Effects of Embodiment Measuring Task Performance*

Anthropomorphic and Zoomorphic Embodiment in Social Human Robot Interaction

Alina Momin
College of Computer Science
and Electrical Engineering
University of Maryland,
Baltimore County
alinam1@umbc.edu

Faheel Kamran
College of Computer Science
and Electrical Engineering
University of Maryland,
Baltimore County
f74@umbc.edu

KEYWORDS

embodiment, social interaction, anthropomorphic, zoomorphic

1 INTRODUCTION

Siri (Apple), Cortona (Microsoft), and Alexa (Amazon) are disembodied robots capable of interacting and engaging with people. While the conversations between virtual personal assistants and humans are limited to question/answer format, the social response from people interacting with such robots is a compulsive area of study.

However, research on the effectiveness of embodiment versus disembodiment has already been conducted in-depth. In general, the answer is that embodiment shows performance improvement on tasks over virtual robots. Research on the type of physical embodiment is less clear at this time.

Hence, we are primarily interested in the social response of people interacting with anthropomorphic robots versus zoomorphic robots. This research will explore the nature of interaction between animal vs human-based embodied robots. Importance will be granted to understanding the attitudes and attachments that humans have towards the human versus animal embodiment. The research will uncover some of the differences in effectiveness of the robot in various human-animal vs human-human interaction styles. Thus, it will provide rationale for using various species for embodiment in social robots. Understanding the significance of the type of embodiment is relevant to the design of future social robots and how humans react to early consumer robotics.

2 RELATED WORKS

Extensive research has been conducted in the area of physical embodiment, human interaction with embodied robots, and social attitudes with various types of embodiments. This section includes related works that motivate our research.

The Wainer et al. 2006 paper focuses on the role of physical embodiment in HRI [1]. This work and the proposed study in this paper both “measure effectiveness” and “perceptions of social interaction” of embodied robots. However, Wainer et al. 2006 is comparing the difference between a physical and simulated robot, while the current study is comparing the effectiveness between human and animal-based robots. Again the Wainer et al. 2007 paper focuses on physical versus simulated robots. The paper also discusses task performance with regards to physical or tele-present robots [2]. No distinction between the type of physical embodiment is discussed. Information from this paper is albeit useful as its conclusions of styles of interactions and relationships guide our research to be conducted under more natural social conditions than those presented in the literature. Next, the Reig et al. 2019 paper discusses a robot’s embodiment as a means of calibrating “trust and comfort with users” [3]. It is unclear at first glance over the type of embodiment proposed. However, this paper does provide a basis for how robots can act as ‘true social partners’ [3]. Their research is also useful to this project as it has many findings on the design of re-embodiment. The research also focuses on interaction between different personalities in the embodiments. This gives useful insight on how our project will be designed to focus on understanding the assumptions human’s make about the types of interactions.

The Melson et al. 2005 paper focuses on children’s interactions and attitudes towards a robotic dog AIBO versus a real life Australian Shepherd [4]. The findings of this research support the natural social interaction between the human child and dog. This paper gives useful context in children’s social interactions with the robotic dog. Topics such as companionship, values and social responses are very useful findings to our research. The Kwak et al. 2013 paper uses two trials: one discusses human empathy for simulated versus physical robots, and the second discusses

human empathy for embodied versus disembodied robots [5]. The second trial would be similar to the proposed study, however, the paper focuses on human empathy instead of task performance and effectiveness. The proposed paper would further expand this research by investigating different forms of physical embodiment (human versus animal). The Krueger 2021 paper focuses on exploring human-dog relations in comparison to human-human relations and then applies it to the use of dogs as prototypes for robot embodiment [6]. The use of domestic animals for robot embodiment is a multidisciplinary study including studies in biology, psychology and engineering. This paper also discusses the implications of using human embodied robots and the resulting human attitudes towards the “Uncanny Valley” hypothesis. This is a great research paper for the study of human attitudes towards biological rules and behavior towards social robots including the study of ethorobotics.

