TU WIEN Informatics

Reaktionen auf und Interaktionen mit einem nonverbalen Roboter

Eine Fallstudie über die Interpretation und Wahrnehmung von Menschen bei der Interaktion mit einem nicht-humanoiden und nicht-verbalen Roboter

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medieninformatik und Visual Computing

eingereicht von

Alin Munteanu-Calen, BSc
Matrikelnummer 01528098

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Assistant Prof. Dr. Astrid Weiss

Wien, 24. Februar 2023

Alin Munteanu-Calen | Astrid Weiss

Reactions to and Interactions with a Non-verbal Robot

A Case Study on People's Interpretation of and Perception of Interacting with a Non-Humanoid and Non-Verbal Robot

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Media Informatics and Visual Computing

by

Alin Munteanu-Calen, BSc
Registration Number 01528098

to the Faculty of Informatics

at the TU Wien

Advisor: Assistant Prof. Dr. Astrid Weiss

Vienna, 24th February, 2023

Alin Munteanu-Calen   Astrid Weiss

Erklärung zur Verfassung der Arbeit

Alin Munteanu-Calen, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 23. Februar 2023

Alin Munteanu-Calen

Danksagung

Ich möchte meiner Betreuerin für die unermüdliche Unterstützung und Hilfe danken, die sie mir trotz ihrer vielen anderen Aufgaben und vollen Terminkalender zuteil werden ließ. Außerdem möchte ich mich bei allen Versuchsteilnehmer*innen für ihre Zeit, ihr Engagement und ihr Feedback bedanken. Ohne den Rückhalt meiner Familie wäre ein solches Unterfangen jedoch kaum zu bewältigen gewesen.

Acknowledgement

I want to thank my supervisor for the unceasing support and help she has shown and provided me in spite of her many other tasks and busy schedules. Furthermore, I wish to reach out to all of the experiment participants and thank them for their time, involvement and feedback. Nevertheless, without the backing of my family, such an endeavour would have been hardly achievable.

Kurzfassung

Humanoide soziale Roboter werden in Bereichen wie Gesundheitswesen, Einkaufen und Telekommunikation erforscht und eingesetzt. Sie sollen das menschliche Verhalten, die Sprache und die Art der Interaktion bis zu einem gewissen Grad nachahmen, um besser in die Rolle zu passen, für die sie gedacht sind. Andere erfolgreiche Beispiele sozialer Roboter wurden gestaltet, indem das Verhalten von Tieren und ausgewählter sozialer Hinweise imitiert wurde. Das Maß an Komfort, Vertrauen, positiver Wahrnehmung und Zuverlässigkeit, das ein Roboter auf eine Person ausstrahlt, hängt mit dessen Aussehen und Interaktionsverhalten zusammen. Abhängig von der Erfahrung, der Vergangenheit und dem Wissen des Menschen sowie der biologischen Klassifizierung werden diese Werte auch unabhängig vom Roboter beeinflusst.
Dies hat zu vielen Studien über menschenähnliche Mensch-Roboter-Interaktion (MRI) geführt, die nur schwer mit hoher Wiedergabetreue durchgeführt werden können und deren Ergebnisse nicht immer reproduzierbar sind. Nicht-humanoide Roboter bieten mehr Möglichkeiten für Innovationen und roboterspezifische Kommunikations- und Interaktionsmerkmale, da sie nicht die Verhaltensmuster von Lebewesen imitieren. Was die nonverbale Kommunikation anbelangt, verwenden diese soziale Hinweisreize, die vom motorischen Möglichkeitsraum, Posen und Blickbewegungen oder Licht- und Tonsignalen abhängen, um wahrnehmbar, lesbar, vorhersehbar und sozial akzeptabel zu sein. Bei so viel Interpretationsspielraum erweist sich die Replikation von Studien mit nichthumanoiden Robotern in nonverbalen Interaktionsszenarien jedoch ebenfalls als herausfordernd. Motivation und Ziel dieser Diplomarbeit ist es daher, die beiden bisher vorgestellten Hauptthemen, nämlich die nonverbale Kommunikation und Interaktion zwischen Menschen und nicht-humanoiden Robotern, unter Berücksichtigung der Grenzen bisheriger Studien und des Phänomens der Replikationskrise in diesem Forschungsbereich zu behandeln, zu bewerten und zu erweitern.
Die Hauptanliegen dieser Forschungsarbeit besteht darin, eine frühere Laborstudie zu wiederholen und deren Design und Ergebnisse zu erweitern. Im ursprünglichen Experiment wurden junge Erwachsene gebeten, mit nicht-humanoiden Robotern, genannt Sphero, zu interagieren, die nur über Lichter, Geräusche und Drehbewegungen verfügen und sich den Teilnehmer*innen verständlich machen mussten, wenn sie diese aufforderten, ihnen zu einer Belohnung zu folgen. Dieses ursprüngliche Szenario wird erweitert, indem auch ältere Erwachsene in das Experiment einbezogen und die Ergebnisse verglichen werden, um Verzerrungen zu verringern. Um ein breiteres Spektrum der Mensch-Roboter-Interaktion mit einem nonverbalen nicht-humanoiden Roboter zu erforschen, wurde außerdem ein zweites Szenario integriert, in dem die Reaktionen der Teilnehmer*innen auf einen plötzlich zu Musik tanzenden Roboter untersucht werden.
Die wichtigsten Ergebnisse dieser Arbeit sind zweifacher Natur. Einerseits zeigt sie, dass es unmöglich war, während des ersten Szenarios deckungsgleiche Ergebnisse zwischen den jungen und älteren Teilnehmer*innen zu erzielen, wenn man die Einschränkungen durch das Ursprungsmaterial, die Umgebung und der Teilnehmer*innen berücksichtigt, was die Replikationskrise als ein existierendes und relevantes Phänomen in der MRI bestätigt. Auf der anderen Seite beleuchten die Ergebnisse des zusätzlichen zweiten Szenarios, in dem der Roboter tanzt, die Einstellung der Teilnehmer*innen gegenüber dem Roboter als companion (Begleiter) und nicht als care-giver (Betreuer) und zeigen die Auswirkungen der Neuartigkeit und ihren Einfluss auf die verschiedenen Altersgruppen sowie die emotionale Reaktion, die zwischen den Generationen variieren; abgesehen von den Unterschieden zwischen den Individuen selbst. Darüber hinaus wurden etablierte und innovative Wege der nonverbalen Kommunikation, die einem nicht-humanoiden Roboter zugeschrieben werden, erforscht und getestet und bieten eine Grundlage für zukünftige Forschungen zu nicht-humanoiden Roboter im Allgemeinen.
Diese Arbeit ist insofern relevant, als sie den Bereich der nonverbalen Kommunikation unter Berücksichtigung von nicht-humanoiden Robotern mit eingeschränkten mechanischen und funktionalen Fähigkeiten untersucht und erweitert. Darüber hinaus hat sie einen Schritt in Richtung der Bewertung, Bearbeitung und Lösung der Replikationskrise in der MRI getan, indem sie ihr Auftreten in einem Studienbeispiel untersucht hat. Nicht zuletzt trägt die Arbeit dazu bei, den menschlichen Interaktionspartner im Austausch mit nicht-humanoiden sozialen Robotern besser zu verstehen, indem sie Erkenntnisse über das Design und Verhalten solcher Roboter generierte, sowie gängige Missverständnisse und unpräzise Aussagen über Personen, die nach Alter, Geschlecht und Erfahrung klassifiziert werden, reflektierte.

Abstract

Humanoid social robots have been researched and deployed in fields such as healthcare, shopping, and telecommunication. These are all meant to mimic human behaviour, speech and interaction type to some degree in order to better fit in the role they are taking on. Other successful examples of social robots have been constructed by copying animal behaviour and selected social cues instead. The levels of comfort, trust, positive perception, and reliability a robot emits to a person is correlated with its appearance and actions. Depending on the human's experience, past, and knowledge, as well as biological classification, the levels are furthermore impacted independently of the robot.

This has lead to many taken-for-granted studies on human-like HRI, which are hard to replicate at a high fidelity and findings are not always reproducible. Moreover, non-humanoid robots have more options for innovation and robot-specific communication and interaction features design, as they are freed of the constraint requiring them to copy existing living beings. Additionally, considering non-verbal communication, they are forced to rely on social cues depending on motor skills, poses and gaze or light and sound signaling to make themselves perceivable, legible, predictable and socially acceptable. Otherwise, having this much freedom of interpretation, replicating studies involving non-humanoid robots in non-verbal constrained scenarios proves to be very delicate. The motivation and scope of this master thesis is therefore to address, evaluate, and expand upon the two main topics presented so far, namely the non-verbal communication and interaction between humans and non-humanoid robots while considering the limitations of past studies and the phenomena of the replication crisis in this field of study.

The main work of this research consists of replicating a previous laboratory study and expanding upon its design and findings. In the original experiment, young adults were asked to interact with non-humanoid robots, called Sphero, which, having only lights, sounds and rotatory movements at their disposal, had to make themselves understood to the participants when asking them to follow it to a reward. Furthermore, this scenario is expanded upon by including also older adults in the experiment and comparing the results to reduce bias. Moreover, to explore a broader range of human-robot interaction (HRI) with a non-verbal non-humanoid robot, a second scenario was also integrated, where the reactions of participants to the robot suddenly bursting out in a dance to music are sought after.

The key findings of this thesis are of a dual nature. On the one hand, it shows that obtaining the close to the same results with both the young and the old participants individually and combined, during the first scenario run-through of the original study, was impossible, considering the limitations of the materials, environment and personas.Thus, the replication crisis is supported as an existing and occurring phenomena in HRI. On the other hand, the results from the added second scenario incorporating dancing further shed light on the attitude of participants towards the robot as a companion versus a care-taker, prove the effect of novelty and its influence on the different age groups, as well as the emotional response and implication, varying from one generation to another, beyond differences between individuals themselves. Moreover, established and innovative ways of non-verbal communication attributed to a non-humanoid robot were researched, tested and offer a basis for future research considering non-humanoid robots at large.

This thesis is relevant as it investigated and innovated the area of non-verbal communication, considering non-humanoid robots with reduced and constricted mechanical and functional capabilities. Furthermore, it took one step into the direction of assessing, processing and solving the replication crisis in HRI, by exploring its occurrence in a consecrated study example. Lastly, but not least, it helped to better understand the human interlocutor in interaction contexts involving non-humanoid social robots, validating learnings on the design of robots and their behaviour, as well as on the common misconceptions and staple statements regarding individuals classified by age, gender, and experience.

Schlüsselwörter

Mensch-Roboter-Interaktion, Nonverbale Kommunikation, Verhalten, nicht-humanoide Roboter, Wahrnehmung, Replikationskrise, Altersunterschiede, Tanz, HRI-Design, Sphero

Keywords

Human-Robot-Interaction, Non-Verbal Communication, Behavior, Non-Humanoid Robots, Perception, Replication Crisis, Age Differences, Dance, HRI-Design, Sphero

Contents

Chapter 1

Introduction

Since the Czech term "robot" was coined for the first time, robots, and above all social robots, have taken over both the fiction and non-fiction fascination of humans, making their appearance in books, movies and games, as well as proving to be a driving force in art, sciences and technology. The later one particularly gained more traction in the last years, due to the increasingly aging population and the call for non-humanoid robots in care. [1]

Often presented as human-like, physically, cognitively, as well as emotionally; many studies have shown that robots can be perceived as anthropomorphized objects, surpassing our physical and moral constraints, aiming at perfection, the greater good and reasoning. [2] There are many indications that humans have a tendency to affiliate with nature, and with other living beings, including non-human species. [3] By anthropomorphizing robots, humans apply social norms and expectations to technical devices, treating robots as social actors. [4] Our mind perception alters and we form emotional attachment to robots. We project our desires onto them and that leads us to questioning ourselves who we are - creations or creators. The more a robot looks, behaves, and feels human-like the more unrest and unsettling feeling it will emit, according to the Uncanny Valley. [2]

However, robots come in many different shapes and sizes, with drastically varying designs and capabilities suited to their tasks needs. In fact, the very design, appearance, and handling of robots is what influences the perception of the human interlocutors, labelling them as more or less agreeable, reliable, trustworthy, and comfortable to be around. [5] Non-humanoid robots are often more machine-like in appearance, hence people are likely to have different expectations towards them than towards other humanoid robots. Most such robots are constrained in their physical aspect, possessing communication modalities that are significantly more limited than those employed by humans, considering also the fact that these should not interfere with their tasks and purposes. [6]

The robot ideals presented in science fiction expect them to be able to speak to us and interact with us on a human-like level, using elaborate language and a proficient vocabulary. Human-human interactions and communication are very complex since they are generally symmetric, develop since birth and are based on the use of language. [7]

Yet, robots mimicking human behaviour will never be perfect [7], so why do robots have to copy human behaviour and capabilities in the first place? [8] Around 60% of human communication is non-verbal, making a translation of it to robots crucial for an intuitive engagement with humans. [9]

Robots are expected to execute various tasks in our daily life, most of which are difficult to be perfectly automated; however, they could be executed in cooperation with a human. [10] Collaborative tasks between the human and robot rely on each collaborator's ability to effectively convey their mental state while accurately estimating and interpreting their partner's knowledge, intent, and actions. [11] Thus, robots need not be only predictable, but also legible in their actions and intents. [12] [13] [10]

Non-humanoid and non-verbal robots are required to acquire and implement new and innovative means of communication to successfully integrate amongst us. The robot communication is constrained by the limited motor capabilities of the robot, but it can also take advantage of expressive options that are not available to humans. [14] Ideally, they would also borrow from animals, copy these biologically, to emit social signs to ease communication translation, e.g. mimicking raised ears to signal attention and alert or surprise, confusion. [15] Robots can just like animals express emotions through their bodies. Humans are sensitive to robot movement since they are psychologically affected by it. [16] [17]

Previous research has proven that the display of emotions is sufficient to generate empathy-like reactions and user acceptance as social agents, easing the integration of the robot in the social contexts. [18] Unfortunately, very little prior work has addressed the opportunities and challenges of creating an emotionally expressive body language for non-humanoid robots, with the aspect of non-verbal communication stacking up the difficulty. [19] It is not yet clear enough, if robots are as effective as humans in communicating this way. [20]

The area of human-robot interaction (HRI), is a young and interdisciplinary field of research, looking to understand the different interconnected aspects of how humans perceive, engage and are to be approached by robots. [21] Researching robots and people's reactions towards them can also offer us insights into human behaviour. The deployment of social robots in care of the elderly or demented, for children with autism or as telerobotics shows that robots are treated similarly to other humans on a regular basis (anthropomorphism), with service robots seeing potential in education, as shopping mall guides and many other applications. [2]

Many studies have been carried out in HRI research, studies which then have been taken for granted and used as a staple in referencing to and basing future work on. This is where the replication crisis revolution in psychology also affects HRI. [22] There are close ties between HRI and social and cognitive psychology - replication crisis is applicable in both fields and translates as failing to achieve the same results as the original study. [23]

The replication crisis in social sciences, particularly in psychology, has indicated

that between 1/2 and 2/3 of past studies cannot be replicated faithfully, yielding the exact same accurate results, due to publication bias, socio-demographic changes in the population and the impossibility to reproduce the exact same context, circumstances and persons with their given states of mind. [24] [23]
Nonetheless, replication is vital to science. Scientists must avoid placing too much faith in single studies, as they are particularly vulnerable overlooking undiscovered errors. A call has been made, to form a collective intention to replicate results across a range of different robots in HRI. [25]
Just to give an example, to contradict my earlier referenced statement regarding the trust we place in a robot based of its appearance, another paper concluded that the robot appearance, as well as the participant gender, make no difference in the distance people keep towards robots, it rather being the novelty factor of the interaction. [23]

The motivation and scope of this master thesis is therefor to address, evaluate, and expand upon the two main topics presented so far, namely the non-verbal communication and interaction between humans and non-humanoid robots while considering the limitations of past studies and the phenomena of replication crisis in this field of study.
To do so, a previous study employing a Sphero SPRK+ ball-robot, shortly named Sphero in this thesis, in an attempt to non-verbally communicate to young participants to follow it, as well as related literature and work, are replicated, analysed and described in detail in the following chapters. The results and findings are analysed empirically and interpreted, aiming at validating the original conclusion of the previous study. Whether exact, similar or different, the outcome determines whether the first study is a trustworthy one to build further research upon, as well as if the replication crisis is that common in the field of human-centered human-robot interaction. The pictures presented in *Figure 1.1* and *Figure 1.2* portray the original Sphero robots used during the original study, as well as a glimpse of the test lab environment setup.

