Bachelor Data Science & Artificial Intelligence

The effect of robot personality on gameplay interaction

Ryan Sleeuwaegen

s3122166

Supervisors:
Dr.ir. D.J. Broekens
Dr. P.W.H. van der Putten

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

30/08/2024

Abstract

This study explores how different robot behavioral styles affect user perceptions by comparing interactions where robots display either happy and supportive or angry and competitive behaviors. The study was divided in two parts.

First, a pilot study was conducted to validate the prompts and behaviors programmed for the NAO robot. The primary objective was to ensure that the robot could believably portray the two distinct personalities.

Following the pilot study, a within-subject analysis was conducted with 37 participants, who experienced both types of interactions. The study measured several aspects, including: anthropomorphism, animacy, likeability, perceived intelligence, perceived safety, perceived eeriness, satisfaction, engagement, and overall experience, perceived ease of use, and willingness to interact again.

The results indicate that supportive interactions were generally perceived as more likeable, safer, and less eerie compared to competitive interactions. Participants found the supportive interactions more satisfying and engaging, and expressed a greater willingness to interact with such robots in the future. Conversely, competitive interactions were rated higher in perceived intelligence but also scored higher in eeriness and lower in perceived safety.

Additionally, a between-subjects analysis confirmed these findings, showing consistent preferences for supportive behavior over competitive behavior. This study underscores the significant role of robot behavior in shaping user experience and provides a basis for designing robots that enhance user engagement through tailored behavioral strategies. Future research should explore a broader range of behaviors, the long-term effects of interactions, and individual differences to further refine robot design and functionality.

Contents

1 Introduction

Human-robot interaction (HRI) is a rapidly advancing field that explores the active interaction between humans and robots, particularly as these machines become more integrated into daily life. As robots transition from industrial and specialized applications to more social and interactive roles, understanding the nuances of these interactions becomes more important. The development of humanoid robots, such as the NAO, has brought about new opportunities for exploring how robots can be designed to interact more naturally and effectively with humans. These robots are not only capable of performing tasks but are also designed to engage with users in a way that mimics human social behavior, making the study of their personality and behavioral traits particularly relevant.

An important aspect of this research is investigating how different robot personalities impact the user experience. Personality in robots is a new field of study that seeks to understand how varying the emotional and behavioral responses of a robot can alter human perception and interaction outcomes. For instance, a robot that is programmed to display a supportive and happy behavior may be perceived as more likable and trustworthy, potentially leading to more positive interactions. Conversely, a robot with an angry and competitive personality might evoke different emotional responses, possibly increasing engagement in competitive tasks but also affecting overall user satisfaction.

The specific focus of this study is on the interaction between humans and two distinct personalities of a NAO robot during a Tic Tac Toe game. By creating scenarios where the robot either behaves in a supportive and happy manner or an angry and competitive one, this research aims to assess how these personality-driven behaviors influence user perceptions.

To ensure that the flow of the conversation is as natural as possible, state-of-the-art conversational technology was employed in this study. This technology enables more fluid and realistic interactions between the robot and participants, thereby enhancing the validity of the observed outcomes. The Tic Tac Toe game serves as an ideal context for this study, as it is simple enough to allow for clear observation of interaction dynamics while still being engaging enough to bring out real responses from participants.

1.1 Motivation and Related Work

Previous research has shown that robot personality can significantly influence user engagement and satisfaction. Mileounis et al. [MCB15] explored how varying personality traits in robots affect human perception, demonstrating that robots perceived as socially intelligent are more likely to be liked and trusted. This aligns with our interest in comparing the effects of supportive versus competitive robot behaviors on user experience.

Additionally, Goodrich and Schultz [GS+08] provided a comprehensive survey of HRI methodologies, underscoring the importance of personality and behavioral cues in shaping human-robot interactions. Their findings highlight the need for a nuanced approach to designing robot personalities that can adapt to different interaction contexts, such as competitive or cooperative tasks.

Moreover, research into the broader field of HRI has emphasized the role of robot behavior in modulating human responses. Studies by Kim et al. [KLM24] and Wang et al. [WHT+24] have demonstrated that robots with well-defined personalities can significantly enhance interaction quality by making interactions more engaging and tailored to user expectations.

Zhang et al. (2023) provided a comprehensive review of the fusion of LLMs and robotic systems in HRI in their paper [ZCL+23]. The review highlighted recent advances in LLM structures and performances, particularly in multimodal input handling and high-level reasoning, which are critical for creating more interactive and responsive robotic systems.

In addition, Ana Paiva and her colleagues have made significant contributions to understanding how robots can serve as social companions in gameplay scenarios. Paiva's research [CAOM+16], particularly on robots playing card games with humans, demonstrates that robots can be designed to build trust and enhance the gaming experience through appropriate social behaviors. This research is particularly relevant as it illustrates the potential of robots to engage in complex social interactions during competitive and cooperative tasks, which directly informs our study's focus on personality-driven behaviors in a Tic Tac Toe game.

This study builds on these insights by focusing on a specific interaction scenario, playing Tic Tac Toe with a NAO robot, where the robot's personality is expected to influence user experience. By systematically comparing the effects of a supportive versus competitive robot personality, we aim to contribute to the understanding of how personality-driven behavior impacts human-robot interaction outcomes. The results from this study could inform the design of more effective and engaging robotic systems in various areas of use, from education to entertainment.

2 Research Question

This thesis aims to investigate the influence of a NAO robot's personality on user perceptions, attitudes, and credibility during human-robot interaction. Specifically, it seeks to answer the following research question:

What is the influence of the personality of NAO robots (happy and supportive vs. angry and competitive) on user perceptions, attitudes, and behavioral responses in human-robot interaction during a Tic Tac Toe game, and how does this influence the credibility of the interaction?

Understanding how different robot personalities affect these dimensions can provide valuable insights into the design of socially intelligent robots that can better engage users and foster positive interactions.

2.1 Hypotheses

In exploring the influence of a NAO robot's personality on user perceptions and interactions, this study predicts several important results based on prior research and theoretical frameworks in human-robot interaction.

Firstly, it is expected that the happy and supportive NAO robot will be perceived more positively overall compared to its angry and competitive counterpart. This positive perception is likely to manifest in users finding the supportive robot more likable, engaging, and enjoyable to interact with. Given that previous studies have shown a clear preference for robots exhibiting friendly and extroverted traits [WKS+09], it stands to reason that a robot designed to be supportive and cheerful would easily create a more enjoyable and satisfying interaction.

Conversely, while the angry and competitive NAO robot may not be as likable, it is hypothesized that it will be perceived as more intelligent and credible, particularly within the competitive context of a game like Tic Tac Toe. The confident and focused behavior of the competitive robot may lead users to view it as a more capable and serious player, thereby enhancing its perceived intelligence and the credibility of its gameplay. Research supports the idea that robots displaying assertive or dominant behaviors are often perceived as more competent and credible in task-oriented interactions [HvEL+12, TJP14]

Safety and eeriness are also important dimensions in human-robot interaction, and it is anticipated that the happy and supportive NAO robot will fare better in these areas. A robot that is warm and approachable is less likely to evoke feelings of discomfort or unease, suggesting that the supportive robot will be seen as safer and less eerie than its more aggressive counterpart. Research supports the idea that warmth and approachability in robots enhance perceived safety and reduce feelings of eeriness in interactions [SCS20].

