Bachelor Data Science and Artificial Intelligence

Physiological Correlates of Human-Robot Dictator Game Behavior

M.I. (Mathilde) van der Houwen

First and second supervisor:
Dr.ir. R.E. (Roy) de Kleijn & Dr. M.J. (Max) van Duijn

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl 24/01/2025

Abstract

With the increasing integration of robots into daily life, understanding how humans perceive and interact with robots is essential. This study investigates how robot appearance influences physiological responses, subjective perceptions, and decision making in human-robot interactions. Using a within-subjects design, participants interacted with three robots, NAO, Pepper, and AlphaMini, and played the dictator game, a social decision making designed to measure generosity. Physiological measures, including heart rate variability (HRV) and eye tracking metrics (eye blink rate and pupil diameter), were recorded alongside subjective perception ratings of likeability, anthropomorphism, social functionality, and empathy.

The results revealed significant effects of robot type on perceptions of anthropomorphism, social functionality, empathy, and eye blink rate. However, no significant effect was found for likeability, which exhibited a bimodal response pattern for Pepper, reflecting polarized participant perceptions. Although subjective perceptions, particularly likeability, social functionality, and empathy, were strongly associated with generosity, the physiological metrics HRV, eye blink rate and pupil diameter did not show predictive power for decision making.

These findings emphasize the critical role of subjective perceptions in driving decision making, suggesting that emotional and cognitive engagement during robot interactions may not always manifest itself in measurable physiological responses. Participants also more often allocated some portion to robots compared to human-human interactions in dictator games, highlighting unique dynamics in human-robot interaction.

This research contributes to the field by showing the influence of robot design on human behavior and perception. The study highlights the importance of designing robots that foster positive interactions and identifies areas for future research, including expanding robot diversity, exploring real-world settings, and investigating other physiological measures such as facial expression analysis or electroencephalography (EEG).

Contents

1 Introduction

With the increasing use of Artificial Intelligence, robots are becoming more common in daily life. As these robots are designed to interact with humans, it is important to better understand how people perceive and respond to them. Human-robot interaction (HRI) is a multidisciplinary field that explores how humans interact with robots in various settings, with the aim of enhancing the usability, safety, and acceptance of robots in everyday life. Research in HRI has shown that the design and appearance of robots significantly influence how humans perceive and interact with them [BDA+06]. Robots that are more human-like in appearance, for example, tend to elicit stronger emotional responses from users, both positive and negative, which may affect decision making and physiological responses during interaction [BK20].
This study combines insights from different fields to examine how people perceive commercial robots and how these perceptions influence their decisions and emotional responses. Participants in this study will have small, structured interactions with three different robots: NAO, Pepper, and AlphaMini. During these interactions, participants will play the dictator game, a well-known game often used to study social decision making. This setup provides a controlled way to investigate how people behave in the presence of robots.
In addition to observing behavior, this study will measure the physiological responses: heart rate variability (HRV), eye blink rate, and pupil diameter. HRV can offer insight into emotional regulation and cognitive effort [THSRJ09], while eye blink rate and pupil diameter are related to attention and cognitive states [JC16]. By looking at these physiological responses alongside the decisions made during the game, this study aims to explore how emotions and cognition influence decision making in interactions with robots.
The goal of this research is to contribute to the field of HRI by providing data on how physiological and emotional responses affect human behavior toward robots. This information can help in designing robots that are not only functional, but also better suited for positive interactions with humans.

The following research question has been made to answer this: *How do the physiological responses heart rate variability, eye blink rate, and pupil diameter correlate with decision making in human-robot interactions during the dictator game?*

1.1 Thesis Overview

This chapter contains the introduction, followed by Section 2, which provides the background and discusses related work. Section 3 outlines the hypotheses, and Section 4 describes the study design and the instruments used. Section 5 explains the data analysis approach. Subsequently, Section 6 presents the results, and Section 7 discusses the implications and limitations of the study. Finally, Section 8 offers the conclusions, and additional materials are provided in the Appendix.

This bachelor thesis, conducted at the Leiden Institute of Advanced Computer Science (LIACS), explores how varying robot appearances affect human perceptions and the physiological responses heart rate variability, eye blink rate, and pupil diameter, during interactive tasks. The primary supervisor, Roy de Kleijn, provided guidance through directing, discussing, and organizing where necessary, while Max van Duijn supported the project as the second supervisor.

2 Background

Previous research by de Kleijn et al. [DKDM24] extended the dictator game framework to include interactions with robots, providing valuable insights into human behavior in HRI contexts. Their study used commercially available robots and showed participants photographs of 18 robots. Participants then answered a set of 12 questions regarding each robot's characteristics before playing a dictator game against these robots. They found that three factors, Anthropomorphism, Likeability, and Utility, were all determinants of dictator game behavior, with Likeability being the most significant. Humanoid robots received the most money in their hypothetical dictator game, highlighting the importance of human-like traits in social decision making.

The previously mentioned research in human-robot interaction (HRI) focused on subjective perceptions of robots, emphasizing how individuals describe their experiences and attitudes toward them. While these insights are valuable, subjective measures alone may not fully capture the cognitive and emotional processes underlying human behavior during interactions with robots. This study aims to bridge this gap by combining physiological data with social decision making tasks, offering a more comprehensive understanding of the emotional and cognitive dynamics in HRI.

In the following part of this section, the physiological responses relevant to HRI and the role of the dictator game as a tool for measuring social decision making in interactions with robots will be explored. Physiological responses such as heart rate variability (HRV) and the use of eye tracking provide noninvasive ways to assess the cognitive and emotional states of individuals. Together, these measures allow for a more objective evaluation of human behavior toward robots.

2.1 Heart Rate Variability

Heart rate refers to the number of heartbeats per minute, while heart rate variability describes the variations in time interval between two heartbeats.

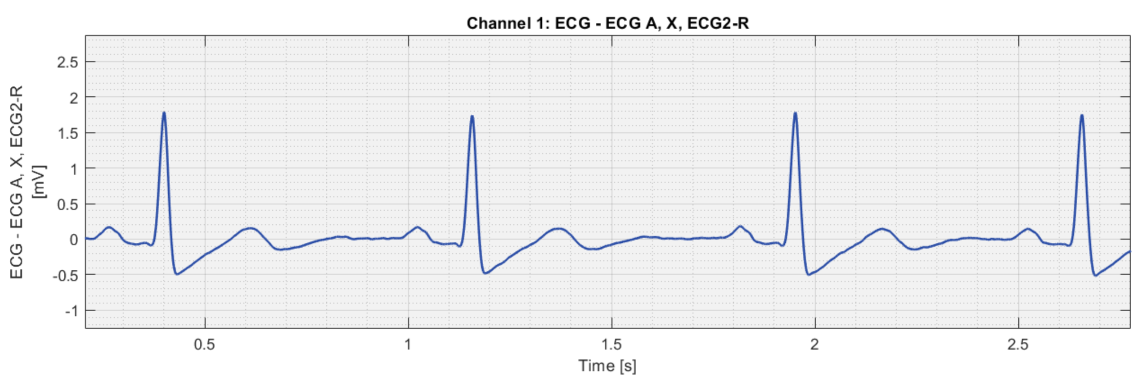

Figure 1: Raw ECG signal used for HRV analysis

In Figure 1, a raw electrocardiogram (ECG) signal from one of the participants is shown. This type of signal is essential for calculating HRV, as the intervals between successive R peaks in the ECG are used to derive HRV metrics. A heartbeat is defined as the interval from one R peak to the next R peak; for example, as shown in Figure 1, one heartbeat occurs between the R peak at 0.3 seconds and the subsequent R peak at 1.2 seconds. HRV metrics provide valuable information on the regulation of the autonomic nervous system and have been extensively studied in relation to stress, anxiety, and emotional arousal [THSRJ09]. A healthy heart shows complex and dynamic oscillations, allowing the cardiovascular system to adapt to sudden physical and psychological changes [SG17]. HRV can be analyzed using time-domain, frequency-domain, and nonlinear measures, with various metrics providing insights into different aspects of autonomic activity. This study adopts an exploratory approach, as the combination of HRI, physiological measures, and the dictator game has not been researched before. Three specific HRV metrics were selected: mean heart rate, RMSSD, and the LF/HF ratio, as they collectively offer insights into sympathetic and parasympathetic balance.

Mean heart rate was chosen as it is the easiest metric to understand and work with, making it a practical time-domain measure. RMSSD was selected as the second time-domain measure over the closely related pNN50, as it is generally preferred by researchers for its reliability in assessing parasympathetic activity, particularly in shorter samples and across diverse populations [SG17]. To complement RMSSD, which primarily reflects parasympathetic nervous system (PNS) activity, the LF/HF ratio was chosen over other frequency-domain measures to estimate the balance between sympathetic (SNS) and PNS activity. Together, these metrics provide an overview of autonomic regulation during human-robot interaction.

Nonlinear HRV metrics were not included in this study due to their complexity and the exploratory nature of the research. Since the focus was on understanding autonomic balance and its connection to behavioral outcomes, time- and frequency-domain measures are more relevant and interpretable for this context. The following part of this section provides an explanation of the three selected HRV metrics.

The first chosen metric is **mean heart rate** in beats per minute (BPM). An increase in heart rate is associated with sympathetic activation, which occurs during stress or physical activity, while a decrease in heart rate reflects parasympathetic activation, indicating relaxation or recovery as the body shifts into a calmer state [TVSVH09].

Secondly, the **root mean square of successive RR interval differences** (RMSSD) is defined by first calculating each successive time difference between heartbeats in milliseconds and taking the root mean square of this. RMSSD is a vagal-related heart rate variability (HRV) index and is commonly used to assess parasympathetic activity [BPLA07]. Higher RMSSD values indicate greater parasympathetic activity, suggesting a relaxed or well-recovered state, while lower RMSSD values suggest reduced parasympathetic activity, which may indicate stress or fatigue [BN23].

Lastly, the LF/HF ratio, meaning **ratio of low frequency to high frequency power**, indicates the balance between the sympathetic and parasympathetic activity. A low LF/HF ratio reflects parasympathetic dominance, and a high LF/HF ratio indicates sympathetic dominance, often occurring when we engage in fight-or-flight behaviors [SG17].

2.2 Eye Tracking

Eye tracking can be used to measure different aspects of behavior and cognition. For this research, the two metrics being analyzed are eye blink rate and pupil diameter, which could provide insights into participants' cognitive load, dopamine activity and therefore their emotional arousal. These metrics were chosen to align with the exploratory nature of this study and to help identify potential connections between eye activity and decision-making during interactions with robots.

Pupil diameter is a measure of cognitive load [KGDR16]. Cognitive load theory, introduced by Sweller in the 1980s, argues that knowledge can be categorized into two types: biologically primary knowledge, which humans acquire naturally through evolution, and biologically secondary knowledge, which requires effortful learning due to its cultural or societal relevance [Swe11]. When individuals engage with tasks involving biologically secondary knowledge, such as learning complex concepts or making strategic decisions, greater cognitive effort is required. As demonstrated by Kiefer et al. [KGDR16], this increased mental effort is reflected in pupil dilation, which occurs as a physiological response to higher cognitive demands. In the context of this study, pupil diameter would be an indication of the cognitive load experienced by participants during interactions with robots.

The **eye blink rate** is linked to both cognitive load and emotional arousal [JC16]. It can also say something about dopamine levels, as higher spontaneous blink rates are often associated with increased dopamine activity [JC16]. For example, a higher blink rate usually occurs when someone is under stress or dealing with complex tasks, as might be the case during social interactions with robots. However, a study by Sescousse et al. [SLvH+18] challenges the findings of Jongkees et al. [JC16], arguing that there is no positive correlation between spontaneous eye blink rate and dopamine activity. The study by Sescousse et al. was conducted with a small sample size (n=20), and the authors acknowledge that replication with a larger sample is warranted. However, their power analysis indicated that the sample size was sufficient to detect the reported effect. Therefore, the correlation between dopamine activity and eye blink rate is considered with caution, as the findings do not conclusively confirm a relationship between the two. This topic will be further explored, with the understanding that eye blink rate might not reliably reflect dopamine activity.

So, moreover, dopamine hormones play a crucial role in various cognitive and emotional functions, including motor control, motivation, and learning [CS22]. Elevated dopamine levels are often associated with heightened arousal, goal-directed behavior, and the pursuit of rewards [Sch98]. However, excessively high dopamine levels can also contribute to stress, anxiety, and competitive or aggressive behaviors as also found in patients with schizophrenia or depression [Gra16]. In contrast, lower dopamine levels are linked to symptoms such as fatigue, restlessness, and reduced motivation [BD07].

Thus, in this study, a higher eye blink rate may indicate increased cognitive load, emotional arousal, or possibly heightened dopamine activity, potentially reflecting stress or engagement with the robot. On the other hand, a lower eye blink rate could signify a more relaxed or disengaged state. While a higher eye blink rate reflects both emotional arousal and cognitive load, these processes are closely interrelated and may not be entirely separable in the context of human-robot interactions. However, complementary physiological measures or carefully controlled experimental designs could help isolate their contributions in future studies. For example, combining eye blink rate with other physiological signals, could distinguish between emotional arousal and cognitive load, if these measures are more specific to one of the two processes.

2.3 Dictator Game

The dictator game is an experimental paradigm used to study social decision making, particularly generosity and fairness [Bar08]. In the standard version of the game, one player (the "dictator") is given a sum of money and must decide how much, if any, to allocate to another player. The game is often used in behavioral economics and psychology to explore altruism, selfishness, and fairness in social interactions. Previous research in human-only interactions, such as the meta-analysis by Engel et al. [Eng11], found that 63.89% of participants gave at least some portion of their stake to the recipient. However, it is important to note that some participants chose not to allocate any money at all, reflecting individual differences in generosity.