Beyond replicating the proposed study done by Faria et. al. [26] and analyzing

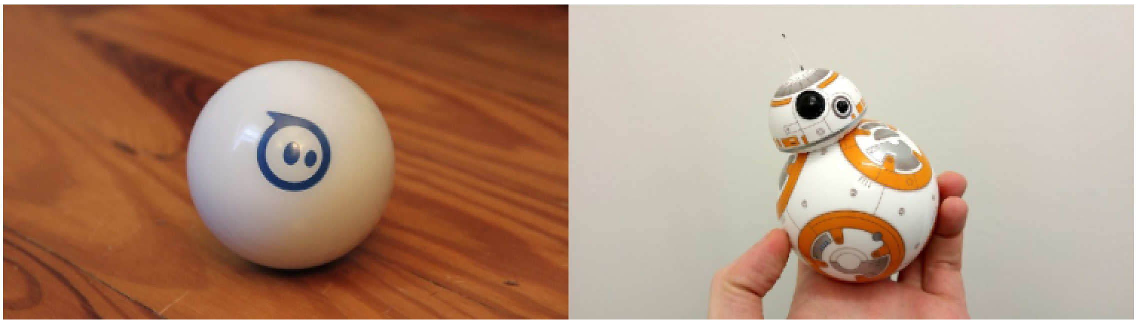


Figure 1.1: The Sphero SPRK+ (left) and BB8 (right) from the original study [26]

the outcome, the work in the frame of this master thesis aims to expand the aforementioned research by an additional scenario involving Sphero dancing, where the careful design of non-verbal communication proves crucial in establishing a legible

Figure 1.2: A picture taken of a corner of the original test lab environment [26]

interaction between the robot and the participant and a clear, positive interpretation of the latter one. Last but not least, the study is also expanded to include an older participant group as well, to better form an image of HRI between younger and older generations.

To offer a clear overview over the work presented in this master thesis, as well as a structure for future work or replications of this research, the thesis consists of the following chapters:

1. **Related Work**: This chapter discusses the related work regarding the topics from the introduction, going into more detail on the correlated topics of human-pets interaction, non-humanoid robots, non-verbal communication and replication crisis exemplified in the human-robot interaction field, as well as highlighting the interdisciplinary nature of this research. Each main field of interest is gathered under its name-like sub-chapter.
   In order to achieve this, a summary of the articles, papers, conference papers and miscellaneous links and sources are presented; all of which are listed in the bibliography.
   The human-robot interaction in particular is explained in detail from the perspective of a human-centered approach, social robotics, empathy, trust, social acceptance, and robot design, perception, behaviour and capabilities, as these are important for grounding this thesis.

2. **Methodological Approach**: This chapter showcases the research questions as well as the procedures undertaken in order to formulate answers and solutions for each question individually. Moreover, the respective methods employed in the research of and for this thesis as well as in the replication of the study are enunciated and conveyed in an appropriate manner.
   Furthermore, the study plan and the information on participants are also elaborately included, not neglecting the thought process behind the reiteration of the original study.

This chapter, moreover, contains a detailed explanation on participants, materials, scoring, the preparations and work that led to the planned experiments, e.g. how the questionnaires, interviews, experiment lab room and scenarios were chosen, how the decoration and robot interaction was designed or how the participants have been recruited, instructed and included in the experiments.
All these aspects are documented entirely from the initial version and iterations leading up to the final one respectively.

3. **Findings**: A clear overview on the results themselves, both numerical quantitative as well as interpretative qualitative ones, is comprised in the body of this chapter of the thesis.
Accompanying these are supporting graphics and written statements of how each experiment unfolded, what similarities and contrasts have been observed, as well as an aggregation, classification and categorization of the findings obtained.

4. **Discussion**: The evaluation of the findings, considering the thorough analysis based on the methods presented in the previous chapter, are exposed and debated accordingly.
This chapter is meant to unify all the findings gained through the analysis of the results from the questionnaires, the observations and the unstructured interviews, and by this to extract answers, solutions and conclusions for each of the research questions. Thus, of much more importance hereby is the reflection of the results in the light of related work. Namely, what of the findings goes in line with previous work or contradicts it.
Future work and shortcomings of the limitations of the study replication, as well as lessons learned out of this undertaking, are also mentioned and discussed here.

5. **Conclusion**: The conclusion summarizes everything which has been conducted within this thesis. It aims at offering a concluding result of the research carried out.
Thoughts, limitations, improvements and future research directions, expanding on or spanning out of this thesis are also presented herein.

Other chapters building up this thesis are listed as follows:

- *Abstract*: A general overview synthesizing the broad area and specific focus of this work, the "but/however" gap that set up the research, the statement of the methods used, the initial findings based on assumptions made for each research question, and the relevance of the entire endeavour.
- *Keywords*: An enumeration of the most defining aspects, as well as most appropriate search identifier terms.
- *Acknowledgements*: A small thank you from the researcher.
- *Table of Contents*: An overview of all chapters and sections in this thesis.

- *References*: The bibliography containing all documents used for the research of the topics related to this work and as a basis for the literature review in *Related Work*.
- *Appendix*: Contains all templates, documents used and resulting files in the recruitment of participants, analysis and transcripts of their filmed interaction and reaction during the experiment, and of the questionnaires and interviews at the end of it.

Chapter 2

Related Work

The scope of this chapter is to offer a compact, detailed and extensive overview on the topics touched by the presented research, such as non-humanoid robots, non-verbal behavior and communication, human-robot, human-animal and human-human interaction, pets, dancing, cognition, care, healthcare and social robots, and the replication crisis. The literature review is the first method, naturally employed when confronted with a new subject to learn or topic to research. Many articles, papers, books, websites and miscellaneous have been investigated, analysed and summarised here, based of linked appliances and theories which influence or are of relevance for the study replication and its expansion.
Based on all the references collected for this work, the *Related Work* chapter aims to offer the thesis background information, arguments and proof related to interaction and non-verbal communication with non-humanoid robots, as well as a short overview of the replication crisis. This helps preparing the reader for the applied knowledge and methods in the replication study by laying the foundation, providing sufficient information to understand both the domain of human-robot-interaction and non-verbal communication and the setting, unfolding and goal of the lab experiment. Furthermore, this information is expanded upon by offering interconnected details and knowledge gains from robot pets, social robots, robots in healthcare as well as robot dance partners and therapy.

To offer a compact, detailed and extensive overview on the topics of non-humanoid robots, non-verbal communication and the replication crisis, many sources have been investigated. The process of initial broad research, narrowed down through filters afterwards to linked appliances and theories which will have found their way in the study replication and expansion, is roughly portrayed in *Figure 2.1*.
The entire literature research presented here serves the purpose of detailing all aspects involved in the study replication hereby, namely that of Faria et al. in "Follow me: Communicating Intentions with a Spherical Robot". [26]

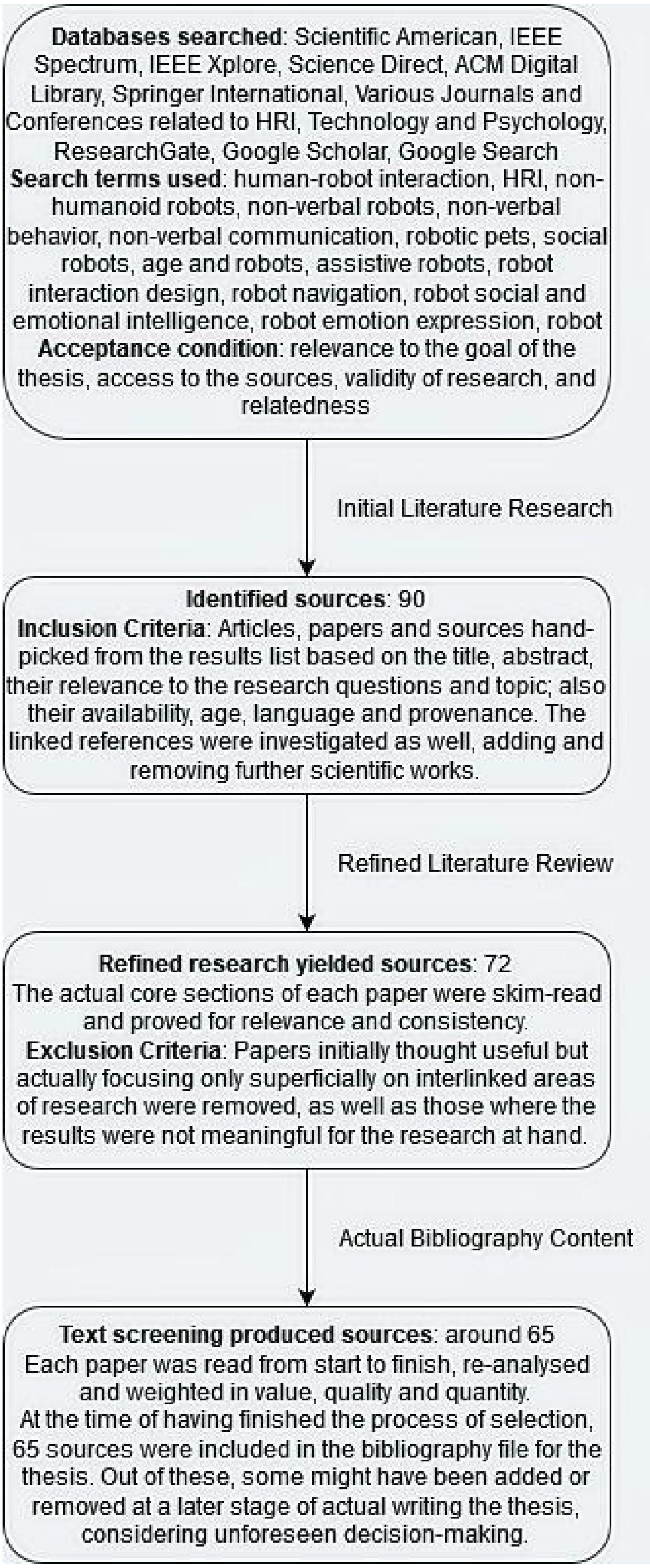


Figure 2.1: Literature Research Process

2.1 Non-verbal Behavior and Communication

Person–robot communication, according to Libin et al., is based upon three basic principles, namely interactivity, equifinality, and multimodality. [27] Around 60% of human communication is non-verbal, which is crucial for robots' engagement with humans as it allows for intuitive interaction between humans and robots. [9]
Furthermore, from Libin et al. perspective, a robots has to adjust its form of communication, and social and emotional intelligence [28], based on whether it plays the role of "[...] an interactive device for training and development of certain skills; a "smart"' tool for guided physical and mental stimulation; a human companion in unusual situations and difficult life circumstances [...]". [27, p. 1792]
Following up, from a robopsychologist's point of view, as in a mediator in human-to-robot communication, artificial entities are to be divided into two assisting robots groups, "[...] which are oriented toward industrial, military, research, medical, and service activities [...]" [27, p. 1789], and interactive stimulation robots, "[...] which are designed for social, educational, rehabilitation, therapeutic, and entertainment purposes [...]".
Thus, the relationships between humans and robots are to be placed into a socio-psychological context instead of a technological one. [27]
There are numerous communication challenges associated with required levels of coordination and collaboration between humans and robots, so that the latter are not only capable of navigating and manipulating in a given environments but also coordinating and collaborating with humans as well [28], "[...] The building of a thorough understanding of how different design choices affect the complex variety of human responses towards robots, such as cognitive, affective, emotional, and relational ones, is a broader challenge than initially expected by the research community. [...]" [22, p. 19]
The activity context and environment condition determines a person's ability to perceive the meaning of a robot's behaviour. This can not be understood separate from the task the robot is performing, or the collaborative activity in which it is engaged. [19] Since non-humanoid robots are often more machine-like in appearance, people are likely to have different expectations than of humanoid robots. [28] However, robots can also operate as autonomous, intelligent agents that act in surprising and unpredictable ways. Therefore, it is vital that robots are capable of accurately conveying their knowledge and capabilities in ways that humans can easily interpret and respond to. [29] [28]

The two scenarios presented in this master thesis' replication study rely on implicit and explicit communication in a laboratory environment, namely the intuitive dancing of the robot and the pursue of it to get the participants to follow it. Understanding how to design and how they were originally imagined, according to each scenario, is important for a successful unfolding.
Non-verbal behaviour and communication are very important in team coordination, whereas a robot adjusted to social cues enhances likeability and perception as well as cooperation. [30] Planning a robot's communication behavior necessitates finding more systematic methods for generating non-verbal signals while maintaining usability. The design of non-verbal signals can also be used to compensate for environmental factors that impact signal perception. [28] Non-facial and non-verbal

methods of affective expression, such as through movement, posture, orientation, color, and sound, are essential for naturalistic social interaction in robots that are designed to be functional and lack expressive faces. [6]
For a successful communication and interaction between these robots and humans, signals have to be created, which are as informative, intuitive and easy to learn as possible. [11] In order to generate intuitive and expressive non-verbal signals, most non-humanoid robots must overcome their constrained appearance and limited communication modalities compared to those employed by humans. [28] [20] [31]
Cha et al. for example [11] proposes using light and sound to signal the human peer for help when collaborating; using features and capabilities beyond those present in humans to get their attention, alert them or put up a request.
An important aspect of the interaction with the Sphero robot used in the aforementioned study is that of non-verbal communication. Sphero is a non-anthropomorphic, non-humanoid, transparent sphere with rolling motion and colourful illumination capabilities. Additionally, the robot is sufficiently robust to withstand moderate physical play, can sense and collect acceleration and angular velocity autonomously and features an array of LED lights to deliver multimodal stimuli. [32] Hence, the way these are made use of and are combined is crucial for the outcome of its communication and interaction with human peers and their interpretation of it. Researching different aspects of non-verbal communication in similar robots and related scientific works helped form an understanding of the depth and background of the replication study as well as of its adjustment and expansion.

