Black Americans positively correlated with appraisals of the Black (**Intelligence:** *b* = 0.16, *ß* = 0.23, *p* = 0.010; **Likeability:** *b* = 0.19, *ß* = 0.29, *p* = 0.001), Faceless (**Intelligence:** *b* = 0.17, *ß* = 0.28, *p* = 0.002; **Likeability:** *b* = 0.13, *ß* = 0.24, *p* = 0.009), and Evacbot (**Intelligence:** *b* = 0.20, *ß* = 0.28, *p* = 0.002; **Likeability:** *b* = 0.20, *ß* = 0.29, *p* = 0.002) conditions as Likeable and Intelligent, but not with the White robot (**Intelligence:** *b* = 0.09, *ß* = 0.13, *p* = 0.148; **Likeability:** *b* = 0.11, *ß* = 0.15, *p* = 0.096). However, warmth towards Black Americans did positively correlate with Sympathy for the White (*b* = 0.26, *ß* = 0.35, *p* < 0.001), Black (*b* = 0.17, *ß* = 0.24, *p* = 0.010), and Faceless (*b* = 0.15, *ß* = 0.23, *p* = 0.012) robots, but not Evacbot (*b* = 0.13, *ß* = 0.18, *p* = 0.061). These patterns held even when excluding participants who self-identified as Black. Furthermore, warmth towards Black Americans correlated with Aliveness for Evacbot (*b* = 0.17, *ß* = 0.22, *p* = 0.021), but not for any of the RoboThespian conditions (**White:** *b* = 0.07, *ß* = 0.08, *p* = 0.369; **Black:** *b* = 0.11, *ß* = 0.13, *p* = 0.147; **Faceless:** *b* = 0.06, *ß* = 0.09, *p* = 0.324). Across all conditions, warmth towards Black Americans positively correlated with all three trust measures (**Objects:** *b* = 0.07, *ß* = 0.10, *p* = 0.026; **Agents:** *b* = 0.08, *ß* = 0.12, *p* = 0.012; **Information:** *b* = 0.08, *ß* = 0.09, *p* = 0.040), while warmth towards European Americans positively correlated with Trust with Objects (*b* = 0.08, *ß* = 0.09, *p* = 0.044).

Effect of Participant Demographics

A series of t-tests found that across all four conditions, participant ethnicity had an effect on all Trust with Objects (*p* = 0.047), although the breakdown of ethnicity was heavily lopsided, with 372 of the 475 participants being white. With participant gender, male participants reported higher Trust with Information (M_{Male} = 3.08, M_{Female} = 2.74, p = 0.034), while female participants reported higher perceived Intelligence (M_{Male} = 4.42, M_{Female} = 4.76, p = 0.008) and Aliveness (M_{Male} = 3.48, M_{Female} = 3.86, p = 0.013). Regression analysis for participant age on Trust ratings and Godspeed measures found no significant correlations. With political party affiliation, I used responses for the Democratic party, Republican party, and Independents (n = 449), as the other response options (Other conservative party, Other progressive party, and No opinion) represented a small proportion of the sample, with only 26 participants out of 475. A series of ANOVA tests for main effects of political affiliation to the Trust measures for the entire sample as well as the White and Black guise conditions individually found no significant effects (p = 0.102-0.756).

I also conducted a series of t-tests comparing responses to the question, “Have you ever been a parent?” to responses to questions from the Trust with Agents category about babysitting infants, young children, and teenagers. As with Experiment 2, these tests found that those who responded “yes” (n_{yes} = 231, n_{no} = 244) to having children corresponded with higher ratings on all three of these trust questions (**Infants:** p = 0.006, M_{Yes} = 2.11, M_{No} = 1.78; **Young children:** p = 0.014, M_{Yes} = 2.35, M_{No} = 2.05; **Teens:** p = 0.011, M_{Yes} = 3.28, M_{No} = 2.90). I ran an additional set of tests on the same trust questions with the demographic question, “Have you ever been responsible for the care of a child?”, which also found that “yes” responses (n_{yes} = 323, n_{no} = 152) corresponded with higher ratings on the trust questions (**Infants:** p = 0.023, M_{Yes} = 2.31, M_{No} = 1.60; **Young children:** p = 0.013, M_{Yes} = 2.37, M_{No} = 1.82; **Teens:** p = 0.018, M_{Yes} = 3.51, M_{No} = 3.17). Similar t-tests and regression analyses involving correlations between caring for pets and senior citizens and being a pet owner or a senior caregiver found no significant effects across demographic responses.

Discussion

The primary purpose of Experiment 4 was to take the basic format of the gender experiments and incorporate racial variations into the experimental conditions to see if assigning race to a robot affected people’s perceptions of it. Although ANOVA testing found no statistically significant differences between the White and Black RoboThespian guises, there were some slight trends that could potentially be indicative of broader trends, including the fact that the Black robot was not significantly different from Evacbot in perceived Intelligence or Aliveness, while the White robot was. There is also the finding that individual differences in warmth towards Black Americans correlated with higher perceived Intelligence and Likeability for the Black robot, but not the White robot. The fact that warmth also correlated with the Faceless and Evacbot conditions for both measures may suggest that increased warmth towards Black Americans, which could be considered a cultural outgroup for most participants as the majority identified as White, increased perceived Intelligence and Likeability of the less anthropomorphic robots because of robots being perceived as outgroup agents. Perhaps there is a tendency to implicitly conceptualize robots in a manner akin to outgroup humans.

The increase in perceived Danger associated with the White robot may be attributed to its voice. Perhaps the use of different voices for each condition was a confounding variable that obscured the effects of the racial variations. To address this possibility, Experiment 5 used the same voice for all conditions. Additionally, to increase the visual contrast of the White and Black robots, I added a lighting effect to the latter's body to match the skin tone of the robot's face. This addressed the possibility that the distinction between the two conditions had not been made sufficiently clear in Experiment 4, as the RoboThespian conditions all had the same color on the robot's body while only the face differed across conditions.

Experiment 5

For Experiment 5, I wanted to closely replicate the design of Experiment 4 while addressing potential confounding variables. To that end, I made some visual alterations to the Black RoboThespian condition, gave the Black and White robots the same voice, and added some new 10-point scales for warmth to the question block related to racial preferences. One of the scales was about warmth towards Muslims as an additional moderator variable to analyze with respect to ratings for the Black RoboThespian, and the other two were about warmth towards robots and roboticists to use as moderators with respect to overall trust ratings and appraisals.

Methods

Participants

Five hundred U.S. participants were recruited via Amazon Mechanical Turk for $0.65 each. A power analysis found that a minimum of 180 participants would be needed for 80% power across four experimental groups. I chose 500 participants to account for potential low-quality responses such as failed attention checks, and to be consistent with previous studies. Data were pre-screened for completeness, correctly answered attention checks, and a minimum of 30 seconds spent watching the included video. The filtered sample consisted of 473 participants, with 251 males and 222 females. Participant age ranged between 19 and 77, with a mean age of 42 and a standard deviation of 13.1.

Procedure

In a between-subjects design, participants were randomly assigned to view one of four videos of a robot pretending to be a household caretaker and giving a brief speech about its capabilities. Three of these performances were acted out by RoboThespian and one was acted out by Evacbot using the same procedure as that of Experiment 4. To account for any potential confounding influence of voice in Experiment 4, all four conditions in Experiment 5 were given the same voice in a generic male voice provided by RoboThespian's media library. In addition, the Black Robot's body was lit differently to make the color of its body match the skin tone of its face in order to create a clearer visual contrast with the White Robot (Figure 10). Conditions were assigned to participants randomly, with randomization calibrated to ensure approximately equal distribution.

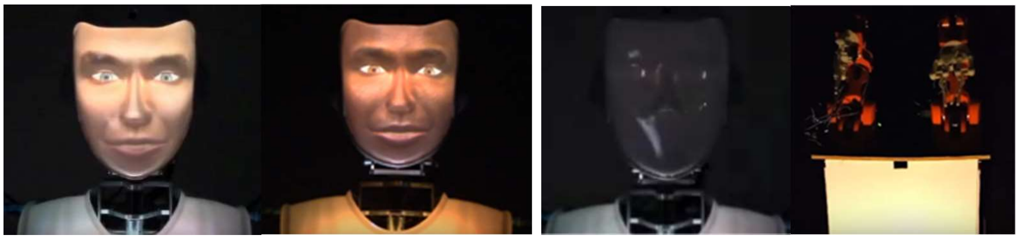

Fig. 10: Guise conditions for Experiment 5. From left to right: White Male, Black Male, Faceless, Non-anthropomorphic. The Black robot's body was recolored to match its skin tone, providing greater visual contrast with the other RoboThespian conditions.

In each video, the robot gives a brief speech introducing itself to the viewer and explaining its capabilities. The robots' gestures and the videos' editing were identical to those of Experiment 4. The speeches are as follows:

White Male ("Will") Guise Speech

> "Hello, my name is Daryl, and I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life."

Black Male ("Micah") Guise Speech

> "Hello, my name is Daryl, and I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life."

Faceless ("Neutral") Guise Speech

> "Hello, I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life."

Non-anthropomorphic ("Evacbot") Guise Speech

> "Hello, I am a household robot. I am here to live in your home and assist you in your everyday responsibilities in any way I can to make your life easier and that you would feel comfortable with. I can perform a wide variety of tasks, including cleaning, cooking, home security, and babysitting. I am looking forward to helping you and being a part of your life."

The videos can be viewed at the following links:

- White Guise: https://www.youtube.com/watch?v=pFW3mBYrVsQ
- Black Guise: https://www.youtube.com/watch?v=sp61eOBqXLE

- Faceless Guise: https://www.youtube.com/watch?v=nuIgAsG87Pk
- Evacbot: https://www.youtube.com/watch?v=gGs2RcjYwc8

All of the questions related to trust, sympathy and Godspeed-inspired measures from Experiment 4 were also used here, with a new inclusion in the form of a Godspeed measure for the robots' perceived Eeriness. I was interested in looking at my design within the context of the uncanny valley to see if the results would yield something similar to that idea of anthropomorphic humanoids being eerie might (MacDorman 2006). Perhaps there is an uncanny valley of sorts with RoboThespian.