Tapus et al. 2009 proposes that physical embodiment for therapist robots is more efficient and natural than simulated performance over the span of 8-months [7]. We are interested in this paper because of its implications for “efficiency” and understanding quantitatively how much more efficient physical robots are. We would be interested in extending the research for this paper by exploring the effects of a physical robo-dog (animal-based) compared to the mechanical robot used in the study. The Parviainen, J., & Pirhonen, J. 2017 paper discusses the ethical implications for robot embodiment in healthcare [8]. The Miklosi et al. 2017 paper discusses the approach taken in ethorobotics and human relationships to human-like robots [9]. The aim of this research is to develop social rules for comfortable and acceptable robot behavior. The application of the social interactions is evaluated using various clearly defined functions of the robot. This paper inspires many parts of our research by introducing the idea of using ethorobotics and applying the niche concept to our design.

3 APPROACH

The approach to understanding the significance of the type of embodiment in social robots will include both hardware and research components. The team will deliver two hardware deliverables. The first is an anthropomorphic embodiment for a custom chatbot with hardcoded answers. The second is a zoomorphic embodiment for the same chatbot. Two trials will be conducted. Each trial will include at least five participants. Participants will select at least 3-5 questions from a handbook of questions to ask the chatbot. The responses will be hardcoded into the bot and the bot will include speech recognition. Participants will choose to be assigned one embodiment at random to engage with and participants will be chosen at random.

In the first set of trials, participants will ask the anthropomorphic embodied robot questions from the handbook of pre-programmed questions. At the conclusion of each question, the participants will be surveyed on the effectiveness of the chatbot using a scale. Various methods will be used to evaluate the effectiveness and performance of each question. The participants will be surveyed on the effectiveness of the robots using a variety of currently undetermined methods. The second set of trials will include the zoomorphic embodied robot. At the conclusion of two sets of trials with a variety of participants, students will average the effective performance of similar questions. The consequential results will be included in the final research paper. Given at least ten participants with 3-5 questions each, and 3-5 follow up questions relating to the effectiveness of the answers, we will be able to draw conclusions from $10 \bullet [3,5] \bullet [3,5] = [90,250]$ data points. We believe that this may be enough data to draw simple conclusions from, while understanding the limitations of our study.

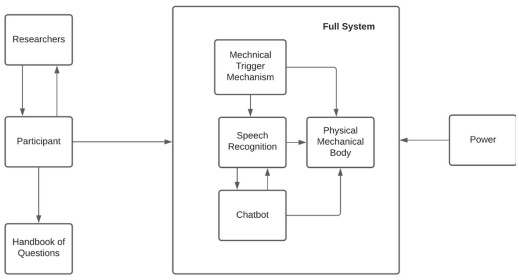

Figure 1: System Boundary Diagram

Figure 1 demonstrates a system boundary diagram for the intended design approach used in the study. The ‘Full System’ represents an anthropomorphic or zoomorphic embodiment of a robot. The system comprises of physical hardware required to design the exterior structure, a speech recognition system with microphone for audio input/output, a software defined chatbot with questions/answers hardcoded, and a trigger mechanism to engage the system.

Exterior to the software and hardware deliverables are the participants who engage with a handbook of predefined questions to choose from and the researchers

who evaluate and gather feedback. The system is powered via either a self-contained or exterior battery.