Typically non-humanoid robots are goal-driven, only taking action when they have a specified task or goal activated. How a robot performs an action is just as important as what it does. [33] Satake et al. [34] proposes that people should be helped by being provided information on the robot they are to interact with or during the interaction. In particular, when people hesitate to ask for help, a proactive way of serving of the robot would be very welcome.
When considering the navigation of a robot, this is a simple, straightforward task carried out by it, however a very complex one considering the perception of human peers and unspoken social rules of our society. [35] [33] Robots move unnaturally and weird. [16] This is further complicated by people's presence and it violates social rules and norms if humans are treated as objects or mere goals. [35] [33] [34]
Moreover, while communication through motion can be highly expressive and natural, it can also interfere with a robot's functional task. [19]
Both the robot and the human must move in a way that takes the other into account, in order to efficiently navigate around each other. Otherwise the perceived quality of the interaction will be lowered and, delaying both of them, it can cause them to be less efficient in whatever task they are trying to achieve. [33] [34]
"[...] Robots need to navigate in a socially acceptable, human-friendly way when they share spaces with humans. Inappropriate distances, that is, spatial invasion, can lead to negative outcomes such as discomfort and consequent patterns of withdrawal or avoidance. [...]". [23, p. 1]
The way humans move carries interpretative emotional meaning, and moving with a robot is perceived likewise. [36] Not to be overlooked, the distance kept between individuals signals one of the following three states of relationship between these, according to Bethel et al. [6]: social, personal and intimate.

Satake et al. [34] suggest the approach of the robot to start a social distance and to move on from there, depending on the goal, context, interaction and driving speed, assuming that it moves slower than people; all while focusing on making it predictable and legible. Adding great value and guidance to classifying the study replication's outcome using Sphero, Satake et al. distinguishes and defines the following scenario outcomes for when a robot is to approach a human being:

- *unreachable* - the robot did not get close to target person;
- *unaware* - the person either did not look at robot or did not listen to it;
- *unsure* - the person recognized its presence and reacted (e.g., checked its reactions); but the robot did not respond correctly in time;
- *rejected* - the person recognized its presence and its behavior, but did not start a conversation.

The "Follow me" study to be replicated in this master thesis [26], ignored the concept of personal space and comfortable distance, including intimacy, by having Sphero bump into the participant's foot as a form of communication. Nonetheless, the study did at a later point in any given trial ask the participants to rate their closeness to the robot, view anthropomorphically as a social entity, by employing a proxemics-based inspired Godspeed questionnaire. [37] The proxemics are classification of distance between individuals, ranging from intimate, followed by personal, then social, and up to public, and help define the closeness between aforementioned entities. [23] [33] There are many other factors other than proxemics, that can be incorporated in the movement of a robot, such as haptics, kinesisc and chronemics [9], but these are not to be focused upon in the scope of the study replication.
Humans understand intuitively and unspoken the existence of and need for each one's own personal space and social distance, e.g. when walking together they do it side by side if given enough space, respecting each other's private spheres. [28] [34] Engaging in conversations happens in the same posture, if walking, as a face-to-face stand is not possible. Hence, a robot in a given scenario, such as in a guidance application, would feel more natural, relatable and less intrusive if it were to walk side-by-side with the person it engages, either leading or following them, depending on the case. It is important to understand, that humans are sensitive to robot movement since they are affected psychologically by it, as well as how they are impacted. [16] [31] [34]
According to Avrunin et al., following an experiment considering goal-oriented and socially-approached movement trajectories and velocities, "[...] We find that the resulting social approaches are rated at least as good as simple goal-directed behavior in cases where the whole approach can be seen by the target person, and significantly better where the robot's starting location is out of sight. [...]". [35, p. 1042] Thus, one should not be approached from behind as the person becomes startled.

In continuation, humans apply social norms and expectations to technical devices. [4] Hence, if the robot is aware of the social context, environment and location parameters of the human peer, by adjusting to the expected social norms, its actions would be easier to interpret and ease the person being approached by it. This helps the robot feel more alive and organic than mechanic and task-oriented.

Related to such studies, Leichtmann et al. surprisingly found that robot appearance and participant gender and age make no difference in the distance towards robots. [23] Wagner-Hartl et al. also concluded from their research on gender and age differences in human-robot interaction, that neutral to negative attitude and acceptance towards robots in care did not depend on the age group at all. [38]
Morales et al. propose the method of self anticipation, when attributing the robot a human partner. [28] Summarised, this refers to "[...] projecting the future position of the partner based on the utility model and then planning an appropriate path for an autonomous mobile robot by using the same utility model to mimic human behavior, provided better results than simpler planning mechanism [...]". [28, p.308] This could be just the beginning, as more cases are to be classified, analysed and implemented, e.g. when a person is around a corner and the robot could either move straight toward a goal or turning towards another goal, based on how people walk, the assumed utility value and the possible prediction of the direction that the person will take. [28]
However, Lu et al. reached conclusion than the argumentation above, to show that the human-robot interaction field (HRI) is conflicted and prone to the replication crisis [30] [31]: "[...] Depending on the desired priorities people will interact with the different algorithms with different levels of effectiveness despite the fact that there is no statistically significant difference in their opinion of the robot. This suggests that people are not consciously aware of many of these subtle social cues, thus prioritising these as secondary objectives in a human-robot interaction scenario, regardless of the effect they have on the interaction. [...]". [33, p. 1713]
The examples regarding robot navigation are just a fraction of a bigger topic, critical of robot integration and acceptance, namely that social intelligence in robots is necessary. Dautenhahn et al. calls for the establishment of a *robotiquette*, more precisely a social rules set for robot behaviour. [39]

Closely related to social intelligence lies also the concept of politeness. Kumar et al. [40] conducted a multi-experiment study using non-verbal robots to track and analyse the politeness in human-robot interaction. According to their results, participants reported higher levels of enjoyment, satisfaction, and trust when they interacted with the politest behavior of the robot, being also more communicative with it. Furthermore, the work of Kumar et al. references the definition of politeness, as it being "[...] a system of interpersonal relations designed to facilitate interaction by minimizing the potential for conflict and confrontation inherent in all human interchange [...]". [40, p. 1]
The concept of a "face‘" has many implications for politeness rules. When referring to human–robot interactions, it becomes a difficult challenge to fit it to robots: a face is emitting too strong feelings, based on verbal communication, and it is susceptible to cultural variations. Therefore, in the undertaken study replication with Sphero, it is good that is only disposes of a drawn face, to exclude such parameters and invariance from the results. Nonetheless, politeness rules are to be considered when preparing a study between a human and a robot, hence the following apply [40]:

- *Do not impose one's actions or views on the participants, at least without first asking for permission to do so.*

- *Participants need to be given options, in order for them to make their own decisions.*
- *Be friendly while interacting with other people, in the sense of producing at a sense of equality between the parties.*

Bethel et al., as a further positive example, found that humans were calmer with robots that exhibited non-facial and non-verbal affective expressions. [6] Beyond politeness but strictly correlated to it, one can identify the sense of humour, as researched by Press et al. in their attempt of designing non-verbal humorous gestures for non-humanoid robots. [41] As Press et al. stated, humour is good for human-robot interaction and should be applied as often as suitable and possible, to increase the sympathy, acceptance and liveliness of the robot viewed through the human prism. Press et al. conducted a study to test the gestures, resulting in participants perceiving the non-verbal humorous behavior to be more entertaining than the baseline, "robotic" condition but also less reliable, due to it not inspiring professionalism and thus, trust; or just by being unexpected for participants to interact with such a robot. If humour could be translated into Sphero's movement and lights flashing and colours, it could help ease the study pressure of the participant and exempt it from invalid results due to participants feeling observed. [20]

Another important piece of knowledge for carrying out the study successfully is that the robot needs to make its intentions clear to its human collaborator [34]: "[...]Predictable and legible motion can be correlated. For example, in an unambiguous situation, where an actor's observed motion matches what is expected for a given intent (i.e. is predictable), then this intent can be used to explain the motion. If this is the only intent which explains the motion, the observer can immediately infer the actor's intent, meaning that the motion is also legible". [12, p. 301]
Interactions between the human and robot rely on each collaborator's ability to effectively convey their mental state while accurately estimating and interpreting their partner's knowledge, intent, and actions. [11]
Predictable motions match the observer's expectation, while legible motions enable them to predict their goal. Considering the implications of the replication study using a robot based mainly on motion, understanding the differences between how to design its actions in a legible way is in the given case more important than its predictability. The human participants confronted to interact with the robot can not predict its outcome as they do not understand its purpose, given no background information on its purpose and functionality prior to the study, yet they must be able to interpret Sphero's actions quick and with confidence. In order to make humans feel comfortable, the robot's movement must tell its intention. [16]
Further research regarding predictability and legibility has been undertaken by Hoffman et al. in a number of projects employing anticipatory actions for robots in human-robot collaboration scenarios. [13]
According to them, reactive agents apply their decision function on the currently perceived state, while anticipatory ones act on a combination of the existing state and a probabilistic view of the temporal activity of the human teammate. It has been thus observed, that humans use visual contact to synchronise with a robot, for example when playing music together. [13]
Anticipatory actions help improve the fluency of the interaction and their legibility,

liven up the communication and overall improve the interaction between robots and humans. [16] [28]

The communication at a human-robot level can be defined as a hierarchically organized configuration of signal exchanges mediated by both the interlocutor's technological and social environments. [27] Its multimodality results in an integration effect of the unique match between a robot's behavior and a person's individual profile. Understanding who is in control in a human-robot relationship, and whether they are aware of each other is critical for building the human part's trust, as most people are inexperienced with robots or autonomous agents. [25]
When robots operating autonomously, laypeople can not assume that it is aware of anyone else in the environment. Thus, it is required that the same types of cues humans give off amongst themselves, such as gaze or gesture, are present, used and understood by the robot, to enable coordination and facilitate the human perception. [11] [30] [31]
Non-humanoid, non-verbal robots should make use of their available capabilities for communication, such as body language in the case of Sphero, in order to display various emotions, hence to generate more empathy and social acceptance, present themselves as believable, and display a predictable behavior. [18] [7] [36] Different expressive channels are needed and as such multimodal expression can be more effective than uni-modal. [...]". [15, p. 252] However, according to Kakazu et al., one should avoid symbolic displays, as these are not easy to interpret. [16]
Despite this, according to Saunderson et al. [9], multimodal displays of nonverbal communication have shown to have higher emotional recognition rates than unimodal nonverbal communication.
According to Rosenthal-von der Pütten et al. in their study of affective non-verbal communication [20], a robot can not possess as effective a non-verbal behavior as human. Attempts have observed an increased self-disclosure but a problematic decreased animacy at the robot. Nonetheless, robots ought to be used to enhance communication between people, adjusting to their needs and expectations, rather than replace them. [42]
Beck et al., having looked at the different communication means of various humanoid and non-humanoid robots, proposes the creation of an "Affect Space", blending emotions to create new ones, based on the parameters of valence and arousal, stretching from the most positive, aroused posture to the most negative, non-aroused posture, to be used to facilitate human-robot interactions. [18] [41] [3] The emotional feedback provided by a robot could be intuitively used by humans to establish whether or not an interaction was successful. [19] [20]
It is easier to decrease a perceived valence of an expression by constraining it to a short duration or to high carry-out speeds, by increasing the frequency of the movement of the limbs, or by expressing avoidance. [19] After all, according to Lakatos et al.: "[...] it has been the portrayal of emotions that has given the Disney characters the illusion of life [...]". [7, p. 2]

Humans can gesture or subconsciously alter their posture and movement to reflect their emotional state; by being attuned to emotional expression in body movement, they can furthermore identify many emotions just by analysing it. [36] [20]
By widening the range of emotional expressions, especially physically induced, of

the robot, it will help human partners not only understand it, but also interact with the robot intuitively. Deciding for key poses and movements, to be used by Sphero during the study, such as specific move patterns, turns and their predefined rhythm, helps improve its expressiveness towards the participants.
Boccanfuso et al. has researched the emotion generating capabilities of Sphero in a study presented in the paper "Emotional Robot to Examine Different Play Patterns and Affective Responses of Children with and without ASD", offering a viable example that Sphero can actually be deployed to make use of behaviour and physique to facilitate non-verbal communication. [32]
Robins et al. also researched the contribution of robots as companions for children with autism and came to the conclusion that repeated exposure to an interactive robot would increase basic social interaction skills. [43] The robot became from a companion a mediator, following the human contact who attributed significance and emotional, subjective meaning to the experiences with it. Robins et al. goes to show, that long studies are of higher benefit and gain when researching HRI and people's responses towards it, as the participants became more spontaneous, proactive and unconstrained as time passed with the robot. It researched the therapeutic outcome of robots in child-robot physical and social interaction, considering the autism spectrum disorder.
An important learning from their study is, that the researchers concluded the experiments successfully, resulting in the combined effect of multicolor flashing and fading LEDs [36], custom-produced music and movement which contributed to the overall conveyance of emotion, agency and intentionality. This supports the use of lights, movement and music in the extended replication study presented in this thesis. Moreover, the use of lights, their meaning and characteristics, coincide with the blue for happy and red for angry presented in the replication study, and inspiring the original study. [26]
In addition to this, a color-changing robot, such as Sphero, is also extremely effective in communication [36], and a good candidate for the study, as, by following arguments presented by Kakazu et al. in his book *Intelligent Autonomous Systems* [16], the longing for tactile communication and eye contact are stirred by Sphero in the human peer through his own ball design and drawn-up face.