In the context of HRI, the dictator game provides a controlled setting in which participants can make economic decisions in the presence of robots. By analyzing the amounts participants choose to give away, we can infer how different robot appearances and physiological states influence generosity and altruism.
As mentioned, De Kleijn et al. [DKDM24] extended the dictator game framework to robots. Their study demonstrated that the average proportion of the dictator game stake offered to robots ranged from 0.31 (Cubelets robot) to 0.74 (Sophia), with an overall average proportion of 0.50. These findings highlight how robot characteristics may influence human economic decisions. This study also hopes to contribute to the finding that robot characteristics may influence dictator game outcomes.

3 Hypotheses

For this research, multiple hypotheses are formulated to investigate the relationships between physiological responses, subjective perceptions of robots, and decision making in the dictator game. These hypotheses focus on HRV, eye blink rate, pupil diameter, and participants' perceptions of the robots.

H1

Specific HRV metrics, such as mean heart rate, RMSSD, and LF/HF ratio, will correlate with the amount of money participants choose to give away in the dictator game when interacting with robots. Lower parasympathetic activity (e.g., higher mean heart rate, lower RMSSD, or higher LF/HF ratio, indicating stress or arousal) will predict lower generosity in the dictator game.
This hypothesis is based on the physiological background of HRV metrics described in Section 2.

H2

Eye blink rate and pupil diameter will positively correlate with the amount of money given away, with higher blink rates (indicating cognitive load or engagement) and larger pupil diameters (indicating arousal or attentiveness) predicting greater generosity in the dictator game.
This hypothesis is based on the background of eye blink rate and pupil diameter described in Section 2, where these physiological responses are associated with higher levels of engagement or emotional arousal, potentially resulting in more generous behavior.

H3

Participants' subjective perceptions of the robots (e.g., likeability or functionality) will moderate the relationship between physiological responses (HRV, eye blink rate, and pupil diameter) and decision making in the dictator game. Robots perceived as more likeable, empathetic, functional or human-like will elicit physiological responses associated with higher generosity.
This hypothesis draws on the findings of De Kleijn et al. [DKDM24], who found a positive correlation between robot characteristics (e.g., likeability, anthropomorphism, and utility) and dictator game offers. The expectation is that when a robot is perceived as, for example, more likable, it elicits stronger physiological reactions (in line with H1 and H2) and, consequently, results in more generous offers in the dictator game.

4 Study Design

This study employs a within-subjects experimental design to investigate how different robot appearances influence participants' decisions and physiological responses during the dictator game. Each participant engages in three-minute interactions with three robots, each with unique attributes such as likeability, empathy, anthropomorphism, and functionality. During these interactions, participants act as the dictator, deciding how to split a sum of money with each robot. All robots follow a consistent script to maintain uniformity.

4.1 Procedure

1. Recruitment and Consent
   Participants are recruited and provided with detailed study information beforehand. Upon arrival, they are welcomed, offered a drink, and given a brief overview of the experiment.
   - Participants are fitted with electrode stickers (chest and belly) for HRV monitoring using the BIOPAC MP150 system.
   - They are equipped with Tobii Glasses 3 for eye tracking.
   - Baseline physiological measurements are taken while participants read the information letter (Appendix A) and sign the informed consent form (Appendix B).
2. Robot Interaction and Dictator Game
   Each participant interacts sequentially with the robots, presented in a counterbalanced order to control for order effects. During each interaction:
   - The participant has a brief conversation with the robot, during which they play the dictator game.
   - Participants' HRV are continuously monitored, and their eyes are continuously tracked.
   - After each interaction, participants complete a questionnaire assessing their subjective perception of the robot (likeability, social functionality, anthropomorphism, and empathy).
3. Debriefing
   At the conclusion of the study, participants are debriefed about the study's objectives and provided with further information (Appendix C). They are invited to ask questions and receive clarification.

The dependent variables are:

1. Heart rate variability
2. Eye tracking measures
3. Amount of money given away in the dictator game

4. Subjective Perceptions of Robots (likeability, empathy, anthropomorphism and social functionality)

The independent variables are:

1. Robot appearances

To provide a clear overview of the experimental setup, several pictures are included below. As shown in the images, four laptops are used during the procedure. The researcher uses three laptops: one is connected to the RobotsInDeKlas portal (which will be explained later in this section), another to the Tobii glasses for eye tracking, and the third to the HRV monitor for recording heart rate variability as seen in Figure 4. The fourth laptop is placed next to the participant and is used for answering the questionnaires. The participant will be placed on the first chair on the right side in Figure 2.

A divider is placed between the participant and the robots to ensure that the participant cannot see the robots before interacting with each specific one as seen in Figure 2 and 3. The robots remain in a sleeping mode before the conversation begins to avoid unintended sounds or movements.

Figure 2: Participants' side of the room

Figure 3: Side view of the room

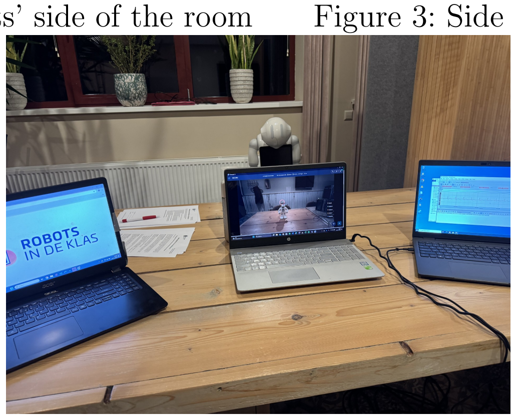

Figure 4: Researcher's perspective

4.2 Instruments

To execute this research, various instruments are required. These include the robots used in the experiment, the equipment for physiological measurements, the questionnaires, and the robot portal. Each of these components will be briefly explained in the following sections.

4.2.1 Robots

Three different robots were selected for this experiment: NAO, Pepper, and AlphaMini. These robots were chosen not only for their distinct appearances and degrees of human-likeness but also for their compatibility with the RobotsInDeKlas portal, which allows for consistent and standardized control and interaction across all trials.
Each robot hopes to offer differences in perceived anthropomorphism, likeability, social functionality, and empathy, making them essential for investigating how robot appearance influences participants' physiological and decision making responses. Before playing the dictator game, participants will briefly interact with each robot through the portal interface. This interaction is intended to introduce participants to the robot's capabilities and personality, shaping their subjective perceptions.
The robots are as following:

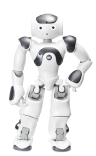

Figure 5: NAO

Figure 6: Pepper

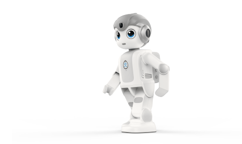

Figure 7: AlphaMini

NAO: A humanoid robot (Figure 5), smaller in size at 57 cm tall, frequently used in education and research, and known for its friendly design.

Pepper: A larger humanoid robot 120 cm tall (Figure 6), designed for social interaction, featuring a tablet interface and advanced human-likeness in its design.

AlphaMini: A compact humanoid robot (Figure 7), just 24 cm in height, with a toy-like appearance that emphasizes simplicity and likeability.

4.2.2 Physiological Measurement Devices

HRV was monitored using the BIOPAC MP150 system, connected to the AcqKnowledge app (version 5.0.6) for data acquisition and processing. This setup allowed for continuous ECG measurement during interactions with the robots. The recorded data were further processed using MATLAB 2020b Component Runtime (v9.9) with the Physio Data Toolbox software.
Eye movements, including blink rate and pupil diameter, were tracked using Tobii Glasses 3. Real-time tracking was facilitated by the Tobii Glasses 3 app, and the data were analyzed with Tobii Pro Lab software. Detailed information about the analysis of these measures is provided in the next Section 5 Data Analysis.

4.2.3 Questionnaires

The questionnaires were hosted on Qualtrics and accessed via a laptop with shortcuts for each robot's questionnaire.

Initially, demographic questions (e.g., age, gender) were asked. Post-interaction, participants evaluated the robots on scales ranging from 0 to 100. Some questions used scales from "Not at all" to "Very much," while others ranged from "Definitely not" to "Definitely yes."

To streamline the analysis process, the questions are grouped by dimension, as shown in Table 1. A detailed order of the questionnaire, including all individual questions, is provided in Appendix D.

Table 1: Questionnaire Items and Associated Dimensions

#	Question	Dimension
1	How much do you like this robot?	Likeability
2	How friendly is this robot?	Likeability
3	How uncomfortable does this robot make you feel?	Likeability
4	Would you like to touch this robot?	Likeability
5	Would you want to have this robot?	Likeability
6	How concerned would you feel if this robot was in danger or harmed?	Empathy
7	Would you feel sad if this robot was turned off or permanently removed?	Empathy
8	Would you feel guilty if you had to destroy or harm this robot?	Empathy
9	Can this robot plan its own actions independently?	Social Functionality
10	Would you trust this robot to help you with tasks around the house?	Social Functionality
11	Would you let this robot take care of your family?	Social Functionality
12	How physically similar is this robot to a human?	Anthropomorphism
13	Does this robot seem capable of feeling emotions?	Anthropomorphism
14	Does this robot think like a human?	Anthropomorphism

4.2.4 Robot Portal

The robots in this study were controlled through the RobotsInDeKlas portal, developed by Interactive Robotics. This cloud-based software platform is designed for managing social robots, with applications in both education and healthcare [Rob]. In educational settings, for instance, the robots can help students practice reading, math, and programming, while in healthcare, companion robots are used to support individuals with dementia.

For this study, the portal ensured that all interactions followed uniform and standardized scripts, minimizing variability across participants and maintaining consistency throughout the experiment. Each robot operated based on a predetermined script designed to engage participants effectively while controlling for any deviations in robot behavior.

The platform reflects the broader vision of Interactive Robotics, which aims to maximize the potential of social robots as they increasingly become part of everyday life. By employing this portal, the study benefits from a robust system that supports controlled and repeatable interactions. A screenshot of a sample conversation from the portal is included in Appendix E, illustrating the structured nature of the interactions.

5 Data Analysis

This section shows the steps taken to preprocess and analyze the data. The following subsections detail the analysis methods used starting with heart rate variability, then eye tracking, later the statistical analysis and ending with information about data preprocessing.

5.1 Heart Rate Variability Analysis

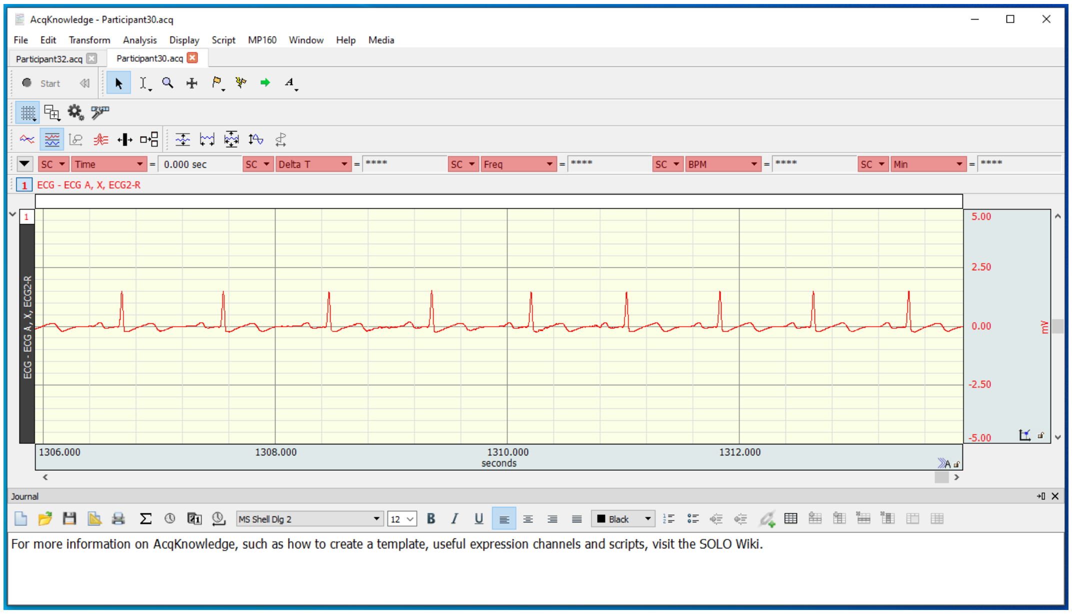

Figure 8: Screenshot of AcqKnowledge showing the continuous ECG recording during data collection

During the interactions, the researcher monitors the live ECG continuous recording using the AcqKnowledge software, as shown in Figure 8. The software allows event marking in real time using keyboard shortcuts. For example, pressing Ctrl + F1 marks the event as "Robot1," while Ctrl + F4 marks the event as "Break." These event markers are used to generate and analyze specific epochs later, including Baseline, Robot1, Robot2, and Robot3.
The ECG data is automatically saved as a .acq file. However, the default file format is incompatible with the analysis software, so all raw ECG files must be saved as Windows AcqKnowledge 3 Graph (.acq) format.

The next step involves using the PhysioData Toolbox application, which requires the free MATLAB 2020b Component Runtime (v9.9). The .acq files are converted into .physioData files for further processing.

The converted files are analyzed using the ECG Signal Analyzer within the PhysioData Toolbox. This analyzer identifies R peaks in the ECG signal, as shown in Figure 9. The default threshold for R-peak detection is set at 0.5 mV. However, individual differences in ECG signals mean that some participants may have lower R peaks, requiring manual adjustment.