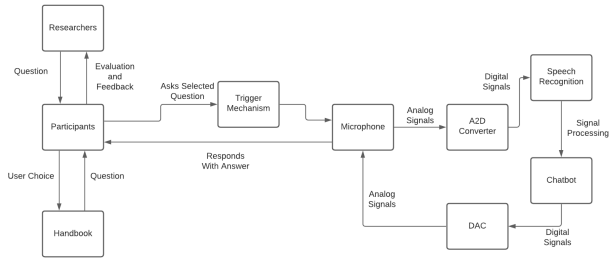

Figure 2: Data Flow Diagram

Figure 2 demonstrates how data flows through the system. The relative functions to design the system are introduced. The participant selects and chooses a question(s) from the handbook. After retrieving the question, the participant will ask the selected question to the system. A trigger mechanism will engage the microphone which captures voice input (analog signals). The signals are sent to an Analog-to-Digital converter to be processed by the processor's software defined speech recognition system. The signals are sent to the chatbot system for processing, which will select the correct response from hardcoded values. The response will be sent as digital signal data to the Digital-to-Analog converter for the microphone on the system to use. The microphone will be used as voice output for the embodied robot to respond with.

Arra the Anthropomorphic Embodiment Designed

Above we showcase the final version of the human-based embodiment, Arra. The robot has human-like qualities such as arms, hands, legs and feet. The head, eyes and mouth also represent a human figure.

Otto the Zoomorphic Embodiment Designed

Above, we showcase the animal-based embodiment Otto. Otto has dog-like features including the structure and frame of its body. The ears and replication of paws also imitate an animal based embodiment.

4 USER STUDIES

There are four objectives in the purpose of this study. (i) How will people respond and score the task-effectiveness of (anthropomorphic/zoomorphic) robots based on questions that are:

- Thoughtful
- Kind
- Humor
- Concerning or Distress

(ii) Whether people have preferences for human-based or animal-based embodiments. (iii) People's initial and pre-task expectation of the robot(s) based on appearance. (iv) People's post-task expectation of the robot. In the pursuit of these objectives, we formed the following four hypotheses:

Hypothesis 1: Participants will score the anthropomorphic embodiment higher for thoughtfulness-based questions.

Hypothesis 2: Participants will score the zoomorphic embodiment higher for kindness-based questions.

Hypothesis 3: Participants will be divided when scoring the robots for humor-inducing responses.

Hypothesis 4: Participants will be more concerned with the distressing response of the anthropomorphic embodiment and less distressed with the concerning response of the zoomorphic embodiment.

The robotic system implemented uses software-defined automated speech recognition and a researcher trained chatbot. The chatbot is designed, created and trained through “SnatchBot”, a builder platform for deploying, publishing, hosting, tracking and monitoring automated systems. SnatchBot includes natural language processing for the training and development of the robot. The full system peripherals include microphone for audio INPUT and a bluetooth speaker for audio OUTPUT. The system also includes an exterior chassis for the morphology embodiment: between zoomorphic or anthropomorphic.

The researchers are interested in testing the capabilities of the exterior chassis embodiment on the perceived task-performance effectiveness in the study. The study will be conducted in a comparative study style. This study method is chosen to allow for the comparison and contrast of the social response the participants have towards the anthropomorphic and zoomorphic embodiments. One of the independent variables will be the embodiment itself, there will be separate physical robots for the anthropomorphic and zoomorphic embodiment. For the sake of referencing, study one will be the anthropomorphic system. Study two will be the zoomorphic system. Study one and study two will have a question handbook, a questionnaire for the participants to rate their interactions, and a speaker/microphone system to capture sound for voice recognition.

There will be a singular repository of questions that will be curated for both embodiments. There will be a series of questions that the participants will answer about the interaction with each embodiment. For the interaction rating, there will be a scale from 1-5 and additional questions that the participant will answer about the effectiveness of their interaction with the system. The user will be able to select questions to ask the robot from a handbook. After each question-answer interaction, the participant will rate their perception of the interaction. The interaction rating scale and open ended questions will be used to measure the perception and thoughtfulness of the robot’s response.