Human-human interactions are very complex since they are generally symmetric, develop since birth and are based on the use of language. [7]
Cass et al. states that robots in HRI studies so far have been focused on borrowing and mimicking human non-verbal behavior, omitting other possibilities of innovation and new designs. [14] Dautenhahn et al. also mention that: "[...] instead of a realistic replication of one or very few human activities, replicating the flexibility and adaptability of human intelligence became a big challenge. [...]". [39, p. 679]
However, robots can also borrow cues and gestures from animals to ease communication translation, nonetheless depending on the robot to e.g. raise its ears. [14] Dogs for one are a good reference model since they adapted to the human environment during domestication, becoming more relatable and more integrated in human's lives [7], making their behavior easily recognizable. [36] [3] Faria et al. [26] proposes one particular approach to be mimicked by the robot, regarding the communication interaction of visual signals, which can be divided into four stages as follows:

1. The sender produces signals to initiate the interaction.

2. It recognizes that the receiver is in a state of attention.

3. The sender sends further signals.

4. The sender might receive a response from the receiver.

"[...] In practice, dogs often use such attention-seeking behaviors: a dog performs movements to and away from a person as an invitation to play or to be followed, or can bark to get attention [...]". [26, p. 666] Pets are often considered part of the family, and emotionally bound to their owners, e.g. dogs and children activate common brain regions in mothers drawing on the hormone oxytocin. [44] [3]
The main advantage of implementing animal-like behaviour, such as a dog's one, into robots is that it is simple enough to be easily realized technically. Many non-humanoid robots seen in HRI research, more precisely in health care, take the forms of commonly recognized animals, e.g. Paro as a baby seal, Sony's Aibo as a dog, and even the extinct Innvo Lab's Pleo, the dinosaur. [36] [44] [42]
Fawcette et al. affirms that humans do benefit significantly from maintaining relationships with non-human animals, in support of the search for robotic animals. [3] Gazing at pets has been especially recorded as positive in pet-owner relationship and interaction. [9]
Given physical and psychological benefits derived through humans' interaction with non-human animals, it seems only logical to propose that psychological interventions ought to incorporate exposure to natural non-human elements. [3]

However, different evidence was also brought by Schumann et al. [45], who, pursuing the replication crisis in the field of human-dog communication and ontogeny, proved that domestication actually did not play a critical role in shaping the ability of dogs to follow human orders.
Instead, it is regarded that human exposure had/has no major effect on dogs' ability to use human-given communicative cues, and assumed that this skill therefore represents a special adaptation in dogs, present from early age, which was bred indirectly or directly, through the early humans acknowledging it, in the waves of domestication; e.g. "[...] puppies use the human pointing gesture they do so by actually following the pointing and not by using simpler mechanisms, e.g. approaching the human's hand and then the cup closest to it [...]". [45, p. 1013] This goes to show, that before designing and creating animal-like robots, humans still have to fully comprehend the human-pet connection.
Related to this, research done by Lakatos et al., tested a scenario with a robot replicating dog behavior. According to the results, "[...] people were able to recognize if the robot transgressed on the basis of its greeting behaviour [...] the findings showed that dog-inspired behaviour is a suitable medium for making people attribute emotional states to a non-humanoid robot [...]". [7, p. 28]
However, by replacing pets with companion robots, Rault et al. notices some challenges and open questions. [44] For once, live pets could become scarcer, a luxury commodity and a symbol, sign of status. For the other, there are also ethical aspects at risk of being dismissed, as how a given pre-programmed robot would be suitable to replace each and every species available as a pet nowadays, and beyond. "[...] The responsibility that we feel for each (robot or animal) may differ, as suggests the difference between the lifetime of a Tamagotchi and that of a live pet. At present,

artificial pets can be described as mediocre substitutes for live counterparts [...]". [44, p. 2]
This is an important aspect to note, when and if implementing animal-like behavior and communication in a robot, so as to not completely turn it into a replacement of the specific animal.
Additionally, interaction with pets has been proven beneficial for the mentally ill, subjected to a prolonged therapy, and for the mental health at large. [17] Specific features of pets, like expression of affection, responsiveness, or willingness to interact with the owner, prove to be utterly important in forming a close relationship between the two. [7] [20]

Overall, there are many indications that humans tend to enjoy the nature, and being close to it and living beings, including non-human species; as they form an emotional connection with these. As Fawcett et al. states that "[...] Humans' affiliation with nature is also reflected in their expressed enjoyment in making contact with or viewing other species [...]", more so than with all sport events combined. [3, p. 125]
A similarity found here between the pet therapy and the outcome of the original study, detailed under the chapter *Findings*, was that verbal social interaction increased drastically, when participants experienced novelty in an interaction, be it an elderly with a dog or a student with a Sphero robot. Past studies in HRI have shown that behavioral and social signals are able to be quantified and analysed, being the most noticeable of the engagement of a person in a human-robot interaction scenario. [46] This suggests that the verbal social reaction supports deep engagement, hence robots could become more accepted and friendly towards by people if they were to copy animal behavior. Especially the elderly were more fond of and invested than the young, when interacting with animals or robotic pets substitutes. [27]
The literature review researched result has not been without contradiction, as Lakatos et al. claims "[...] that the behavioural repertoire of pet-like social robots is very limited and for this reason, compared to animal pets, they proved to be less successful in maintaining humans' interest in long term [...]". [7, p. 3] Finding opposing results and claims displays how conflicted HRI is and how the replication crisis can accentuate it. [30]
Collins et al. cites that on the other hand "[...] Robots that are biomimetic in their morphology, in the way they move, and that have expressive faces are immediately and intuitively engaging, owing to our familiarity with mammalian channels for conveying emotion and intent [...]", leading to casual (naive), elder, demented or pet-accustomed persons to be more warm and open towards robots displaying such characteristic. [15, p. 244]
Non-humanoid robots should try to copy animals biologically in order to emit more familiar and understandable social signs. With a different morphology, such a robot is forced by the physical differences to generate movements—postures, (speech-accompanying) gestures [47] or whole-body movements—that are different. [36]
One such case presented in the paper by Colling et al., shows that modelling the affective state using body posture [16] and navigation, head movement, and patterns of pulsating lights can evoke reliable perceptions of affect in naive participants, as well as positive behaviors and expressions [17] [48] [49], as listed previously: "[...] Work with humanoid robots has shown that affect can be communicated well using

body language (gesture) based directly on that observed in humans.
In favour of the body language is the major difference between facial and bodily expressions, namely that the latter can be recognized from a much bigger distance than the other. [19] On top of that, some emotions are more powerfully expressed and easier conveyed using a body than using a face, e.g. joy, fear and sadness. Colour and lights are also an option to express emotions, when designing interaction tailored for Sphero.
However, Rosenthal-von der Pütten et al. question the validity of such non-verbal behavior, as "[...] there is no consensus on which emotion is represented best by which color, nor whether one should use accompanying variations (like frequency or wave form/sharpness of flashing lights) or how these variations exactly influence perception yet [...]" [20, p. 571]; hence rising replication issues to the original "Follow me" study. [26]

The robot behaviors are constrained by the limited motor capabilities of the robot, but they can also take advantage of expressive options that are not available to humans. Robot's body movement and orientation attract attention of and initiate interaction with the human peer. [14] [28]
Gesture in itself is an important feature of social interaction [47], despite lacking thorough research in the HRI area. [31] It is frequently used by human speakers to illustrate and substitute for what speech alone can not communicate, as in to convey referential, spatial or iconic information. Explicit participant attention towards the robot greatly improves the non-verbal communication between them and the robot. Movements can be considered to be gestures as well, although definitions and categorizations of shape and properties of it vary widely among papers researched by Salem et al. [47] [31]
Gestures and, in the case of Sphero, colours, facilitate the expression of its mental state. According to Cass et al., there is a high need for metaphorical signals development in human-robot non-verbal communication. [14] [28]
Lakatos et al. argue that robots should necessarily appear neither human-like nor pet-like but rather fitting and functional with regard to the human community they find themselves in. Additionally, their social behaviour should mirror this function as well, taking into account the existing embodiment and given technological constraints. [7] [28]

Keunwook et al. approaches their own robotic creations, the Post-plants, with innovative ideas and concepts, not relying on either human-like, nor pet-like influences. [8] Robots do not have to mimic human behaviour or language with as high a fidelity as possible, when they can actually instead respond to touch in their own physical way. [7] Adequate timing and natural appearance of the robot's body movements are essential to create the impression of the robot's liveliness. [31]
In Keunwook et al. [8] designs of future HRI scenarios and robot interactions, robot liveliness is simulated, with a focus on non-anthropomorphic robots, alien-looking plants married with inorganic elements alike. The aim is to offer a new concept of relaxation, to convey emotions and facilitate information feedback. The robot's main body behaviour is inspired by plants, signaling a lack of care through a physical state - hopefully achieved by Sphero with the dancing initiative.
Inspired by the theories of Laban [16] [19] and psychology study of Russel psychology,

humans read information and emotions from movement, such as arousal (awareness), valence (positive, negative) and stance (open-closed attitude). [8] These theories and work done by Keunwook et al. is important in the design of the scenarios and interaction with Sphero.

2.2 Human-Robot Interaction

HRI is a young and highly interdisciplinary informatics field. [21] "[...] A human-robot interaction setting (is best described) as a technology probe, (a one-time sample of a study involving interacting with a given technology), that offers opportunities to learn about the relationship of people, data, and technology beyond learning about the system itself. [...]". [22, p. 19]
Robots, especially social ones, are often presented as human-like and human-like-abled, both physically, mentally and emotionally, nowadays common, cheap and increasingly more often deployed as companions and interlocutors. However, these are only simulations, always on, better than their human peers at repetitive jobs.[2] [20]
Having mentioned social robots, Dautenhahn et al. lists some defining characteristics that a social robot should adhere to in order to belong to, namely [39]:

1. *evocative* - relying on the human tendency to anthropomorphize and capitalize on feelings evoked;
2. *situated* - reacting to and perceiving their given social environment;
3. *sociable* - proactively engaging their human peers;
4. *intelligent* - showing burrowed aspects from human-style social intelligence;
5. *interactive* - different from conventional HRI, interaction hereby plays a key role in the peer-to-peer relation.

Researching robots helps research human behaviour as well, especially considering social robots in care of the elderly or demented, for children with autism or as telerobotics. Socially interactive robots have been designed and built for the purposes of exploration, e.g. for scientific experiments about human behavior or cognition; service, e.g. healthcare, assistance and communication; and influence, e.g. education, therapy and entertainment. [50] [38]
Research including diverse robots in different areas has shown that robots are treated the same as other humans, in spite of their anthropomorphy or lack of it. [2] [15] [38][46] Most of these tasks in these domains are difficult to be automated perfectly, however, they could be executed in cooperation with a human/humans. [10]

Acceptance of any technology in a society is highly dependent on functional and social aspects. Hence, the social behavior of robots is considered influential on humans' willingness to interact with them. [40] Humans even attribute gender and intelligence when interacting with an anticipatory robot. [13]
By applying social rules to robots, humans get to reflect on their personality and personas, how they perceive the world through human filter, namely anthropomorphism. Interactions with robots makes humans question who they are and project

their desires onto them, as well as their mind perception and emotional attachment, despite robots being just physical embodiment of technologies; distorting the reality in a sociable desirable direction. [2] [5] [31] [4] Additionally, the successful interpretation of non-humanoid robot body language is dependant on its anthropomorphization by the human peer. [36]
People feel engaged when interacting with a robot and attribute human-like traits basing these off facial expressions, posture, speech and laughter; with the robots appearance and features influencing furthermore its reliability, perception and trust. [46] [40][5] [5] [20] [42] Mismatches between the appearance and the behaviour results in the human's prior - often unrealistic – expectations, mostly based on the robot's appearance, being violated resulting in a feeling of unease. [7] Non-humanoid robots can take any form so long as it lacks the morphology of a human or humanoid. [36]
Human-robot interaction is a specific form of inter-specific social interaction, analog to human-human and human-pet interaction, as the robot is being anthropomorphized by the human individual. [7]

Analysing the Sphero robot [26] [51], one could classify it as an entertainment robot with educational purpose, based on its researched usage and advertising. Thanks to its playful design, it assumes a strategic role for promoting robot acceptability and adoption in the future years. [52]
Lupetti et al. hereby describes robots as "[...] glue between physical world and digital universe [...] (enabling) valuable interaction modalities that should reduce sedentary behaviors (in humans) [...]". [52, p. 631–632]
Of course, when referring to HRI one does not focus only on the human but also on the robot. As such, the field can be split in two approaches, according to Dautenhan et al. [39] On one hand, there is the robot centered HRI - "[...] an autonomous entity that is pursuing its own goals based on its motivations, drives and emotions, whereby interaction with people serves to fulfil some of its 'needs , e.g. social needs are fulfilled in the interaction, even if the interaction does not involve any particular task [...]". [39, p. 684] Hereby, the robot possesses skills that enable it to "survive in the environment" deployed to or otherwise fulfil predefined "internal needs".
On the other hand, there is the widespread and mostly focused on human-centered HRI - which is primarily concerned with how a robot can fulfil its task specification in a manner that is acceptable and comfortable to humans. From Dautenhahn et al. work: "[...] research studies how people react to and interpret a robot's appearance and/or behaviour, regardless of its behavioural robot architecture and the cognitive processes that might happen inside the robot [...]". [39, p. 684] The increased focus on human- instead of mechanic-oriented, transformed the very purpose of robots' existence, together with their engineering design and physical appearance. [27]
For an accomplished human-centered HRI approach, the following rendered challenges are extracted from the work previously cited:

1. finding a balanced and consistent design between robot behaviour and appearance;
2. designing socially acceptable behaviour;
3. developing new methods and methodologies for HRI studies and evaluation of HRIs;

4. identifying the needs of individuals and groups of subjects to which a robot could adapt and respond;

5. avoiding the so-called 'uncanny valley' - the effect of feelings of discomfort and dread instilled to humans by robots that look too much the first alike. [20]

Furthermore, when analysing relationships in HRI, there are two paradigms to be distinguished. First, there is the caretaker paradigm, which implies that the human is responsible for the robot's needs and happiness, showing similar behavior as towards babies or infants, invoking emotional and psychological investment. This is an important one, as it was proven, that e.g. species displaying such nurturing instincts were more accepted and sympathetic to humans than the others otherwise. [39] [53] The human, hereby being an enabler, is the one with the knowledge, skills, abilities and powers to enable the functioning and operation of the robot. [53] However, as enablers may suggest superiority, and the situation of conflict between an elderly and an enabled robot is still unclear, tension arises when having to choose between enablement and autonomy, unless both the robot and the person are positioned on the same hierarchical level, of companions.. [53]
Then, there is the companion paradigm, which implies the opposite, where robots are the caretakers and companions of humans, highly applicable in the case of elderly or sick persons' assistance, which highlights that robots need to be proactive and non-intrusive, but also to not overreact with interfering, manipulate or control the person they are supposed to accompany and look after. [39] [54] [55] [53]
Such a companion is expected to imitate real-life behavior, display motoric, emotional and cognitive animal- or human-like behavior, posses skills for various communication levels, and influence the human cognition, emotional and behavioral response. [9] [27]
In addition to that, Libin et al. adds that "[...] robots with various levels of artificial intelligence and types of sensory feedback can benefit people with different combinations of skills and abilities in a very broad context of individual situations [...]". [27, p. 1791] This is highly important for social robots in healthcare and care.