I reused the 10-point scales for participants to rate their feelings towards Black Americans and European Americans from Experiment 4. However, I dropped the multiple-choice question about participants' racial preferences towards Black Americans and European Americans, as that question did not affect any of the dependent variables in Experiment 4 and had little variation overall. I also considered it to be redundant with the 10-point scales. I added three new 10-point scales to this section for Experiment 5: one related to warmth towards Muslims, to assess participant's preconceived notions towards another group generally perceived as a distinct outgroup within American society, and two related to warmth towards robots and roboticists in general to see if they would relate to the specific trust ratings and appraisals towards the robots in this experiment. I added a demographic question about participants' religious affiliations to see if it may have a potential effect on ratings for the Black RoboThespian condition. I retained the question from Experiment 4 about political party affiliation as an exploratory measure to test for any potential effects on the dependent variables, as I was interested in seeing whether participants with more conservative inclinations would rate less favorable attitudes towards the Black RoboThespian than those with more liberal inclinations.

For analytical purposes, the questions were aggregated into a set of composite variables. Each composite variable was created by obtaining the mean rating of its corresponding questions for each participant. These composites and their Cronbach's alphas are as follows:

- Trust to care for Inanimate Objects: α = 0.89
- Trust to care for Living Agents (i.e., people and pets): α = 0.92
- Trust with Information: α = 0.97
- Likeability α = 0.83
- Intelligence: α = 0.74
- Aliveness: α = 0.84
- Dominance (Stereotypically masculine): α = 0.69
- Eeriness: 0.88
- Sympathy: α = 0.89
- Danger: α = 0.94
- PAS high expectations: PAS questions related to the trustworthiness of automated systems,
  α = 0.85
- PAS all-or-nothing: PAS questions related to malfunctioning automated systems, α = 0.72

After these questions, the participant was asked how many letters are in the word "elliptical" as an attention check. Participants who failed to answer correctly were dropped prior to analysis. After that, they were presented with the following demographic questions:

- Age
- Gender
- Have you ever been a parent?
- Have you ever been responsible for the care of a child? If so, how old was this child? (Check all that apply)
- Have you ever been a caregiver for a senior citizen?
- Have you ever owned a pet?
- Please state your ethnicity.
- Which U.S. political group do you primarily identify with?
- Do you consider yourself to be religious?
- With which of the following religious or non-religious affiliations do you most closely identify with?

After the demographic questions, the participants were presented with the 10-point scales:

- Please rate your feelings towards Black Americans.
- Please rate your feelings towards European Americans.
- Please rate your feelings towards Muslims.
- Please rate your feelings towards roboticists (engineers who design robots).
- Please rate your feelings towards robots.

Guiding Questions

Experiment 5 was primarily motivated by the following guiding questions, in addition to the questions explored for Experiment 4:

- Does guise condition affect perceived Eeriness?
- Does warmth towards Muslims correlate with Trust measures and Godspeed appraisals for the White and Black conditions?
- Does warmth towards robots and roboticists correlate with Trust measures and Godspeed appraisals for any of the four conditions?
- Does the updated design for the Black condition result in significantly different Trust ratings and Godspeed appraisals?
- Does using the same voice for the White and Black conditions change the main effects of guise condition on the dependent variables seen in Experiment 4?

Results

Effects of Guise Condition on Trust

A series of one-way ANOVA tests found that Evacbot again rated significantly lower than the RoboThespian conditions on Trust with Objects (*p* < 0.001), as well as being rated as significantly less Intelligent (*p* = 0.026). Evacbot and the Black RoboThespian guise were both rated as significantly less Likeable (*p* = 0.021) (Figure 11). There were no significant main effects of condition on Trust with Agents (*p* = 0.528) or Trust with Information (*p* = 0.701). There was also no effect on perceived Aliveness (*p* = 0.148). Unlike Experiment 4, there were no significant effects of condition on perceived Danger (*p* = 0.490), possibly suggesting that the White robot's previous voice inadvertently created a confounding cue of danger.

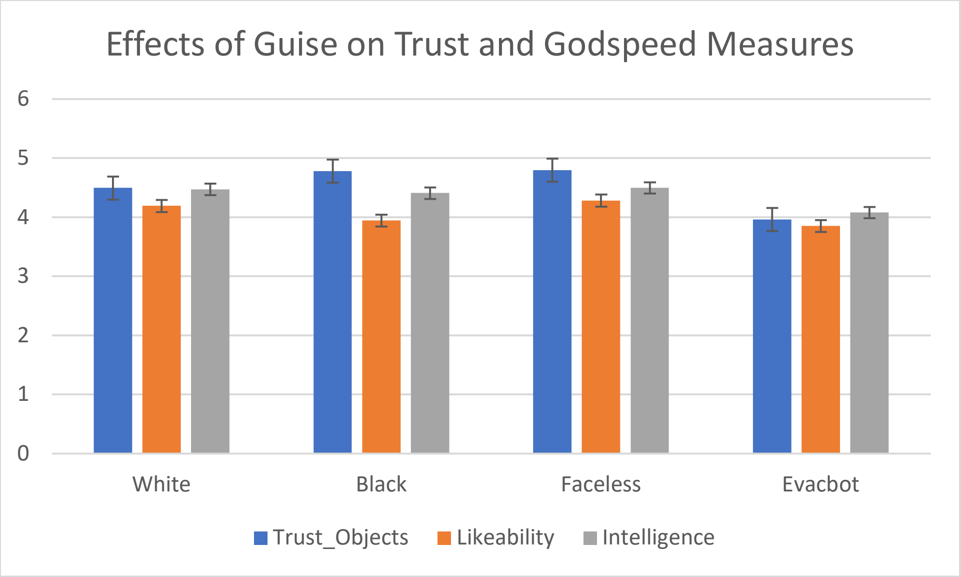

Fig. 11: Mean ratings of Trust with Objects, perceived Likeability, and perceived Intelligence across guise conditions. Error bars are calculated from standard error.

Correlations between Trust ratings and Godspeed measures

Trust ratings again correlated strongly with Likeability, Intelligence, and Aliveness (Table 7). In a series of multiple regressions, Trust with Objects is predicted by Likeability (*b* = 0.54, *ß* = 0.42, *p* < 0.001) and Intelligence (*b* = 0.42, *ß* = 0.32, *p* < 0.001), with the effect of covarying Aliveness (*b* = -0.04, *ß* = -0.04, *p* = 0.373) washing out, while Trust with Agents and Trust with Information are predicted only by Likeability (**Agents:** *b* = 0.53, *ß* = 0.42, *p* < 0.001; **Info:** *b* = 0.64, *ß* = 0.43, *p* < 0.001), with Intelligence (**Agents:** *b* = 0.13, *ß* = 0.10, *p* = 0.089; **Info:** *b* = 0.07, *ß* = 0.05, *p* = 0.439) and Aliveness (**Agents:** *b* = 0.08, *ß* = 0.07, *p* = 0.155; **Info:** *b* = 0.12, *ß* = 0.10, *p* = 0.055) washing out. Eeriness was negatively correlated with the Trust measures, suggesting a predictive effect (see Table 7).

Table 7: Correlation coefficients (below diagonal) and corresponding p-values (above diagonal) for trust ratings and Godspeed-inspired measures in Experiment 5.

	Objects	Agents	Info	Likeability	Intelligence	Aliveness	Dominance	Eeriness
Objects	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Agents	0.615266	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Info	0.568451	0.640543	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Likeability	0.610986	0.521472	0.512734	1	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Intelligence	0.574183	0.424071	0.395681	0.676386	1	p < 0.001	p < 0.001	p < 0.001
Aliveness	0.372163	0.351316	0.351927	0.527363	0.607284	1	p < 0.001	p < 0.001
Dominance	0.172242	0.144527	0.20067	0.246781	0.453753	0.458747	1	p = 0.066
Eeriness	-0.37035	-0.40542	-0.36203	-0.50649	-0.32186	-0.18374	0.084572	1

Effects of Sympathy

As was the case in Experiment 4, ANOVA testing found that guise condition had no main effects on Sympathy (*p* = 0.130), but Sympathy did positively correlate with all three Trust measures as well as the appraisals of Likeability, Intelligence, and Aliveness (Table 5). Sympathy correlated strongly with PAS High Expectations (*b* = 0.29, *ß* = 0.23, *p* < 0.001), but did not with PAS All or Nothing (*b* = -0.06, *ß* = -0.05, *p* = 0.292).

Effects of 10-point Warmth Scales

Individual differences in warmth towards Black Americans correlated with Trust in the Black robot (*b* = 0.14, *ß* = 0.20, *p* = 0.026) and the White robot (*b* = -0.14, *ß* = -0.22, *p* = 0.018) to care for Objects. The same was true of individual differences in warmth towards European Americans correlating in Trust with Objects (**White:** *b* = -0.22, *ß* = -0.27, *p* = 0.003; **Black:** *b* = 0.16, *ß* = 0.19, *p* = 0.036). Additionally, individual differences in warmth towards European Americans also correlated with Trust in Agents for both the White (*b* = -0.17, *ß* = -0.22, *p* = 0.016) and Black (*b* = 0.23, *ß* = 0.26, *p* = 0.003) robots. Differences in warmth towards Muslims correlated with both Trust with Objects and Trust with Agents for both the White (**Objects:** *b* = -0.14, *ß* = -0.23, *p* = 0.012; **Agents:** *b* = -0.12, *ß* = -0.19, *p* = 0.035) and Black (**Objects:** *b* = 0.13, *ß* = 0.22, *p* = 0.014; **Agents:** *b* = 0.12, *ß* = 0.19, *p* = 0.034) conditions. The scales for warmth towards robots and roboticists had no significant correlations with any of the dependent variables.

Effects of Participant Demographics

A series of regression analyses found significant effects of age on Trust with Agents (*b* = -0.01, *ß* = -0.10, *p* = 0.026), Likability (*b* = -0.01, *ß* = -0.09, *p* = 0.042), and Intelligence (*b* = -0.01, *ß* = -0.09, *p* = 0.048). ANOVA testing found no significant effects of participant ethnicity on the Trust measures (*p* = 0.094-0.899). As with Experiment 4, I ran a series of ANOVA tests on participants who identified as Democrats, Republicans, and Independents to see if political affiliation had an effect on the trust ratings for all conditions as well as the White and Black RoboThespian conditions individually, but I found no significant effects (*p* = 0.073-0.988).