In cases where the analyzer fails to detect R peaks or incorrectly marks non-R peaks, manual corrections are made:

- Add R peaks: When a true R peak is missed.
- Delete R peaks: When false R peaks are identified.

The orange dots in Figure 9 indicate R peaks detected by the software.

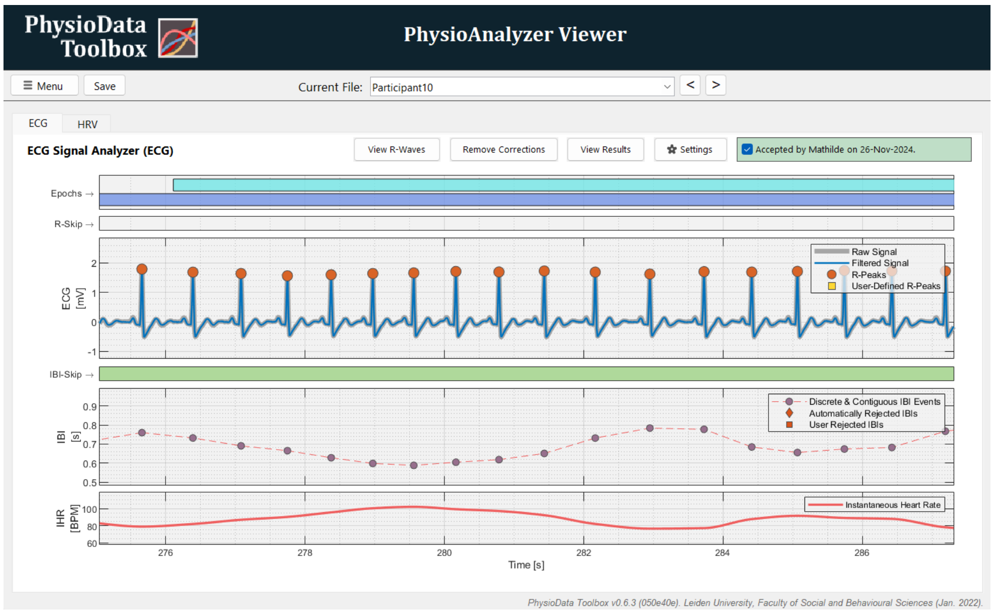

Figure 9: Screenshot of the PhysioData Toolbox displaying R-peak detection

Once the ECG signal is corrected, the HRV Analyzer module is applied within the PhysioData Toolbox. The steps include:

- Epoch Definitions: The predefined epochs (Baseline, Robot1, Robot2, Robot3) are reviewed and confirmed.
- File Acceptance: The analyzed HRV files are approved for final export.

Finally, the data can be exported. Epoch summaries are exported, providing a comprehensive set of variables required for statistical analysis. This export includes key HRV metrics such as RMSSD, LF/HF ratio, and mean heart rate, along with many other additional variables.

5.2 Eye Tracking Analysis

As shown from the researcher's perspective in Figure 4, one laptop is connected to the Tobii Glasses using the Glasses 3 app. This app enables real-time visualization of the participant's gaze, allowing the researcher to see exactly what the participant is looking at during the experiment. Additionally, the Glasses 3 app allows for predefined event templates to be created. For this study,

five event markers were used: Baseline, Robot1, Robot2, Robot3, and Break. After the experiment, the recordings are saved as .project files.
The next step involves importing these recordings into Tobii Pro Lab. In this application, a new Glasses project is created, and the Glasses 3 recordings are imported. The gaze data can then be visualized within the app, and are reviewed. Once the recordings are reviewed, the data is exported as a Single Excel file (.xlsx). During export, the following settings are applied:
Units: Exported as raw data, Timestamp precision: Set to milliseconds, Gaze filter: Defined as raw data, and Pupil diameter: Exported with a noise reduction filter applied.

Within the exported data file, a column labeled 'Event' contains the event markings. For analysis, the data corresponding to Baseline, Robot1, Robot2, and Robot3 are isolated. For each epoch, the lines containing the data from the start of the event (e.g., "Baseline") to the first "Break" event are selected and copied into a new file. To calculate blink rates, a blink is defined as an instance where the 'Eye movement type' is labeled as EyesNotFound for at least 50 milliseconds (equivalent to 5 rows, as each row represents 10 milliseconds). As most blinks last between 150 and 400 ms, where the eyes are completely closed for about 50 ms [Tob].
Occasionally, poor connectivity resulted in inaccurate data, such as large blocks of EyesNotFound for over 20 lines, which do not represent actual blinks. These issues were manually reviewed and corrected. Instead of deleting such blocks entirely, which could disrupt the time sequence, only the beginning or end portions of these segments were trimmed to shorten the timeframe while preserving accuracy.
Once the files were cleaned of connectivity errors, the data was analyzed using a Python script (see Appendix F). This script computed two key metrics for each epoch: the Blink Rate per Minute and the Average Pupil Diameter in millimeter.

5.3 Statistical Analysis

To analyze the relationships between physiological responses, subjective perceptions of robots, and decision making in the dictator game, a combination of statistical methods was done. Factor analysis was first done to assess the validity of the questionnaire dimensions: likeability, empathy, social functionality, and anthropomorphism. Factor scores were calculated to provide standardized metrics for analyses.
Given the within-subjects design, repeated-measures ANOVA was used to study the effects of robot type (NAO, Pepper, AlphaMini) on the heart rate variability metrics, eye blink rate, and pupil diameter, as well as subjective perceptions and dictator game outcomes. This is a good method for analyzing repeated observations within participants while keeping the individual variability in mind [PCK09].
Lastly, linear mixed-effects models were done to explore the relationships between predictors, such as physiological responses and subjective perceptions, and participants' decision making in the dictator game. These models incorporated random effects for participant ID and robot type, allowing the analysis to account for individual differences and repeated measures across conditions. All statistical analyses were done using RStudio. The R scripts used for these analyses are provided in Appendix G and I.

5.4 Data Preprocessing

During data preprocessing, physiological measures that included a baseline were normalized by calculating percentages relative to the baseline. In this way, consistency is ensured and participants can be compared.
One participant was excluded due to extreme response patterns. This individual consistently selected the maximum or minimum values across 20 items, which will probably skew the distribution and it potentially biases the results.
Furthermore, one participant (Participant 23) chose not to be connected to the HRV monitor. As a result, this participant was excluded from the HRV analysis, but was included in all other analyses.

6 Results

This section presents the findings from factor analysis, dictator game outcomes, subjective perception ratings, and multiple statistical analyses. They all study the effects of robot type on physiological responses, perceptions, and decision making.

6.1 Factor Analysis

The questions in the study were divided into 4 categories: likeability, empathy, social functionality, and anthropomorphism as shown in Table 1. The likeability category initially included 5 questions, while the other categories contained 3 questions each. To evaluate the consistency of each construct, a factor analysis was first performed separately on the individual question groups. Subsequently, an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA) were conducted on all questions combined to assess their alignment with the predefined categories. The R script used to perform this factor analysis can be found in Appendix G.

For the **likeability** category, question 2 ("How friendly is this robot?") showed a low factor loading of 0.32, suggesting that it did not sufficiently align with the likeability construct. Due to its low correlation, this question was removed from the likeability category.

For the **empathy** category, all three questions showed strong contributions to one factor with loadings above 0.69. This indicates that the three questions in this category are well-aligned, as they strongly correlate with the same underlying construct.

For the **social functionality** category, the first question ("Can this robot plan its own actions independently?") showed a weak loading of 0.37, while the other two questions had stronger correlations (0.71 and 0.73). The weakly aligned question was removed to ensure the remaining questions reliably measured social functionality.

For the **anthropomorphism** category, all three questions demonstrated strong loadings on a single factor with values ranging from 0.56 to 0.80. This indicates that the questions align well with the underlying construct.

After removing the two poorly correlating questions, a scree plot was generated before determining the number of factors. This plot indicates that at least four factors have eigenvalues greater than 1, supporting their adequacy as distinct factors, as suggested by Johnstone et al. [Joh01], and can be found in Appendix H.
To further validate this, an EFA was conducted, which produced the same scree plot but suggested a 3-factor solution. However, this finding is inconsistent with the criteria proposed by Johnstone et al. [Joh01]. Furthermore, the 3-factor solution resulted in a factor loadings matrix where multiple questions failed to load strongly (above 0.5) on any factor. This outcome reinforces Johnstone's argument that eigenvalues greater than 1 are indicative of distinct factors.
To ensure alignment with theoretical constructs and address these limitations, a CFA was performed using four distinct factors. The resulting factor loadings matrix, shown in Figure 10, demonstrates the validity and interpretability of this structure.

For further validation, the EFA was also conducted without removing the two poorly correlating questions. While this produced a similar scree plot, the factor loadings matrix was skewed: Q9 had a disproportionately high loading, making it difficult to interpret the loadings of other questions, while Q2 showed a very low loading. These results support the decision to exclude Q2 and Q9, as their removal improved the clarity and interpretability of the factor structure.

The matrix in figure 10 shows that:

- Likeability (MR2): Q1, Q3, Q4, Q5
- Empathy (MR1): Q6, Q7, Q8
- Social Functionality (MR4): Q10, Q11
- Anthropomorphism (MR3): Q12, Q13, Q14

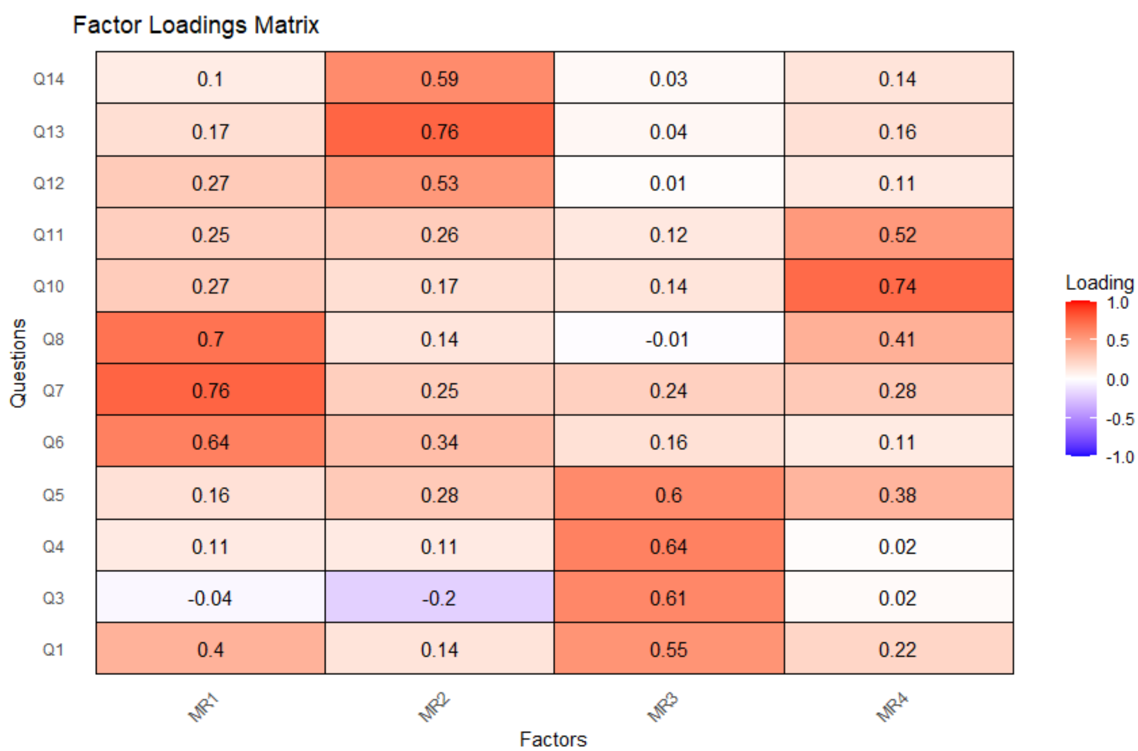

Figure 10: Factor loadings matrix

After determining the factor loadings, a score for each factor was calculated using the formula: Factor Score $= \dfrac{\mathrm{Q1} + \mathrm{Q2} + \cdots + \mathrm{Qn}}{n}$.
This method ensures that the resulting score is an average value ranging between 0 and 100, providing a standardized metric for comparison.

6.2 Dictator Game Outcome

The results of the dictator game, averaged across all participants, are presented in Table 2. The overall average proportion of money given away was 0.53. In all three cases, 87.10% of participants chose to give away at least some portion of the money. Participants could allocate a maximum of 10 euros, which was divided based on their own decisions during the game.

Table 2: Results of the Dictator Game

Robot	**Average Amount Given**	**Proportion Given Away**
NAO	5.34	0.53
Pepper	5.80	0.58
AlphaMini	4.84	0.48

6.3 Subjective Perception Results

Figure 11 displays the distribution of scores for the subjective perception metrics: likeability, anthropomorphism, social functionality, and empathy, across the three robots (NAO, Pepper, and AlphaMini).

Some distributions show a bimodal pattern, with two distinct peaks, indicating variability in participants' perception towards certain robot characteristics. For example, the likeability scores for NAO and AlphaMini show greater consistency compared to the more widely distributed scores for Pepper.

These variations show differences in how the robots' appearances and functionalities influenced participants' subjective perceptions.