While the original"Follow me" study [26] was carried out in a biased way, only at a university with young students between 18 and 26 years of age, an important group, that of the previous generations and elderly was left out, as demographically speaking humanity is and will be dealing with an aging population. [1][1]
As more and more elderly will live alone, and with dementia on the rise, robots are intended as companions and social workers, as in social robots, to fill in the lacking manpower and vacant positions. [55] It is expected for different age populations to use social robots for different purposes; researchers foresee especially older adults to be more frequent users of social robots. [55]
Looking to also design a further HRI scenario, not only add another target group, the topic of dancing came across the research results, standing out as a clear form of entertainment, easily understood by every human being. [38]
Dancing, can be seen as a an art form, representing imitation and patterns, the coordinated production of sound and movement, that humans find pleasurable in social interactions scenarios. Moreover, beyond dance activities having numerous health benefits, such as activating one's body and mind, exercising and memorising at the same time, it is also a means of communication, as well as a psychological

and physical therapy, contributing to robotic applications research and helping with diagnosing characteristic of pathologies such as autism and schizophrenia. [50] [56]
Further motivation to include dancing in my thesis came from the fact that attention to the rhythmic characteristics of non-verbal interaction has not yet been widely adopted in social robotics research, according to Robins et al. and their attempt at analysing and testing dancing between non-verbal robots and humans. [50] An additional challenge lies in the synchronisation of the robot and the human, facilitating a fluid and enjoyable dancing experience; subject which will be discussed in the third chapter *Methodological Approach*.
To add further arguments to the stated motivation, also very little prior work has addressed the opportunities and challenges of creating an emotionally expressive body language for non-humanoid non-verbal robots. [19]
As physical mobility decreases with age by around 87%, the concept of healthy aging has arisen to combat it, namely being active and enjoying one's own mobility while taking care of it. [54] [55] [57] [53] Hereby, social companion robots are to assist elderly people to live comfortably and independently in a non-invasive manner, thus being more accessible [53] and adapting somewhat like pets in homes; [57] while keeping in mind that most elderly find being taken care of as degrading themselves, especially by robots. [30]
Based on Sawami et al. research on dementia and dancing [58], it was proven that dance in itself stimulates brain training thus increasing the results correctness in cognitive test afterwards.
Kosuge et al. found furthermore in their study on the HRI involving a dance-designated, mobile platform with humanoid upper-body, named *MS DanceR*, that the more knowledge it had of dances and the more skillfully and elaborate the movements were, the more the human counterparts experienced it. [10] Dancing in itself is an imitation of each partner's moves, an action and a reaction building the interaction in the pair. [49] Robots would ideally communicate and move actively based on their human peer intentions to effectively carry out a task requiring physical interaction, in this case when dancing. Therefor each person has to estimate the intention of the partner and move based on their intention, information of the environment, their knowledge of the dance, and so on and so forth. [10] Kroma et al. has carried out a similar experiment yet with non-humanoid and non-verbal robots, reaching similar conclusions and outcomes. [48]
Quan et al. carried out a dancing study using a Sphero robot with the scope of promoting more often home exercises for the elderly and analysing its benefits, such as improved physical abilities; providing a source of inspiration for the design of a related scenario involving dancing. [55]
Furthermore, the latter scenario was also favored, based on statements found in literature. "[...] Robots have the potential to help older adults perform healthy activities, which could lead to improved health and greater independence. [...]", and "[...] Partner dance between humans has been shown to be an effective form of exercise to improve physical function in older adults. [...]"; according to Chen et al. [54, p. 1, 7]
Zuckerman et al. even stated in their work about robot companionship and its relation to robot's function, that non-humanoid robots, using non-verbal communication and other functions than companionship, such as dancing, should be explored more, as these seem very promising for HRI; they make social experience and emotions

possible, even without mimicking human behaviour. [30]
Sawari et al. observed a positive interaction of the elderly with the dancing therapy robot, feeling happy, revitalized in body and mind, healed and socially engaged after every session; and even remembering specific songs and learnt choreography with music. [58] This also improved the positive perspective of the robot and its sympathy as perceived by the participants. [49]
Potnuru et al. observed in an own study including a 3D printed dancing humanoid companion robot, that the ballet dance moves were very well accepted by all types of audience from young to old. [57] This is an important finding to consider when designing the dancing scenario for Sphero, considering what most people take for entertainment and which music genre would be wider accepted considering the participants, their background and the location of the study.
Motivational and emotional benefits were observed in Chen et al. study of the interaction between and acceptance of elderly persons and a dance robot, nonetheless a mobile platform with a body and arms, opposite of Sphero. [54]
However, as Sphero is less humanoid it could actually improve the acceptance if the interaction is designed correctly. All the adults in Chen's trial were accepting of the robot, touching it, and found the dancing enjoyable and user-friendly. They also enjoyed leading the robot, which correlated to the aspect of healthy aging and assistive technology that people do not like to feel helpless, but in control of their surroundings and smart assistants. [1]
"[...] The decision of cognitively unimpaired but less able people to take assistive technology into their homes is a decision to live with a certain loss of privacy in return for effective emergency or care response or more effective management of long-term medical conditions. [...]", as cited by Sorell et al. [53, p. 186]

Moreover, another aspect to regard in the methodology and scenario design of the replication study, that arose from Chen et al. paper, was how to not make the activity too easy, boring, predictable or unclear. In order to keep humans interested, as Tanaka et al. suggests, it is indicated to make use of complex, yet also unpredictable moves, when designing the robot behavior. [49]
This is based off the concepts of above mentioned predictability and legibility, with the latter having priority to the first one. Since the robot is the one initiating the dance, it is important to understand the taxonomy of robotic dance, as described by Peng et al. in their work "Robotic Dance in Social Robotics - A Taxonomy". [56]
Robotic dance is classified into four categories: cooperative human–robot dance, imitation of human dance motions, synchronization for music, and creation of robotic choreography; with the later one being the most important in the case of Sphero starting to dance, followed by it adapting to the human peer. Designing this scenario meant creating a default choreography for the robot. The creation of robotic choreography reflects that robots "behave" autonomously and intelligently. Good robotic choreography preserves characteristics of the human dance, innovates, and is executed in accordance with human aesthetics. [56]
However, human dances cannot be reproduced just by employing robots as mobile platforms on wheels, as these can still not accurately imitate human dance in its entirety. Thus, human-like dance by robots is hard to achieve, especially if the physical structure of robots limits them from doing so. With regard to imitating the human peer, Peng et al. also mention, that "[...] it is unclear when to begin

to imitate, how to imitate, and what should and should not be imitated [...]". [56, p. 290]

Following up on the age differences of the participants, it is important to acknowledge that age is a relevant factor in HRI. [40]
Young people start with higher curiosity, openness and acceptance of robots than the elderly until the novelty effect dissipates. [1] [17] The elderly find the robots to be more useful than the youth, suffering from less anxiety and showing an increased ease of use after the novelty effect of the interaction with them disappears, which is a crucial fact, considering that the elderly especially are the biggest user group of social robots, given that they become increasingly lonely people, thus being more prone to depression and having a decreased quality of life. [42] [53]
Explicit attitude is typically positive, measured by capturing a person's controllable, deliberate, and conscious perceptions or judgements. Conversely, implicit attitude usually functions without one's full awareness or control, leading to a more negative impact. [1] Young people are already used to technology, expecting it to be increasingly interactive [42], thus do not correlate it with something bad as much as the elderly do; showing a more explicit attitude than implicit towards it as opposed to the elderly and their use of other computing technologies. [30]
Nonetheless, the elderly have less knowledge on limitations and failures of robots and thus an increased trust, once obtained, being more open and impacted by the novelty effect. [40] [42]
To further detail the concept of trust, as it is so important to HRI, it can be classified in two categories. [36] On the one side there is the *performance-based* trust, assessed by the robot's ability to complete a given task. Hereby, non-verbal communication is crucial for a positive influence of the robot. [9] [42] On the other side, more relevant to the study replication undertaken in this master thesis, there is the *relation-based* trust, formed in the human individual by superficial cues rather than actual capabilities of a robot.
The elderly hold different views and opinions on anthropomorphized objects, such as companion robots, they prefer different designs and hold different, more reserved and stable attitudes towards robots. As such, they behave differently when interacting with them: the higher the mind perception attributed to a robot, the better the engagement with it, and the elderly tend to do this more often than the young. [46]

Besides, as the perspective of not only the young compared to the elderly, but also of males compared to females differ, it is important to distribute the participants among these parameters as equally as possible. [40]
Although, Lakatos et al. and Wagner-Hartl et al. found in their HRI experiments that age and gender surprisingly did not matter as much, whereas the main observations were that women, especially young ones, rated the robots livelier and dog owners had less negative attitudes towards the robot presented with in the experiment. [7] [38]
Wagner-Hartl et al. assumes that young women in particular attach greater importance to personal care and are therefore reluctant to pass these topics on to other people or robots. [38]

2.3 Replication Crisis

The replication crisis became a phenomenon when it was observed, that a worryingly large number of research results are difficult to replicate or reproduce, failing to achieve the same results as the original study, thus being discarded as unreliable. Originating in social and experimental psychology and rapidly spreading to other fields of science, it has become prevalent in HRI, as the ties between it and the social and cognitive psychology are very close, due to the main human play in the interaction. [21] [22] [25] [42]
The methods and protocols available in HRI do not exempt the studies from being undermined by the exact same problems experienced in other fields. [24]

Ullman et al. states that scientists must avoid placing too much faith in single studies, as they are particularly vulnerable to errors and replication incoherence. [25] The replication crisis signals an urgent need in HRI to break down study generalisations and re-generalise and aggregate results obtained but with a specific robot, or robots comparable in appearance and abilities, as together with other different robots, in order to eliminate doubting values and lower robot distrust. Therefor, to combat the replication crisis, not only consistency across participants, contexts and methods, but also across robots is vital. Therefore, researchers must form a collective intention to replicate results across a range of different robots in HRI. [25]

According to Irfan et al., 2/3 of investigated psychology studies could not being replicated due to publication bias, sociological changes in the population, impossibility to reproduce the exact same context, circumstances and persons with their given states of mind. [24]
To this end, they state that: "[...] social facilitation or inhibition, like many other psychological effects, may be affected by a combination of several other factors: the observer effect (also known as the Hawthorne effect), demand characteristics, cultural differences and personality. These effects are potential confounds, and adequately accounting for each of these in the experimental design is problematic [...]". [24, p. 6]
To briefly explain the Hawthorne effect, named after its founder, is a social phenomena coined in the first half of the previous century, when Hawthorne underwent an working place assessment by observing the employees in their daily routine. Even though the true intent of the study was not revealed until the debriefing, subjects felt observed whatever the condition and this might have impacted their behaviour, as he later stated. [24]

Leichtmann et al. actually states that the knowledge on personal space in human-robot interaction is in fact very limited, which imposes the question of replication crisis, "[...] Although a wide variety of theories on ultimate and proximate causes exist, these have been neglected so far in the literature, leading to a rather incoherent picture. Furthermore, methodological weaknesses were identified such as small sample size, statistically underpowered designs, overestimated effect sizes, and questionable research practices. Lessons learned from the replicability crisis in psychology need to be addressed more carefully in human-robot interaction research. Theoretically well-grounded pre-registered studies with large sample sizes, high power,

transparent reporting, and open data, will enlighten this field [...]". [23, p. 11]
In another research cooperation with Nitsch [4], Leichtmann et al. validate the replication crisis by studying the *CASA* hypothesis, meaning *"computers as social actors"*. This hypothesis refers to how people subconsciously apply social rules and expectations to technology at large. Therefore, if the true effect sizes are small, which is common for social-psychological phenomena, or the variance of true effects is large, the average sample sizes are too small and thus studies have a low power. "[...] Low powered studies in turn have a low probability of finding these true effects, producing more false positive results. Even worse, effect sizes in underpowered studies are often overestimated, a problem also known as the "winner's curse". [...]". [4, p. 1015]
Furthermore, the replication crisis is also strictly correlated to the social desirability bias. This bias translates to an intensively studied psychological phenomenon and refers to the tendency by participants under specific conditions to answer questions in a more socially desirable direction than they would under other conditions. Such a bias is only increased with technology improving its ability to evoke anthropomorphism. [4]
The following conclusion can be drawn out of Leichtmann et al. work [4], namely that inter- and individual differences of the participants have a greater effect on variance of the outcome than the actual interaction design. They suggest, that instead of simply assuming that computers or robots are treated as social actors, more careful reflection is needed on it occurring in a specific study, as well as on how to design the robot interaction more mechanical.
As not every replicated study has to differ from the original one, an example is offered by Wagner-Hartl et al. study, which achieved results in their care experiment aligned with the original trial and blamed the failures or deviations on bad robot interaction. [38]

There has been numerous research around the replication crisis topic, which together has formed a guideline for future HRI research, in order to avoid past mistakes, ensure the validity of the results and support an irrefutable outcome. As only about half of psychological studies can be accurately reproduced, promoting good research practice is vital in invalidating the replication crisis occurrence. [21]
Irfan et al. [24] conclude their research and study work on the replication crisis, by offering the following suggestions for future work in the HRI field:

1. *Replicate and reproduce* - first replicate a social psychology effect with humans before moving to robots, as people, times and methods change and errors are easy to overlook.

2. *Null-results are interesting* - do not discredit studies done in the past just because the results were inconclusive or negative;

3. avoid questionable research practices - such as selective data reporting or selective data usage in order to support a particular hypothesis.

4. *Register your study* - hand in the protocol before beginning with a clinical study.

5. *Avoid the Hawthorne effect* - as described above, make sure that the participants is not aware of being watched or feels comfortable about it.

6. *Come up with HRI reference tasks* - "[...] while there is merit in attempting to reproduce effects from social psychology with robots instead of people, it might be worth identifying new effects and tasks relevant to Human-Robot Interaction and its applications look at interaction between humans and robots through new lenses [...]". [24, p. 7]

Furthermore, according to Belpaeme et al. [21], the following should be considered when carrying out a study in order to omit important mistakes and bad practices observed in an analysis of many previous HRI studies and experiments:

- If one does not have many *resources* available for a study, they should opt for carrying it out in a *lab environment* instead of at a specific real, non-lab, location. Public *wild* spaces are dynamic, uncontrollable and partially predictable environments.[14] Almost three quarters of HRI studies are lab-based. [21] However, in order to support the validity of the obtained results, the living lab, a semi-naturalistic environment in which conditions of natural environments are replicated, would rather be the ideal decision. "[...] These environments allow for complex technological setups, while offering a perhaps more relaxed environment in which the behaviour of the user can be more natural. [...]". [21, p. 3]

- *"Wizard-of-Oz"* is the more favored method of carrying out a study, with around 60% of the studies employing it. This translated to the robot "faking" autonomy by it being remotely controlled by a person, referred to as *"wizard"*. This helps provide for lacking perceptual and cognitive capabilities of the robot, making up for missing features and allowing for faster and earlier prototyping. It also eases the observation of the experiment by allowing the "wizard" to be in control of it. Nonetheless, one must not forget the Hawthorne effect [24], implying that people being observed will react differently than usual, likely unnatural: "[...] In the context of HRI research, the Hawthorne effect becomes acutely relevant. In research studies participants are often aware they are taking part in a study, for example through being recruited or through signing a consent form. Even if no experimenters or video equipment is visibly present during the study, the mere fact of taking part in an experiment will already change the participant's natural behaviour and responses. This often leads to unexpected results or the changed behaviour of the participants washes out small effects. [...]". [21, p. 10]

- By using a robot on-screen in a video, one can not really interact with it themselves, but rather just assess how others have or might have interacted with it, depending on the material displayed. It is advised to use *robots in person* whenever possible, in order to obtain more useful results.

- More than 80% of HRI studies draw on university students for their participants' pools, a convenient but biased resources, not necessarily providing good and viable data. Ideally one uses *large sample sizes in the wild*, as in non-lab environments, where balancing and a lack of bias naturally occur, due to participants varying in age, gender, sex, expertise and background.

- Only one in 15 studies offers the participants the chance to interact with the robot more than once, which leads to profound implications. The *first* *contact* of a person with a robot is always influenced by its novelty effect, which can lead to biased and misleading results. As HRI concerns itself with how the user behaviour and perception will evolve over time, not the short-term reward of singular interactions, there is a strong necessity for *long-term interaction*. Adding to this, Bethel et al. states that "[...] humans calibrated their responses (in her studies) to (non-verbal and non-humanoid) robots based on their first robot encounter (and its novelty effect) [...]" [6, p. xii] Thus, it is interesting to see how and if the participants perspective and interaction changes from one scenario to the other in the replication study at hand.