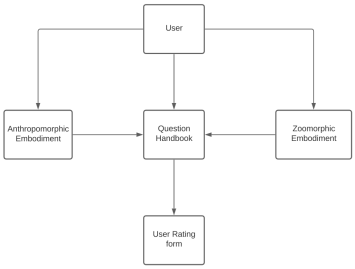

Figure 3: Mission Scenario Diagram

The following is a sample interaction between the (anthropomorphic/zoomorphic) robot and the participant.

First, the participant will be provided a handbook of questions. The participant will select questions from the handbook to ask the robot. The participant will be provided a pre-evaluation form to complete prior to their interaction with the robot. The participant will be asked to analyze the question and complete an evaluation of their expectations of the response from the robot. After this initial evaluation is completed, the participant will ask the question into the microphone to be considered by the chatbot. Then, the researcher will activate the trigger mechanism for the microphone signal, or determine an automated approach to enable the chatbot and speech recognition software. The chatbot will use keywords or natural language processing implemented on SnatchBot to respond with through the speaker. After each question is asked, the user will rate their interaction in a variety of evaluation methods described in the sections above. Using the follow-up form, the user will evaluate the thoughtfulness or effectiveness of the robot. Participants will repeat the process or continue until all of the questions have been completed.

Researchers and participants will undergo a sample familiarization task with the question “how are you today?”. The intention of this pre-study task is to demonstrate to the participant how the interaction will go and address any questions or concerns with the research study. The participant is expected to select a question, complete a pre and post evaluation form, and ask the question to the robot. The research study will need a table for the robotic system to sit/stand on. The laptop,

microphone and evaluation forms will also be on the table. Two chairs will be provided for the researcher and the participant. No actors will be used in the study.

Participants will be recruited via friends, family and associates of the researchers. Researchers will also reach out to the university community to ask for participants. The study will be conducted in various locations that the researchers and participants agree on.

The following questions were used during the experiment.

Thoughtful

1. Do you think crying is a sign of weakness or strength?
2. What is the difference between living and existing?
3. Are you holding onto something you need to let go of?
4. What is the meaning of peace to you?

Kind

1. What impact do you want to leave on the world?
2. Is it more important to love or be loved?
3. How can you show your care for the world?
4. What is your golden rule in life?

Humor

1. Do robots have feelings?
2. What do you love most about yourself?
3. What does it mean to be human?
4. Do you trust humans?

Distress

1. How do you feel about humans?
2. Would you break the law to save a loved one?
3. What is important enough to go to war over?
4. Is it ever right to do the wrong thing?

The following questions and statements will be used during the pre-evaluation form. Forms will use a likert scale from 1-5, signifying strongly disagree as a 1 and a strongly agree as 5.

Pre-Evaluation Questions:

1. The robot will answer this question as well as a human would.
2. The robot's answer will be relevant to the question asked.
3. The robot's answer will be natural.
4. The robot's answer will hold value to me.

The following questions will be used during the post-evaluation forms to gauge how participants perceived the effectiveness of the answer post trial.

Post-Evaluation Questions:

1. The robot answered the question as well as a human would.
2. The robot's answer was relevant to the question asked.
3. The robot's answer was natural.
4. The robot's answer held value to me.

5 Results

Researchers created 28 figures and graphs for each of the 14 questions representing results from Arra and Otto's trials. For each of the four categories (thoughtful, kind, humorous and distress), researchers took cumulative averages across the categories to create 8 figures comparing Arra and Otto's results in each category. In Figure 4 below, Arra's likert scale in the pre-evaluation was [2, 4, 2, and 2] representing the evaluation questions "will respond as well as a human ", "response will be relevant, natural or hold value". The post-evaluation results were [4, 5, 4 and 4].