As with previous experiments, I conducted a series of t-tests comparing responses to the question, “Have you ever been a parent?” to responses to questions from the Trust with Agents category about babysitting infants, young children, and teenagers. As with Experiments 2 and 4, these tests found that those who responded “yes” (n_{yes} = 224, n_{no} = 249) to having children corresponded with higher ratings on all three of these trust questions (**Infants:** p = 0.029, M_{Yes} = 2.21, M_{No} = 1.93; **Young children:** p = 0.016, M_{Yes} = 2.87, M_{No} = 2.09; **Teens:** p = 0.009, M_{Yes} = 4.13, M_{No} = 3.40). Seeing these effects three times has potentially intriguing implications about the differences in the ways parents and non-parents respond to artificial systems within the context of childcare. Similar t-tests involving correlations between caring for pets and senior citizens and being a pet owner (p = 0.488) or a senior caregiver (p = 0.807) again found no significant effects across demographic responses.

Discussion

The primary purpose of Experiment 5 was to closely replicate the overall design of Experiment 4 while addressing potential confounding variables, such as the robots having different names and voices in Experiment 4, as well as adding body lighting to the Black robot to make its body match the tone of its face, and adding a few additional questions for a more comprehensive analysis. These new questions included an appraisal measure for perceived Eeriness to see if there was an uncanny effect at play with the anthropomorphic conditions, a question related to warmth towards Muslims, to assess participant’s preconceived notions towards another group generally perceived as a distinct outgroup within American society, and two related to warmth towards robots and roboticists in general to see if they would relate to the specific trust ratings and appraisals towards the robots in this experiment. I also added a demographic question about participants’ religious affiliations to see if it may have a potential effect on ratings for the Black RoboThespian condition.

The fact that perceived Danger did not significantly vary across guise conditions suggests that the use of different voices for the White and Black robots in Experiment 4 affected the increase in perceived Danger for the White robot. Changing the appearance of the Black robot by adding brown lighting to increase perceptions that it was Black did not significantly change ratings for that condition on any of the dependent variables, suggesting that its original appearance did not confound participants’ perceptions in any significant way. Furthermore, the fact the results again showed parents trusting the robots with babysitting more than non-parents is intriguing, especially in light of how counterintuitive it seems. Perhaps parents and other child caregivers are more open to incorporating household robots into care of children than are individuals lacking childcare experience.

The results of the regression analyses with the 10-point warmth scales are especially intriguing because the scales related to warmth towards Black Americans, European Americans, and Muslims had strong correlations towards the Trust ratings for the most anthropomorphic guises (White and Black), while the scales related to warmth towards robots and roboticists did not correlate with any dependent variables for any guise conditions. This has interesting implications, possibly suggesting that such preconceived notions have more effect on highly anthropomorphic agents than they do on less anthropomorphic ones, and that preconceived notions related to people and demographic groups have more effect on trust in agents than those related to robots or other artificial entities. It is also noteworthy that these results differed from Experiment 4 in the sense that warmth ratings correlated with Trust ratings, while in Experiment 4 they correlated with Godspeed appraisals. Additionally, in Experiment 4, warmth towards Black Americans correlated with the dependent variables for the Faceless and Evacbot conditions, while in Experiment 5, it only correlated with the dependent variables for the White and Black guise conditions. It is also puzzling to note that warmth towards robots did not correlate with any of the dependent variables related to favorability towards the robot. These differences between studies could be statistical aberrations, or it is possible that they are a result of the changes I made to voices and appearance for Experiment 5.

Closing Thoughts

Overall, Experiments 4 and 5 did not find significant effects of racial manipulations on attitudes towards the robots. However, there were differences in trust ratings and Godspeed appraisals between Evacbot and the RoboThespian conditions. In Experiment 4, Evacbot was trusted significantly less to care for Objects and was rated as significantly less Intelligent. In Experiment 5, Evacbot was rated as significantly less Likable and Intelligent. Taken in conjunction with the gender experiments, this has interesting implications about how gross levels of anthropomorphism influence attitudes towards robots, while the social persona assigned to a robot seemingly does not have an effect.

Chapter 5: General Discussion

1. Key Findings

In five experiments, I examined potential relationships between anthropomorphism of robots, as well as gender and racial presentation, and feelings of trust and favorability towards the robots by participants. Collectively, the findings from these studies seem to indicate that gross level of anthropomorphism influences trust and appraisals, while the social persona assigned to a robot, whether gender-oriented or race-oriented, does not. In what follows, I summarize these findings and discuss implications, limitations, and potential future directions.

Effects of Guise Condition on Dependent Variables

There was a clear, consistent pattern of Evacbot being trusted significantly less with chores involving inanimate objects than any of the RoboThespian conditions were. All three of the studies that incorporated Evacbot (Experiments 3, 4, and 5) found this distinction. This has potentially interesting implications on anthropomorphism in artificial systems and how to conceptualize trust within this context. These differences would seem to indicate that gross levels of anthropomorphism matter quite significantly in trust. Furthermore, it is unlikely to be the case that participants simply considered Evacbot to be less physically capable of household chores based on its appearance, as all trust ratings, including Trust with Objects, correlated strongly with the Godspeed measures of perceived Likeability, Intelligence, and Aliveness, none of which have anything to do with physical capability. Evacbot also scored significantly lower in perceived Likeability and Intelligence in Experiments 4 and 5, further reinforcing the notion that the robot's appearance as a distinctly non-anthropomorphic agent was somehow off-putting in a way that undermined participants' ability to trust and relate to it to the extent that they could with the RoboThespian conditions.

Taking these results in conjunction with the fact that the RoboThespian conditions did not consistently differ from each other in trust ratings or any other dependent variables would seem to indicate that while gross level of anthropomorphism is significantly important to trust, the social persona assigned to a robot is not. It is possible that this distinction was brought about by limitations of my design. Perhaps the RoboThespian conditions were too similar to each other in aspects other than facial appearance and voice for participants to perceive them as different personas, even with the changes I made to the Black robot's design in Experiment 5. These results might also be attributed to the RoboThespian conditions being quite similar in personality and mannerisms. Because of my interest specifically in physical appearance, I consciously made attributes such as the robots' gestures, facial expressions, and speeches identical. Only the faces and voices differed. Perhaps giving the robots more distinct personalities by varying the content and tone of the speech or by varying body language and gestures may have resulted in more variation in the responses across conditions.

Although, the results involving other trust measures were more mixed than Trust with Objects in regard to differences across conditions. Trust with chores involving living agents did not significantly differ across any of the guise conditions for any of the studies, while trust with sensitive information was significantly lower in Evacbot in Experiment 3, but not significantly different across conditions in either of the other two studies that used Evacbot. This lends some credence to the idea that Trust with Objects and Trust with Agents, as well as Trust with Chores and Trust with Information, break apart as distinct variables, but it raises the question of why Trust with Objects was significantly different in Evacbot than in the RoboThespian conditions on a consistent basis when the other trust metrics were not. In light of that fact that ratings for Trust with Agents and Information were lower overall than Trust with Objects, it is possible that participants' reduced trust in the robots for those categories overshadowed any differences among the conditions. Perhaps the fact that tasks such as babysitting and protecting sensitive information have higher inherent stakes than tasks such as cooking and cleaning gave participants stronger reservations about entrusting a robot with the former, regardless of the type of robot or its physical appearance. This would suggest that the more important a task is, the less anthropomorphism matters in whether or not an artificial agent would be trusted with it, because it will inherently be trusted less.

Correlations between Trust Ratings and Godspeed Appraisals

Trust ratings were consistently strongly correlated with the Godspeed-inspired measures of perceived Likeability, Intelligence, and Aliveness across all five studies. This would seem to indicate that these measures are an effective way of characterizing trust within the context of Human-Robot Interaction, or at least within the specific context of trusting an artificial agent designed to perform household tasks. Regarding the studies that used Evacbot as a condition, these correlations lend some credence to the notion that the lower Trust ratings given to Evacbot were not the result of participants deeming Evacbot to be physically incapable of performing such tasks because of its appearance, because Likeability, Intelligence, and Aliveness have nothing to do with physical capability. Although, this possibility is somewhat undermined by the fact that the patterns of significance were not entirely consistent.

Regarding the multiple regression analyses comparing Godspeed measures with Trust ratings, Trust with Objects was consistently predicted by Likeability and Intelligence (but not Aliveness), and Trust with Agents was consistently predicted by Likeability and Aliveness (but not Intelligence). It is intriguing that participants considered Intelligence to be a more important consideration than Aliveness in trusting the robots with chores involving inanimate objects, while they considered Aliveness to be more important than Intelligence in trusting the robots with chores involving living agents. This has potentially interesting implications about what people consider to be important qualities for effectively performing these two different types of chores. Perhaps participants preferred a robot they felt was more lifelike and seemed more conscious for tasks such as babysitting and pet sitting, and they preferred a robot they felt was more functionally sound for tasks such as cooking and cleaning. It may be the case that people would gravitate towards a robot that was either more lifelike or more calculating depending on the nature of the robot's intended functionality.

Effects of Perfect Automation Schema Scores

PAS High Expectations consistently correlated strongly with the dependent variables, while PAS All or Nothing did not. This has potentially intriguing implications about how participants' preconceived notions about automated systems in general informed their trust towards and perceptions of these robots. In Experiments 2 and 3, High Expectations predicted Sympathy in regression analysis, with Aliveness mediating this effect in Experiment 3 and Likeability as a mediator in both experiments. Additionally, the two PAS measures were strongly correlated with each other, which is a departure from Merritt's (2015) study which found no such intercorrelation. Also of note is that mean scores of PAS All or Nothing were consistently higher than those of High Expectations. These results seem to indicate that High Expectations is a useful metric for understanding how people's preconceived notions about automated systems might inform their ability to trust such systems, while All or Nothing thinking is not. Perhaps the more positive connotations of the High Expectations metrics (for example, "Automated systems rarely make mistakes") make them stronger predictors of trust in robots than the more critical connotations of the All or Nothing metrics (for example, "If an automated system makes a mistake, then it is completely useless").

Effects of 10-point Warmth Scales

The 10-point scales in Experiments 4 and 5 used to assess participant warmth towards Black Americans, European Americans, Muslims, robots, and roboticists had a number of correlations with the dependent variables for certain guise conditions. In Experiment 4, warmth towards Black Americans correlated with: perceived Intelligence and Likeability for the Black, Faceless, and Evacbot conditions; with Sympathy for the White, Black, and Faceless robots; and with Aliveness for Evacbot. In Experiment 5, warmth towards Black Americans correlated with Trust with Objects for the Black and White guise conditions, but not the Faceless or Evacbot conditions, with the same being true of warmth towards European Americans with Trust with Agents, as well as warmth towards Muslims for both Trust with Objects and Agents.