Weiss contributes to these guidelines in her habilitation treatise: "[...] HRI findings obtained by one research group need to be replicated by other groups to further build on them. However, singularly focusing on reliability, replicability, and representativness as the main criteria for assessing quality in HRI research fails to address research that operates outside of positivist paradigms. I propose rigour as a high-level criterion that emphasizes the different qualities stemming from relativist, interpretivist, and critical perspectives in addition to positivist research. [...]" [22, p. 33]
Based on her research and suggestions, the following set of quality criteria is inspired by the already rigorous one developed in the field of visualisation design studies [22]:

1. *informed* - existing knowledge and design guidelines inform the design and take the context of the situation into account in order to facilitate new interpretations;

2. *reflexive* - the subjectivity of the persona that is the researcher is adequately presented and reflected upon;

3. *abundant* - complexity is accounted for by involving many different perspectives, as the diversity in human responses towards robots requires the respective studies to be complex, flexible, and multifaceted.

4. *plausible* - knowledge claims are evidenced, appropriate, and persuasive, relying heavily on the methodological conduct and the researcher's subjective perspective;

5. *resonant* - the research inspires understanding and invites action and future work, contributing and motivation future researchers to take up the work where it was left and continue expanding upon it;

6. *transparent* - the reporting invites scrutiny, instead researchers are expected to provide all the details of what, how and, most importantly, why something was done;

7. *ethical* - the developed design follows ethical conduct, as designing, developing and testing robots and human-robot interaction impacts more societal aspects, beyond the robot's usability and functionality initially focused on.

2.4 Relevant Related Work for Study Replication

In this chapter, a lot of information has been offered, both on the robot and human aspect of HRI, as well as on studies carried out in this field and the challenges with replicating them. Referring to age differences in perception, reaction and interaction with a robot, as well as the robot perception paradigms, non-verbal communication capabilities and social cues integration is crucial for the design and implementation of the methods in the study replication, as well as for the undertaking and unfolding of the experiments. Furthermore, understanding how the replication crisis impacts the field of HRI and following grounded guidelines, while acknowledging both their weaknesses and strengths, is also of great value for the analysis and discussion of the findings.

Regarding the "Follow me" scenario, robot navigation is complicated because of people's presence, potential violation of social norms if humans are treated as objects approached in a straight-line, particularly from behind, instead of socially. Actions are easier to interpret and people feel more comfortable otherwise. [35] [36] [9] [42]
Social cues and non-verbal communication are as important for the target scenario as is the display of emotions in team coordination. Social cues enhance likeability and perception as well as cooperation. Moreover, implicit communication and explicit communication are key aspects in designing the "Dance with me" and "Follow me" scenarios. [29] [23] [20] [47]

Another paper adds more arguments to the aspect of robot navigation, human obstacles and navigation based on instinct and social rules. Failures lead to lower scores of perceived quality, interaction initiation occurrence and efficiency in cooperation. [33] [6]
How a robot performs an action is just as important as what it does, thus a look into social navigation behaviours is important. [33]
Robots have to predict the walking behavior of people based on a study that offers insight into how to correctly label human reactions in HRI. Positive aspects of robots as interaction facilitators are also enunciated here. [34]

The necessary questionnaires for the replication are also looked into, with the addition of an extra one looking to evaluate participant impact and feelings. [**Panas**] [37] [59]
Other related work looks into the aspects of robots displaying emotions to generate empathy and be socially accepted. Hence, analysing different methods for non-humanoid and non-verbal robots to be agreeably perceived and integrated. [18] [15] [7] [3] [19] [56] [42] [30]
Furthermore, this leads to investigating robots in roles as dance partners, during leisure time or dance therapy [58], considering the benefits, outcome and design practices for the realisation of the scenario incorporating dancing. [10] [48] [57] [50]
The study itself is not without limitations, as the method of "Wizard-of-Oz", for controlling a robot from afar to simulate its proactive independence and cognition is flawed by the Hawthorne effect - people change their behaviour because they feel observed nonetheless. Many other tips and best practices are collected from similar

papers supporting this evidence. [21] [36]

Thereafter, papers regarding usage of robots in healthcare are analysed to offer more insight into the current robots, their capabilities, interaction and experiences with them. [32] [55] [53] These aspects are further analysed from the prism of past studies and research regarding the human perception on and interaction with robots. [2] They also offer insight into previous results regarding different young and old human groups and their perception of robots. [54] [1] [27]
The novelty effect is also crucial in understanding previous and expected results when facing individuals with robots for the first time in their experience. [1] [16] [40]
The robot behaviors are constrained by the limited motor capabilities of the robot, but they can also take advantage of expressive options that are not available to humans. Mimicking animal and human behaviour and signal cues for a successful non-verbal communication are investigated in these papers. [44] [45] Social intelligence in particular correlated with human-robot interaction in robots is also investigated as part of better understanding how robots are perceived and act, focusing on the legibility, predictability and perceivability of their action. [14] [11] [15] [39] [12] [17] [13] [16] [8] [48] [4] [3] [41]

The close ties between HRI and social and cognitive psychology are tangible when considering the replication crisis. This phenomena will be researched in the experience of previous HRI studies, as well as studies coming from the social sciences field, where it is originating from. Best practices are also assessed to better setup the experiments based on the replication. [24] [25] [38] Considering how psychology impacts HRI research, the principles of anthropomorphism, social cues and behaviours regarding humans and robots are also investigated in particular papers. [46] [57] [49] [22] Person-robot communication, with all its challenges, forms it takes and aspects, is also highlighted as a key research point in the design of the scenarios for this paper. [27] [3] [6]
To provide references and arguments to each of the topics presented in this work, supporting evidence from previous studies and research is provided from among numerous papers. [40] [7]
Examples of social robots, their physical digital expression (phygital) and experimental projects are offered in every paper, with some sources serving mainly as impacting examples of it. [52] [8] [48] [6] [57] [50] [3] [19] [43] [31]

Chapter 3

Methodological Approach

As previously stated, the aim and scope of this master thesis is to address, evaluate, and expand upon the non-verbal communication and interaction between humans and non-humanoid robots while considering the limitations of past studies and the phenomena of replication crisis in this field of study. To this extent, the research questions involved in investigating the goals, as well as the methods, materials and samples used, are described in detail in the following sections.

3.1 Research Questions

To reach the aims of this master thesis, the following research questions are proposed and explained in relation to underlying assumptions and outcome expectations:

- How accurate are the results of a reproduced study in comparison with the original one, by Faria et al.?
Replicating studies is assumed to be a complex tentative, requiring a cloning of the setup and setting in order to maintain the fidelity to the original attempt and its outcome, to obtain similar results pointing towards the same conclusion. Nonetheless, the key aspects influencing a study are in actual variables, encapsulating factors such as participant background, characteristics and details, robot similarity, type, design, action and reaction, as well as the environment and context assimilation. Such an approach is hard to accomplish, due to irreproducible exact interaction, communication, individuals, context, environment and bias. Thus, the expected outcome is a validating one for the replication crisis phenomena as an acknowledged by and present in the area of HRI, as well as similar but deviating from the original findings and their analysis.

- How valid and trustworthy is the original study by Faria et al. considering a comparison with its replication?
Following up on the previous question and its assumption, one can further assume that even if the results were similar, if not the same, what guarantees their validity, considering that the participants, interaction and setting are not exactly the same. This opens them up for investigations on causality, coincidence, validity and interpretation. The participants are not the same, the robot is in itself expected to act similar but showing the influence of the researcher at the same time; it is hard to narrow down the actual results to support the

original study claim, if there are many other narratives and conclusions to be drawn out of the replication. Following, the original study is to be visualised as a foundation, stepping stone, onto which, if attested correctly, future work can be implemented in a plausible and respected way.

- How well can a different non-verbal scenario with the Sphero robot be interpreted, replicating a previous study?
  Considering the reproducibility of the initial scenario with a target group, assuming that a second scenario and a different target group are added, it is expected to yield similar to equal results, as both scenarios interaction and approach are fundamentally novel to the recruited participants, as it has been in the original study for the original participants. Moreover, the outcome is expected to describe, fund and prove the efficacy of non-verbal communication and interaction between humans and non-humanoid robots, given a considerate and adequate interaction, behavior and communication, both in design theories and practice processes.

3.2 Participants

The original study [26] recruited 31 student participants from the university it was carried out at, as 90% of them were students and 10% student workers, with a prevalence in the area of science and technology, namely 84%. The recruited gender pool was biased, with only 11 identifying as females, while 20 as males. Their age varied in the span of 18 to 26 years old, with the relative young mean of 22 years of age.
Out of these participants, a total of 17, 12 males and 5 females, interacted with Sphero, while the other 14 recruited interacted with another Sphero product, namely the Star Wars inspired BB8. [60] The latter one is a legacy product, no longer officially distributed and supported by the company.
Hence, only the 17 participants and the outcome following their recorded interaction and analysed results are of importance to the replication of the study.
This master thesis aimed to not only replicate the original study and research human-robot interaction and design, but also provide neutral, valid and trustworthy results, eligible as a foundation for future work and undertakings. Thus, bias had to be removed from the original study, the participants were to be as balanced and diverse as possible.
In this regard, an initial participant cap of 20 persons was set as the goal for the study, wrapping it up at 30 instead. These were composed in equal measure of males and females, 15 of each. The background of the selected participants was varied, with only 13.4% of them identifiable as full-time students, stemming from the Technical University of Vienna.
Given the age differences between myself and the original researchers, as well as of the participants, two groups were defined, in order to faithfully analyse the results compared to the original study, as well as expand their target group application and gather information on an older group of participants. However, in the end of the planning and undertaking of the replication study, the following two age groups were defined accordingly, based on the classification of generations [61] offered by Michael Dimock et al.:

1. One young group, close to the original one, with an age span ranging from 18 to 27, with their mean of 25 years of age. These are part of the Generation Z, starting with people born in 1997 and onwards, yet with an allowed margin of 2 to 3 years of consideration, starting with 1994, due to the blurred line between generations not being such a strict boundary.

2. One older group, not elderly but belonging to the previous generation, namely the Milllenials from 1981 to 1996, given or taken 2 to 3 years as well, considering their age span to be between 28 and 43 years of age. In order to keep the results as unbiased as possible, professors, associates, doctors or lectures from the institute or the university at large were not part of the target participant group, hence not invited to partake in this study, for as their related specialisation to not influence the outcome.

The methods of recruiting the participants were as varied as proven successfully. Official invitations were posted on social media platforms, e-mail distributions and forums, both private and official, belonging to different universities and work places. Individuals were sometimes physically recruited in the hallways outside the testing lab, or in other areas of everyday life. Word of mouth helped spread the invitation to further participants, as the proposal of interacting with a robot proved to be very attractive.
To explain the latter affirmation, copying the original study likewise, the participants were only informed that they would be interacting with a robot in the scope of observing, capturing and analysing an unfolding of human-robot-interaction, leaving the other details to add up to the novelty effect and the first impact of the Sphero robot and them.
Each participant agreed traveling to the conveyed test lab location and partake in the study. They were offered an information sheet, as well as a consent and data privacy and protection declaration, to which extent they are to remain anonymous and their data stored securely.
Only the recorded video footage, audio files and written observations, as well as the analysis of their participation and interaction was required and disclosed as part of this master thesis. This ensured heightened comfort, ease of participation, trust between the researcher and the participant, as well as a relaxed atmosphere over all, to facilitate natural and unique reactions, both physically and verbally, in the unfolding of their interaction with Sphero.

3.3 Materials

The various materials used in the study replication and research include digital and printed documents, chocolate bars, hardware, as in a laptop, Sphero robots and the camera used for recording, and software, as in the app to control Sphero with, the browser, a stable internet connection and university access to scientific articles in the research of the bibliography, YouTube for music and various social platforms and chat providers for recruitment of and contact with the participants.
Two Sphero SPRK+ Robots were provided for the study replication (*Figure 3.1*), to be able to carry out the experiments without interruptions of the need to abort the trial, due to low battery. This helped with tightening the schedule, being able to

plan more participants in a single day in rapid successions of replaying the original study. These are a newer version than the original Sphero 1.0 used in the paper; although there is no difference between them except for the exterior transparency.

Figure 3.1: The Sphero robots used in the study replication

In order to make use of "Wizard-of-Oz" technique, the official Sphero Edu Application was downloaded and used to control the movement, lights, colours and flashing of Sphero. The version used was the official one found in the Windows Microsoft Store through their official webpage. [62]
The laptop was crucial not only for controlling the robot seamlessly, without the participants noticing, but also for researching the literature online, using the Technical University of Vienna's access to diverse libraries, as well as for playing the music during the study, preparing all the documents necessary, and recording and analysing the footage from during the unfolding of each study experiment.
The ideal choice for this undertaking was a 13" inch laptop with touchscreen, in order to best mask and hide any control of the robot and multi-tasking during the experiment. The song was played on it from the online platform YouTube, while the external meetings camera was connected to the laptop and controlled through it's own Camera application.
Besides the equipment listed so far, chocolate was procured and served, for to each participant, either during the experiment or after it, depending on the outcome. It is not known what the original study offered it's participants to thank them for participating, so Kinder line of chocolate brands was selected as fitting to thank the participants for their time, due to it's availability, packaging, good taste and positive reputation.
Worth mentioning as well, the necessary information sheets, consent giving and data privacy and protection, also the forms to be filled in by the participant and the questions of the semi-formal interview had to be created and printed for each study run, derived or inspired of course by the original one's drafted for the initial study. This required additional German translation of every document in comparison to the original study, as the participants would prefer German over English

and vice versa, with the exception of the "general data protection and regulation", abbreviated as GDPR, information sheets, which were printed only in English and presented in case of questions. Each participant was handed their own participation content and information sheet, in addition to the kept by the researcher.
This amounted to 10 sheets per participant, with the final folder accumulating over 300 sheets and being stored safely at the university.
The recorded footage of the study is likewise kept privately, in accord with the GDPR rules, in a private repository of the university.
The implicit use of stationery, other common input and output devices, as well as objects not worth mentioning, is not being described here, as it is not important for future replications of this master thesis.

A more detailed explanation on the conception of each document, the design of the scenarios and choices made can be found in the following section.
A digital version of the documents and sheets templates created for the study recording and analysis can be found in the *Appendix*, both in both English and German.

3.4 Procedure

In order to carry out the study replication, a broad research was carried out, studying specific articles and papers in depth to gain a better understanding on the related topics and interdisciplinary field of HRI.
Afterwards, the original paper "Follow me" [26] was read through multiple times, extracting, sorting and classifying all the information relevant for the lab experiment. Unfortunately, the original paper was not comprehensive at all in its procedure documentation; many pieces of information had to be extracted from different sections of the paper, while substantial details were completely left out and supposedly taken for granted and understood by any reader, not having partaken in the original study. This hardened the task of understanding what was done a lot, leaving many aspects to open interpretation and self decision, all discussed under the included section *Limitations* to follow.
Having gone through breaking the study apart, two target participant groups and two scenarios were identified, respectively added, to the replication attempt. The design of each of these scenarios is described in detail, adding the justification for the choices made, in later subsections.
Afterwards, a test lab environment was decided upon, arranging the entire space as seen fit and preparing the equipment accordingly.
The documents were meanwhile prepared and printed for a trial run. The goal was to copy as close as possible the original work done, hence the original documents, while expanding on these with additional questionnaires and informal interview sessions as seen appropriate, in order to gain a more meaningful result, both qualitative and quantitative.
The scoring itself of all the data gathered follows the tables, figures and findings from the original paper, adding notes, supplementary knowledge and even new gains along the analysis of the study replication's outcome.
The entire study replication tentative underwent an initial pilot testing phase, where two trial study sessions were carried out with volunteers from the institute of "Human Computer Interaction Research Group" at the Technical University of Vienna.