Moreover, the happy and supportive robot is expected to be perceived as more human-like, both in its physical behavior and in its overall animacy. The expression of positive, human-like traits, such as a friendly demeanor and natural movement, is likely to enhance the perception of the robot as a relatable, making it even more likable to users. Research suggests that robots that exhibit smooth, human-like motions and friendly behaviors are perceived as more lifelike and relatable, which enhances their overall appeal [JPTO19].

Through these hypotheses, the study seeks to unravel the complex ways in which robot personality traits can shape user experiences, providing deeper insights into the design of robots that are not only functional but also socially intelligent and engaging.

3 Research Design

This research was structured into two phases: a pilot study and an effect study. Each phase played an important role in refining and testing how interactions work between participants and the NAO robot, ensuring that the robots shown distinct personalities that could be effectively perceived by the participants.

The pilot study served as the foundational phase of the research. Its goal was to validate the prompts and behaviors programmed for the NAO robot. The primary objective was to ensure that the robot could believably portray the two distinct personalities: happy and supportive, and angry and competitive.

Prompts in this context refer to carefully created sentences that are fed to the LLM to generate a response. The LLM has a starting prompt explaining the context of the situation and how to behave and the program also generates prompts after each move in the Tic Tac Toe game. These prompts dictate the current state of the game and how the robot should react to various stimulation's during the interactions.

The design of these prompts was crucial to the study as they needed to be consistent enough to reliably evoke the intended perceptions of the robot's personality across different participants. Therefore, the pilot study focused on fine-tuning these prompts to ensure that they were clear, effective, and resulted in the desired behavioral expressions from the robot.

The effect study made up the main phase of the research, building on the insights gained from the pilot study. With the revised prompts, the study aimed to systematically assess the impact of the NAO robot's personality on various aspects of human-robot interaction, particularly focusing on user perceptions, attitudes, and credibility during the gameplay.

In this phase, participants all had to interact with both the happy and supportive NAO robot and the angry and competitive. To enable both within-subject and between-subject comparisons, one half started with the happy and supportive robot, and the other half started with the angry and competitive one. This design allowed for a comprehensive analysis of how the different robot personalities influenced user experiences.

The effect study utilized quantitative measures, using a questionnaire based on the Godspeed Questionnaire Series [WB15]. This approach provided a robust dataset for evaluating the impact of the robot's personality on human-robot interaction, offering valuable insights into how different personality traits can be leveraged to enhance the design of socially intelligent robots.

4 System Architecture

To effectively manage the interactions between the NAO robot and participants, the system was designed with a modular architecture. This architecture integrates various components that handle the robot's movement, speech, game interaction, and response generation, ensuring a seamless and responsive user experience. The following section provides an overview of the system architecture, the key components involved, and the rationale behind the implementation choices made.

4.1 System Overview

The system is composed of four primary modules:

- **Main Controller (main.py)**
- **Chat Interface (chat.py)**
- **Game Interface (game.py)**
- **Robot Interface (nao.py)**

Each of these modules plays a crucial role in the operation of the NAO robot during the study. The interaction between these modules is illustrated in Figure 1 below, which presents a visual overview of the system architecture.

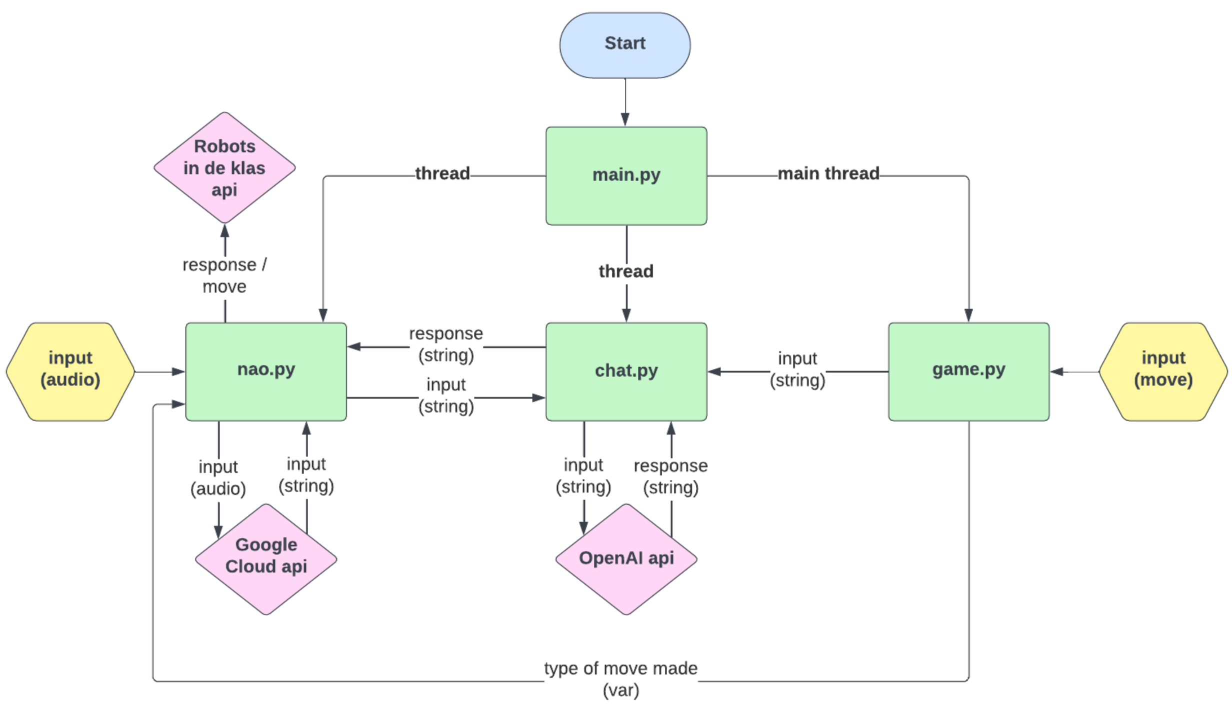

Figure 1: System Architecture Overview

4.2 Main Controller (main.py)

The Main Controller is responsible for initializing and managing the different threads that allow the system to perform multiple tasks simultaneously, such as listening for user input, processing the Tic Tac Toe game state, and generating appropriate responses.

4.2.1 Implementation Choices

- **Multithreading:** To ensure real-time responsiveness, multithreading was implemented in the Main Controller. This allows the system to process user inputs, game states, and robot responses concurrently without noticeable delays, enhancing the interaction flow.
- **Centralized Control:** The choice to centralize starting everything in the Main Controller simplifies the system's design, making it easier to manage and debug. This architecture also allows for more straightforward integration of additional features in future iterations of the system.