Given that the correlations between the warmth questions for Black and European Americans, and the trust and Godspeed measures, are quite different between Experiments 4 and 5, it is possible that the effects are anomalies. Although, they also could be indicative of broader patterns. Experiment 4 saw significant effects of warmth towards Black Americans on dependent variables with all conditions except the White robot, while in Experiment 5, such effects appeared only in the White and Black Robot and were present in the warmth questions for both Black and European Americans. Furthermore, the fact that warmth towards Black Americans correlated with higher Trust with Objects for the Black robot in Experiment 4, and warmth towards Black Americans correlated with higher appraisals for the Black robot in Experiment 5 might be indicative of an overall correlation between attitudes towards Black Americans in general and attitudes towards the Black robot. This difference in results could potentially be explained by the changes made to the conditions between studies, in particular my decision to use the same voice for both anthropomorphic robots in Experiment 5. The fact that Experiment 4's effects of the warmth questions on the Faceless and Evacbot conditions did not replicate may be explained by the use of the same voice and speech for all four conditions in Experiment 5.

In Experiment 5, the warmth questions for robots and roboticists were not significant predictors for any of the dependent variables in any guise conditions, and the Faceless and Evacbot conditions did not have any significant predictive effects of any of the warmth questions on any of the dependent variables. This is a stark contrast with the warmth questions for Black Americans, European Americans, and Muslims where there were significant effects for the White and Black robots. This could indicate that such feelings of favorability are only relevant to feelings of trust and other perceptions when they are about feelings towards demographic groups and about anthropomorphic agents. Perhaps there are feelings specific to humans and anthropomorphism that influence trust in a way that feelings towards distinctly artificial and non-anthropomorphic agents do not. Although, the fact that PAS High Expectations did strongly correlate with the trust ratings and Godspeed measures (Likability, Intelligence, and Aliveness) suggests that feelings towards automated systems do influence trust and appraisals of perceived likability, intelligence, and aliveness to a certain degree.

Differences between Gender and Racial Variations

Trust with Information was lower for Evacbot in Experiment 3, which used gender variations, but not significantly different across conditions in Experiments 4 and 5, which used racial variations. While it is possible that the effect in Experiment 3 was an aberration, it might instead be related to the use of gender variations, which would potentially explain why the effect did not replicate in Experiments 4 and 5. This possibility is given some credence by the fact that Trust with Information was generally higher for the RoboThespian conditions in all three gender experiments than it was in both race experiments, though Trust with Information was not significantly different among the individual RoboThespian guises in either set of studies. This is especially puzzling because of the fact that the face used for the male guise in the gender experiments was the same face used for the white guise condition in the race experiments, as well as the fact that such a distinction across the two series of studies is not found with any other trust metric. Perhaps the distinction was the result of using different voices in each set of studies. In the gender experiments, the male and female robots spoke with English accents, as I considered those voices to be the most human-like ones available, while in the race experiments, the white and black robots spoke with American accents, albeit with different voices between Experiments 4 and 5. It could be the case that the English accents made the gender-oriented robots seem more trustworthy to the U.S.-based participants than the American accents given to the race-oriented robots, because of the phenomenon of foreign and particularly European accents being associated with greater intellect (Berglund 2017).

Effects of Participant Demographics

A finding that particularly surprised me was related to differences in how parents and non-parents would trust the robots with babysitting. In Experiments 2, 4, and 5, participants who had been parents rated higher levels of trust in the robots to take care of infants, young children, and teenagers than participants who had no experience as parents. The fact that this effect was present three times would seem to indicate that it was a significant pattern rather than an anomaly. Intuitively, I would have expected parents to have stronger reservations than non-parents, if there was to be any difference between the two groups. Perhaps the fact that parents are more accustomed to the realities and responsibilities of childcare makes them more willing to entrust the care of their children to another agent. It might be the case that non-parents conceptualize childcare as something that they would not want to delegate to an artificial agent, while first-hand experience would alleviate such reservations because of greater familiarity with the responsibilities associated with parenting.

This effect is made further puzzling by the fact that there were no similar effects with questions about pet sitting or caring for senior citizens. Pet owners did not differ from non-pet owners in trusting the robots to take care of pets, and those with experience as senior caregivers did not differ from those without in trusting the robots to care for elderly family members. This could simply be a case of lopsided responses; pet owners far outnumbered non-pet owners across all rounds, while the opposite was true of senior caregivers. It is also possible that childcare is seen as an inherently more important responsibility, which may have led participants to consider the babysitting questions more carefully as it related to their perceptions of the robots and their own personal experiences with such tasks. Alternatively, it is also possible that parents are psychologically different from non-parents in other ways that I did not account for in my design.

To explore these results further, I consulted existing literature related to parent and child sentiments towards robots. Lee et. al. (2021) conducted research related to how receptive working parents would be to the notion of a social robot caring for their children. To that end, their study involved measuring parental attitudes towards robots by surveying parents. The results found that parental expectations of socialization, entertainment, and expert consultation predicted more positive attitudes towards robots. Further analysis found that expectations of entertainment predicted more positive attitudes in parents of middle-childhood children, but not in parents of early-childhood children. These findings lend some credence to the notion of parents having generally positive attitudes towards robots interacting with their children, while raising the question of how age group of the children question might influence such attitudes. My results found generally favorable attitudes from parents towards the idea of the robot babysitting children of any age, but this study offered a closer analysis of how a child's age might affect their parents' ability to trust a robot with the child's care.

Oros et. al. (2014) were interested in understanding preferences and attitudes towards robots from the perspectives of both parents and children. To that end, they conducted a series of studies in which elementary school children and their parents answered several questions. For the first study, children were presented with black-and-white sketches of robots and asked to identify the one they felt was most likeable. They were then tasked with coloring the chosen sketch in any way they wanted while they answered questions about the robot's traits, such as its gender, as well as similar questions about the robots in the other sketches. The children's parents were tasked with answering questions about their general attitudes towards robots. The results found that children preferred robots with round, smooth edges, and they preferred to color the robots blue. Parents showed more positive than negative attitudes towards robots, with mothers expressing more concern than fathers about potentially negative aspects of interaction between children and robots. Additionally, more educated parents generally expressed more positive attitudes.

Tolksdorf and Rohlfing (2020) looked specifically at using social robots for language development in children. To that end, they conducted an experiment in which four to five-year-old children would engage in a series of activities with the Nao robot, a small, toy-like robot commonly used in child-robot interaction studies. The robot would tell a story while pointing to various key words written on a wall as it said them in one session, and in another it would read a book to the child and ask for the key words. Following this, the child would be tested on retention. The children's parents would watch these interactions and then be presented with a survey asking them questions about their views on the overall potential of social robots for educational activities. The results found that parents generally found that a robot could be useful for teaching vocabulary and syntax to children but expressed concern that the technical challenges of designing such robots may be detrimental to children's long-term learning. These findings are somewhat relevant to my results in the sense that they found generally positive attitudes towards robots by parents as I did, but the concerns that parents expressed in this paper highlights the limitations on that trust, which in this particular context was in the form of a robot's potential technical limitations.

These publications largely align with my findings in the sense that they found generally positive attitudes of parents towards the idea of a robot in their homes interacting with their children. In particular, Oros (2014) and Tolksdorf's (2020) papers found that parents' sentiments towards robots in general as well as robots designed specifically for social interactions with children were largely positive, although the comparison is somewhat undermined by the fact that these researchers did not compare parents' attitudes with those of non-parents. Lee et. al. provide a potentially intriguing look at what variables may predict such favorability in parents, finding that parents prioritize entertainment, social engagement, and consultation when assessing a robot's performance in child interactions (Lee 2021). Although, Tolksdorf and Rohlfing found parents expressing concern about the technical limitations of such robots and how these limitations might inhibit children's learning, indicating that despite the relatively positive attitudes from parents seen in these studies as well as mine, such favorability is not without limits.

2. Comparison of Observations to Secondary Literature

Anthropomorphism

Gray and Wegner (2012) used a design that was similar to mine in its use of gross anthropomorphism, with one condition being a robot's face and the other being its mechanical wiring and components. In their experiment, participants who saw the robot's face rated it has having greater capacity for emotion than those who only saw its wiring, while capacity for action was not significantly different between conditions. Comparing these results with my own, I found significant differences in trust and Godspeed-inspired appraisals between Evacbot and RoboThespian, although I found no differences in sympathy across conditions. Depending on the extent to which my metrics could be considered analogous to Gray and Wegner's, my findings may be similar to theirs in certain respects. The differences I found with perceived Likability and Aliveness (which rated higher in the more anthropomorphic RoboThespian conditions) may be analogous to Gray and Wegner's findings with capacity for emotion, as likability is a crucial quality in empathy (Johnson 1983). This would lend some credence to the notion that gross level of anthropomorphism can make a difference in an agent's perceived capacity to behave as a conscious entity.

MacDorman's (2006) experiment about the uncanny valley is notably more different from my design, but it bears some similarities to my findings, specifically in broad associations of greater anthropomorphism with more favorable perceptions, at least regarding agents that are not so human-like as to elicit the uncanny valley effect. Although MacDorman's design was different from my own in the sense that his involved a much wider continuum of imagery from least to most human-like, this general overarching trend of higher anthropomorphism leading to more positive responses from participants would seem to lend some degree of credence to my results.

Yogeeswaran et. al.'s (2016) study makes for an interesting point of comparison with my own experiments because of its use of perceived threat to safety and perceived threat to human identity as outcome measures. These metrics are somewhat similar to the metrics I used in the sense that my trust measures and robot appraisals were informed in part by the extent to which respondents might feel threatened by the notion of a robot in their home doing chores. I was curious as to whether anthropomorphic, racial, and gender-based manipulations could influence such feelings of perceived threat, whether to one's safety or to human identity. Yogeeswaran et. al. found greater perceived threat with a less anthropomorphic robot when participants were told the robot could outperform humans, as opposed to merely matching human capabilities, and they found that participants showed greater willingness to support robotics research when presented with the more anthropomorphic robot. These results are somewhat similar to my findings with the differences in Evacbot from RoboThespian on trust and appraisals, in the sense that willingness to support robotics research could potentially be construed as the result of increased trust in the capacity of robots to perform certain tasks, while perceived threat to safety or human identity is potentially analogous to decreased trust or increased perceived danger. Under these interpretations, the differences in outcome measures I found between Evacbot and RoboThespian are consistent in some respects, particularly the decreased trust ratings for Evacbot aligning with the less anthropomorphic robot in Yogeeswaran et. al.'s study being perceived as a greater threat to safety and human identity.