Lastly but not least, all known deviations, limitations and differences in the actual study compared to the initial one, recorded from the beginning on and accumulating as the study unfolded, were gathered together, synthesised and reproduced in the following section of this chapter.
The following sections detail all the steps mentioned above accordingly, aiming at recreating the missing picture from the original paper and, together with it, as a validating and corrective iteration of it, offer a solid and trusting foundation for future research based on these attempts and their findings.

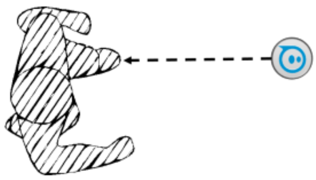
3.4.1 Study Replication and Expansion

The original "Follow me" study is compared in detail to its reiteration under the section *Limitations*. Nonetheless, to in/validate the crisis replication, the details from the paper had to be extracted, pieced together and the holes filled in, in order to obtain a trustworthy recreation and outcome. To this extent, besides adding the second target group for a wider range of response and analysis, two scenarios were fleshed out.

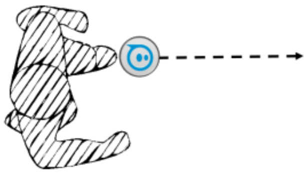
First, is the scenario of the robot asking the participant to follow it, as presented in the work of Faria et al. [26]. Here, the robot makes use of physical touch and lights, in order to convey its intention to the participant, as it is summarised in *Figure 3.2*. The robot's communication is designed as it moving towards the person to start the interaction, bumping into them if required to get attention, followed by walking away. The replication expanded upon the language conveyed by the illumination, filling in beyond the emotions of happy and sad, as mentioned in the original paper. The robot is flashing to signalise emotions, varying the colours accordingly. These were rudimentary presented in the paper without a specific reason. However these were used as a inspiration for the study redesign and redefinition, being attributed more legible meanings. Originally:

- Blue meant attracting attention when flashing it, but it could be also used when the participant reacts accordingly or the robot is greeting them correspondingly or the participant understands Sphero or to more clearly communicate joy, related to the attention received.
- White or very bright yellow signaled happy, whereas, due to the similarity with white, it would be used as rather a neutral colour, to be used when the robot is simply interacting, moving about or being neither happy, nor upset.
- Red translated to sadness, that the robot dislikes something that the participant does, or that it has an error combined with flashing, example giving being picked up, kicked, blocked or just not followed by the participant as intended; a pattern which was reimplemented as such.

With the participant initially seated by a table with one of the researchers, if they did indeed stand up, interact with and follow the robot, once confronted with Sphero, they would be as a following led to a sweet reward, signaling that the experiment had ended. They would be then offered some questionnaires to fill in and would be then free to leave.



(a) The robot goes towards the person.



(b) The robot bumps into the person and walks away.

(c) The robot stops.

(d) The robot flashes/nods its head.

Figure 3.2: Robot Behaviour to ask a Person to follow it [26]

The second scenario was added on the wish of extending the work carried out, looking for a second interaction chance, going beyond short-term interactions, the novelty effect, the Hawthorne effect and the bias that the participants might initially have. As reasoned in the chapter *Related Work*, a scenario incorporating dancing was decided upon to allow the participants to interact in a different scenario, where the robot was playing the role of an assistant, an entertainment social robot, switching the paradigm from the initial companion one to the care-taker one. This scenario is refered to as "Dance with me".
Participants were now superior, as they did not have to understand the intention of the robot, and predict its actions accordingly. Instead, the legibility would be tested. Hierarchically speaking, the participants were now superior to the robot compared to the previous scenario, where they were situated rather as inferior in relation to it; thus they would have to understand what the robot tried communicating with no specific intention behind it.
The robot would again make use of its body and illumination expression to convey what it intends, but the rules will have changed. Whereas in the "Follow me" scenario they would move, turn and be driven around by a remotely hidden researcher, this time its movement has to fit the music, leaving the design of the dance to the interpretation of creativity of the researcher controlling the robot, as in the "wizard".
Its movement pattern was thought out to be more fluid and pattern-oriented, to symbolise a learnt choreography. Circling wide or spinning in place, as well as approaching the participant in a more playful way, moving in between their legs and adjusting itself to their involvement, rather than just bumping into them, were added with the aim of pretending Sphero's appearance as an entertainment assistive robot as best as possible. The lights would change from just blue, red and white, employing a full rainbow palette, making use of every colour available, such as green, yellow, orange and pink. Their meaning was the effect of excitement, party and dance mood, disco atmosphere. Although it would still flash red if picked up or

confined in place, red would be displayed also while dancing otherwise, also easy to be mistaken for under a similar pinkish nuance.
The song replayed during the "Dance with me" scenario was chosen at random, basing the decision on the popularity and accessibility of the genre it represented, namely upbeat, dance-inviting, modern, electronic funk. [63]
This song had been heard in various content created on YouTube and other trending platforms in the past, making it possibly a recognisable choice. The rhythm and melody were found suitable for the physique of Sphero, allowing it to take advantage of its ball shape, in performing various turns, spins and fluid movements, fitting naturally to this type of song its appropriate dance style, while also enhancing the feeling of throwing a party, being at a discotheque and overall feel-good, dance mood with its illumination capabilities. All in all it was expected for it to add to the excitement and immersion during the study.

Having two scenarios at disposal to test, yet looking to replicate the original questionnaires, for the sake of a well-founded quantitative and qualitative analysis and knowledge gain, as well as inn order to avoid bias during the study experiments, the participants were split into two groups, composed of a balanced mix of age. One group would start with the original "Follow me" scenario and the other with the "Dance with me" scenario. This way, the research would faithfully and flawlessly replicate the study while also expanding and improving on its outcome.
The segments of the experiment were allocated time based on the following planning:

- Starting with the introduction, including signing the consent and filling out the personal information, 3 minutes were allocated.
- Then, for the "Follow me" scenario, 2 to 3 minutes were foreseen for Sphero to try thrice to suggest the participants to follow it.
- Alternatively, for the "Dance with me" scenario, a fixed aim of 1 minute and 30 seconds was determined, so as for the song to play through it's first verse and chorus, before starting to repeat itself.
- To fill in the first questionnaire post the first scenario, as well as discuss the first questions related to it, 6 to 10 minutes was the intended duration.
- For the final questionnaire and the last questions another 3 to 5 minutes were planned.

3.4.2 Laboratory Experiment

The laboratory experiment is looking to recreate the original study setting [26], while expanding it by one more scenario and studying it with participants with a broader age range than in the original publication. To successfully carry out the experiment reproduction as a singular researcher, the following methods will be applied:

- "Wizard-of-Oz" - manipulating the robot to simulate its liveliness, spontaneity and motion fluidity.
- Video and audio recorded observations - in order to focus on being the "wizard", the entire interaction will have to be captured without the participants

knowing forehand that they are being observed, in order to avoid the Hawthorn effect.

- Questionnaires and interviews - both from the original paper as well as new ones added to, detailed in the next section.

In order to replicate the study, a test laboratory, similar to the one presented in the original paper, was required. This was described by Faria et al. [26] as being similar to a living-room, interior design wise, offering a homey feeling. The only available suitable space at the institute of the "Human Computer Interaction Research Group" was the onsite library. Fortunately, looking at the picture of the room provided in the original paper, this library offered the same furniture, as in chairs, tables and books shelves, with plenty of open space in the middle; validating the use of the library as being adequate for the replication attempt. The room can be viewed as captured through the camera on the laptop screen in *Figure 3.3*.

The test lab was prepared before the start of the study session on a given day,

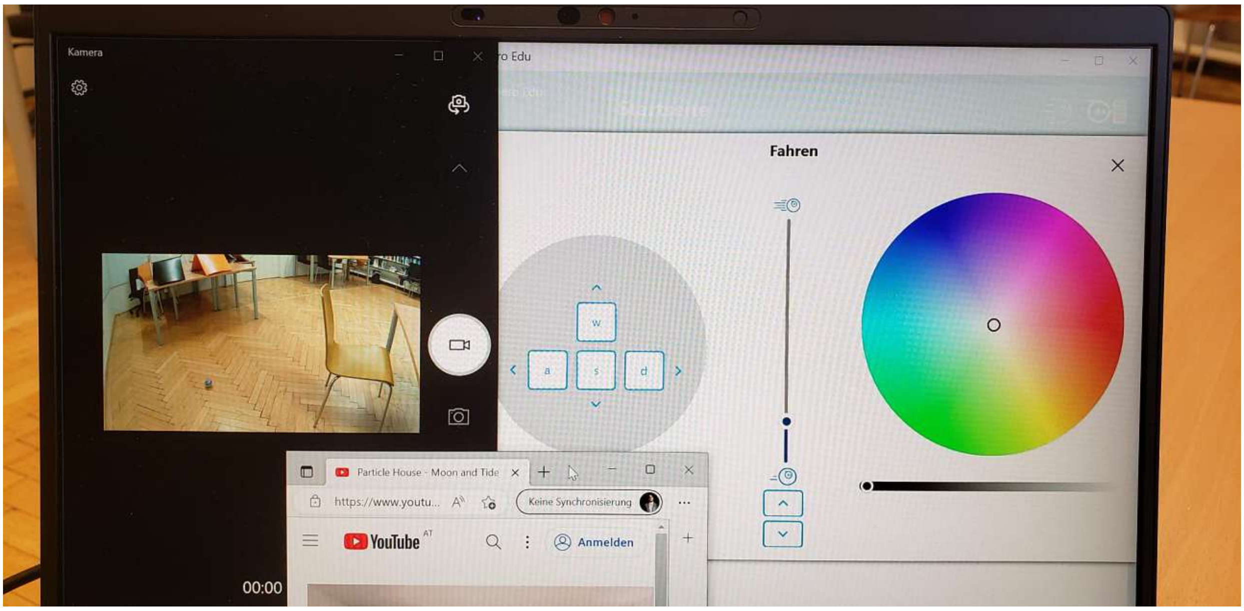


Figure 3.3: Experiment Lab from the Researcher Perspective

taking around 30 minutes to prepare the equipment. The recording camera was positioned in the corner of the room, next to the researcher table where the participants were also seated initially and during them answering the questionnaires and interview questions, kept at hand on another chair next to the participant, as presented in the *Figure 3.4*.

The Windows Camera application was used to capture the widest possible image of the entire interaction space, resuming it more to the left side of the room than the right one. The Sphero Edu App was opened next to the camera to control the robot through arrow keys, while using a finger to change the brightness and colour simultaneously. Thus the aspect of "Wizard-of-Oz" should not be as detectable as using a mouse and hearing clicks. Sphero would approach the seated

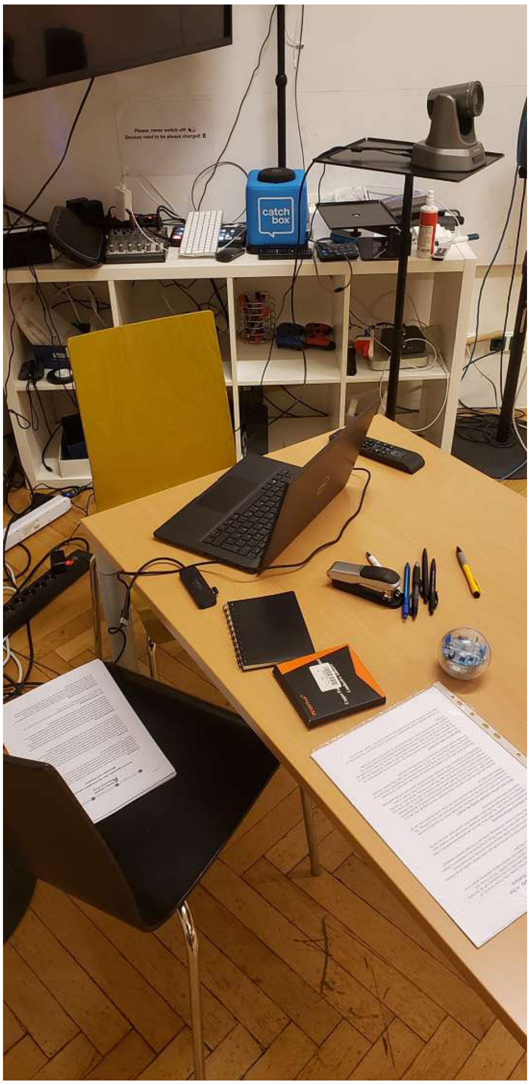


Figure 3.4: Researcher-Participant Corner

participant, see chair in *Figure 3.3*, from the tables behind them. YouTube was also opened and ready, with the song for the "Dance with me" scenario readily loaded, just to be started when necessary and then hidden in the background again. The researcher and participant corner, with the whole setup, can be observed in the *Figure 3.4*. The library room reserved for the study offered a space of 30 square meters, posing enough space for the interaction with Sphero to unfold, as well as somehow to recreate the original experiment room. Lacking a wall to create two rooms, this was improvised using chairs and tables available in the room, behind which the end sign for the "Follow me" scenario and the sweet reward were hidden, see *Figure 3.5*.

Other than that, the other chairs and tables available, except for the one in the corner where the participant and researcher are seated at, were pushed to the other corners in the right side of the room, by the bookshelves. This again copies the original setup presented in the pictures from the paper very faithfully. The original setup has been presented in the introductory chapter, namely in *Figure 1.2*.

To offer more information on the two scenarios tested, these are namely a "Fol-

Figure 3.5: Reward Corner for "Follow me" Scenario

low me" and the new "Dance with me" previously mentioned. This study aimed to collect data of at least 20 different individuals. To this extent 30 participants were successfully recruited, leading to a well-founded quantitative and qualitative analysis and consistent knowledge gain. In order to avoid bias during the study experiments, the participants are be split into two groups, with a balanced mix of gender and age. One group will start with the first "Follow me" scenario and the other with the "Dance with me" scenario. This way, the research incorporate less to no bias in the findings and their analysis, compared to the original study.