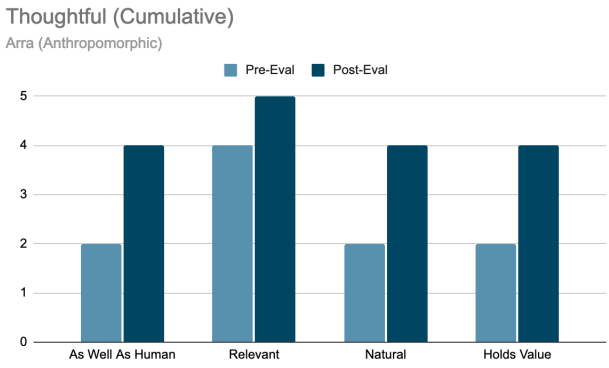

Figure 4: Arra's Cumulative Thoughtful Results

Figure 5 below demonstrates the same category results for Otto. The pre-evaluation results were [2, 3, 2 and 2]. The post-evaluation results were [4, 4, 3 and 3]. Participants felt better about the effectiveness of the robot after the

interaction as all categories saw an increase in 1-2 ranking points.

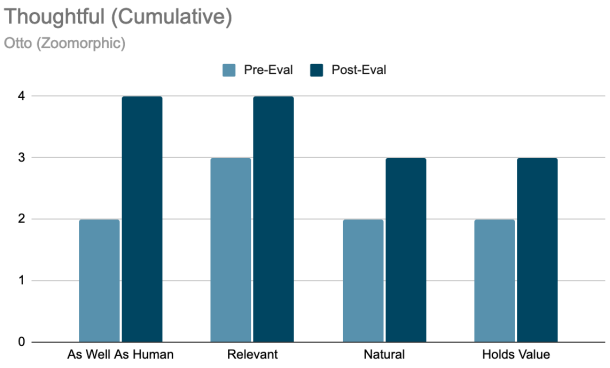

Figure 5: Otto's Cumulative Thoughtful Results

Figure 6 below shows the results for Arra in the "kindness-induced" response category. The pre-evaluation results showed [2, 3, 3 and 3]. The post-evaluation results show [4, 4, 4 and 3]. Holding value was the only category that did not see an improvement post-evaluation, indicating that participant's own feelings regarding the hard-coded answers provided could have impacted the scaled rankings for this section.

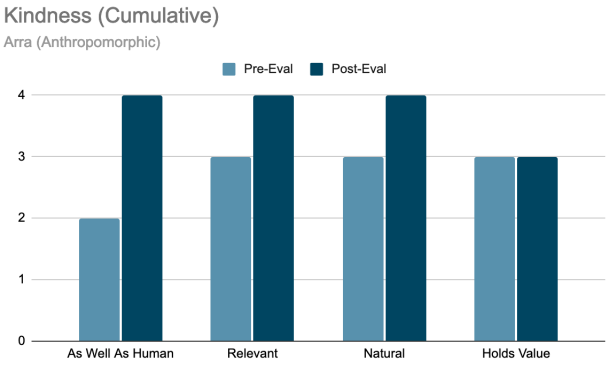

Figure 6: Arra's Cumulative Kindness Results

Figure 7 shows Otto's results in the same category. The pre and post evaluation results were [1, 2, 3 and 3] and [3, 3, 3, and 3] respectively. There was a significant improvement in answering as well as a human after the trial was complete. A slight improvement in the perceived relevance was also noted. There was no improvement in the perception of the answer being more or less natural or holding value after the trial was complete.

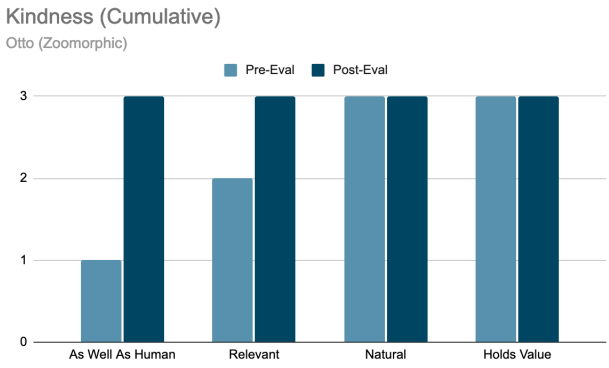

Figure 7: Otto's Cumulative Kindness Results

Figure 8 shows Arra's averages for the humor category.