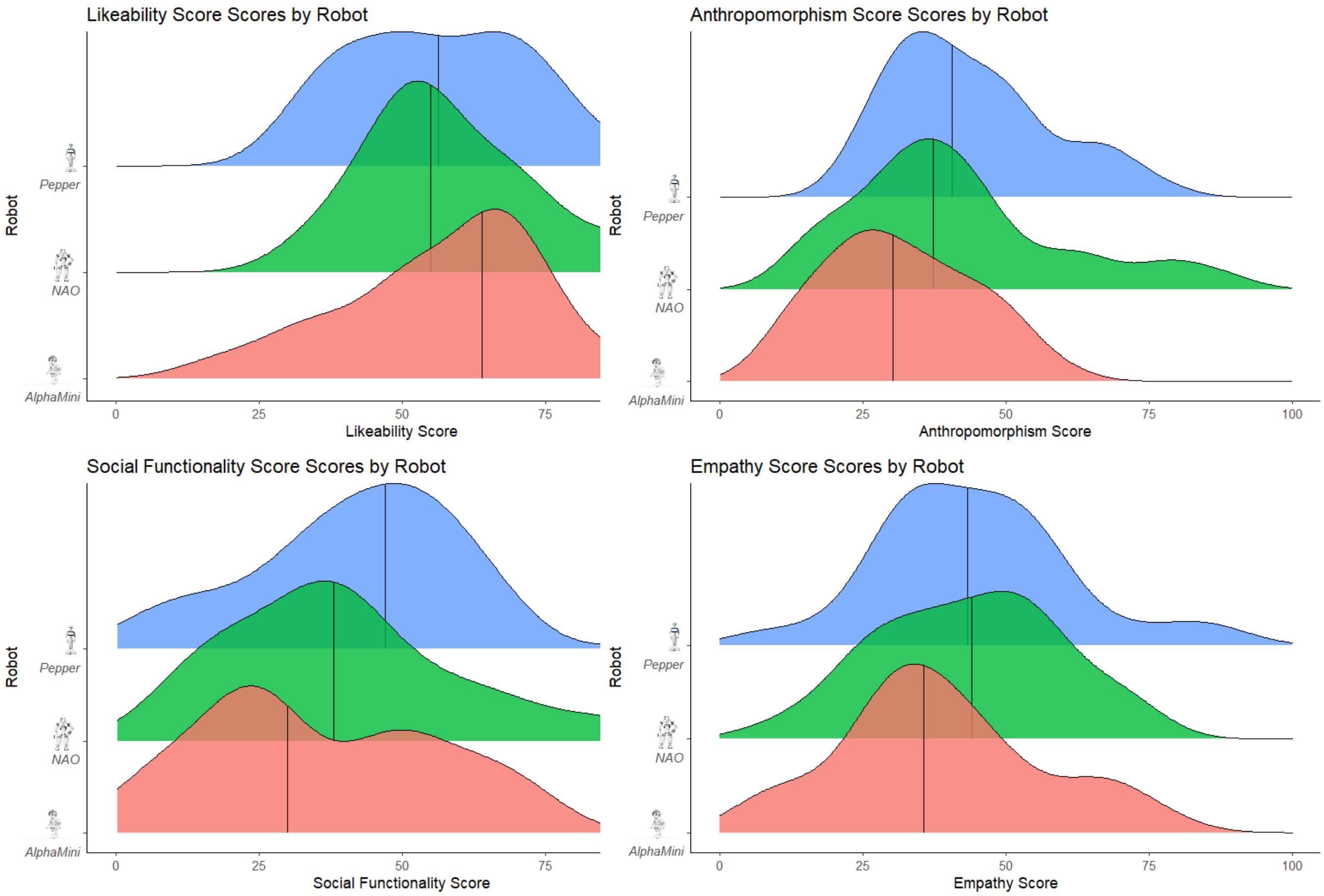

Figure 11: Comparison of robot scores across different metrics: Likeability, Anthropomorphism, Social Functionality, and Empathy.

6.4 Repeated-Measures ANOVA

The results of the repeated-measures ANOVA examining the effect of RobotType (NAO, Pepper, AlphaMini) on physiological measures, decision making, and subjective perceptions are presented below. The R script to do this analysis can be found in Appendix I.

Subjective Perceptions

- **Likeability**: No significant differences were found in likeability ratings across robot types, $F(2, 60) = 0.053, p = 0.949, \eta_g^2 = 0.001$. Participants rated all robots as equally likable. This will later be discussed in the section Discussion 7.
- **Anthropomorphism**: A significant effect of RobotType on perceived anthropomorphism was observed, $F(2, 60) = 13.943, p < 0.001, \eta_g^2 = 0.111$. This suggests that participants perceived the robots differently in terms of human-likeness.
- **Social Functionality**: RobotType significantly influenced perceptions of social functionality, $F(2, 60) = 3.793, p = 0.028, \eta_g^2 = 0.024$. This indicates that the robots were perceived differently in their ability to perform social tasks.

- **Empathy**: Perceived empathy towards the robot also varied significantly across robot types, $F(2,60) = 3.818, p = 0.027, \eta_g^2 = 0.024$. This indicates that participants felt a different level of empathy towards the different robots.

Heart Rate Variability Measures

- **HR Mean**: No significant differences were found in HRMean across robot types, $F(2,58) = 0.402, p = 0.671, \eta_g^2 = 0.005$. This indicates that participants' mean heart rate was not influenced by the type of robot they interacted with.

- **RMSSD**: Similarly, no significant effect of robot type was observed for RMSSD, $F(2,58) = 1.019, p = 0.367, \eta_g^2 = 0.011$. This suggests that heart rate variability (as measured by RMSSD) remained consistent across robots.

- **LF/HF Ratio**: The low-frequency to high-frequency ratio (LFHFRatio) also showed no significant differences between robot types, $F(2,58) = 0.490, p = 0.615, \eta_g^2 = 0.004$.

Eye Tracking Measures

- **Eye Blink Rate**: A significant effect of RobotType on eye blink rate was found, $F(2,60) = 6.214, p = 0.004, \eta_g^2 = 0.011$. This indicates that interactions with different robots elicited different physiological responses, potentially reflecting emotional arousal differences. These differences are visualized in Figure 12.

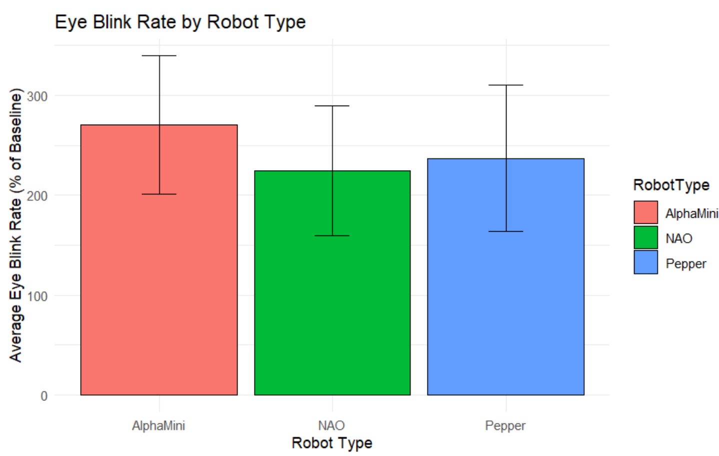

Figure 12: Bar plot of eye blink rate with 95% confidence intervals for each robot type

- **Pupil Diameter**: No significant differences in pupil diameter were observed across robot types, $F(2,60) = 1.233, p = 0.299, \eta_g^2 = 0.006$, suggesting that cognitive load was not influenced by the robots.

Dictator Game

- **Dictator Game Offers**: A significant effect of RobotType was found on the amount of money offered in the dictator game, $F(2, 60) = 3.235, p = 0.046, \eta_g^2 = 0.016$. This indicates that participants allocated their money differently across the robots. These differences are visualized in Figure 13.

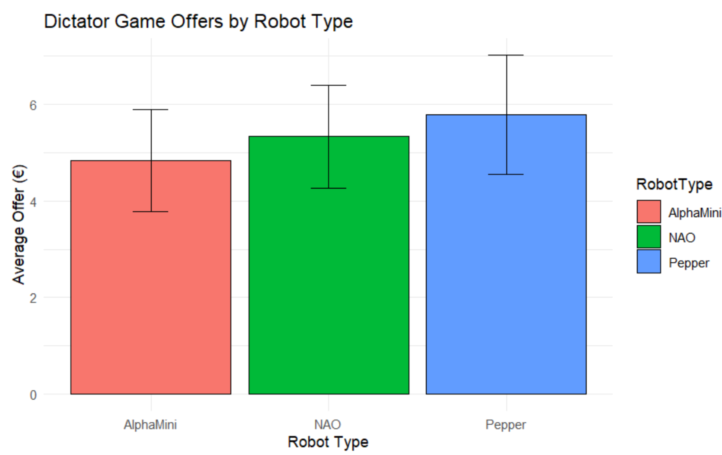

Figure 13: Bar plot of dictator game with 95% confidence intervals for each robot type

So, RobotType significantly influenced decision making (dictator game offers), perceived anthropomorphism, social functionality, empathy, and eye blink rate.

6.5 Linear Model Analysis

The mixed-effects models analyzed the relationship between various predictors and participants' dictator game offers. The models included random effects for both robot type and participant ID to account for variability across these groups. The results are summarized below. The R script to do this analysis can be found in Appendix I.

Significant Predictors

- **Likeability** ($\beta = 0.0502, p = 0.016$): Participants gave higher offers in the dictator game when they rated the robot as more likeable. This will be discussed further in the section Discussion 7.

- **Social Functionality** ($\beta = 0.0424, p = 0.009$): Robots perceived as more socially functional elicited higher offers.

- **Empathy** ($\beta = 0.0569, p = 0.016$): Robots that were rated as more empathetic were associated with increased generosity in the dictator game.

Non-Significant Predictors

- **Anthropomorphism** ($\beta = 0.0225, p = 0.194$): No significant effect was found for anthropomorphism.
- **HR Mean** ($\beta = -0.9760, p = 0.340$): The mean heart rate showed no significant relationship with dictator game offers.
- **RMSSD** ($\beta = 0.0017, p = 0.873$): The heart rate variability metric RMSSD showed no significant relationship with dictator game offers.
- **LF/HF Ratio** ($\beta = 0.0012, p = 0.630$): No significant effect was observed for the low-frequency to high-frequency ratio of heart rate variability.
- **Eye Blink Rate** ($\beta = 0.0034, p = 0.246$): Eye blink rate did not significantly predict offers.
- **Pupil Diameter** ($\beta = -0.0564, p = 0.100$): There was no significant effect found for pupil diameter, however since $p = 0.100$ it might be associated with dictator game offers. Further research is needed to confirm this potential effect.

This same linear model analysis also examined whether subjective perceptions of robots (e.g., Likeability, Empathy, Social Functionality, Anthropomorphism) moderate the relationship between physiological responses (HRV metrics, Eye Blink Rate, and Pupil Diameter). The results showed no significant correlations between HRV metrics (HRMean, RMSSD, and LF/HF Ratio) and participants' subjective perceptions of the robots, nor between subjective perceptions and attention-related physiological responses (eye blink rate and pupil diameter). These findings, summarized in Table 3, indicate that subjective perceptions of robots did not significantly influence participants' physiological responses during interactions.

Table 3: P-values for physiological metrics and subjective perceptions of robots

Metric	**Likeability** (p)	**Social Functionality** (p)	**Anthropomorphism** (p)	**Empathy** (p)
Eye Blink Rate	0.425	0.418	0.447	0.399
Pupil Diameter	0.593	0.628	0.710	0.498
HRMean	0.468	0.474	0.982	0.485
RMSSD	0.290	0.092	0.693	0.729
LF/HF Ratio	0.969	0.794	0.354	0.325

7 Discussion

This section discusses the study findings and provides information on the results and their implications. First, factor analysis and dictator game outcomes are explored. Then, the hypotheses are evaluated based on the rest of the results.

7.1 Factor Analysis

The factor loadings matrix in Figure 10 indicates that some questions contributed slightly to multiple factors. For example, Q1 (”How much do you like this robot?”) is part of the likeability construct, but also scored 0.4 on empathy. Similarly, Q5 (”Would you want to have this robot?”) was categorized as likeability, but had a loading of 0.38 on social functionality. Lastly, Q8 (”Would you feel guilty if you had to destroy or harm this robot?”) is part of the empathy construct but also scored 0.41 on social functionality.

These overlapping scores suggest that certain questions capture elements of multiple constructs. This could indicate the multidimensional nature of these perceptions, where participants associate characteristics like empathy and social functionality with how much they like the robot. Future studies might consider refining these constructs or including additional items to better capture these overlaps.

7.2 Dictator Game Outcome

The results of the dictator game showed that 87.10% of the participants chose to give at least some portion of their money to robots, which is noticeably higher than the 63.89% reported in human-human interactions by the meta-study by Engel et al. [Eng11]. This suggests that interactions with robots may lead to different behaviors compared to interactions with humans.

There are several possible reasons as to why participants in this study were more likely to give something away to robots. One reason could be that they were fascinated or amazed by the robots, which led them to give more. Additionally, robots might be seen as less judgmental or more innocent, which could make people feel more generous toward them. Unlike humans, robots can not take advantage of someone’s kindness, and there is no social comparison involved, which might make giving feel more genuine.

These findings provide useful insights into how people behave when sharing with robots and highlight the need for further exploration of the psychological mechanisms behind this behavior.

7.3 Hypothesis Evaluation

Three hypotheses were formulated for this study, as outlined in Section 3, and each is discussed below.