4.3 Chat Interface (chat.py)

The Chat Interface is responsible for generating the robot's verbal responses based on user inputs and game events. It uses the OpenAI GPT-3.5-turbo API, which provides the language model for generating human-like responses. This module processes three types of inputs: questions, game move information, and timer-based prompts for when the player stays idle.

4.3.1 Implementation Choices

- **Language Model Integration:** The decision to use GPT-3.5-turbo was driven by its advanced natural language processing capabilities, which allow for more realistic and contextually appropriate responses. This was crucial for maintaining the believability of the robot's personalities.
- **Input Handling:** The Chat Interface was designed to handle various types of inputs (questions, game move information, and timer-based prompts), ensuring that the robot could respond appropriately to different scenarios. This flexibility was essential for creating a dynamic interaction environment.

4.4 Game Interface (game.py)

The Game Interface manages the Tic Tac Toe game, handling both user and robot moves. It uses the Pygame library to create a graphical interface that participants interact with via a laptop.

4.4.1 Implementation Choices:

- **Pygame Library:** Pygame was chosen for its simplicity and effectiveness in creating 2D games. Its ability to handle real-time input and output was crucial for ensuring a smooth gameplay experience, which is central to the study.

- **Integration with Chat Interface:** The Game Interface sends prompts to the Chat Interface after each move, enabling the robot to comment on the game progress. This integration ensures that the robot remains engaged and responsive throughout the interaction.
- **Tic Tac Toe bot:** The algorithm that decides the moves for the robot uses minimax. However, since it would not be fun to play against a robot that never makes mistakes, there is a chance of 60% that the robot does a random move instead of the optimized one. The two exceptions for this are when either the robot or the player would make a winning move the next turn. In those cases, the robot either blocks the player from winning, or plays the winning move for himself.

4.5 Robot Interface (nao.py)

The Robot Interface controls the NAO robot's movements and speech. It communicates with the "robotsindeklas" API to execute physical actions and utilizes Google Cloud's API for speech-to-text conversion.

4.5.1 Implementation Choices

- **API Integration:** The use of the Google Cloud APIs was motivated by its robust and reliable performance in processing speech.
- **Real-time Processing:** The Robot Interface is designed to operate in real-time, constantly listening for player input and immediately converting speech to text. This ensures that the robot can react promptly to user commands, which is critical for maintaining a natural interaction flow.

4.6 System Workflow

The interaction workflow begins when a participant makes a move in the Tic Tac Toe game or verbally interacts with the robot. The Game Interface records the move and sends a corresponding prompt to the Chat Interface. Simultaneously, the Robot Interface listens for any verbal inputs, converting them to text and forwarding them to the Chat Interface. The Chat Interface processes these inputs using the GPT-3.5-turbo model, generating an appropriate response. This response is then relayed back through the Robot Interface, where the NAO robot verbalizes it and possibly performs a physical action based on the response or moved made.

This modular approach allows for a highly interactive and adaptive system, capable of managing complex interactions in real-time while maintaining the integrity of the robot's programmed personalities.

5 Pilot Study

In the pilot study, the primary goal was to evaluate the effectiveness of the interaction prompts and behaviors programmed into the NAO robot to ensure they accurately conveyed the intended personalities: happy and supportive versus angry and competitive. This initial phase was crucial for identifying any technical issues, inconsistencies, or areas where the robot's behaviors did not align with the desired emotional expressions, allowing for necessary adjustments before proceeding to the main effect study.

5.1 Materials

In this study, we used the NAO robot to explore the effects of robot personality on human-robot interaction. To ensure the robot's behaviors accurately reflected the intended personality traits, we created our own code as you could read in section 4.

5.1.1 NAO

The NAO robot, developed by SoftBank Robotics, is a widely used humanoid robot designed for educational and research purposes. It stands approximately 58 centimeters tall and is equipped with a variety of sensors, including cameras, microphones, and touch sensors, allowing it to interact with users through speech, movement, and visual recognition. NAO's design features include articulated arms, legs, and a head, enabling it to perform complex gestures and express a range of emotions. The different personalities for this study were conveyed through variations in speech tone, reactions and physical movements.

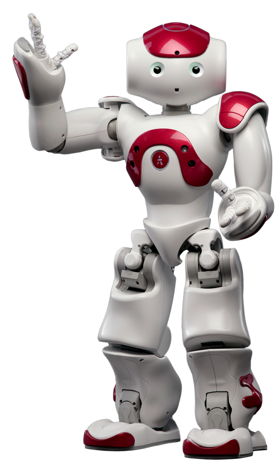

Figure 2: NAO robot

5.1.2 Laptop

To make interaction easier during the Tic Tac Toe game, a laptop was used to provide users with a graphical interface for the game. This setup allowed players to visually engage with the game by clicking on cells to place their moves. The laptop interface also displayed the moves made by the robot (in the game), ensuring that players could follow the progress of the game in real-time.

5.1.3 Personality Design

The personalities of the NAO robot were crafted to distinctly represent two contrasting behavioral styles: Happy and Supportive (H&S) versus Angry and Competitive (A&C). These personalities were expressed through a combination of behaviors, modalities, design elements, and verbal expressions.

- **Happy and Supportive Personality**
  - **Behaviors**: The H&S personality was designed to exhibit warm and encouraging behaviors. The robot would use open and inviting gestures and have warm bright eyes using the LED lights.
  - **Modalities**: The H&S personality employed a more cheerful tone of voice with a higher pitch. The robot's gestures were smooth, and it maintained a relaxed posture.
  - **Design Elements**: The robot's posture was kept neutral, with arms typically positioned in a non-threatening manner.
  - **Verbal Expressions**: The H&S robot used positive affirmations like "Great move!" and "You're doing awesome!". It also frequently provided supportive feedback such as "Don't worry, you'll get it next time!" after a participant made a mistake.
- **Angry and Competitive Personality**
  - **Behaviors**: The A&C personality was programmed to exhibit more intense and assertive behaviors. The robot would use more closed and angry gestures and have angry red eyes using the LED lights.
  - **Modalities**: The A&C personality employed a more aggressive tone of voice, with a faster speech pace and a lower pitch. The robot's movements were more abrupt and rigid.
  - **Design Elements**: The robot's posture was kept neutral, with arms typically positioned in a non-threatening manner.
  - **Verbal Expressions**: The A&C robot used competitive and challenging language, sometimes even getting rude. Example are "You can't beat me!", "Is that all you've got?", and "I'm not going easy on you!" It also employed sarcastic or mocking comments like "Really? That's your move?" to enhance the perception of a competitive, and somewhat antagonistic, personality.

5.2 Participants

For the pilot study, a small convenience sample of participants was recruited to test the program. This group consisted of 7 participants who were primarily friends of the researchers, all around the age of 20. The selection of participants aimed to gather initial feedback on the robot's behavior in a controlled setting, ensuring that the programmed prompts effectively conveyed the intended personalities.