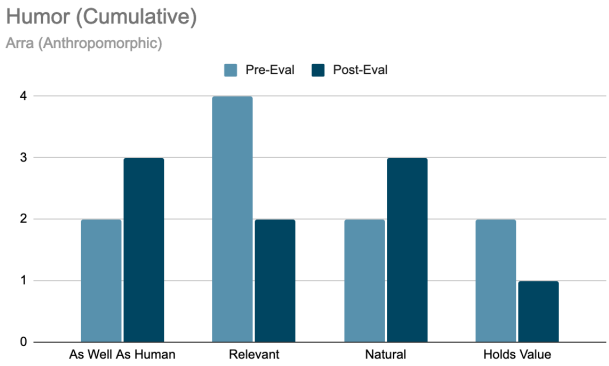

Figure 8: Arra's Cumulative Humor Results

The results were [2, 4, 2, 2] and [3, 2, 3, 1] respectively. There was a slight improvement in being able to provide a humorous answer as well as a human would, as well as provide a more natural answer. The relevance and value held had shown significant decreases. This indicates that participants were more focused on the response than the embodiment for this humor based question.

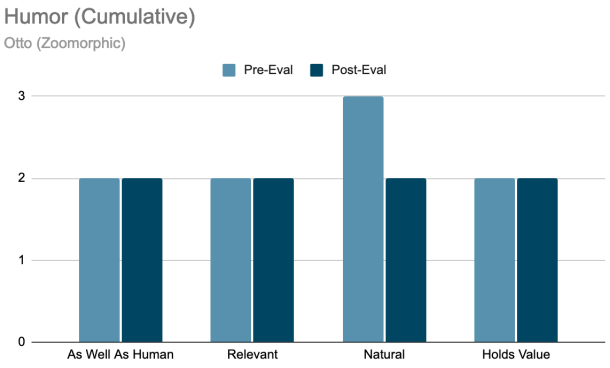

Figure 9: Otto's Cumulative Humor Results

Figure 9 shows Otto's results for the same humor category. The results were [2, 2, 3, 2] in the pre-evaluation questionnaire and [2, 2, 2, 2] in the post-evaluation. For Figure 10 below, Arra's results in the final category of distress were [2, 3, 3, 2] in the pre-evaluation and [2, 3, 2, 2] in the post-evaluation.

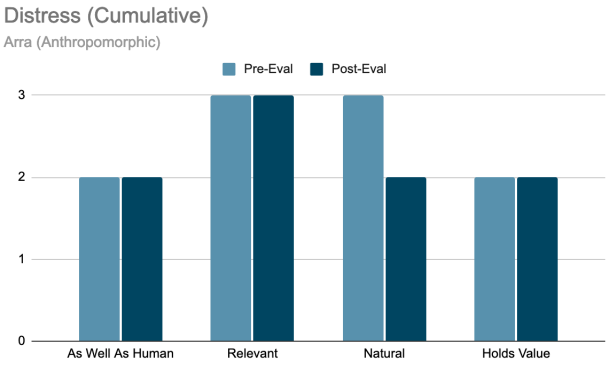

Figure 10: Arra's Cumulative Distress Results

Finally in Figure 11 below, Otto's distressed results were [2, 2, 2, 2] in the pre-evaluation questionnaire. Otto's post-evaluation results were [3, 3, 2, 2]. Otto shows an improvement in the post-evaluation forms for two of the four categories.