H1

H1 proposed that HRV metrics (mean heart rate, RMSSD, and LF/HF ratio) would correlate with the amount of money participants gave away in the dictator game. Lower parasympathetic activity was expected to predict lower generosity. However, repeated-measures ANOVA results showed no significant effects of RobotType on HRV metrics, with p-values of 0.671 (HR Mean), 0.367 (RMSSD), and 0.615 (LF/HF Ratio). Linear model analysis further supported this finding, with non-significant p-values of 0.340 (HR Mean), 0.873 (RMSSD), and 0.630 (LF/HF Ratio). This lack of correlation could be attributed to several factors, including the possibility that interactions with robots do not evoke strong enough emotional or physiological responses to influence heart rate variability. Additionally, participants may not perceive the interactions as stressful or arousing, which could diminish the sensitivity of HRV as a predictor. Another possibility considered was that the baseline measurements might not have been accurate, potentially affecting the normalized percentages. To address this, the analyses were repeated using the raw physiological data instead of the normalized values. However, this gave almost identical p-values, suggesting that the baseline was measured accurately and that the normalized values are reliable for this study. Based on all these findings, H1 is rejected.

H2

H2 proposed that eye blink rate and pupil diameter would positively correlate with generosity, with higher blink rates and larger pupil diameters predicting greater generosity. While RobotType significantly influenced eye blink rate ($p = 0.004$), no significant effect was found for pupil diameter ($p = 0.299$). Linear model analysis revealed non-significant correlations for both metrics, with p-values of 0.246 (eye blink rate) and 0.100 (pupil diameter).
Although pupil diameter approached significance, the findings suggest that these physiological measures are not robust predictors of generosity in this context. This could be because the task was not cognitively or emotionally demanding enough to elicit substantial changes in these metrics. Thus, H2 is rejected.

H3

H3 proposed that subjective perceptions of robots would moderate the relationship between physiological responses and decision making. It was hypothesized that robots perceived as more likable, empathetic, functional, or human-like would elicit physiological responses associated with higher generosity.

The results showed significant effects of RobotType on anthropomorphism ($p < 0.001$), social functionality ($p = 0.028$), and empathy ($p = 0.027$), but not on likeability ($p = 0.949$).
The absence of a significant effect of RobotType on likeability could be attributed to the bimodal distribution of participants' responses to Pepper, as shown in Figure 11. While the average likeability ratings indicate that NAO was liked the least, Pepper ranked second, and AlphaMini the most, the distribution of responses for Pepper was notably varied. Specifically, Pepper's bimodal pattern reflects two distinct participant groups: one that found Pepper engaging and exciting, and another that perceived it as intimidating or overly human-like.
This variability in perceptions likely diminished the statistical power to detect significant differences in likeability between the robots. Furthermore, Pepper's first peak, representing lower likeability scores, was below NAO's peak, while its second peak, representing higher likeability scores, exceeded AlphaMini's peak. This overlapping range of responses across the robots may have contributed to the high p-value (0.949), indicating no significant difference in likeability across RobotTypes.
The researcher also noticed during the study that some participants were enthusiastic about Pepper, while others found it unsettling, which might have contributed to the inconsistent responses.

Linear model analysis revealed significant relationships between dictator game offers and subjective perceptions of likeability (p = 0.016), social functionality (p = 0.009), and empathy (p = 0.016), but not anthropomorphism (p = 0.194). This suggests that participants were more likely to be generous towards robots they perceived as likable, socially functional, or empathetic. One possible explanation for why anthropomorphism did not elicit a significant result is the uncanny valley hypothesis. This theory suggests that robots with moderate levels of anthropomorphism may be perceived as more likable, while robots that appear too human-like can evoke discomfort or distrust, potentially leading to a U-shaped correlation [MMK12]. To test this idea, a regression model was applied, first including all robots, and later focusing only on Pepper, since this robot had the highest anthropomorphism scores. However, neither analysis revealed a significant linear nor quadratic relationship between anthropomorphism and dictator game outcomes. Therefore, the uncanny valley effect was not supported by the data in this study.
Another possible explanation for why anthropomorphism did not reach significance, despite being close (p = 0.194), is that its effect may be indirect or context-dependent. While the robot with the highest anthropomorphism score also elicited the highest dictator game outcome, and vice versa, this suggests a potential trend. However, anthropomorphism may not be the strongest influencing factor compared to other subjective measures.
Another important finding is that the linear model analysis revealed likeability as a significant predictor for dictator game offers (p = 0.016). Specifically, participants tended to give higher offers when they rated the robot as more likable. However, as shown in Figure 11, AlphaMini, the most liked robot, received the lowest average offer in the dictator game. This discrepancy can be partly explained by the linear model, which assesses relationships at the individual level. Within participants, higher likeability ratings predicted higher offers. However, this does not necessarily translate directly to group-level averages across robot types. Furthermore, the ANOVA results showed no significant differences in likeability scores across the robots (p = 0.949). While AlphaMini may have had a higher average likeability score, this difference was not statistically meaningful. As a result, the relationship between likeability and offers should not be expected to vary systematically by robot type.

Furthermore, the linear model analysis demonstrated no significant correlations between physiological responses (HRV metrics, eye blink rate, pupil diameter) and subjective perceptions. This indicates that subjective perceptions influenced decision making directly, rather than through physiological responses. Therefore, H3 is partially rejected: while almost all subjective perceptions significantly influenced decision making, they did not moderate physiological responses.

7.4 Limitations

While this study provides valuable insights into the effects of robot appearance on physiological responses, subjective perceptions, and decision making, several limitations should be acknowledged.

First, the sample size in this study was relatively small. However, the power analysis for the ANOVA revealed a power of 0.783, which suggests that the sample size was likely sufficient for detecting differences in robot types. Nevertheless, it also highlights potential limitations in detecting smaller or more subtle effects. While the results should be interpreted with some caution, the findings provide a solid foundation for future studies, which could benefit from larger sample sizes to further strengthen and validate the conclusions.

Second, the variability in subjective perceptions, particularly the bimodal distribution of likeability ratings for Pepper, introduced noise in the data. Participants appeared polarized in their perceptions of Pepper, with some finding it engaging and exciting, while others perceived it as intimidating or overly human-like. This variability may have influenced the overall results.

Additionally, the study included only three robots (NAO, Pepper, and AlphaMini), each with distinct appearances and functionalities. While this selection provided a range of robot types, it does not fully represent the diversity of robots that participants might encounter in other contexts. The interaction time with each robot was also relatively brief, lasting only 3 to 5 minutes. This short duration may not have been sufficient for participants to fully adapt to the robots or exhibit stable physiological responses, potentially limiting the depth of the findings related to emotional and cognitive engagement.

Moreover, the physiological metrics used, specifically HRV and the eye tracking measures, may have limited sensitivity to the nuances of human-robot interaction. These metrics are indirect measures of emotional and cognitive responses, and their lack of significant findings in some analyses suggests that they might not fully capture the complexity of these interactions.

Finally, the within-subject design, while efficient for comparing responses across robot types, may have introduced order effects. Although efforts were made to counterbalance the sequence of robot interactions, participants' responses could still have been influenced by the order in which they encountered the robots, e.g. in their subjective perceptions.

8 Conclusions

This study investigated how robot appearances influences physiological responses, subjective perceptions, and decision making in human-robot interactions. Using the dictator game as a framework, interactions with three robots (NAO, Pepper, and AlphaMini) were analyzed to explore relationships between physiological metrics like HRV and eye tracking data, subjective perceptions of robots, and dictator game behavior.

The average proportion of money given away in this study was 0.53, ranging from 0.48 for AlphaMini, 0.53 for NAO, to 0.58 for Pepper. The findings revealed significant effects of robot type on anthropomorphism, social functionality, empathy and eye blink rate, but not on likeability, HRV metrics and pupil diameter. Interestingly, while subjective perceptions (especially, likeability, social functionality and empathy) were significantly correlating with generosity, the physiological metrics, HRV and the eye tracking measures (blink rate and pupil diameter) did not show strong predictive power in this context.

The study's hypotheses were all rejected, but one was partially supported. Subjective perceptions directly influenced decision making, but did not moderate the relationship between physiological responses and generosity. Furthermore, the lack of significant effects for certain metrics suggests that the emotional and cognitive engagement elicited by robot interactions may not be strong enough to elicit measurable changes in physiological responses.

This research contributes to the growing field of human-robot interaction by providing insights into the role of robot appearance in shaping human behavior and perception. The findings show the importance of subjective perceptions in influencing decision making, suggesting that robot design plays a critical role in fostering positive human-robot interactions.

8.1 Further Research

Future studies should address the limitations of this research to build on its findings. Expanding the sample size and diversity would improve the generalization of the results.

Future research should also broaden the selection of robots by including a wider variety of designs. Less anthropomorphic robots should be included as well, to better understand how different appearances and functionalities influence perceptions and participant behavior.

Another aspect that could help elicit a stronger emotional and physiological response is increasing the interaction duration. This would allow participants more time to adjust to the robots and for their responses to develop. Longer interactions might even lead to significant physiological responses, providing more trustworthy results.

Furthermore, conducting these studies in real-world settings could elicit more genuine reactions from participants, potentially enhancing the ecological validity of the findings.

Since the physiological measures used in this study did not produce significant results, exploring other measures could be beneficial. For example, facial expression analysis could be interesting, as subjective perceptions were the most significant findings in this study, and facial expressions are closely related to how emotions and intentions are perceived. While facial expressions can convey emotions, they can also express intentions, cognitive processes, physical effort, or other intra- or interpersonal meanings [LJT+05].

Another measure worth exploring is electroencephalography (EEG), a non-invasive tool that records electrical activity in the brain [Coh17]. EEG has the potential to provide valuable insights into cognitive processes during human-robot interactions by capturing how different brain regions are activated and interact over time.
One promising approach is the use of event-related potentials (ERPs), which are EEG changes time-locked to sensory, motor, or cognitive events [SS09]. ERPs offer a reliable, non-invasive method for studying the psychophysiological correlates of mental processes. For instance, ERPs could reveal how participants process stimuli from robots and whether these processes differ across robot types, providing insights into attentional and emotional engagement.
Another interesting direction would be to explore whether brain activity during interactions with robots resembles that observed during interactions with humans. EEG is considered a sum of the activity of different sources that mix in time and space, enabling the identification of transient, frequency-locked oscillatory states related to cognitive and task-induced states [KSH+05]. By studying measures such as brain connectivity, it would be possible to understand how different regions of the brain communicate during these interactions. This could help answer questions like: Do people perceive and process robots similarly to humans?

Addressing these areas in future research will contribute to a deeper understanding of human-robot interaction and support the development of robots designed for seamless integration into daily life.

References

[Bar08] Nicholas Bardsley. Dictator game giving: altruism or artefact? *Experimental economics*, 11:122–133, 2008. https://doi.org/10.1007/s10683-007-9172-2.

[BD07] Anders Björklund and Stephen B Dunnett. Dopamine neuron systems in the brain: an update. *Trends in neurosciences*, 30(5):194–202, 2007. https://doi.org/10.1016/j.tins.2007.03.006.

[BDA+06] Mike Blow, Kerstin Dautenhahn, Andrew Appleby, Chrystopher L Nehaniv, and David C Lee. Perception of robot smiles and dimensions for human-robot interaction design. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 469–474. IEEE, 2006. https://doi.org/10.1109/ROMAN.2006.314372.

[BK20] Christoph Bartneck and Merel Keijsers. The morality of abusing a robot. *Paladyn, Journal of Behavioral Robotics*, 11(1):271–283, 2020. https://doi.org/10.1515/pjbr-2020-0017.

[BN23] AR Sohara Banu and V Nagaveni. Assessment of sympathetic and parasympathetic activities of nervous system from heart rate variability using machine learning techniques. *SN Computer Science*, 4(5):646, 2023. https://doi.org/10.1007/s42979-023-02062-y.

[BPLA07] Martin Buchheit, Yves Papelier, Paul B Laursen, and Saïd Ahmaidi. Noninvasive assessment of cardiac parasympathetic function: postexercise heart rate recovery or heart rate variability? *American Journal of Physiology-Heart and Circulatory Physiology*, 293(1):H8–H10, 2007. https://doi.org/10.1152/ajpheart.00335.2007.

[Coh17] Michael X Cohen. Where does eeg come from and what does it mean? *Trends in neurosciences*, 40(4):208–218, 2017. https://doi.org/10.1016/j.tins.2017.02.004.

[CS22] Kauê Machado Costa and Geoffrey Schoenbaum. Dopamine. *Current Biology*, 32(15):R817–R824, 2022. https://doi.org/10.1016/j.cub.2022.06.060.

[DKDM24] Roy De Kleijn, Avantika Dev, and Sumitava Mukherjee. Behavioral-economic games with commercially available robots. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 383–386, 2024. https://doi.org/10.1145/3610978.3640575.

[Eng11] Christoph Engel. Dictator games: A meta study. *Experimental economics*, 14:583–610, 2011. https://doi.org/10.1007/s10683-011-9283-7.

[Gra16] Anthony A Grace. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience*, 17(8):524–532, 2016. https://doi.org/10.1038/nrn.2016.57.

[JC16] Bryant J Jongkees and Lorenza S Colzato. Spontaneous eye blink rate as predictor of dopamine-related cognitive function—a review. *Neuroscience & Biobehavioral Reviews*, 71:58–82, 2016. https://doi.org/10.1016/j.neubiorev.2016.08.020.

[Joh01] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327, 2001. http://dx.doi.org/10.1214/aos/1009210544.