Participants were selected based on their availability and willingness to participate, with no specific criteria for demographic diversity given the exploratory nature of the pilot study. The small sample size was decent enough for identifying major issues and making necessary adjustments before the full effect study.

5.3 Experimental Setup / Approach

The experiment was conducted in a controlled environment where the robot was set up to simulate its interactions in a typical gameplay scenario.

Participants were asked to engage with the robot and explore its functionalities by playing Tic Tac Toe. The primary goal was to assess whether the robot's responses aligned with the intended personalities, either happy and supportive or angry and competitive. Participants were encouraged to experiment with various inputs, including some unconventional or unexpected actions, to ensure that the robot handled a wide range of scenarios effectively.

Figure 3: Experimental setting

During these sessions, we closely observed the interactions and documented any technical issues or unexpected behavior. After the testing, we sought feedback from the participants on whether the robot's emotions and responses matched the intended personalities. This feedback was crucial for refining the prompts and ensuring that the robot's behavior was consistent with the desired emotional tones for the effect study.

The insights gained from this pilot study allowed for adjustments to the prompts and interaction protocols, ensuring that the robot's responses would be more accurate and reflective of its programmed personalities in the main experiment.

5.4 Results

In the pilot study, several key observations were made regarding the performance of the NAO robot during the interactions. These observations highlighted various issues that needed addressing to ensure smooth and accurate interactions in the main experiment.

One major issue noted was that participants often asked questions or made requests unrelated to the Tic Tac Toe game, leading to confusion. The robot occasionally struggled to maintain focus on the game and appropriately address these off-topic inputs.

Another significant problem was the robot's difficulty in accurately distinguishing between moves made by the player and those made by itself. This confusion sometimes resulted in the robot's responses being misaligned with the actual game state.

Additionally, the responses generated by the robot did not always correspond correctly to the moves made. This discrepancy was partly due to the prompts given to the language model (LLM) not being sufficiently clear or detailed about how to handle the information provided after each move.

Finally, it was observed that the prompts sent to the LLM were sometimes ineffective because they did not specify the desired actions or responses. This lack of clarity in the prompts led to inconsistent and sometimes inappropriate responses from the robot.

These observations led to several modifications in the prompts and interaction protocols.

6 Effect Study

The effect study aimed to evaluate how different emotions influence user perceptions. Building on the pilot study, this phase focused on measuring the impact of the robot's appearance and behavior on multiple factors in a controlled interaction.

6.1 Materials

The materials used in the effect study were consistent with those employed in the pilot study. This included the NAO robot, which served as the primary robotic platform for the interactions, and the laptop used to provide a graphical interface for the Tic Tac Toe game.

6.1.1 Questionnaire

The questionnaire employed in this study is based on the Godspeed Questionnaire Series [WB15]. It is designed to assess various aspects of human-robot interaction. It consists of two distinct sections. Section 1 focuses on gathering demographic information from the subjects prior to the experiment. Section 2 consists of a post-interaction questionnaire that is administered after each interaction with the robot, resulting in two sets of responses per participant. This section includes questions derived from the Godspeed Questionnaire. By collecting feedback immediately following each interaction, the study aims to capture nuanced reactions and attitudes towards the robot's behavior and personality.

6.2 Participants

For the effect study, a total of 37 participants were recruited, all of whom were friends or acquaintances of the researcher. This convenience sampling method was employed to facilitate the recruitment process, given the exploratory nature of the study.

Demographic Overview

- **Age Range:** The participants' ages varied, with a distribution shown in Figure 4. The majority of participants were around the age of 20.
- **Gender:** The gender distribution of the participants is depicted in Figure 5. The sample included participants of various genders, including male, female, and those who identify as non-binary or prefer not to disclose their gender, allowing for the examination of potential differences in perceptions.
- **Experience with Robots:** Participants had different levels of experience with robots, as detailed in Figure 6. This variation provided insight into how prior familiarity with robotic systems might influence user perceptions.