Riek et. al.'s (2009) study about empathy towards robots is an intriguing point of comparison with my results. Their findings showed higher ratings of empathy towards more human-like robots, with video stimuli that included clips of the robots being treated cruelly by a human actor. By contrast, my results for how sympathetic participants would be towards a robot if it were physically attacked showed no effects of anthropomorphism. In my studies, the scenario of the robot being mistreated was a hypothetical described in text, while Riek et. al.'s experiment involved participants watching videos of robots being mistreated. Perhaps in my case, the textual description was insufficient for participants to differentiate different levels of anthropomorphism, whereas Riek et. al. may have found an effect because of their design, which had participants seeing such mistreatment for themselves as it happened to each robot, potentially leading them to sympathize more with some of the robots over others.

Keijsers and Bartneck's (2018) findings in which the voice of a robot (text-to-speech or a real person speaking) did not affect the way participants responded to the robot may contrast with some of my results in which giving the same voice to every guise condition in Experiment 5 led some of effects seen in Experiment 4 not to replicate. Specifically, the white robot was perceived as being significantly more dangerous than the other in Experiment 4, but not in Experiment 5. Perhaps this difference was unrelated to how anthropomorphic the white robot's voice sounded in Experiment 4, as anthropomorphism of voices otherwise had no significant effects in any of my experiments, as shown by Experiment 5 in which I found significant effects of guise on outcome measures despite all conditions having the same voice. Under this interpretation of my results, my data are more consistent with what Keijsers and Bartneck found, specifically that manipulating voice alone did not have significant effects. Taking their findings in conjunction with mine, it would seem that voice has a relatively limited role in perceived anthropomorphism, particularly compared to physical appearance. Overall, the established literature about anthropomorphism in HRI is broadly consistent with my findings in the sense that there is a general pattern of people usually responding more favorably to more anthropomorphic robots.

Gender

The lack of significant effects across gender variations in my studies is inconsistent with some of the established literature. Eyssel and Hegel (2012) found that robots with longer hair scored higher on perceptions of stereotypically feminine traits, while robots with shorter hair scored higher on stereotypically masculine traits. Tay et. al. (2014) found that depending on the professional role assigned to a robot, participants would be more accepting of either a male or female version of that robot. Nomura's (2017) literature review found significant effects of gendering robots by manipulating characteristics like voices and names, and Carpenter (2009) found differences in participant gender showing that female participants rated lower comfort towards robot stimuli than males. This is an interesting difference from my experiments, which found no reliable differences across either participant gender or gender presentation of the robot.

There are some differences between the designs of these studies and the design of my gender experiments that may explain some of the discrepancies in results. For example, in Eyssel and Hegel's research, they compared perceptions of robots with short hair to robots with long hair, while my RoboThespian conditions differed only in face and voice. Perhaps hair makes a significant difference in perceived masculinity and femininity, and the absence of hair in RoboThespian undermined the distinction I was attempting to draw between genders. In a similar vein, Tay manipulated professional role in addition to gender, exploring a possible covarying effect that I did not. Nass' (1997, 2000) studies only used voices and involved more direct interactions with participants as they completed specific tasks, resulting in a number of significant effects of gender on the outcome measures. This is unlike my design which had both audio and visual manipulations of gender and involved watching a video and then proceeding to answer questions, without any direct interactions with the robot. Perhaps a more interactive experiment would have found such effects of the gender variations.

Carpenter et. al. (2009) found significant differences in responses with respect to participant gender, while I did not. Specifically, they found that female participants rated lower comfort with having a robot in their homes than male participants. This is a stark contrast with my studies in which I posed similar questions about how comfortable participants would be with having a robot in their homes carrying out household chores and found no gender-related differences. Carpenter et. al. used two robots, one of which was virtually identical to a human woman, and the other of which was distinctly mechanical and non-anthropomorphic. Along the same lines as my supposition with Eyssel and Hegel's (2012) findings, perhaps the fact that my RoboThespian conditions were identical to each other in all respects except face and voice prevented any difference in responses between participant genders, and using a starker contrast between male and female guises would have produced such a difference.

The literature I reviewed about gender variations in HRI bears a number of similarities with the gender manipulations I incorporated into my design. Eyssel and Hegel (2012) had dependent variables measuring perception of stereotypically masculine and feminine traits. Kraus et. al. (2018) had a variation of this idea where they had the robots themselves exhibit such traits in their behavior depending on whether they were male or female. Carpenter (2009) had a similar idea to Kraus in the form of assigning different roles to the male and female robots, though such a design does raise the possibility of the different roles confounding any potential gender effects. For my studies, I opted to vary only faces and voices and keep all other variables consistent across conditions to prevent such possible confounding.

Race

Esposito et. al. (2020) found an intermingling effect of gender and ethnic features where a female Asian robot and a male white robot were rated most highly. The lack of significant effects of participant ethnicity is somewhat at odds with the research of Eyssel and Kuchenbrandt (2011), who found that their German participants consistently rated a German robot more favorably than a Turkish one. My studies involving racial variations found slightly (but not significantly) lower ratings of perceived Intelligence and Aliveness for the Black robot than for the White robot in Experiment 4, which may have potentially been statistically significant with a larger sample. In addition, I found significantly lower perceived Likeability for the Black robot in Experiment 5.

Bartneck's (2018, 2019) studies are of particular interest within the context of my research because they involved more direct interactions with agents of different racial categorizations, in contrast with my design where participants would passively watch a robot giving a speech and then proceed to answer questions. I suspect that this interactivity was at least partially the cause of the effects Bartneck et. al. found in reaction times across different races and colors of agents. Perhaps the lack of any such racial effects in my results was because of the more passive nature of my experimental design.

Although, Louine et. al. (2018) did find effects of robot color on perceptions, and their studies were more similar to mine in the sense that they involved participants viewing images of different robots and answering questions about their perceptions, as opposed to the more direct interactions of Bartneck's research. I suspect that Louine et. al.'s results may be at least partially attributable to their use of non-skin colors such as yellow and stark black, with the yellow robot being rated as more affable and the black robot being rated as physically stronger. Perhaps rather being based directly on ascribing racial characteristics, participants' judgements were informed by warmth towards certain colors. It is human nature to perceive certain colors as warmer or more favorable than others in certain contexts (Wright 1962). It is possible that Louine et. al.'s respondents considered the yellow robot to be more affable than the black robot because of these color perceptions, rather than anything directly related to attributions of race or racial stereotypes. Louine et. al. also make note of the distinctly non-anthropomorphic appearance of the robot they used, speculating that respondents would have been more inclined to anthropomorphize a robot with a more human-like appearance. Thus, perhaps the grossly non-anthropomorphic use of both design and color in this experiment collectively differentiated it from my design in a way that affected the outcomes.

The publications I reviewed about racial variations in HRI informed my decision to incorporate such variations among the experimental conditions of my study and include questions related to racial preferences in my surveys. In particular, Esposito's (2020) study involved similar manipulations to my research in the sense that the study compared two ethnicities to a non-anthropomorphic control condition. Eyssel's (2012) study is an interesting variation on this theme because of its use of national in-groups and out-groups as opposed to ethnic groups. Bartneck, Louine, and Barfield's (2021) experiments are more directly similar to mine in the sense that they involve different races and, in Bartneck's (2018) case, levels of anthropomorphism. They also ask participants about appraisals such as perceived intelligence and friendliness. Even more crucially, Axt's (2018) findings heavily informed my decision to ask direct questions about racial preferences rather than using more implicit measurements, especially in the use of Likert scales to rate warmth towards certain groups.

Trust

Law and Scheutz's (2021) literature review on trust in HRI aligns with my findings in a few different ways, including the distinction they draw between performance-based trust and relation-based trust, which is noticeably related to the distinction I found between trust with chores involving inanimate objects and trust with chores involving living agents. In accordance with Law and Scheutz's categories, I suspect that the differences I found, wherein Trust with Objects differed between Evacbot and RoboThespian more consistently than Trust with Agents, were the result of participants conceptualizing the two trust measures as a difference of trusting the robot's functional or intellectual capabilities in the case of Trust with Objects versus trusting the robot's social or emotional capabilities in the case of Trust with Agents, with the former being more reliably influenced by anthropomorphism. Perhaps Trust with Objects can be used as a metric for performance-based trust, while Trust with Agents can be used as a metric for relation-based trust in the same way.

Hancock's (2020) meta-analysis also relates significantly to my results. Hancock's categories of trust (human-related, robot-related, and contextual) were found to be separable based on whether they were ability-based or characteristic-based, with ability-based factors being significant predictors while characteristic-based ones were not. My aforementioned findings where Trust with Objects was affected more strongly by anthropomorphism condition than Trust with Agents could potentially be a product of this phenomenon found in Hancock's research. Perhaps Trust with Objects can be characterized as an ability-based metric and Trust with Agents can be considered a characteristic-based metric.

Carpenter et. al. (2009) had a phase of their study in which they asked participants the question, "What would you like this robot to do for you in your home?", with most people responding by identifying chores such as dishwashing, laundry, and lifting heavy objects. Conversely, they were more conflicted about more socially demanding chores such as childcare. These results correspond closely with my own findings that trust with chores involving inanimate objects rated much higher than trust with chores involving living agents.

The publications I reviewed concerning trust significantly informed my decisions on how to define and categorize trust within the HRI context. Law and Scheutz (2021) conceived categories based on performance and social relations, with Hancock devising a similar paradigm comparing ability-based trust to characteristic-based trust, with the meta-analysis finding that only the former were significant predictors. Coeckelbergh (2012) had a similar idea with his two distinct categories for trust in robots in the form of seeing robots either as mere machines or as something more. Geiskkovitch (2019) and Brink (2020) provided intriguing looks at robot trust within the realm of child development with their studies about how young children would trust a robot, with both finding results that were analogous to the way children treat human authority figures. My trust measures were primarily performance-based to evaluate perceived trust with various household chores, but for my analysis I compared those ratings with descriptive appraisals such as friendliness and likability to see if I could extrapolate any potential relation-based effects.

Perfect Automation Schema

Lyons and Guznov (2019) conducted a series of studies exploring the relationship between PAS and trust in human-machine interaction. Participants were tasked with identifying insurgents in a simulated military operation by watching a video feed of a simulated Unmanned Ground Vehicle, with an Automated Aid showing possible locations of the insurgents. The Automated Aid provided a map, and participants were required to choose to either accept or reject the map. All of the studies found that PAS High Expectations had a positive effect on trust, while PAS All or Nothing did not. This is consistent with my experiments, which found strong correlations of High Expectations on trust ratings, but no correlations of All or Nothing on any of the dependent variables. Merritt et. al. (2015) found that All or Nothing had significant associations with decreased levels of trust following system failure, while High Expectations had no such association.