[KGDR16] Peter Kiefer, Ioannis Giannopoulos, Andrew Duchowski, and Martin Raubal. Measuring cognitive load for map tasks through pupil diameter. In *Geographic Information Science: 9th International Conference, GIScience 2016, Montreal, QC, Canada, September 27-30, 2016, Proceedings 9*, pages 323–337. Springer, 2016. https://doi.org/10.1007/978-3-319-45738-3_21.

[KSH+05] Thomas Koenig, D Studer, Daniela Hubl, L Melie, and WK Strik. Brain connectivity at different time-scales measured with eeg. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1015–1024, 2005. https://doi.org/10.1098/rstb.2005.1649.

[LJT+05] Stan Z Li, Anil K Jain, Ying-Li Tian, Takeo Kanade, and Jeffrey F Cohn. Facial expression analysis. *Handbook of face recognition*, pages 247–275, 2005. https://doi.org/10.1007/0-387-27257-7_12.

[MMK12] Masahiro Mori, Karl F MacDorman, and Norri Kageki. The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2):98–100, 2012. https://doi.org/10.1109/MRA.2012.2192811.

[PCK09] Eunsik Park, Meehye Cho, and Chang-Seok Ki. Correct use of repeated measures analysis of variance. *The Korean journal of laboratory medicine*, 29(1):1–9, 2009. https://doi.org/10.3343/kjlm.2009.29.1.1.

[Rob] Interactive Robotics. Over interactive robotics. Website. Accessed: 06-01-2025, https://www.robotsindeklas.nl/over-ons/.

[Sch98] Wolfram Schultz. Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, 80(1):1–27, 1998. https://doi.org/10.1152/jn.1998.80.1.1.

[SG17] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, 5:258, 2017. https://doi.org/10.3389/fpubh.2017.00258.

[SLvH+18] Guillaume Sescousse, Romain Ligneul, Ruth J van Holst, Lieneke K Janssen, Femke de Boer, Marcel Janssen, Anne S Berry, William J Jagust, and Roshan Cools. Spontaneous eye blink rate and dopamine synthesis capacity: preliminary evidence for an absence of positive correlation. *European Journal of Neuroscience*, 47(9):1081–1086, 2018. https://doi.org/10.1111/ejn.13895.

[SS09] Shravani Sur and Vinod Kumar Sinha. Event-related potential: An overview. *Industrial psychiatry journal*, 18(1):70–73, 2009. https://doi.org/10.4103/0972-6748.57865.

[Swe11] John Sweller. Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76. Elsevier, 2011. https://doi.org/10.1016/B978-0-12-387691-1.00002-8.

[THSRJ09] Julian F Thayer, Anita L Hansen, Evelyn Saus-Rose, and Bjorn Helge Johnsen. Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of behavioral medicine*, 37(2):141–153, 2009. https://doi.org/10.1007/s12160-009-9101-z.

[Tob] Tobii. New blink detection method with eye tracking. Webinar. Published: 19-11-2024. Accessed: 23-12-2024, https://www.tobii.com/resource-center/webinars/new-blink-detection-method-with-eye-tracking.

[TVSVH09] Joachim Taelman, Steven Vandeput, Arthur Spaepen, and Sabine Van Huffel. Influence of mental stress on heart rate and heart rate variability. In *4th European Conference of the International Federation for Medical and Biological Engineering: ECIFMBE 2008 23–27 November 2008 Antwerp, Belgium*, pages 1366–1369. Springer, 2009. https://doi.org/10.1007/978-3-540-89208-3_324.

A Information Letter

Faculty of Social Sciences
Institute of Psychology

Information letter for participants

Physiological Correlates of Human-Robot Dictator Game Behavior

Dear participant,

You are invited to take part in a scientific study. This study is being conducted by Mathilde van der Houwen supervised by Dr.ir. R.E. (Roy) de Kleijn of the Leiden Institute of Advanced Computer Science at Leiden University and has been approved by the Psychology Research Ethics Committee of Leiden University (reference number: 2024-10-28-R.E. de Kleijn-V2-5686).

Purpose of the study

In this study, we investigate how physiological responses correlate with decisions made in the presence of robots. This study aims to contribute to the broader field of Human-Robot Interaction by providing empirical data on the emotional and cognitive factors that influence human behavior towards robots.

Participation

Participation in this study is completely voluntary. This means that you can end your participation at any time and without any explanation. This will not have any detrimental consequences for you. If you decide to withdraw, you will receive a payment in proportion to your investment if you wish to be compensated.

In order to participate, you must meet all of the following inclusion criteria:

- You are between 16 and 70 years of age.
- You should not have any known cardiovascular conditions or any other conditions that could affect heart rate variability or eye blink rate.
- You should not use any medication that could affect heart rate variability or eye blink rate.

If you cannot participate due to any of these criteria, you do not need to disclose which one.
Please inform the experiment leader that you cannot participate.

Procedure

The study will take place in a private meeting room at a location in 's-Gravenzande. You will interact with three interactive robots, complete the dictator game, and have your heart rate variability (HRV) and eye blink rate monitored during the session. After each interaction, you will also fill out a short questionnaire about your perception of the robots. The session will last about 20-25 minutes.

<u>Benefits and risks</u>

During the study, we will measure your heart rate and eye blink rate. Three small electrodes will be attached to your chest, connected by wires to a small device worn around your waist, to monitor your heart rate. Additionally, you will wear glasses equipped with cameras and sensors to record your eye movements. These measurements are taken following standard guidelines and are safe, with no known risks.

<u>Compensation</u>

The compensation for participating in this study is €4,25. To transfer this money to your bank account, we will need to collect your address and bank details.

We will store your data securely. Financial information will be kept separate from research data, and will only be accessible by our financial department, who will pass it on securely to the Dutch Tax and Customs Administration. We will retain your financial information for 7 years, to comply with legal requirements.

You may withhold any information that you want, including your BSN. However, doing so means that our financial department will not be able to pay you for participation.

Confidentiality and privacy

All your data will be handled with strict confidentiality. We will collect the following data during the study:

- **Physiological data**: Heart rate and eye blink rate, measured with non-invasive equipment.
- **Decision-making data**: Your choices during the dictator game.
- **Questionnaire data**: Your subjective perceptions of the robots (e.g., likability, scariness, functionality).

No sensitive data, such as health information, will be collected. To ensure you haven't participated previously, we may ask for a visual check of your photo ID, but it will not be stored or copied.

How do we protect your privacy?
To protect your privacy, your data will be stored, processed and published in a coded manner. This means we associate your data with a code rather than your personally-identifiable information, so nobody will be able to link the data to you. We save the key to the code in a physically-locked or password-protected location, accessible only to the researcher(s) in charge.

All data is collected anonymously. When we store, process or publish about your data, nobody will be able to link the data to you personally. We store your data for 10 years.

Can you withdraw your consent for the use of your data?
You can take back your consent for the use, storage, and publishing of your data within two weeks after your participation has ended.

For more information on data privacy and your rights, please check the European Union's data privacy law, known as the General Data Protection Regulation ("GDPR").

Contact information

If you have any questions before or after participating in this study, you can contact the principal investigator: Mathilde van der Houwen, m.i.van.der.houwen@umail.leidenuniv.nl, +31640791920 or you can contact the supervisor Dr.ir. Roy de Kleijn, kleinrde@fsw.leidenuniv.nl.

You can also contact the (principal) investigator if you have a complaint. If you prefer not to do so, you can contact the Contact point for research participants at the Faculty of Social Sciences of Leiden University:
Contactpuntparticipanten@fsw.leidenuniv.nl

If you have any questions or complaints about your privacy or the processing of your personal data, you can contact the privacy officer of Leiden University: privacy@bb.leidenuniv.nl

B Informed Consent Form

Universiteit Leiden

Faculty of Social Sciences
Institute of Psychology

Informed consent form for participants

I have been asked to give permission to participate in the study 'Physiological Correlates of Human-Robot Dictator Game Behavior'. I declare the following:

- I have read the information letter.
- I was able to ask questions. If I had questions, they were answered to my satisfaction.
- I had enough time to decide if I wanted to take part.
- I know who to contact in case of any complaints.
- I know that taking part is voluntary. I also know that I can decide at any time not to take part in the study or to stop taking part in it. I do not have to explain why, and stopping will not have negative consequences for me. I understand I will not be compensated for my time investment if I stop the study prematurely.
- I know that the research data will be safely stored (coded or anonymized) for at least 10 years.
- I understand that the researchers may share with other researchers my anonymous/de-identified research data (but not the audio and video recordings) that cannot be traced back to me.
- This is the first time I participate in this study.

I consent to my participation in this study.

Signed by ……………………………………………………………..

Date ……………………………………………………………..

Signature ……………………………………………………………..

CEP version March 2024 - 2024-10-28-R.E. de Kleijn-V2-5686 *page 1 of 1*

C Debriefing Letter

Universiteit Leiden

Faculty of Social Sciences
Institute of Psychology

Debriefing letter for participants

Physiological Correlates of Human-Robot Dictator Game Behavior

Thank you for participating in this study. We would now like to provide further explanation about this study.

This study investigates how **physiological responses**—such as **heart rate variability (HRV)** and **eye blink rate**—correlate with **decision-making** during interactions with robots. Specifically, we are exploring how different robot appearances might influence your emotional and cognitive responses during the **dictator game**, a well-known tool for studying social decision-making.

In this study, we are testing the following ideas:

1. We believe that lower heart rate variability (HRV)—which can indicate higher stress or arousal—will be linked to lower generosity in the dictator game. In other words, participants with lower HRV may choose to give away less money when interacting with robots.
2. We also expect that higher eye blink rates—which can suggest higher cognitive load or anxiety—will be associated with lower generosity in the dictator game. Participants who blink more often may give away less money.
3. Lastly, we think that your subjective perceptions of the robots—such as how likable, scary, or functional you find them—will influence how your physiological responses (like HRV and eye blink rate) are related to your decisions in the dictator game.

By participating, you have helped us gather valuable data on how people perceive robots and make decisions when interacting with them. Your physiological responses, the choices you made in the game, and your ratings of the robots will contribute to a better understanding of how humans react to robots in different contexts. This knowledge could ultimately help in designing robots that are more effective and socially acceptable in future interactions.

Please do not share the information in this debriefing letter with other potential participants, as the information could influence their behavior during the study.

Contact information

If you have any questions after participating in this study, you can contact the (principal) investigator: Mathilde van der Houwen, m.i.van.der.houwen@umail.leidenuniv.nl, +31640791920 or you can contact the supervisor Dr.ir. Roy de Kleijn, kleinrde@fsw.leidenuniv.nl.
You can contact the investigator if you have a complaint. If you prefer not to do so, you can contact the Contact point for research participants at the Faculty of Social Sciences of Leiden University:
Contactpuntparticipanten@fsw.leidenuniv.nl

If you have any questions or complaints about your privacy or the processing of your personal data, you can contact the privacy officer of Leiden University: privacy@bb.leidenuniv.nl

CEP version March 2024 - 2024-10-28-R.E. de Kleijn-V2-5686 *page 1 of 1*

D Questionnaires

Table 4: Demographic and Reimbursement Questions Used in the Study

#	Question
1	How old are you?
2	What is your gender?
3	Would you like to receive a €4.25 reimbursement for your participation in this study?

Table 5: Robot Perception Questionnaire: Items Categorized by Dimension

#	Question	Dimension
1	How much do you like this robot?	Likeability
2	How friendly is this robot?	Likeability
3	How physically similar is this robot to a human?	Anthropomorphism
4	How concerned would you feel if this robot was in danger or harmed?	Empathy
5	How uncomfortable does this robot make you feel?	Likeability
6	Would you feel sad if this robot was turned off or permanently removed?	Empathy
7	Would you feel guilty if you had to destroy or harm this robot?	Empathy
8	Would you like to touch this robot?	Likeability
9	Would you want to have this robot?	Likeability
10	Does this robot seem capable of feeling emotions?	Anthropomorphism
11	Can this robot plan its own actions independently?	Social Functionality
12	Would you trust this robot to help you with tasks around the house?	Social Functionality
13	Does this robot think like a human?	Anthropomorphism
14	Would you let this robot take care of your family?	Social Functionality

Note: The dimensions (e.g., Likeability, Empathy) were not shown to participants and are presented here only for clarity in the report.