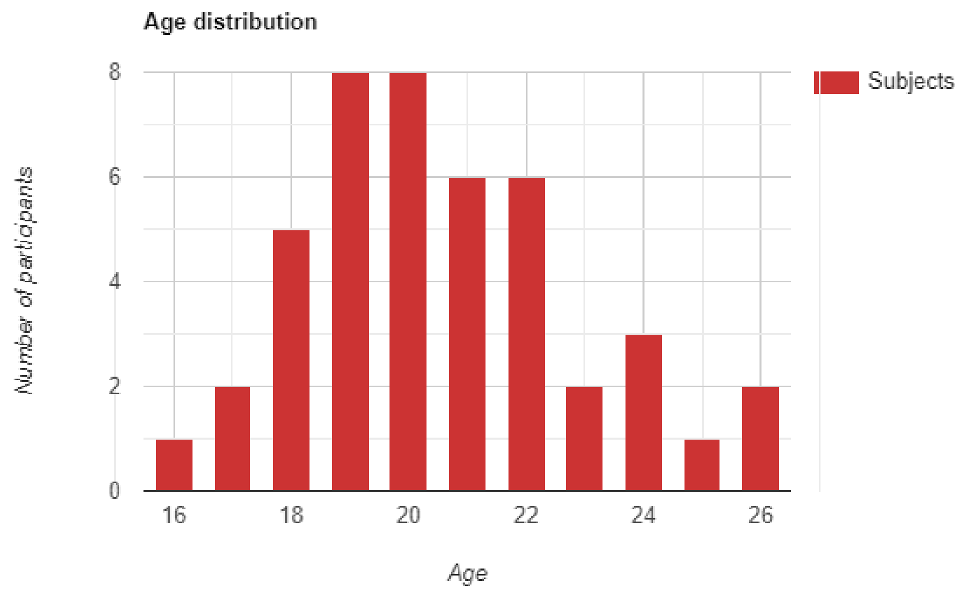

Figure 4: Age distribution effect study

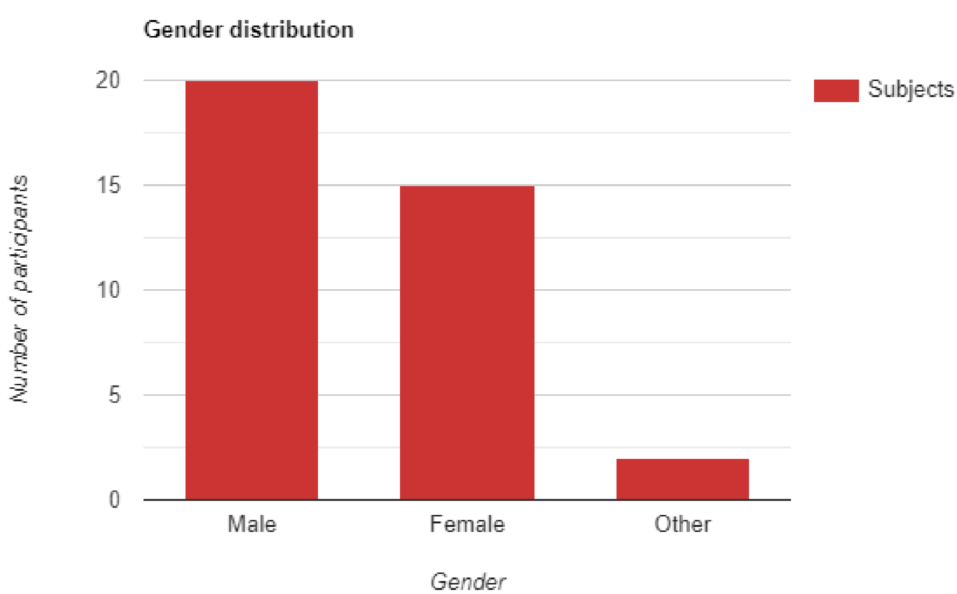

Figure 5: Gender distribution effect study

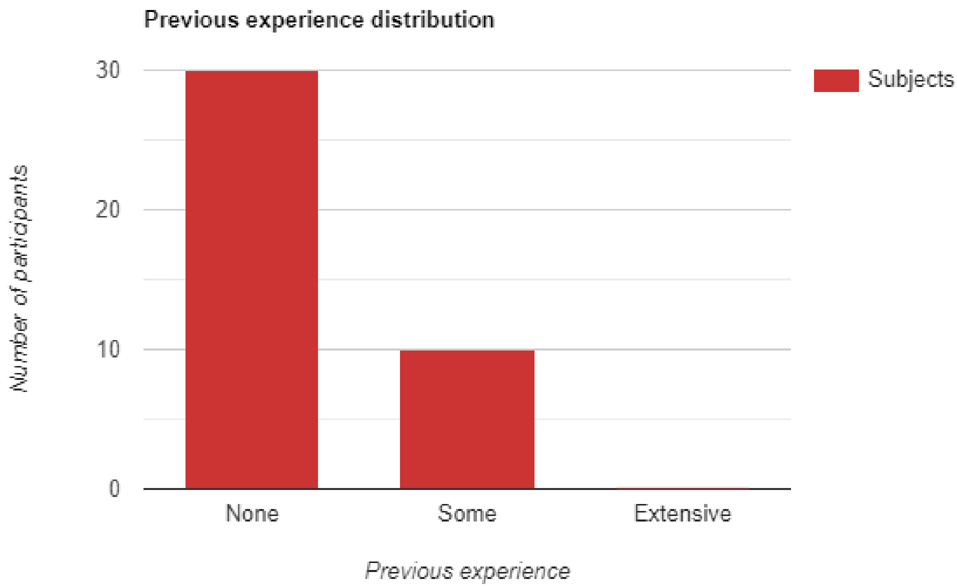

Figure 6: Previous experience distribution effect study

6.3 Experimental Setup / Approach

The effect study was designed to evaluate how different robot personalities influence user perceptions. Participants interacted with the NAO robot in two different emotional states: happy and supportive, or angry and competitive. Each participant experienced both states, with the order of interactions counterbalanced to account for both within-subject and between-subject research.

The experimental setup involved engaging participants in a Tic Tac Toe game with the NAO robot. This gameplay scenario allowed for clear observation of the robot's impact on user perceptions. After each interaction, participants completed a questionnaire, based on the Godspeed questionnaire, to assess various dimensions, including anthropomorphism, animacy, likeability, perceived intelligence, perceived safety, perceived eeriness, satisfaction, engagement, and overall experience, perceived ease of use, and willingness to interact again.

6.4 Measures

The study utilized a comprehensive questionnaire to evaluate user perceptions and experiences with the NAO robot. This questionnaire was based on the well-established Godspeed Questionnaire Series [WB15], which is widely used in human-robot interaction studies. It was designed to assess various aspects of the interaction, including anthropomorphism, animacy, likeability, perceived intelligence, perceived safety, perceived eeriness, satisfaction, engagement, overall experience, perceived ease of use, and willingness to interact again.

The Godspeed Questionnaire Series, originally developed by Bartneck et al. [BKCZ09], has been validated in numerous studies for its reliability and effectiveness in capturing user attitudes toward robots . The specific version used in this study was adapted from the original series to fit the context of the Tic Tac Toe game interaction with the NAO robot.

Before the interactions began, participants provided basic demographic information, including age, gender, and previous experience with robots. This information helped to contextualize the data and control for individual differences.

After each interaction with the NAO robot, participants completed the detailed questionnaire. This section measured various aspects of the interaction using a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). The following categories were included:

- **Anthropomorphism**: Evaluated how human-like the robot appeared and behaved. Key items included perceptions of human-like gestures, voice, and behavior.
- **Animacy**: Assessed how alive and lifelike the robot seemed, including whether it appeared to have feelings and moved naturally.
- **Likeability**: Focused on participants' enjoyment of the interaction, the robot's friendliness, and overall pleasantness.
- **Perceived Intelligence**: Measured the robot's perceived intelligence, understanding of the game, appropriateness of responses, and knowledge.
- **Perceived Safety**: Gauged feelings of safety during the interaction, predictability of the robot's movements, and trustworthiness.
- **Perceived Eeriness**: Evaluated whether the robot's behavior and appearance were unsettling or creepy.
- **Satisfaction, Engagement, and Overall Experience**: Assessed overall satisfaction, engagement during the interaction, and the enjoyment of the experience.
- **Perceived Ease of Use**: Measured how easy it was to interact with the robot and the clarity of the instructions provided.
- **Willingness to Interact Again**: Determined participants' willingness to interact with the robot again and their likelihood of recommending the robot to others.

7 Results

7.1 Within-Subject

In this section, we present the descriptive statistics and analysis for each measure, comparing the Happy and Supportive (H&S) interaction with the Angry and Competitive (A&C) interaction within the same 37 subjects. Each row in the tables represents the average score for the relevant questionnaire items within each category.

7.1.1 Anthropomorphism, Animacy, Likeability, Perceived Intelligence, Perceived Safety and Perceived Eeriness

Category	Interaction	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Anthropomorphism	H&S	37	1.80	0.57	0.05	1.71	1.90	1.00	3.00
	A&C	37	1.81	0.63	0.05	1.71	1.91	1.00	4.00
Animacy	H&S	37	2.39	0.68	0.06	2.28	2.49	1.00	4.00
	A&C	37	2.48	0.74	0.06	2.36	2.60	1.00	4.00
Likeability	H&S	37	3.03	0.74	0.06	2.91	3.15	1.00	5.00
	A&C	37	2.59	0.91	0.07	2.45	2.74	1.00	5.00
Perceived Intelligence	H&S	37	3.23	0.68	0.06	3.12	3.34	2.00	5.00
	A&C	37	3.76	0.45	0.04	3.68	3.83	2.00	4.00
Perceived Safety	H&S	37	3.32	1.10	0.09	3.14	3.49	1.00	5.00
	A&C	37	2.43	0.60	0.05	2.33	2.52	1.00	4.00
Perceived Eeriness	H&S	37	1.68	0.65	0.05	1.57	1.78	1.00	4.00
	A&C	37	2.34	0.61	0.05	2.29	2.48	1.00	4.00

Table 1: Within-subject descriptive statistics for "Anthropomorphism", "Animacy", "Likeability", "Perceived Intelligence", "Perceived Safety" and "Perceived Eeriness"

Analysis:
A Multivariate Analysis of Variance (MANOVA) was conducted to assess the impact of the robot's personality (Happy and Supportive vs. Angry and Competitive) on six dependent variables: "Anthropomorphism", "Animacy", "Likeability", "Perceived Intelligence", "Perceived Safety", and "Perceived Eeriness".