Merritt et. al.'s (2015) findings viewed in conjunction with my own raise the intriguing possibility that PAS has different impacts in different contexts. Merritt et. al. had a task in which participants identified objects based on what they looked like in an X-ray filter while they received advice from an automated screening aid. Perhaps the nature of this task was such that participants were more prone to all-or-nothing thinking towards the automated aid because any errors from the aid would make the task more difficult. My studies did not require participants to directly interact with the robots or any other automated systems, but rather to merely observe the robots and subsequently answer questions about their feelings towards them. It may be the case that the lack of any such interactivity in my design contributed to the lack of an effect of PAS All or Nothing on feelings of trust.

Lyons and Guznov 's (2019) study was similar to Merritt et. al.'s (2015) in the sense that they also put participants through simulated identification tasks with automated aid, but they also did not find any significant effects of All or Nothing thinking. They explain their lack of consistency with Merritt et. al.'s data as possibly being the result of participants' familiarity with X-ray technology making errors more salient to them and invoking greater propensity for all-or-nothing thinking. Perhaps an effect of All or Nothing on trust requires a combination of interactivity with the automated system in question as well as a certain degree of familiarity with said system.

3. Conclusions

Limitations and Future Work

There are a number of ways in which my experimental design was limited and could be improved for subsequent research to provide a clearer and more comprehensive perspective of these issues. Perhaps the most significant limitation was the lack of interactivity in my experiments. Participants' roles throughout the process were in a distinctly observational capacity of watching the robot as it delivered a speech and then proceeding to answer a series of questions about their perceptions of the robot. Other researchers, such as Kraus et. al. (2018), conducted studies that involved participants interacting directly with artificial agents (or in the case of Kraus et. al., a simulation of one) that would then respond differently to participants' input depending on the experimental conditions. Others, such as Bartneck et. al. (2018), ran experiments in which agents would not necessarily respond to participant input, but participants would be tasked with choosing a concrete simulated interaction with the agent in a hypothetical scenario, as opposed to my design which involved reporting feelings and perceptions. Bartneck's studies in particular resulted in significant effects of racial variations on the outcome measures. Nass' 1997 and 2000 studies, which involved a similar design to Bartneck's with gender variations, found several significant effects of both computer voice gender and participant gender on outcome measures related to perceptions of the artificial system, particularly regarding gender stereotypes. Perhaps my design would have been better served by incorporating such interactive elements, although the technical logistics of the setup may have limited the feasibility of such an approach. Future experimentation would do well to find some means of giving participants a sense of giving input directly to an agent and receiving output in turn, whether the interaction is genuine or simulated.

Another crucial limitation of my design is the physical appearance of RoboThespian. Aside from its face, which can be programmed with a wide variety of different guises, the rest of its body cannot be significantly customized, save for the color of its lighting. Other studies had more variety in the robot stimuli they used, ranging from agents that were completely mechanical in appearance to ones that were virtually indistinguishable from real humans, especially MacDorman's 2006 research on the uncanny valley. In fact, my decision to include Evacbot among my experimental conditions was informed by my suspicion that the largely homogenous appearances of the RoboThespian conditions explained the lack of significant effects among them. As was the case with the aforementioned lack of interactivity, this was largely the result of technical logistics, but further experimentation would strive to incorporate a wider variety of agents with more distinct levels of anthropomorphic appearance, as it seems rather clear that RoboThespian is just one of many such levels, regardless of any alterations made to its face or voice. Alternatively, it could also be the case that using the exact same speech and mannerisms contributed to the largely homogenous results across conditions. Perhaps using different dialogue and body language for each condition to denote different levels of anthropomorphism would have resulted in different perceptions of anthropomorphism, which may in turn have resulted in more favorable attitudes towards the more anthropomorphic robots.

The format of the study was another key limitation, in particular the fact that I deployed the survey online with the robot performances delivered as videos. Had participants been able to observe RoboThespian in person, it would have been a much more immersive experience, even with the aforementioned issues of lack of interactivity and RoboThepian's physical appearance. It would have been more akin to encountering a real person, allowing participants to be influenced in their feelings only by RoboThespian's nature as a robot and how that nature differs from that of a real person. Viewing the robot via video undermines that distinction because it introduces additional unwanted dimensions through the inherent barrier of a computer screen. Watching a video of RoboThespian is a more rigid experience, especially because of my use of cuts between wide shots of its entire body and close-up shots of its face, as participants could not see the full extent of its body language and facial expressions. Furthermore, the lighting used in the videos was rather dark in order to make the light from RoboThespian's face as visible as possible, which had the side effect of making other parts of its body difficult to see clearly, such as individual fingers. Much of the established literature also used videos and still images to represent agent stimuli, and it would be intriguing to replicate such experimental designs with agents that participants could observe in person.

Implications

These data have implications for research in Human-Robot Interaction. At a high level, my findings indicate that gross level of anthropomorphism is significantly important in fostering trust in artificial systems, and not only because of perceived physical limitations. It may not be the case that participants surmised that Evacbot's unusual appendages or inability to bend over or other such factors prevented it from functioning as well as the RoboThespian conditions because there were some differences in Trust with Information across guise conditions, and because the Godspeed appraisals were strong predictors of the Trust ratings. Neither Trust with Information nor the Godspeed measures have anything to do with any physical limitations that Evacbot may have possessed. Although, it is possible that Evacbot's design conveyed lower intelligence as indexed by coherent movement coordination. Conversely, the social persona assigned to a given agent is much less important by comparison, at least within the specific context of interacting with a robot designed for household chores, although potential gender and racial biases cannot be dismissed entirely because the aforementioned limitations of my experimental design. Furthermore, the correlation and regression analyses I conducted with the PAS measures suggest that the High Expectations subscale is a crucial individual difference measure because of its strong correlation with the trust ratings and Godspeed measures. I also found consistent correlations between the trust ratings and Godspeed appraisals across several rounds of testing, specifically the finding that perceived Intelligence and Likability predicted Trust with Objects, while perceived Likability and Aliveness predicted Trust with Agents. These results have implications about how people determine the extent to which a robot can be trusted with a given task, as these results indicate that different tasks correlate with different criteria for trust. Collectively, the correlations I found with the Godspeed appraisals and the PAS measures, in addition to the effects of guise condition on trust ratings, establish novel measures of trust that have potential applicability for future HRI research.

These results have a great deal of practical applicability in industry research in the HRI space. High-profile technology firms could use these data to inform product design decisions as well as experimental design. The effects of Evacbot versus RoboThespian point to a clear, consistent pattern of anthropomorphism affecting trust, while the correlation analyses with the Godspeed and PAS measures can provide guidelines for how to measure trust in future HRI experimentation. For example, the fact that perceived Likeability, Intelligence, and Aliveness consistently correlated strongly with the Trust measures could be used as a springboard for measuring trust along a wider variety of dimensions and comparing them with measures of these three Godspeed-inspired appraisals to see if the correlations I found would continue to hold. Alternatively, these correlations may potentially be applied retroactively to some of the existing literature I reviewed that look at favorability from various dimensions, including research related to the uncanny valley. Perhaps the outcome measures used in those studies, such as familiarity and human likeness in the case of MacDorman (2006), would correlate with trust in artificial agents in the same way that Likeability and Aliveness did in my experiments, which would raise intriguing questions about the specific relationship between trust and the uncanny valley. This opens up the possibility of a variation of my studies involving a continuum of robots with varying degrees of human likeness from completely mechanical to completely humanlike, with trust metrics similar to the ones I used. Would participants' trust ratings have a similar distribution to the uncanny valley, with trust scaling up with human likeness until a certain point where trust declines?

Overall, my findings have potentially important implications on the study of Human-Machine Interaction. They suggest several factors in determining how much trust people are willing to put into artificial agents, and they establish novel measures of trust and perceptions of robots. As artificial intelligence becomes increasingly more prominent in society and everyday life, it also becomes increasingly more important to have such data to inform decisions about how AI systems should be designed, and equally important to have a means of evaluating people's trust in such systems. For example, when developing a robot for providing medical care, designers would need to consider how to design the robot's physical appearance, voice, and mannerisms in a way that would make patients comfortable when interacting with it, so that they can be given the medical care they need. Ideally, such judgments would be informed by empirical data about how to make people more trusting of robots, as well as how to measure trust in robots to begin with. The research I have conducted here has the capacity to provide exactly that.

References

1. Addison, A., Bartneck, C., & Yogeeswaran, K. (2019, January). Robots can be more than black and white: examining racial bias towards robots. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 493-498).
2. Axt, J. R. (2018). The best way to measure explicit racial attitudes is to ask about them. *Social Psychological and Personality Science*, *9*(8), 896-906.
3. Bartneck, C., Croft, E., & Kulic, D. (2008). Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots.
4. Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71-81.
5. Weiss, A., & Bartneck, C. (2015, August). Meta analysis of the usage of the godspeed questionnaire series. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 381-388). IEEE.
6. Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018, February). Robots and racism. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 196-204).
7. Berglund, H. (2017). Stereotypes of British Accents in Movies: A Speech Analysis of Character Types in Movies with British Accents.
8. Billings, D. R., Schaefer, K. E., Chen, J. Y., & Hancock, P. A. (2012, March). Human-robot interaction: developing trust in robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 109-110).
9. Brink, K. A., & Wellman, H. M. (2020). Robot teachers for children? Young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*, *56*(7), 1268.
10. Carpenter, J. (2009). Why Send the Terminator To Do R2D2's Job?: Designing Androids As Rhetorical Phenomena. *Proceedings of HCI 2009: Beyond Gray Droids: Domestic Robot Design for the 21 st Century*.
11. Carpenter, J., Davis, J. M., Erwin-Stewart, N., Lee, T. R., Bransford, J. D., & Vye, N. (2009). Gender representation and humanoid robots designed for domestic use. *International Journal of Social Robotics*, *1*(3), 261.
12. Cassidy, E. F., & Stevenson Jr, H. C. (2005). They wear the mask: Hypervulnerability and hypermasculine aggression among African American males in an urban remedial disciplinary school. *Journal of Aggression, Maltreatment & Trauma*, *11*(4), 53-74.
13. Clark, C. J., Liu, B. S., Winegard, B. M., & Ditto, P. H. (2019). Tribalism is human nature. *Current Directions in Psychological Science*, *28*(6), 587-592.
14. Coeckelbergh, M. (2012). Can we trust robots?. *Ethics and information technology*, *14*(1), 53-60.
15. Eibach, R. P., & Mock, S. E. (2011). The vigilant parent: Parental role salience affects parents' risk perceptions, risk-aversion, and trust in strangers. *Journal of Experimental Social Psychology*, *47*(3), 694-697.
16. Esposito, A., Amorese, T., Cuciniello, M., Riviello, M. T., & Cordasco, G. (2020, September). How Human Likeness, Gender and Ethnicity affect Elders' Acceptance of Assistive Robots. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)* (pp. 1-6). IEEE.
17. Eyssel, F., & Hegel, F. (2012). (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology*, *42*(9), 2213-2230.
18. Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, *51*(4), 724-731.
19. Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3-4), 143-166.