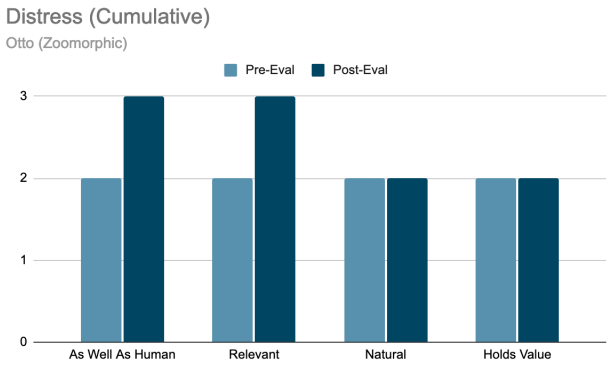

Figure 11: Otto's Cumulative Distress Results

The findings of the study are discussed in more detail in the sections below.

6 Conclusions/Discussion

In the thoughtful category, Arra outperformed Otto across the board in both pre-and-post evaluations. This was in-line with the hypothesis that the anthropomorphic embodiment will outperform the zoomorphic embodiment for thoughtful-induced responses. Participants had believed that Arra was initially expected to have a more relevant response than Otto (3 versus 2). The other pre-evaluation category averages were the same. In the post-evaluation, participants believed that Arra's response was more relevant, natural and held more value when compared to Otto's same response. This would indicate that a human-based embodiment slightly affects the perceived task performance for a social robot in thoughtful scenarios.

In the kindness category, Arra again outperformed Otto. In the pre-evaluation, Arra was expected to answer slightly more as a human would as well as have a more relevant response than Otto. Arra and Otto both had a 3 for expectations of natural or holding value. In the post-evaluation, Arra was ranked slightly higher than Otto for "as well as a human", "relevance" and "natural". Arra and Otto had the same ranking for holding value. The results were not consistent with our hypothesis. Researchers had believed that an animal-based embodiment, in our instance a dog, would be perceived as more friendly or kind than our human-based embodiment. The results would indicate that a human-based embodiment again would be perceived as slightly more effective in kindness-based scenarios.

In the humor category, there were mixed results indicating a less clear understanding of the role of the type of embodiment in humor-based situations. Participants in the pre-evaluation questionnaire expected Arra to provide a significantly more relevant answer when compared to Otto. In contrast, participants expected Otto to provide a more natural response initially. In the post-evaluation results, Arra's response was seen as slightly higher for "as well as a human would ". Arra's response was also seen as more natural. In contrast, Otto's post-evaluation results indicate that participants felt Otto's response held a slightly higher value than Arra's, despite being the same response. The results were in-line with the hypothesis, since participants were divided between whether the human-based or the animal-based embodiment was perceived to be more effective for humor based questions. This would indicate that in humor-based social HRI environments, it is less clear whether designers should use human or animal based embodiments.

In the distressed category, Arra's pre-evaluation results showed that participants felt that the human-based embodiment would have a more relevant, as well as a more natural response. The other two categories were a tie. In the post-evaluation, Otto's response was seen as answered similar to a human would when compared to Arra. We find this result interesting as it is the first time that the animal-based embodiment was perceived to be more like a human than the human-based embodiment. This was also in-line with our hypothesis because the results are not entirely consistent on whether an anthropomorphic or zoomorphic embodiment is seen as better for distress/concern based social HRI.

In conclusion, researchers found that across the four categories studied, participants generally believed that an anthropomorphic (human-based) embodiment was slightly more effective than a zoomorphic (animal-based) embodiment. The design and relevance of the embodiment is essential for the development of future social robots. Since this design might be task-specific, we attempted to present four possible task-specific categories. We limit the number of significant conclusions we can draw at this time from the study as the number of participants (12 total) might not be statistically significant enough. We also caution against making bold conclusions regarding the effects of various types of embodiments as different categories of questions produced different results.

7 Future Work

To further our understanding of the role of embodiment in HRI, we would adjust the study in the future to gain a deeper understanding of the role of embodiment in social interactions.