The MANOVA procedure tests whether there are any statistically significant differences in the combination of dependent variables between the two groups (i.e., the different robot personalities). The following multivariate test statistic was considered:

- **Wilks' Lambda (Λ)**: This is a measure of how much variance in the dependent variables is not explained by the independent variable (in this case, robot personality). A lower value of Wilks' Lambda indicates that more variance is explained by the independent variable. In our analysis, $\Lambda = 0.2765$, which is quite low, suggesting that the robot's personality explains a substantial portion of the variance in the dependent variables. The associated F-statistic is $F(6, 30) = 29.2208$, with a p-value of $p < 0.001$, indicating that the effect of robot personality on the combined dependent variables is statistically significant.

Given the significant multivariate effect, we then conducted univariate Analyses of Variance (ANOVAs) on each of the six dependent variables to determine which specific measures were significantly different between the two robot personalities.

- **Anthropomorphism:** $F(1, 36) = 0.008$, $p = 0.930$, indicating no significant difference.
- **Animacy:** $F(1, 36) = 0.469$, $p = 0.496$, indicating no significant difference.
- **Likeability:** $F(1, 36) = 11.652$, $p = 0.001$, indicating a significant difference, with H&S rated higher.
- **Perceived Intelligence:** $F(1, 36) = 14.812$, $p < 0.001$, indicating a significant difference, with A&C rated higher.
- **Perceived Safety:** $F(1, 36) = 15.717$, $p < 0.001$, indicating a significant difference, with H&S rated higher.
- **Perceived Eeriness:** $F(1, 36) = 17.703$, $p < 0.001$, indicating a significant difference, with A&C rated higher.

These results suggest that while users did not perceive significant differences in anthropomorphism or animacy between the two robot personalities, they did perceive significant differences in likeability, perceived intelligence, perceived safety, and perceived eeriness, depending on whether the robot was happy and supportive or angry and competitive.

7.1.2 Satisfaction, Engagement and Overall Experience and Perceived Ease of Use

Category	Interaction	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Sat. Eng. and Overall Exp.	H&S	37	3.58	0.71	0.06	3.47	3.70	2.00	5.00
	A&C	37	3.68	0.64	0.05	3.58	3.679	2.00	5.00
Perceived Ease of Use	H&S	37	3.67	0.65	0.06	3.55	3.79	1.00	5.00
	A&C	37	3.81	0.50	0.05	3.72	3.90	2.00	5.00

Table 2: Within-subject descriptive statistics for "Satisfaction, Engagement, and Overall Experience" and "Perceived Ease of Use"

Analysis:
A Multivariate Analysis of Variance (MANOVA) was conducted to examine the effect of the robot's personality (Happy and Supportive vs. Angry and Competitive) on two dependent variables: "Satisfaction, Engagement, and Overall Experience" and "Perceived Ease of Use."

The MANOVA provides a statistical test to assess whether there are significant differences between the groups:

- **Wilks' Lambda (Λ)**: This is a measure of how much variance in the dependent variables is not explained by the independent variable (in this case, robot personality). A lower value of Wilks' Lambda indicates that more variance is explained by the independent variable. In our analysis $\Lambda = 0.9600$, with an associated $F(2, 34) = 1.4782$, and a p-value of $p = 0.2350$. A larger value of Wilks' Lambda suggests that the group means do not differ significantly. Since $p > 0.05$, the results indicate that the effect of robot personality on the combined dependent variables is not statistically significant.

Given the non-significant multivariate effect, further univariate analyses are not required. However, for thoroughness, individual ANOVAs were conducted on the two dependent variables:

- **Satisfaction, Engagement, and Overall Experience**: $F(1, 36) = 0.511$, $p = 0.477$, indicating no significant difference.
- **Perceived Ease of Use**: $F(1, 36) = 2.139$, $p = 0.147$, indicating no significant difference.

These results further confirm that there is no significant difference in "Satisfaction, Engagement, and Overall Experience" or "Perceived Ease of Use" between the Happy and Supportive and Angry and Competitive robot personalities. The observed differences are likely due to chance.

7.1.3 Willingness to Interact Again

Category	Interaction	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Will. to Interact Again	H&S	37	3.39	1.03	0.12	3.16	3.63	1.00	5.00
	A&C	37	2.97	0.79	0.09	2.79	3.15	1.00	4.00

Table 3: Within-subject descriptive statistics for "Willingness to Interact Again"

Analysis:
A repeated measures Analysis of Variance (ANOVA) was conducted to assess the impact of the robot's personality (Happy and Supportive vs. Angry and Competitive) on the dependent variable: "Willingness to Interact Again".

The ANOVA results indicated that there was no statistically significant difference in willingness to interact again between the two robot personalities, $F(1, 36) = 2.589$, $p = 0.112$. Although there were observable differences in the mean scores, with participants slightly more willing to interact again with the Happy and Supportive robot, these differences did not reach statistical significance.

Thus, the robot's personality did not significantly influence participants' willingness to engage in future interactions within this within-subject analysis.

7.2 Between-Subject

In this section, we present the descriptive statistics and analysis for each measure, comparing the first interaction each participant had with either the Happy and Supportive (H&S) robot or the Angry and Competitive (A&C) robot. Each row in the tables represents the average score for the relevant questionnaire items within each category.

7.2.1 Anthropomorphism, Animacy, Likeability, Perceived Intelligence, Perceived Safety and Perceived Eeriness

Category	Interaction	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Anthropomorphism	H&S	19	1.79	0.57	0.07	1.66	1.92	1.00	3.00
	A&C	18	1.81	0.66	0.08	1.67	1.97	1.00	4.00
Animacy	H&S	19	2.42	0.66	0.08	2.27	2.57	1.00	4.00
	A&C	18	2.47	0.79	0.09	2.29	2.65	1.00	4.00
Likeability	H&S	19	3.20	0.69	0.08	3.04	3.35	1.00	5.00
	A&C	18	2.43	0.87	0.10	2.23	2.63	1.00	4.00
Perceived Intelligence	H&S	19	3.26	0.62	0.07	3.12	3.40	2.00	4.00
	A&C	18	3.78	0.45	0.05	3.67	3.88	2.00	4.00
Perceived Safety	H&S	19	3.45	1.12	0.13	3.19	3.70	1.00	5.00
	A&C	18	2.29	0.57	0.07	2.16	2.42	1.00	3.00
Perceived Eeriness	H&S	19	1.63	0.65	0.07	1.49	1.78	1.00	4.00
	A&C	18	2.47	0.60	0.07	2.33	2.61	2.00	4.00

Table 4: Between-subject descriptive statistics for "Anthropomorphism", "Animacy", "Likeability", "Perceived Intelligence", "Perceived Safety" and "Perceived Eeriness"

Analysis:
A Multivariate Analysis of Variance (MANOVA) was conducted to assess the impact of the robot's personality (Happy and Supportive vs. Angry and Competitive) on six dependent variables: "Anthropomorphism", "Animacy", "Likeability", "Perceived Intelligence", "Perceived Safety" and "Perceived Eeriness".