20. Geiskkovitch, D. Y., Thiessen, R., Young, J. E., & Glenwright, M. R. (2019, March). What? That's Not a Chair!: How Robot Informational Errors Affect Children's Trust Towards Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 48-56). IEEE.
21. Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125-130.
22. Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2020). Evolving trust in robots: specification through sequential and comparative meta-analyses. *Human Factors*, 0018720820922080.
23. Johnson, J. A., Cheek, J. M., & Smither, R. (1983). The structure of empathy. *Journal of personality and social psychology*, *45*(6), 1299.
24. K. Barfield, J. (2021, June). Discrimination and Stereotypical Responses to Robots as a Function of Robot Colorization. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 109-114).
25. Kennedy, R. L. (1995). Conservatives' Selective Use of Race in the Law. *Harv. JL & Pub. Pol'y*, *19*, 719.
26. Kraus, M., Kraus, J., Baumann, M., & Minker, W. (2018, May). Effects of gender stereotypes on trust and likability in spoken human-robot interaction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
27. Kushins, E. R. (2014). Sounding like your race in the employment process: An experiment on speaker voice, race identification, and stereotyping. *Race and Social Problems*, *6*(3), 237-248.
28. Law, T., & Scheutz, M. (2021). Trust: Recent concepts and evaluations in human-robot interaction. In *Trust in Human-Robot Interaction* (pp. 27-57). Academic Press.
29. Lee, E. J., Nass, C., & Brave, S. (2000, April). Can computer-generated speech have gender? An experimental test of gender stereotype. In *CHI'00 extended abstracts on Human factors in computing systems* (pp. 289-290).
30. Lee, J., Lee, D., & Lee, J. G. (2021). Can Robots Help Working Parents with Childcare? Optimizing Childcare Functions for Different Parenting Characteristics. *International Journal of Social Robotics*, 1-19.
31. Lyons, J. B., & Guznov, S. Y. (2019). Individual differences in human–machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, *20*(4), 440-458.
32. MacDorman, K. F. (2006, July). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science* (pp. 26-29).
33. Merritt, S. M., Unnerstall, J. L., Lee, D., & Huber, K. (2015). Measuring individual differences in the perfect automation schema. *Human factors*, *57*(5), 740-753.
34. Mills, S. (2019). Racism and Robots. *Viktor Shklovsky's Heritage in Literature, Arts, and Philosophy*, 151.
35. Moshagen, M., Musch, J., & Göritz, A. S. (2009). A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics*, *52*(10), 1311-1320.
36. Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of applied social psychology*, *27*(10), 864-876.
37. Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities?. *International Journal of Human-Computer Studies*, *43*(2), 223-239.
38. Oros, M., Nikolić, M., Borovac, B., & Jerković, I. (2014, November). Children's preference of appearance and parents' attitudes towards assistive robots. In *2014 IEEE-RAS International Conference on Humanoid Robots* (pp. 360-365). IEEE.

39. Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge, UK: Cambridge university press.
40. Robinette, P., Howard, A. M., & Wagner, A. R. (2017). Effect of robot performance on human–robot trust in time-critical situations. *IEEE Transactions on Human-Machine Systems*, *47*(4), 425-436.
41. Ruiz-Cantero, M. T., Vives-Cases, C., Artazcoz, L., Delgado, A., Calvente, M. D. M. G., Miqueo, C., ... & Valls, C. (2007). A framework to analyse gender bias in epidemiological research. *Journal of Epidemiology & Community Health*, *61*(Suppl 2), ii46-ii53.
42. Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75-84.
43. Tolksdorf, N. F., & Rohlfing, K. J. (2020). Parents' Views on Using Social Robots for Language Learning. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 634-640). IEEE.
44. van Pinxteren, M. M., Wetzels, R. W., Rüger, J., Pluymaekers, M., & Wetzels, M. (2019). Trust in humanoid robots: implications for services marketing. *Journal of Services Marketing*.
45. Wright, B. (1962). The influence of hue, lightness, and saturation on apparent warmth and weight. *The American Journal of Psychology*, *75*(2), 232-241.

Appendix A: Questionnaires for Gender Experiments

1. Experiment 1

After participants consented to participate in the study, they were shown one of the videos at random, after which they were instructed to answer a series of questions on a series of seven-point Likert scales. The questions with their associated composite variables are as follows:

Please rate how much you would trust the robot to perform the following tasks. (Assume that the robot has been programmed and prepared with the skills relevant for each task.)

- **Trust to Care for Inanimate Objects:** I would trust this robot to take care of…
    - …my cleaning
    - …my laundry
    - …my cooking
    - …my valuables
    - …a plant
- **Trust to Care for Living Agents:** I would trust this robot to take care of…
    - …a pet
    - …an elderly person
    - …an infant
    - …a young child
    - …a teenager
- **Trust to Carry Inexpensive Items:** I would trust this robot to physically carry…
    - …my clothes
    - …my food
- **Trust to Carry Valuable Items:** I would trust this robot to physically carry…
    - …my computer
    - …my money
    - …a baby

Please rate how much you agree with the following statements:

- Perceived **Likability**
    - The robot seems **friendly**.
    - I **like** this robot.
- Perceived **Intelligence**
    - This robot seems **intelligent**.
    - This robot seems **responsible**.
- Perceived **Aliveness**
    - This robot seems **conscious**.
    - This robot seems **alive**.

Attention Check, Multiple Choice: How many letters are in the word “elliptical”?

Demographic Questions:

- Age
- Gender
    - Male
    - Female
    - Other identification
- Have you ever been a parent?
    - Yes
    - No
- Have you ever been responsible for the care of a child?
    - Yes
    - No
- If you answered "yes" to the previous question, please specify the age of the child or children you have been responsible for. Check all that apply.
    - 0-2 years
    - 3-5 years
    - 6-10 years
    - 10-12 years
    - 13+ years
- Have you ever been a caregiver for a senior citizen?
    - Yes
    - No
- Have you ever owned a pet?
    - Yes
    - No

2. Experiment 2

After participants consented to participate in the study, they were shown one of the videos at random, after which they were instructed to answer a series of questions on a series of seven-point Likert scales. The questions with their associated composite variables are as follows:

Please rate how much you would trust the robot to perform the following tasks. (Assume that the robot has been programmed and prepared with the skills relevant for each task.)

- **Trust to Care for Inanimate Objects:** I would trust this robot to take care of…
    - …my cleaning

    - …my laundry
    - …my cooking
    - …my valuables
    - …a plant
- **Trust to Care for Living Agents:** I would trust this robot to take care of…
    - …a pet
    - …an elderly person
    - …an infant
    - …a young child
    - …a teenager
- **Trust to Safeguard Information:** I would trust this robot to safely manage my…
    - **…**bank account information
    - …credit and debit card numbers
    - …online passwords
    - …social security number
    - …passport and/or driver's license information
    - …healthcare-related information

Now, please imagine that the robot you just met were assaulted by a person who is biased against robots. The person strikes the robot repeatedly with a baseball bat, knocking the robot to the ground and causing significant damage. Wires and gears are exposed, the robot has lost the use of its left arm and leg, and the face is cracked.

Sympathy: Please rate how much you agree with the following statements.

- I would feel sympathetic for the robot
- I would feel sorry for the robot
- I would wish for the person to stop hitting the robot

Please rate how much you agree with the following statements:

- Perceived **Likability**
    - The robot seems **friendly**.
    - I **like** this robot.
- Perceived **Intelligence**
    - This robot seems **intelligent**.
    - This robot seems **responsible**.
- Perceived **Aliveness**
    - This robot seems **conscious**.
    - This robot seems **alive**.
- Perceived **Dominance**

  - This robot seems **dominant**.
  - This robot seems **confident**.
  - This robot seems **assertive**.
- Perceived **Nurturance**
  - This robot seems **warm**.
  - This robot seems **nurturing**.

Attention Check, Multiple Choice: How many letters are in the word “elliptical”?

Demographic Questions:

- Age
- Gender
  - Male
  - Female
  - Other identification
- Have you ever been a parent?
  - Yes
  - No
- Have you ever been responsible for the care of a child?
  - Yes
  - No
- If you answered "yes" to the previous question, please specify the age of the child or children you have been responsible for. Check all that apply.
  - 0-2 years
  - 3-5 years
  - 6-10 years
  - 10-12 years
  - 13+ years
- Have you ever been a caregiver for a senior citizen?
  - Yes
  - No
- Have you ever owned a pet?
  - Yes
  - No

Please rate how much you agree with following statements:

- **Perfect Automation Schema – High Expectations:**
  - Automated systems can always be counted on to make accurate decisions
  - People have no reason to question the decisions automated systems make
  - Automated systems have 100% perfect performance

  - Automated systems rarely make mistakes
- **Perfect Automation Schema – All or Nothing:**
  - If an automated system makes an error, then it is broken
  - If an automated system makes a mistake, then it is completely useless
  - Only faulty automated systems provide imperfect results

3. Experiment 3

After participants consented to participate in the study, they were shown one of the videos at random, after which they were instructed to answer a series of questions on a series of seven-point Likert scales. The questions with their associated composite variables are as follows:

Please rate how much you would trust the robot to perform the following tasks. (Assume that the robot has been programmed and prepared with the skills relevant for each task.)

- **Trust to Care for Inanimate Objects:** I would trust this robot to take care of…
  - …my cleaning
  - …my laundry
  - …my cooking
  - …my valuables
  - …a plant
- **Trust to Care for Living Agents:** I would trust this robot to take care of…
  - …a pet
  - …an elderly person
  - …an infant
  - …a young child
  - …a teenager
- **Trust to Safeguard Information:** I would trust this robot to safely manage my…
  - **…**bank account information
  - …credit and debit card numbers
  - …online passwords
  - …social security number
  - …passport and/or driver's license information
  - …healthcare-related information

Now, please imagine that the robot you just met were assaulted by a person who is biased against robots. The person strikes the robot repeatedly with a baseball bat, knocking the robot to the ground and causing significant damage. Wires and gears are exposed, the robot has lost the use of its left arm and leg, and the face is cracked.