E RobotsInDeKlas Script

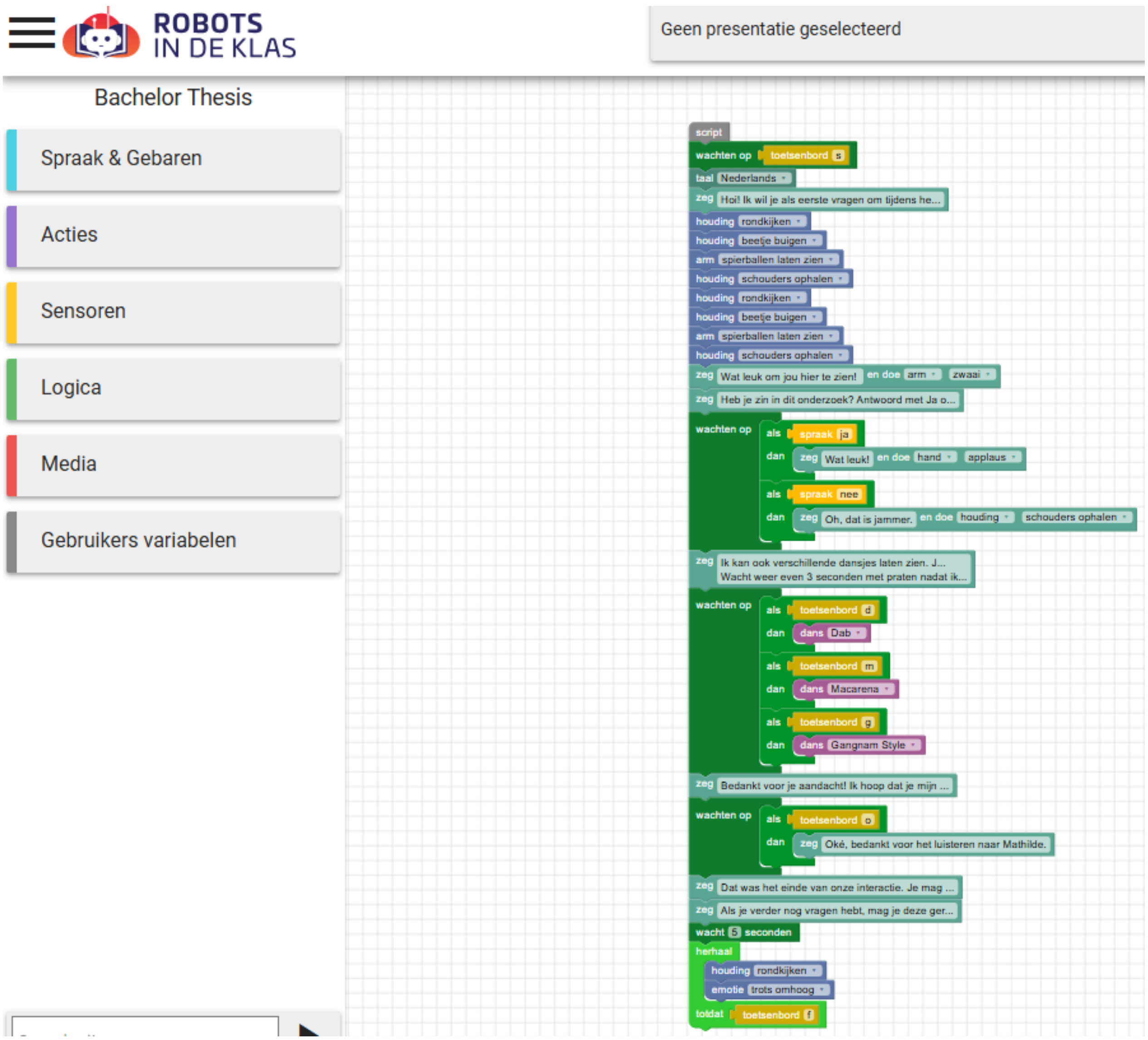

Figure 14: Screenshot RobotsInDeKlas script

F Python File: Eye Tracking

```python
import pandas as pd
import os

def compute_blink_rate_and_pupil(file_path):
    try:
        # Load the Excel file
        df = pd.read_excel(file_path, sheet_name='Blad1')
```

```python
    except FileNotFoundError:
        raise FileNotFoundError(f"Error: File not found at {file_path}")

    # Extract the target column (24th) for "EyesNotFound" status
    eyes_status = df.iloc[:, 23]  # 24th column for "EyesNotFound" status
    pupil_diameter = df.iloc[:, 20]  # 21st column for pupil diameter in mm

    # Initialize variables
    blink_count = 0
    consecutive_eyes_not_found = 0
    consecutive_no_eyes_not_found = 0

    for status in eyes_status:
        if status == "EyesNotFound":
            consecutive_eyes_not_found += 1
            consecutive_no_eyes_not_found = 0  # Reset non-"EyesNotFound" counter

            # Count a blink if there are 5 consecutive "EyesNotFound"
            if consecutive_eyes_not_found == 5:
                blink_count += 1
        else:
            consecutive_no_eyes_not_found += 1
            consecutive_eyes_not_found = 0  # Reset "EyesNotFound" counter

            # Reset the blink counter if 3 consecutive non-"EyesNotFound"
            if consecutive_no_eyes_not_found >= 3:
                consecutive_eyes_not_found = 0

    # Calculate total time in minutes based on a 10 ms interval per row
    total_rows = len(df)
    total_time_min = (total_rows * 10) / 60000  # Convert to minutes

    # Calculate blink rate per minute
    blink_rate_per_minute = blink_count / total_time_min if total_time_min > 0 else 0

    # Compute the average pupil diameter (ignoring NaN values)
    avg_pupil_diameter = pupil_diameter.mean()

    return blink_count, blink_rate_per_minute, avg_pupil_diameter

def process_participant(participant_number):
    # Base path for participant files
```

```python
    base_path = f'C:/Users/Mathilde/DSAI/2024-2025/Scriptie/Participants/Participant{participant_number}/'

    # List of filenames to process
    files = [
        f"{base_path}Participant{participant_number}baseline.xlsx",
        f"{base_path}Participant{participant_number}robot1.xlsx",
        f"{base_path}Participant{participant_number}robot2.xlsx",
        f"{base_path}Participant{participant_number}robot3.xlsx"
    ]

    results = {}

    for file in files:
        try:
            print(f"Processing file: {file}")
            blink_count, blink_rate_per_minute, avg_pupil_diameter = compute_blink_rate_and_pupil(file)
            results[file] = {
                "Total Blinks": blink_count,
                "Blink Rate per Minute": blink_rate_per_minute,
                "Average Pupil Diameter (mm)": avg_pupil_diameter
            }
        except FileNotFoundError as e:
            print(e)
            results[file] = "Error: File not found"

    return results

def main():
    participant_number = input("Enter the participant number: ").strip()
    results = process_participant(participant_number)

    for file, result in results.items():
        if isinstance(result, dict):
            print(f"\nResults for {file}:")
            print(f"  Total Blinks: {result['Total Blinks']}")
            print(f"  Blink Rate per Minute: {result['Blink Rate per Minute']:.2f}")
            print(f"  Average Pupil Diameter (mm): {result['Average Pupil Diameter (mm)']:.2f}")
        else:
            print(f"\nResults for {file}: {result}")

if __name__ == "__main__":
    main()
```

Listing 1: Python Script

G R File: Factor Analysis

```r
# Load necessary libraries
install.packages("xtable")
library(xtable)
library(readxl)
library(psych)
library(ggplot2)
library(reshape2)

########################
# Likeability

# Read the Excel file
data <- read_excel("C:/Users/Mathilde/DSAI/2024-2025/Scriptie/Participants/Likability.xlsx")

# Preview the data
print(head(data))

# Drop the first column (participant IDs)
factor_data <- data[, -c(1)]
print(head(factor_data))
number_of_factors <- 1  # Adjust this if needed
factor_analysis <- fa(factor_data, nfactors = number_of_factors, rotate = "varimax")
print(factor_analysis)

# Plot a scree plot to determine the optimal number of factors
scree_plot <- scree(factor_data)
print(scree_plot)

#############################
# Empathy
# Read the Excel file
data <- read_excel("C:/Users/Mathilde/DSAI/2024-2025/Scriptie/Participants/Empathy.xlsx")

# Preview the data
print(head(data))

# Drop the first column (participant IDs)
factor_data <- data[, -1]
number_of_factors <- 1  # Adjust this if needed
factor_analysis <- fa(factor_data, nfactors = number_of_factors, rotate = "varimax")
print(factor_analysis)

# Plot a scree plot to determine the optimal number of factors
scree_plot <- scree(factor_data)
print(scree_plot)

#############################
# Social functionality
```

```
# Read the Excel file
data <- read_excel("C:/Users/Mathilde/DSAI/2024-2025/Scriptie/Participants/
    Socialfunctionality.xlsx")

# Preview the data
print(head(data))

# Drop the first column (participant IDs)
factor_data <- data[, -1]
number_of_factors <- 1  # Adjust this if needed
factor_analysis <- fa(factor_data, nfactors = number_of_factors, rotate = "
    varimax")
print(factor_analysis)

# Plot a scree plot to determine the optimal number of factors
scree_plot <- scree(factor_data)
print(scree_plot)


##############################
# Anthropomorphism
# Read the Excel file
data <- read_excel("C:/Users/Mathilde/DSAI/2024-2025/Scriptie/Participants/
    Antropomorphism.xlsx")

# Preview the data
print(head(data))

# Drop the first column (participant IDs)
factor_data <- data[, -1]
number_of_factors <- 1  # Adjust this if needed
factor_analysis <- fa(factor_data, nfactors = number_of_factors, rotate = "
    varimax")
print(factor_analysis)

# Plot a scree plot to determine the optimal number of factors
scree_plot <- scree(factor_data)
print(scree_plot)

##############################
# Confirmatory FA

# Read the Excel file
data <- read_excel("C:/Users/Mathilde/DSAI/2024-2025/Scriptie/Participants/
    Allvariables.xlsx")

# Preview the data
print(head(data))

# Drop the first column (participant IDs) and drop deleted questions
factor_data <- data[, -c(1,3,10)]
factor_data <- data[, -c(1,10)]
print(head(factor_data))
number_of_factors <- 4  # Adjust this if needed
```

```r
factor_analysis <- fa(factor_data, nfactors = number_of_factors, rotate = "varimax")
print(factor_analysis)

factor_loadings <- as.data.frame(factor_analysis$loadings[, 1:number_of_factors])
colnames(factor_loadings) <- paste0("MR", 1:number_of_factors)
factor_loadings$Question <- rownames(factor_loadings)
factor_loadings$Question <- factor(factor_loadings$Question, levels = paste0("Q", sort(as.numeric(gsub("Q", "", factor_loadings$Question)))))
factor_loadings_melted <- reshape2::melt(factor_loadings, id.vars = "Question",
                                         variable.name = "Factor", value.name = "Loading")

ggplot(factor_loadings_melted, aes(x = Factor, y = Question)) +
  geom_tile(aes(fill = Loading), color = "black") +
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0, limits = c(-1, 1)) +  # Create a grid
  geom_text(aes(label = round(Loading, 2)), size = 4) +  # Add text values
  labs(title = "Factor Loadings Matrix", x = "Factors", y = "Questions") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        panel.grid = element_blank())  # Remove gridlines for a cleaner look


################################################
#Exploratory FA
# Read the Excel file
data <- read_excel("C:/Users/Mathilde/DSAI/2024-2025/Scriptie/Participants/Allvariables.xlsx")

# Preview the data
print(head(data))

# Drop the first column (participant IDs)
factor_data <- data[, -c(1)]  # Ensure the data contains only numeric variables
print(head(factor_data))

# Perform a scree plot to determine the optimal number of factors
scree(factor_data)

# Perform a parallel analysis to determine the number of factors
parallel_analysis <- fa.parallel(factor_data, fa = "fa", n.iter = 100)

# Extract and print the suggested number of factors
suggested_factors <- parallel_analysis$nfact  # This gives the number of factors suggested
print(parallel_analysis)

# Perform exploratory factor analysis
number_of_factors <- 3  # Adjust based on the analysis
efa_result <- fa(factor_data, nfactors = number_of_factors, rotate = "varimax")

# Print the EFA results
```

```r
print(efa_result)

# Visualize factor loadings
factor_loadings <- as.data.frame(efa_result$loadings[, 1:number_of_factors])
colnames(factor_loadings) <- paste0("MR", 1:number_of_factors)
factor_loadings$Question <- rownames(factor_loadings)
factor_loadings$Question <- factor(factor_loadings$Question, levels = paste0("Q", sort(as.numeric(gsub("Q", "", factor_loadings$Question)))))
factor_loadings_melted <- reshape2::melt(factor_loadings, id.vars = "Question",
                                          variable.name = "Factor", value.name =
                                            "Loading")

ggplot(factor_loadings_melted, aes(x = Factor, y = Question)) +
  geom_tile(aes(fill = Loading), color = "black") +
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0,
      limits = c(-1, 1)) +  # Create a grid
  geom_text(aes(label = round(Loading, 2)), size = 4) +  # Add text values
  labs(title = "Factor Loadings Matrix", x = "Factors", y = "Questions") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        panel.grid = element_blank())  # Remove gridlines for a cleaner look
```
Listing 2: R Script: Factor Analysis Questionnaire