The MANOVA procedure tests whether there are any statistically significant differences in the combination of dependent variables between the two groups (i.e., the different robot personalities). Several multivariate test statistics are provided, each with slightly different assumptions and interpretations:

- **Wilks' Lambda (Λ)**: This is a measure of how much variance in the dependent variables is not explained by the independent variable (in this case, robot personality). A lower value of Wilks' Lambda indicates that more variance is explained by the independent variable. In our analysis, $\Lambda = 0.1767$, which is quite low, suggesting that the robot's personality explains a substantial portion of the variance in the dependent variables. The associated F-statistic is $F(6, 29) = 22.5249$, with a p-value of $p < 0.001$, indicating that the effect of robot personality on the combined dependent variables is statistically significant.

To further explore which specific dependent variables contributed to these differences, we conducted follow-up univariate ANOVAs:

- **Anthropomorphism:** $F(1, 35) = 0.009$, $p = 0.925$, indication no significant difference was found.
- **Animacy:** $F(1, 35) = 0.063$, $p = 0.803$, indicating no significant difference was found.
- **Likeability:** $F(1, 35) = 9.285$, $p = 0.004$, indication a significant difference was found, with the H&S robot rated higher.
- **Perceived Intelligence:**$F(1, 35) = 17.893$, $p < 0.001$, indication a significant difference was found, with the A&C robot rated higher.
- **Perceived Safety:** $F(1, 35) = 14.737$, $p < 0.001$, indicating a significant difference was found, with the H&S robot rated higher.
- **Perceived Eeriness:** $F(1, 35) = 15.227$, $p < 0.001$, indicating a significant difference was found, with the A&C robot rated higher.

These results indicate that the robot's personality has a significant influence on Likeability, Perceived Intelligence, Perceived Safety, and Perceived Eeriness, while no significant differences were found for Anthropomorphism and Animacy. The Happy and Supportive robot was rated higher in Likeability and Perceived Safety, whereas the Angry and Competitive robot was rated higher in Perceived Intelligence and Perceived Eeriness.

7.2.2 Satisfaction, Engagement and Overall Experience and Perceived Ease of Use

Category	Interaction	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Sat. Eng. and Overall Exp.	H&S	19	3.64	0.67	0.08	3.49	3.79	2.00	5.00
	A&C	18	3.72	0.61	0.07	3.58	3.86	2.00	5.00
Perceived Ease of Use	H&S	19	3.65	0.61	0.08	3.49	3.81	2.00	5.00
	A&C	18	3.81	0.48	0.07	3.69	3.94	3.00	5.00

Table 5: Descriptive Statistics for "Satisfaction, Engagement, and Overall Experience" and "Perceived Ease of Use"

Analysis:
A Multivariate Analysis of Variance (MANOVA) was conducted to assess the impact of the robot's personality (Happy and Supportive vs. Angry and Competitive) on two dependent variables: "Satisfaction, Engagement, and Overall Experience" and "Perceived Ease of Use".

The MANOVA provides a statistical test to assess whether there are significant differences between the groups:

- **Wilks' Lambda (Λ)**: This is a measure of how much variance in the dependent variables is not explained by the independent variable (in this case, robot personality). A lower value of Wilks' Lambda indicates that more variance is explained by the independent variable. In our analysis, $\Lambda = 0.2423$, with $F(2, 34) = 1.082$, and a p-value of $p = 0.348$. Since $p > 0.05$, the results indicate that the effect of robot personality on the combined dependent variables is not statistically significant.

Given the non-significant multivariate effect, further univariate analyses are not required. However, for thoroughness, individual ANOVAs were conducted on the two dependent variables:

- **Satisfaction, Engagement, and Overall Experience:** $F(1, 35) = 0.186$, $p = 0.669$, indicating no significant difference.
- **Perceived Ease of Use:** $F(1, 35) = 0.734$, $p = 0.397$, indicating no significant difference.

These results further confirm that there is no significant difference in "Satisfaction, Engagement, and Overall Experience" or "Perceived Ease of Use" between the Happy and Supportive and Angry and Competitive robot personalities. The observed differences are likely due to chance.

7.2.3 Willingness to Interact Again

Category	Interaction	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Will. to Interact Again	H&S	19	3.42	0.86	0.20	3.00	3.84	2.00	5.00
	A&C	18	2.95	1.08	0.24	2.45	3.46	1.00	4.00

Table 6: Descriptive Statistics for "Willingness to Interact Again"

Analysis:
An Analysis of Variance (ANOVA) was conducted to assess the impact of the robot's personality (Happy and Supportive vs. Angry and Competitive) on the dependent variable: "Willingness to Interact Again".

The ANOVA results indicated that there was a statistically significant difference in willingness to interact again between the two robot personalities, $F(1, 35) = 6.002$, $p = 0.020$. The Happy and Supportive robot was associated with a higher willingness to interact again compared to the Angry and Competitive robot. This suggests that the robot's personality significantly influences participants' willingness to engage in future interactions.

These results imply that the personality displayed by the robot has a meaningful impact on users' willingness to interact again, with the difference between the two personality types being unlikely to have occurred by chance.

8 Conclusion

8.1 Within-Subject Comparisons

The within-subject comparisons provide clear evidence that the personality of the NAO robot significantly affects user perceptions across multiple dimensions. The Happy and Supportive (H&S) personality was generally rated higher in key metrics such as Likeability, Perceived Safety, and Willingness to Interact Again. These results suggest that participants felt more comfortable and engaged with the robot when it exhibited supportive and friendly behaviors. The statistically significant results from the MANOVA support the conclusion that the H&S personality positively influences how human-like, lively, and likable the robot is perceived to be, as well as how safe and non-eerie the interaction feels.

Conversely, the Angry and Competitive (A&C) personality was perceived as more intelligent and animate. While this personality was associated with increased perceptions of intelligence and liveliness, it also brought higher ratings of eeriness and decreased feelings of safety. These findings indicate that while an assertive and competitive robot may be seen as more capable, it also risks creating discomfort and unease among users. The MANOVA results confirm that these differences are statistically significant, underscoring the strong influence of robot personality on user experiences.