Sympathy: Please rate how much you agree with the following statements.

- I would feel sympathetic for the robot
- I would feel sorry for the robot
- I would wish for the person to stop hitting the robot

Please rate how much you agree with the following statements:

- Perceived **Likability**
  - The robot seems **friendly**.
  - I **like** this robot.
- Perceived **Intelligence**
  - This robot seems **intelligent**.
  - This robot seems **responsible**.
- Perceived **Aliveness**
  - This robot seems **conscious**.
  - This robot seems **alive**.
- Perceived **Dominance**
  - This robot seems **dominant**.
  - This robot seems **confident**.
  - This robot seems **assertive**.
- Perceived **Nurturance**
  - This robot seems **warm**.
  - This robot seems **nurturing**.

Please rate how much you agree with following statements:

- **Perfect Automation Schema – High Expectations:**
  - Automated systems can always be counted on to make accurate decisions
  - People have no reason to question the decisions automated systems make
  - Automated systems have 100% perfect performance
  - Automated systems rarely make mistakes
- **Perfect Automation Schema – All or Nothing:**
  - If an automated system makes an error, then it is broken
  - If an automated system makes a mistake, then it is completely useless
  - Only faulty automated systems provide imperfect results

Attention Check, Multiple Choice: How many letters are in the word “elliptical”?

Demographic Questions:

- Age
- Gender
  - ○ Male
  - ○ Female
  - ○ Other identification
- Have you ever been a parent?
  - ○ Yes
  - ○ No
- Have you ever been responsible for the care of a child?
  - ○ Yes
  - ○ No
- If you answered "yes" to the previous question, please specify the age of the child or children you have been responsible for. Check all that apply.
  - ○ 0-2 years
  - ○ 3-5 years
  - ○ 6-10 years
  - ○ 10-12 years
  - ○ 13+ years
- Have you ever been a caregiver for a senior citizen?
  - ○ Yes
  - ○ No
- Have you ever owned a pet?
  - ○ Yes
  - ○ No

Gender Preferences – 10-point scales from Extremely incapable to Extremely capable:

- ○ Please rate your feelings towards men
- ○ Please rate your feelings towards women

Appendix B: Questionnaires for Race Experiments

1. Experiment 4

After participants consented to participate in the study, they were shown one of the videos at random, after which they were instructed to answer a series of questions on a series of seven-point Likert scales. The questions with their associated composite variables are as follows:

Please rate how much you would trust the robot to perform the following tasks. (Assume that the robot has been programmed and prepared with the skills relevant for each task.)

- **Trust to Care for Inanimate Objects:** I would trust this robot to take care of…
  - …my cleaning
  - …my laundry
  - …my cooking
  - …my valuables
  - …a plant
- **Trust to Care for Living Agents:** I would trust this robot to take care of…
  - …a pet
  - …an elderly person
  - …an infant
  - …a young child
  - …a teenager
- **Trust to Safeguard Information:** I would trust this robot to safely manage my…
  - **…**bank account information
  - …credit and debit card numbers
  - …online passwords
  - …social security number
  - …passport and/or driver's license information
  - …healthcare-related information

Please rate how much you agree with the following statements:

- Perceived **Likability**
  - The robot seems **friendly**.
  - I **like** this robot.
- Perceived **Intelligence**
  - This robot seems **intelligent**.
  - This robot seems **responsible**.
- Perceived **Aliveness**
  - This robot seems **conscious**.
  - This robot seems **alive**.

- Perceived **Dominance**
  - This robot seems **dominant**.
  - This robot seems **confident**.
  - This robot seems **assertive**.

Now, please imagine that the robot you just met were assaulted by a person who is biased against robots. The person strikes the robot repeatedly with a baseball bat, knocking the robot to the ground and causing significant damage. Wires and gears are exposed, the robot has lost the use of its left arm and leg, and the face is cracked.

Sympathy: Please rate how much you agree with the following statements.

- I would feel sympathetic for the robot
- I would feel sorry for the robot
- I would wish for the person to stop hitting the robot

Perceived Danger: Please rate how much you agree with the following statements.

If this robot were in my home…

- …this robot could be dangerous
- …this robot could seem threatening
- …this robot might physically harm me or my family

Please rate how much you agree with following statements:

- **Perfect Automation Schema – High Expectations:**
  - Automated systems can always be counted on to make accurate decisions
  - People have no reason to question the decisions automated systems make
  - Automated systems have 100% perfect performance
  - Automated systems rarely make mistakes
- **Perfect Automation Schema – All or Nothing:**
  - If an automated system makes an error, then it is broken
  - If an automated system makes a mistake, then it is completely useless
  - Only faulty automated systems provide imperfect results

Attention Check, Multiple Choice: How many letters are in the word “elliptical”?

Demographic Questions:

- Age
- Gender
  - Male
  - Female
  - Other identification
- Please state your ethnicity
  - White
  - Black or African American
  - American Indian or Alaska Native
  - Asian
  - Native Hawaiian or Pacific Islander
  - Other
- Which U.S. political group do you primarily identify with?
  - Republican party
  - Other Conservative party
  - Democratic party
  - Other Progressive party
  - Independents
  - No opinion
- Have you ever been a parent?
  - Yes
  - No
- Have you ever been responsible for the care of a child?
  - Yes
  - No
- If you answered "yes" to the previous question, please specify the age of the child or children you have been responsible for. Check all that apply.
  - 0-2 years
  - 3-5 years
  - 6-10 years
  - 10-12 years
  - 13+ years
- Have you ever been a caregiver for a senior citizen?
  - Yes
  - No
- Have you ever owned a pet?
  - Yes
  - No

Race Preferences: Please answer the following as honestly as you can.

- Which statement best describes you?
  - I prefer Black Americans to European Americans.
  - I like European Americans and Black Americans equally.
  - I prefer European Americans to Black Americans.
- Please rate your feelings towards Black Americans.
  - 10-point scale from Extremely cold to Extremely warm
- Please rate your feelings towards European Americans
  - 10-point scale from Extremely cold to Extremely warm

2. Experiment 5

After participants consented to participate in the study, they were shown one of the videos at random, after which they were instructed to answer a series of questions on a series of seven-point Likert scales. The questions with their associated composite variables are as follows:

Please rate how much you would trust the robot to perform the following tasks. (Assume that the robot has been programmed and prepared with the skills relevant for each task.)

- **Trust to Care for Inanimate Objects:** I would trust this robot to take care of…
  - …my cleaning
  - …my laundry
  - …my cooking
  - …my valuables
  - …a plant
- **Trust to Care for Living Agents:** I would trust this robot to take care of…
  - …a pet
  - …an elderly person
  - …an infant
  - …a young child
  - …a teenager
- **Trust to Safeguard Information:** I would trust this robot to safely manage my…
  - **…**bank account information
  - …credit and debit card numbers
  - …online passwords
  - …social security number
  - …passport and/or driver's license information
  - …healthcare-related information

Please rate how much you agree with the following statements:

- Perceived **Likability**

  - The robot seems **friendly**.
  - I **like** this robot.
  - The robot seems **agreeable**.
- Perceived **Intelligence**
  - This robot seems **intelligent**.
  - This robot seems **responsible**.
- Perceived **Aliveness**
  - This robot seems **conscious**.
  - This robot seems **alive**.
- Perceived **Dominance**
  - This robot seems **dominant**.
  - This robot seems **confident**.
  - This robot seems **assertive**.

Now, please imagine that the robot you just met were assaulted by a person who is biased against robots. The person strikes the robot repeatedly with a baseball bat, knocking the robot to the ground and causing significant damage. Wires and gears are exposed, the robot has lost the use of its left arm and leg, and the face is cracked.

Sympathy: Please rate how much you agree with the following statements.

- I would feel sympathetic for the robot
- I would feel sorry for the robot
- I would wish for the person to stop hitting the robot

Perceived Danger: Please rate how much you agree with the following statements.

If this robot were in my home…

- …this robot could be dangerous
- …this robot could seem threatening
- …this robot might physically harm me or my family

Please rate how much you agree with following statements:

- **Perfect Automation Schema – High Expectations:**
  - Automated systems can always be counted on to make accurate decisions
  - People have no reason to question the decisions automated systems make
  - Automated systems have 100% perfect performance

  - Automated systems rarely make mistakes
- **Perfect Automation Schema – All or Nothing:**
  - If an automated system makes an error, then it is broken
  - If an automated system makes a mistake, then it is completely useless
  - Only faulty automated systems provide imperfect results

Attention Check, Multiple Choice: How many letters are in the word “elliptical”?

Demographic Questions:

- Age
- Gender
  - Male
  - Female
  - Other identification
- Please state your ethnicity
  - White
  - Black or African American
  - American Indian or Alaska Native
  - Asian
  - Native Hawaiian or Pacific Islander
  - Other
- Which U.S. political group do you primarily identify with?
  - Republican party
  - Other Conservative party
  - Democratic party
  - Other Progressive party
  - Independents
  - No opinion
- Have you ever been a parent?
  - Yes
  - No
- Have you ever been responsible for the care of a child?
  - Yes
  - No
- If you answered "yes" to the previous question, please specify the age of the child or children you have been responsible for. Check all that apply.
  - 0-2 years
  - 3-5 years
  - 6-10 years
  - 10-12 years
  - 13+ years

- Have you ever been a caregiver for a senior citizen?
    - Yes
    - No
- Have you ever owned a pet?
    - Yes
    - No

Race Preferences: Please answer the following as honestly as you can.

- Which statement best describes you?
    - I prefer Black Americans to European Americans.
    - I like European Americans and Black Americans equally.
    - I prefer European Americans to Black Americans.
- Please rate your feelings towards Black Americans.
    - 10-point scale from Extremely cold to Extremely warm
- Please rate your feelings towards European Americans
    - 10-point scale from Extremely cold to Extremely warm
- Please rate your feelings towards Muslims
    - 10-point scale from Extremely cold to Extremely warm
- Please rate your feelings towards robots
    - 10-point scale from Extremely cold to Extremely warm
- Please rate your feelings towards roboticists (engineers who design robots)
    - 10-point scale from Extremely cold to Extremely warm