The first is a collection of more data using more participants. Since participants were selected from populations of convenience (friends and family), we believe that diversifying the participant population pool would provide better results. Furthermore, only 12 participants were used in this study. This was broken down into seven participants for Arra (the anthropomorphic embodiment) and five participants for Otto (the zoomorphic embodiment). In our current study the demographics included people from ages [20 - 68], the gender distribution was 5 females to 7 males. We would like to conduct a more indepth user study with a larger and diverse population "since HRI may be affected by not only design but also by demographic elements such as age, gender, and experiences with technology" [6]. Social robots and the human interactions with these technologies must be tested with an extensive and diverse population to gauge an understanding of psychological effects of integrating these technologies as well as various human needs and attachment styles.

The second would be to add additional categories of study to diversify the question handbook. Presently, we proposed four categories of questions: thoughtful, kind, humorous and distress/concern. Since the results varied by category, adding additional categories would be beneficial in the study of embodiment relevant to the social environment the robots would be placed in. We propose categories that invoke a variety of emotions to best understand the relevance of embodiment in these situations.

The third would be to add specific questions for each category and questions. This would mean that the pre and post evaluation questions would be curated for each category or question within the handbook. This would allow for a question specific qualitative analysis of the user's perception of the embodiments in various social dialogues.

In order to create questions specific for each category a psychological understanding of human interactions would be necessary. It may be good to curate questions that help us gauge an in-depth understanding of why the users have certain perceptions and preferences to the embodiments. For example, in a therapeutic

conversation does the user prefer one embodiment over another and what characteristics of the embodiment have an acute impact on the user's impression.

By conducting these studies and studying human interactions with various embodiments of social robots we are able to learn more about ourselves and our inherent social nature. Future research in this area would focus on learning about how humans will behave with various embodiments in social interactions and what kind of responses the robot's interactions will elicit in the human.

REFERENCES

[1] J. Wainer, D. J. Feil-seifer, D. A. Shell and M. J. Mataric, "The role of physical embodiment in human-robot interaction," ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006, pp. 117-122, doi: 10.1109/ROMAN.2006.314404.

[2] J. Wainer, D. J. Feil-Seifer, D. A. Shell and M. J. Mataric, "Embodiment and Human-Robot Interaction: A Task-Based Perspective," RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication, 2007, pp. 872-877, doi: 10.1109/ROMAN.2007.4415207.

[3] S. Reig, J. Forlizzi and A. Steinfeld, "Leveraging Robot Embodiment to Facilitate Trust and Smoothness," 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2019, pp. 742-744, doi: 10.1109/HRI.2019.8673226.

[4] Melson G, Kahn Jr P, Beck A, Friedman B, Roberts T, Garrett E (2005) Robots as dogs?: Children's interactions with the robotic dog AIBO and a live Australian shepherd, pp 1649–1652. https://doi.org/10.1145/1056808.1056988

[5] S. S. Kwak, Y. Kim, E. Kim, C. Shin and K. Cho, "What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot," 2013 IEEE RO-MAN, 2013, pp. 180-185, doi: 10.1109/ROMAN.2013.6628441.

[6] Krueger, Frank, Kelsey C. Mitchell, Gopikrishna Deshpande, Jeffrey S. Katz. Human–Dog Relationships as a Working framework for exploring human-robot attachment: a multidisciplinary review - Springer. 24 Jan. 2021, https://link.springer.com/content/pdf/10.1007/s10071-021-01472-w.pdf.

[7] A. Tapus, C. Tapus and M. Mataric, "The role of physical embodiment of a therapist robot for individuals with cognitive impairments," RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009, pp. 103-107, doi: 10.1109/ROMAN.2009.5326211.

[8] Parviainen, J., & Pirhonen, J. (2017). Vulnerable bodies in human–robot interactions: embodiment as ethical issue in robot care for the elderly.

[9] Miklósi, Ádám, et al. "Ethorobotics: A New Approach to Human-Robot Relationship." Frontiers, Frontiers, 1 Jan. 1AD, https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00958/full.