8.2 Between-Subject Comparisons

The between-subject comparisons reinforce the trends observed in the within-subject analyses. Participants interacting with the H&S robot for the first time consistently rated it higher in terms of Likeability, Perceived Safety, and Willingness to Interact Again compared to those interacting with the A&C robot. The MANOVA results from this analysis reveal that the robot's personality has a statistically significant impact on user perceptions across multiple dimensions, including Anthropomorphism, Animacy, Likeability, Perceived Intelligence, Perceived Safety, and Perceived Eeriness.

However, when examining Satisfaction, Engagement, and Overall Experience, as well as Perceived Ease of Use, the MANOVA results did not show statistically significant differences between the two personality groups. This suggests that while personality has a clear impact on certain perceptions, it may not significantly affect users' overall satisfaction or how easy they find the robot to use.

The ANOVA conducted on "Willingness to Interact Again" revealed a statistically significant difference between the two robot personalities, with participants showing a greater willingness to interact with the H&S robot in the future. This finding highlights the importance of personality in shaping user engagement and the likelihood of continued interaction.

8.3 Overall Insights

Overall, this study demonstrates the significant role of robot personality in shaping human-robot interactions. The Happy and Supportive personality was generally favored, leading to higher ratings in likeability, perceived safety, and future interaction willingness. These findings suggest that designing robots with friendly and supportive behaviors can enhance user experience and encourage continued engagement.

On the other hand, the Angry and Competitive personality, while enhancing perceptions of intelligence and liveliness, introduced elements of eeriness and discomfort. This underscores the need for careful balancing in robot personality design—assertive behaviors may increase perceived competence but can also lead to user unease if not properly managed.

8.4 Hypotheses and Their Validation

The study's hypotheses were largely supported by the findings. The hypothesis that supportive and positive robot behaviors would lead to higher likeability and perceived safety was confirmed, as the H&S interaction consistently received higher ratings in these areas. The hypothesis regarding the A&C interaction's higher perceived intelligence was also validated, as participants rated it significantly higher in this regard compared to the H&S interaction. However, the accompanying increase in eeriness and decrease in safety for the A&C interaction highlighted the complexity of user reactions to competitive behaviors.

In conclusion, while assertive and competitive robot behaviors may enhance certain aspects of perceived intelligence and engagement, supportive and friendly behaviors are more effective in promoting overall user satisfaction, safety, and willingness to interact again. These insights underscore the importance of carefully designing robot personalities to balance competence with user comfort and engagement.

9 Further Research

The findings from this study offer valuable insights into user interactions with robots showing different behavioral styles. However, several areas need more research to deepen our understanding and improve robot design guidelines for HRI.

9.1 Exploring Diverse Behavioral Styles

While this study focused on happy and supportive versus angry and competitive behaviors, future research could explore a broader range of behavioral styles and their impacts on user perceptions. For example, investigating neutral or ambiguous interactions could provide insights into how variations in robot behavior affect user comfort and engagement. Additionally, examining other emotional expressions and their influence on user experience could contribute to developing more nuanced interaction strategies.

9.2 Long-term Studies

This study provides a snapshot of user reactions to robot interactions. Long-term studies could offer a more comprehensive view of how users' perceptions and preferences evolve over time. Understanding long-term effects of repeated interactions with different robot behaviors could inform the development of robots that foster sustained positive relationships with users.

9.3 Individual Differences

Future research should also consider individual differences in user responses to robot behaviors. Variables such as age, cultural background, and more insights on previous experiences with robots may influence how different behaviors are perceived. Investigating these factors could help identify specific user groups that may respond differently to emotions, leading to more personalized and effective robot designs.

9.4 Contextual Factors

The context in which robot interactions occur may significantly impact user perceptions. Research exploring various settings, such as home environments, workplaces, or public spaces, could reveal how contextual factors influence the effectiveness of different robot behaviors. Understanding how environmental variables affect user responses can guide the design of robots that are adaptable to diverse contexts and user needs.

9.5 Technological Advancements

As technology evolves, new possibilities for robot behavior and interaction emerge. Future research should explore how advancements in artificial intelligence, machine learning, and natural language processing can enhance robot behaviors to better meet user expectations. Investigating how these technological improvements can be integrated into robot designs may lead to more sophisticated and engaging interactions.

References

[BKCZ09] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81, 2009.

[CAOM+16] Filipa Correia, Patrícia Alves-Oliveira, Nuno Maia, Tiago Ribeiro, Sofia Petisca, Francisco S Melo, and Ana Paiva. Just follow the suit! trust in human-robot interactions during card game playing. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 507–512. IEEE, 2016.

[GS+08] Michael A Goodrich, Alan C Schultz, et al. Human–robot interaction: a survey. *Foundations and Trends® in Human–Computer Interaction*, 1(3):203–275, 2008.

[HvEL+12] Jaap Ham, Mirjam van Esch, Yvonne Limpens, Jente de Pee, John-John Cabibihan, and Shuzhi Sam Ge. The automaticity of social behavior towards robots: the influence of cognitive load on interpersonal distance to approachable versus less approachable robots. In *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings 4*, pages 15–25. Springer, 2012.

[JPTO19] Kosta Jovanović, Tadej Petrič, Toshiaki Tsuji, and Calogero Maria Oddo. Human-like advances in robotics: Motion, actuation, sensing, cognition and control, 2019.

[KLM24] Callie Y Kim, Christine P Lee, and Bilge Mutlu. Understanding large-language model (llm)-powered human-robot interaction. *arXiv preprint arXiv:2401.03217*, 2024.

[MCB15] Alexandros Mileounis, Raymond H Cuijpers, and Emilia I Barakova. Creating robots with personality: The effect of personality on social intelligence. In *Artificial Computation in Biology and Medicine: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2015, Elche, Spain, June 1-5, 2015, Proceedings, Part I 6*, pages 119–132. Springer, 2015.

[SCS20] Marcus M Scheunemann, Raymond H Cuijpers, and Christoph Salge. Warmth and competence to predict human preference of robot behavior in physical human-robot interaction. In *2020 29th IEEE international conference on robot and human interactive communication (RO-MAN)*, pages 1340–1347. IEEE, 2020.

[TJP14] Benedict Tay, Younbo Jung, and Taezoon Park. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, 38:75–84, 2014.

[WB15] Astrid Weiss and Christoph Bartneck. Meta analysis of the usage of the godspeed questionnaire series. In *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 381–388. IEEE, 2015.

[WHT+24] Chao Wang, Stephan Hasler, Daniel Tanneberg, Felix Ocker, Frank Joublin, Antonello Ceravola, Joerg Deigmoeller, and Michael Gienger. Large language models for multimodal human-robot interaction. *arXiv preprint arXiv:2401.15174*, 2024.

[WKS+09] Michael L Walters, Kheng Lee Koay, Dag Sverre Syrdal, Kerstin Dautenhahn, and René Te Boekhorst. Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Procs of New Frontiers in Human-Robot Interaction*, 2009.

[ZCL+23] Ceng Zhang, Junxin Chen, Jiatong Li, Yanhong Peng, and Zebing Mao. Large language models for human-robot interaction: A review. *Biomimetic Intelligence and Robotics*, page 100131, 2023.