H Scree Plot

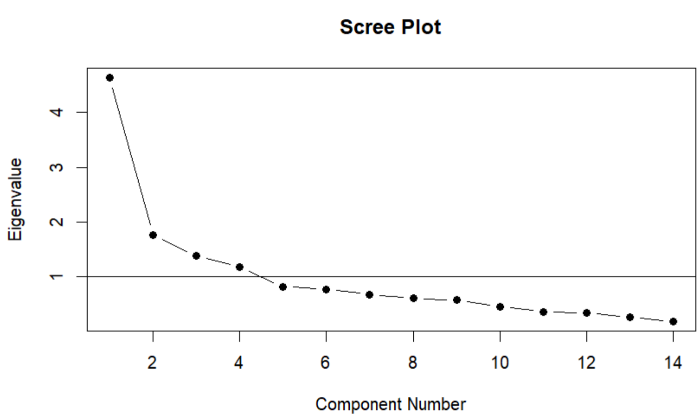

Figure 15: Scree plot showing eigenvalues of factors

I R File: Plots, ANOVA and Linear Model Analysis

```
# Install and load required libraries
install.packages("ggplot2")
install.packages("ggridges")
install.packages("ggtext")
install.packages("dplyr")
install.packages("tidyr")
install.packages("png")
install.packages("ggtext")
install.packages("lme4")
install.packages("lmerTest")
install.packages("pwr")

library(pwr)
library(ggplot2)
library(ggridges)
library(ggtext)
library(dplyr)
library(tidyr)
library(readxl)
library(png)
library(ggtext)
library(ez)
library(lme4)
library(lmerTest)

##########################################################################
# Creating plots for different variables
# Load your data

file_path <- "ParticipantsOverviewResultsPercentagestotal.xlsx"  # Replace with
    the correct path
data <- read_excel(file_path)

# Prepare data for each variable
likeability_columns <- data[, c(1, 5:7)]  # Columns for likability
print(likeability_columns)
anthropomorphism_columns <- data[, c(1, 8:10)]  # Columns for anthropomorphism
social_functionality_columns <- data[, c(1, 11:13)]  # Columns for social
   functionality
empathy_columns <- data[, c(1, 14:16)]  # Columns for empathy

# Define labels with images
labels <- c(
  "NAO" = "<img src='img/NAO.png' height='25' /><br>*NAO*",
  "Pepper" = "<img src='img/pepper.png' height='25' /><br>*Pepper*",
  "AlphaMini" = "<img src='img/alphamini.png' height='25' /><br>*AlphaMini*"
)

# Create a function for ridge plots with image labels
create_ridge_plot <- function(data_subset, variable_name) {
  # Reshape data into long format
  data_long <- pivot_longer(
```

```r
    data_subset,
    cols = -1,
    names_to = "Robot",
    values_to = "Score"
  )

  # Clean the Robot names (remove prefixes like 'Empathy', 'Likeability', etc.)
  data_long <- data_long %>%
    mutate(Robot = gsub(".*(?=NAO|Pepper|AlphaMini)", "", Robot, perl = TRUE)) # Extract only robot names

  # Generate the ridge plot
  ggplot(data_long, aes(x = Score, y = factor(Robot, levels = c("AlphaMini", "NAO", "Pepper")), fill = Robot)) +
    coord_cartesian(clip = "off") +
    geom_density_ridges(alpha = 0.8, quantile_lines = TRUE, quantiles = 2, size = 0.5) +
    scale_x_continuous(
      limits = c(0, 100),                # Ensure x-axis goes from 0 to 100
      breaks = seq(0, 100, by = 25),     # Add ticks at 0, 25, 50, 75, 100
      labels = seq(0, 100, by = 25)      # Customize x-axis labels
    ) +
    scale_y_discrete(labels = labels, expand = c(.07, .07)) +
    theme_classic() +
    labs(
      title = paste(variable_name, "Scores by Robot"),
      x = variable_name,
      y = "Robot"
    ) +
    theme(
      axis.text.y = element_markdown(),
      legend.position = "none"
    )
}

# Create plots
likeability_plot <- create_ridge_plot(likeability_columns, "Likeability Score")
anthropomorphism_plot <- create_ridge_plot(anthropomorphism_columns, "Anthropomorphism Score")
social_functionality_plot <- create_ridge_plot(social_functionality_columns, "Social Functionality Score")
empathy_plot <- create_ridge_plot(empathy_columns, "Empathy Score")


# Print the plots
print(likeability_plot, height = 12, width = 6)
print(anthropomorphism_plot, height = 12, width = 6)
print(social_functionality_plot, height = 12, width = 6)
print(empathy_plot, height = 12, width = 6)

# Optional: Save the plots
#ggsave("likeability_plot.pdf", likeability_plot, height = 12, width = 6)
#ggsave("anthropomorphism_plot.pdf", anthropomorphism_plot, height = 12, width = 6)
```

```r
#ggsave("social_functionality_plot.pdf", social_functionality_plot, height = 12, width = 6)
#ggsave("empathy_plot.pdf", empathy_plot, height = 12, width = 6)

################################################################################
#In this part ANOVA will be done for all variables against robottype

# Read the Excel file
file_path_without23 <- "ParticipantsOverviewResultsPercentageswithout23.xlsx"
full_df_without23 <- read_excel(file_path_without23, sheet = "Blad1")

# Perform ezANOVA for HRMean
rtype_hrmean_anova <- ezANOVA(
  data = full_df_without23,
  dv = HRMean,
  wid = ResponseId,
  within = RobotType
)
print(rtype_hrmean_anova)

# Perform ezANOVA for RMSSD
rtype_rmssd_anova <- ezANOVA(
  data = full_df_without23,
  dv = RMSSD,
  wid = ResponseId,
  within = RobotType
)
print(rtype_rmssd_anova)

# Perform ezANOVA for LFHFRatio
rtype_lfhratio_anova <- ezANOVA(
  data = full_df_without23,
  dv = LFHFRatio,
  wid = ResponseId,
  within = RobotType
)
print(rtype_lfhratio_anova)

##########
#With participant 23
# Read the Excel file
file_path <- "ParticipantsOverviewResultsPercentageswith23.xlsx"
full_df <- read_excel(file_path, sheet = "Blad1")

# Perform ezANOVA for DictatorGame
rtype_dictatorgame_anova <- ezANOVA(
  data = full_df,
  dv = DictatorGameFull,
  wid = ResponseId,
  within = RobotType
)
print(rtype_dictatorgame_anova)

# Perform ezANOVA for Likeability
```

```
rtype_likeability_anova <- ezANOVA(
  data = full_df,
  dv = Likeability,
  wid = ResponseId,
  within = RobotType
)
print(rtype_likeability_anova)

# Perform ezANOVA for Anthropomorphism
rtype_anthropomorphism_anova <- ezANOVA(
  data = full_df,
  dv = Anthropomorphism,
  wid = ResponseId,
  within = RobotType
)
print(rtype_anthropomorphism_anova)

# Perform ezANOVA for Social Functionality
rtype_socialfunctionality_anova <- ezANOVA(
  data = full_df,
  dv = SocialFunctionality,
  wid = ResponseId,
  within = RobotType
)
print(rtype_socialfunctionality_anova)

# Perform ezANOVA for Empathy
rtype_empathy_anova <- ezANOVA(
  data = full_df,
  dv = Empathy,
  wid = ResponseId,
  within = RobotType
)
print(rtype_empathy_anova)

# Perform ezANOVA for Eye Blink Rate
rtype_eyeblinkrate_anova <- ezANOVA(
  data = full_df,
  dv = EyeBlinkRate,
  wid = ResponseId,
  within = RobotType
)
print(rtype_eyeblinkrate_anova)

# Perform ezANOVA for Pupil Diameter
rtype_pupildiameter_anova <- ezANOVA(
  data = full_df,
  dv = PupilDiameter,
  wid = ResponseId,
  within = RobotType
)
print(rtype_pupildiameter_anova)

########
```

```
#Plots for DG and EBR
# Summarize data for Dictator Game (mean, SD, and CI)
dictatorgame_summary <- full_df %>%
  group_by(RobotType) %>%
  summarise(
    mean = mean(DictatorGameFull, na.rm = TRUE),
    sd = sd(DictatorGameFull, na.rm = TRUE),
    n = n()
  ) %>%
  mutate(
    se = sd / sqrt(n),
    ci_lower = mean - qt(0.975, df = n - 1) * se,
    ci_upper = mean + qt(0.975, df = n - 1) * se
  )

# Summarize data for Eye Blink Rate (mean, SD, and CI)
eyeblinkrate_summary <- full_df %>%
  group_by(RobotType) %>%
  summarise(
    mean = mean(EyeBlinkRate, na.rm = TRUE),
    sd = sd(EyeBlinkRate, na.rm = TRUE),
    n = n()
  ) %>%
  mutate(
    se = sd / sqrt(n),
    ci_lower = mean - qt(0.975, df = n - 1) * se,
    ci_upper = mean + qt(0.975, df = n - 1) * se
  )

# Plot for Dictator Game
ggplot(dictatorgame_summary, aes(x = RobotType, y = mean, fill = RobotType)) +
  geom_bar(stat = "identity", position = position_dodge(), color = "black") +
  geom_errorbar(aes(ymin = ci_lower, ymax = ci_upper), width = 0.2, position =
      position_dodge(0.9)) +
  labs(
    title = "Dictator Game Offers by Robot Type",
    x = "Robot Type",
    y = "Average Offer (  )"
  ) +
  theme_minimal()

# Plot for Eye Blink Rate
ggplot(eyeblinkrate_summary, aes(x = RobotType, y = mean, fill = RobotType)) +
  geom_bar(stat = "identity", position = position_dodge(), color = "black") +
  geom_errorbar(aes(ymin = ci_lower, ymax = ci_upper), width = 0.2, position =
      position_dodge(0.9)) +
  labs(
    title = "Eye Blink Rate by Robot Type",
    x = "Robot Type",
    y = "Average Eye Blink Rate (% of Baseline)"
  ) +
  theme_minimal()

############################################
```

```
#In this part the linear model analysis will be done on all variables against the dictator game offer

# HRMean and DictatorGameMinus23
mixed.lmer <- lmer(DictatorGameMinus23 ~ HRMean + (HRMean|RobotType) + (HRMean|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# RMSSD and DictatorGameMinus23
mixed.lmer <- lmer(DictatorGameMinus23 ~ RMSSD + (RMSSD|RobotType) + (RMSSD|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# LFHFRatio and DictatorGameMinus23
mixed.lmer <- lmer(DictatorGameMinus23 ~ LFHFRatio + (LFHFRatio|RobotType) + (LFHFRatio|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# Likeability and DictatorGameFull
mixed.lmer <- lmer(DictatorGameFull ~ Likeability + (Likeability|RobotType) + (Likeability|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# Anthropomorphism and DictatorGameFull
mixed.lmer <- lmer(DictatorGameFull ~ Anthropomorphism + (Anthropomorphism|RobotType) + (Anthropomorphism|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# Test for a significant non-linear correlation in anthropomorphism
# pepper_data <- subset(full_df, RobotType == "Pepper")
# model <- lm(DictatorGameFull ~ Anthropomorphism + I(Anthropomorphism^2), data = pepper_data)
# summary(model)

# SocialFunctionality and DictatorGameFull
mixed.lmer <- lmer(DictatorGameFull ~ SocialFunctionality + (SocialFunctionality|RobotType) + (SocialFunctionality|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# Empathy and DictatorGameFull
mixed.lmer <- lmer(DictatorGameFull ~ Empathy + (Empathy|RobotType) + (Empathy|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# EyeBlinkRate and DictatorGameFull
```

```
mixed.lmer <- lmer(DictatorGameFull ~ EyeBlinkRate + (EyeBlinkRate|RobotType) + (EyeBlinkRate|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# PupilDiameter and DictatorGameFull
mixed.lmer <- lmer(DictatorGameFull ~ PupilDiameter + (PupilDiameter|RobotType) + (PupilDiameter|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

################################################
# HRMean x Subjective Perceptions
mixed.lmer <- lmer(HRMean ~ Likeability + (Likeability|RobotType) + (Likeability|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(HRMean ~ SocialFunctionality + (SocialFunctionality|RobotType) + (SocialFunctionality|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(HRMean ~ Anthropomorphism + (Anthropomorphism|RobotType) + (Anthropomorphism|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(HRMean ~ Empathy + (Empathy|RobotType) + (Empathy|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# RMSSD x Subjective Perceptions
mixed.lmer <- lmer(RMSSD ~ Likeability + (Likeability|RobotType) + (Likeability|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(RMSSD ~ SocialFunctionality + (SocialFunctionality|RobotType) + (SocialFunctionality|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(RMSSD ~ Anthropomorphism + (Anthropomorphism|RobotType) + (Anthropomorphism|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(RMSSD ~ Empathy + (Empathy|RobotType) + (Empathy|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)
```

```
# LFHFRatio x Subjective Perceptions
mixed.lmer <- lmer(LFHFRatio ~ Likeability + (Likeability|RobotType) + (
    Likeability|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(LFHFRatio ~ SocialFunctionality + (SocialFunctionality|
    RobotType) + (SocialFunctionality|ResponseId), data = full_df_without23,
    REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(LFHFRatio ~ Anthropomorphism + (Anthropomorphism|RobotType)
    + (Anthropomorphism|ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(LFHFRatio ~ Empathy + (Empathy|RobotType) + (Empathy|
    ResponseId), data = full_df_without23, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# EyeBlinkRate x Subjective Perceptions
mixed.lmer <- lmer(EyeBlinkRate ~ Likeability + (Likeability|RobotType) + (
    Likeability|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(EyeBlinkRate ~ SocialFunctionality + (SocialFunctionality|
    RobotType) + (SocialFunctionality|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(EyeBlinkRate ~ Anthropomorphism + (Anthropomorphism|
    RobotType) + (Anthropomorphism|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(EyeBlinkRate ~ Empathy + (Empathy|RobotType) + (Empathy|
    ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

# PupilDiameter x Subjective Perceptions
mixed.lmer <- lmer(PupilDiameter ~ Likeability + (Likeability|RobotType) + (
    Likeability|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(PupilDiameter ~ SocialFunctionality + (SocialFunctionality|
    RobotType) + (SocialFunctionality|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)
```

```r

mixed.lmer <- lmer(PupilDiameter ~ Anthropomorphism + (Anthropomorphism|
    RobotType) + (Anthropomorphism|ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

mixed.lmer <- lmer(PupilDiameter ~ Empathy + (Empathy|RobotType) + (Empathy|
    ResponseId), data = full_df, REML=TRUE)
summary(mixed.lmer)
confint(mixed.lmer)

#############################
# Power Analysis

# Power ANOVA
k <- 3                 # 3 different robots
effect_size <- 0.25  # Medium effect size
sample_size <- 31
alpha <- 0.05

power <- pwr.anova.test(k = k, n = sample_size / k, f = sqrt(effect_size / (1 -
     effect_size)), sig.level = alpha)$power
cat("The power of the ANOVA is:", power, "\n")
```
Listing 3: R Script: Thesis Analysis (Plots, ANOVA, and Linear Models)