The "Follow me" scenario was designed according to the scenario sketch available in the paper, to be seen in *Figure 3.6*. Hereby though, Sphero is supposed to approach the participant from behind, to startle them, attract attention and shift it away from the "wizard". Moreover, the participants are asked to stand up and interact with the robot at the beginning of each scenario, to further guide them, signalise the different phases of the experiment and shift focus from the researcher to the robot. Then, the first part of the interaction happens in the open space between the tables, which can be caught on camera. From here on, the robot attempts

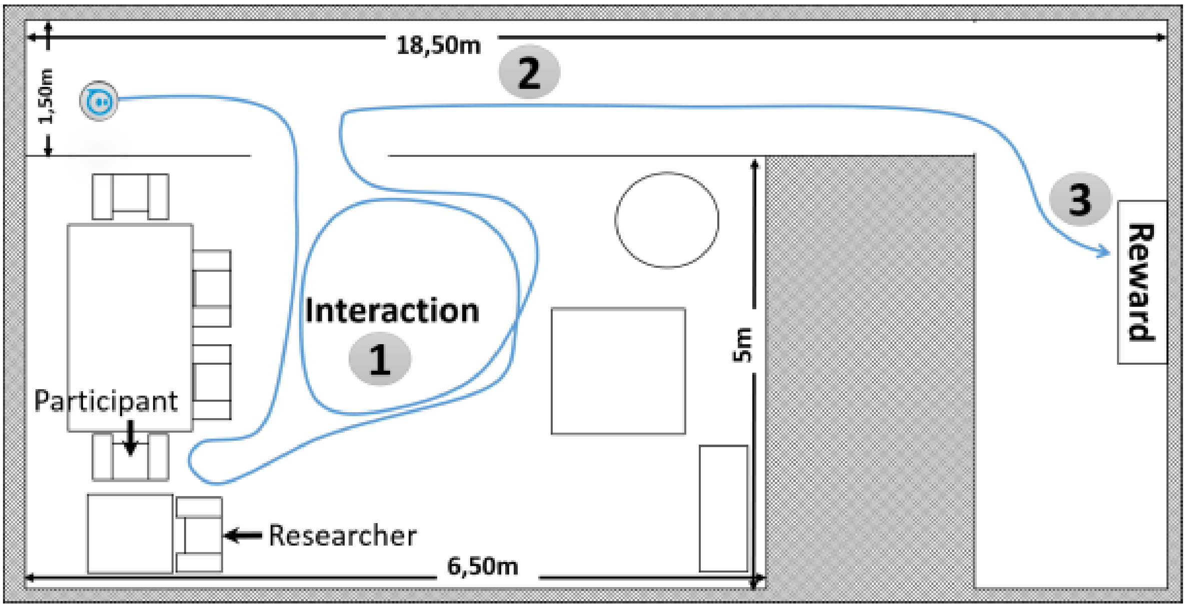


Figure 3.6: Study Experiment Scenario Sketch

to lead the participant, make them follow it, namely step 2, and on to the hidden sweets behind the chairs, as in step 3 in the figure, where the scenario is concluded, Sphero's lights are turned off and the movement is halted, to fake it disconnecting for the moment being, and the participant is asked to return to fill in the first questionnaires and forms.

"Dance with me" sees Sphero manually placed by the researcher nearby from where the participant is seated. Then, the lights are turned on again, while the participant finishes filling in the last form, and the robot starts spinning, rolling around, driving in and around the participant's feet, while changing colours and adjusting it's brightness in rapid succession, to give the impression of a disco globe dancing. This interaction resumes to the area presented by the first step of the interaction under *Figure 3.6*. Then the participant is seated again to fill in the remaining questionnaire and discuss through the short informal interview questions. The same principles are applied when the scenarios are carried out in reverse order, regarding Sphero mimicking being disconnected and being turned on at the given signal.

3.4.3 Questionnaires and Interviews

Prior to and during the experiment, the participants were only informed that they would be interacting with a robot in the scope of observing, capturing and analysing an unfolding of human-robot-interaction, leaving the other details to add up to the novelty effect and the first impact of Sphero on the participant. The other explanations, questions and feedback were left for the debriefing. To this end, many questionnaires where employed both in the original study and added additionally in its replication, to capture the novelty of the interaction with Sphero.

Replicating the original study, the participants will have to be handed out two short questionnaires, one on proximity to the robot based on the *"Inclusion of the Other in the Self Scale"* [59], and the other on the Godspeed's likeability and intelligence perception dimensions. [37] To better gain an understanding of and insight into how the interaction with the robot really went, from a first-person's perspective, the participants will be shortly interviewed at the end of the experiment, allowing for the collection of more qualitative data to analyse the findings and outcomes a posteriori. To this end, an accompanying PANAS [**Panas**][64] scale will be used to collect and asses the interaction feedback and emotions of the participants with the possibility of relating it to the analysis of the interview questions themselves.
In any case, they will be asked to fill in the original questionnaire after their respective first scenario, whichever it may be considering interchanging this to avoid bias, and rate the robot under the novelty effect, as this is the most raw and unbiased data to be collected in this study. The PANAS table listing emotions is presented after each scenario and compared amongst itself, to investigate differences and changes post the novelty effect and a prolonged interaction with the robot, given a scenario, thus context and interaction type, change.
All used documents, including questionnaires and interviews, are attached to the *Appendix* of the thesis in bilingual language, and are described in detail on their structure, reasoning and intent in the following paragraphs.

In the beginning of the experiment, the participants are handed out an information and consent sheet, as well as an optional GDPR document, if requested. These are meant to inform and ask for consent on anonymously using the data collected, as well as them optionally subscribing to receiving a copy of the thesis once finished to view the work they have contributed to.
A one-pager is afterwards handed out, looking at gathering non-sensitive information on their persona, such as age, gender, occupation or profession, study or expertise background, past contact and interaction with robots, and knowledge and experience with Sphero, the latter three enumerated with the added mention of if any.
After their first scenario is completed and Sphero's lights were turned off to simulate it being inactive, they are invited to fill in the first questionnaires:

- Godspeed - aiming at capturing the perspective of the participant towards the robot, its intelligence, likeability and animacy are captured in 16 scales, built out of five possible options on a scale from a negative to a positive opposite feeling, such as e.g. *dislike* - *like*.

- Proximity - making use of a diagram provided in the reference work of Aron et al.[59] This investigates the closeness the participant felt to the robot viewed through the prism of it being a social entity. The circles in the diagram, representing *self* (the participant) and *other* (the robot) are thus drawn beginning with separate in more and more overlapping iterations. The rating values could be structure based on the classification of space into social, namely 1 and 2; personal, namely 3,4 and 5; and intimate, namely 6 and 7. [6] Hereby, the values of 2-3 and 5-6 are border ones, possibly being valid for both of the space classifications it stands between.

- The PANAS table, featuring 20 different emotions and four scales of intensity is handed out for the first scenario. To distinguish which table was filled in for which scenario, the pages are marked at the bottom with an *F* for "Follow me", namely with a *D* for "Dance with me" respectively. Should a feeling not have been sparked in the participant, the row would be left blank and skipped, resulting in feelings and their intensity being marked only if they also occurred. While the original referenced PANAS table is more extensive, removing the fifth column of "not at all" helped reduce both the visual cluster and the workload of the participant, as well as of the researcher in their post-analysis.

To help with the workload, the participants are allowed to enjoy their sweet reward. After this first break the second scenario is started, following the same procedure as the first scenario, with the robot's light turned off and the participant invited to take a seat again. At this point there is only the retake of the PANAS scale and the informal interview left.

- The other remaining PANAS table with the other marking according to the scenario is handed out. This time attention is payed to the differences between the answers now and those of the previous PANAS sheet, kept close-by in sight range. These are investigated and asked upon either after the participant has finished filling it in and/or during the informal interview that follows.

- The interview is meant to substitute the questionnaire, regarding the two questions presented in the paper, of why the participant thinks that the robot bumped into them and flashed. These questions are duplicated for the "Dance with me" scenario, with an added question of what the participant made out from the robot dancing, its intent, perceivability, predictability and legibility. The questions can be observed in the *Appendix*. These were: to "When the robot touched me and moved away was for:", with the expected answer "ask to follow it", and "when the robot flashed was for:", with the expected answer "ask for attention". The questions were posed for both the "Follow me" and "Dance with me", with an added questions on the interpretation of Sphero dancing.

Having progressed through all documents, the experiment is concluded, questions and are answered and feedback is discussed, while also disclosing the real nature of the experiment, thesis aim and "wizard" controlling the robot, faking its consciousness and aliveness.

3.4.4 The Scoring

The data collected during the experiment was of quantitative and qualitative nature and follows the same analysis procedure as the tables and example given in the original paper. As these were not always explained in enough detail, many of them were deduced from the text provided by the researchers; solving the puzzle of putting the pieces together required logical reasoning, close investigation and grouping of the information provided, as well as filling in the gaps and lacking guidance with the best solution reasonable or relatable.
Hence, the topics and data tables to be approached and compared to the replication results include an analysis of participants, unfolding of the study and findings, as follows:

- Participants are extensively investigated and classified, based on their quotas, age, gender, occupation, expertise, previous knowledge and interaction with Sphero, plus an additional questioning of their knowledge and interaction with robots in general.

- The intelligence, likeability and additionally the animacy of the robot from the perspective of the participant, offering an overview of the mean and standard deviation for each of these, as seen in the original paper.

- The proximity felt by the participant towards the robot, considered a "social entity". To this end, the visual representation of the results in the form of a bar chart is represented in the same manner, to stay faithful to the original paper.

- Time stamps set throughout the recorded footage, aggregate to determine a mean time participants needed to decide to interact with Sphero in each given scenario, as well as the duration spent doing so and, in the case of the "Follow me" scenario, the time taken to follow it to the finish sweet bowl.

- It will include a description of how the participants reacted and interacted, focusing on whether the participants followed and/or interacted with Sphero or not, aiming at observing the relatedness between speech connected with physical and speech connected with behavioral reaction.

- The quantitative interpretation and analysis of the answers to the questions regarding the behavior of the robot viewed by the participant, with the added duplicate and extra question on the new "Dance with me" scenario, all arranged under the form of a closing informal interview. Additionally, an in-depth review of the findings and their related conversation is supplemented as well.

- An assessment of the interaction between the participants and the robots, namely the verbal, physical and behavioral reactions that the humans had towards the robot, are extracted from the video footage and written notes and presented summarised in a table, following the design of the one in the paper, to compare the two.

Besides, other aggregated results classified under different names and groups of provenance and relevance are added, respectively too vague and non-correlated aspects are removed, as mentioned below:

- The participants were asked to associate the robot's behavior with intents, given five options. As these options are not made clear anywhere in the original paper, they are presumably represented by the Godspeed questionnaire, featuring scales with ranges containing five options likewise. Additionally, these are supplemented by both the informal interview analysis and the table describing interactions between the participant and the robot through speech, physical and behavior association.

- The PANAS emotional impact tables collected from each scenario are aggregated for each participant group individually, as well as combined, and presented in its own paragraphs and charts.

- The informal interview is presented and discussed from a qualitative analysis perspective, aggregating common affirmation and pointing out unique and outstanding knowledge gained, making use of tables and lists.
- A Binomial Test is mentioned to have been performed in order to test the significance of the results. This, however, was not described in an understandable way to recreate it, hence it is omitted from the findings analysis.
- As the participants are only interacting with one robot, it is not necessary to include the Pearson's chi-square test to attest whether or not there is an association between the robot interacted with and the proximity felt towards it, nor the ANOVA calculation, to spot differences in the Godspeed results between Sphero and the BB8 in the original study.

The presentation form of these results is adjusted accordingly, where left unmentioned, and is displayed under the chapter *Findings*.

3.4.5 Experiment Pilot Trials

Two pilot runs were undertaken one week before the official start of the replication study. These were carried out with invited student volunteers from the same institute this thesis was written at. It helped test each of the scenarios, as well as to sense-check the documents and questionnaires, receiving much appreciated feedback which ultimately iterated and improved the study to its final form, as presented to the actual participants.
To start off with the feedback suggestions and the "Follow me" scenario, ideally participants would be guided through the paperwork and offered some assistance before, during and after the scenarios, should they have any questions or remarks. As a researcher it is expected to not only sit there and carry out the interaction with the participant in a very rigid and formal way, but instead open up, become more friendly and help them feel relaxed, comfortable, thus willing to interact. This led to the assumption being made, that it also helps diminishing the Hawthorne effect, while it drives focus away from the researcher to the robot, which drastically increases the success of the "Wizard-of-Oz" technique.
To this end respectively, with Sphero positioned behind the participants at the start of the experiment, and its lights turned on, although startling, it definitely helped shift the attention 180 degrees, so that the participant would not notice the other person operating the robot.

Not lighting up in the beginning to possibly attract attention was a good idea, one which could be repeated between the scenarios. Sphero was not to be moved so as to not attract attention; instead the participant's focus was then directed at the facilitator and the paper sheets handed out to them. This also hinders them from changing their opinion, having had a very short break from interacting with it. Light is thus used to signalise the awake/sleep states of Sphero.
As Sphero was viewed as a pet toy, or similar to a digital pet, it helped redesign its movements and reactions when being picked up or blocked, while also leading to it being perceived more inaccurately in action intention. By rotating in place, even though unable to move, its intention and dislike became legible and lively. This

however also lead to it being gendered, anthropomorphized by the participants.
As the test run participants would pick it up or interact with it in a similar physical way, the colour scheme of red for negative emotions and blue for positive ones proved effective, however the flashing of the light should happen only after the colour has changed, in order to make the transition clear to the participants.

Moving onwards to the second scenario, "Dance with me", it was observed that participants might rarely feel the ease to start dancing with and to Sphero, which led to the adjustments of the second set of questions in the informal interview, regarding the participant's perception and understanding of the dance. Instead of asking the participants how they interpreted the robot and its actions, they would instead be served directly the information that the robot was dancing, being thus asked what they made of it, hinting more towards the obvious reply of Sphero being a disco globe. The song's duration was also defined at around 1 minute and 30 seconds, depending on the interaction willingness of each participant.
No matter the scenario carry-out order, the second scenario would always have to be announced in order for the participants to know that the break was then over, hence encouraged to leave their seats and interact more freely with Sphero. In the trials both participant were presented with the same order, also defining the remark to remember to alternate scenarios, as proposed in the undertaking's description previously.

Initially having had efficient but complex layouts for the questionnaires, necessitating less paper but more time and cognitive resources to fill them in, proved to be a failed approach. Likewise the idea of integrating the informal interview into the interaction assessment PANAS table and the time-intensive pre-work, of handing out an information sheet, a consent sheet and a substantial GDPR document to each participant turned out not to wrok well.
Instead, it was decieded for the actual experiment, to put the GDPR would be aside, being open to clarify questions regarding and implications. The information and consent sheet would be merged into a single page, reducing the clustering of information and helping the participant to not feel having to learn too much information at once. Actually reading and going with them together through the sheet before signing it further helped speed-up the pre-study meeting. The PANAS survey would now be dealt in two separate tables, one for each scenario, as described under *Questionnaires and Interviews*. The informal interview would then be carried out separately, distinguished from the papers yet intertwined and connected to the context as the situations arose. Hereby, questions and answers could be discussed further in depth, in found necessary or interesting for the results.

The environment itself was well arranged and designed, using chairs and tables to simulate the environment, hiding the sweet reward behind stacked chairs so as to not been seen when entering the room, as well as placing the robot under the tables out of view were marked as positive ideas. What could not be changed was the robot being adjusted at higher speeds, only 16% of it's speed setting was used due to the nature of floor, being very slippery and limiting the speed choices to the slowest possible for it not to slip but at the same time the fastest to keep up with the participants.

Regarding its lights, colour and flashing or blinking, these were adjusted, namely for the "Dance with me" scenario to blink and change randomly and continuously, simulating a disco globe, while when moving to display blue when happy or asking for attention, gray/white when moving around neutrally, red when picked up, not followed, hurt or sad. When switching to red or signaling happiness or attention seeking with blue it would furthermore flash.

Unfortunately, what had slipped the trial experiments, was the fact that the video audio was often distorted or muted in the recording, due to the laptop being tasked with recording the surrounding and playing the music, leading to overlapping sound waves and hence the failure. Nonetheless, all the information possible was extracted from the document, given that the participant's voice was louder than the music, punching through the mix to be captured intelligibly.
Another aspect remarked only later in the study, was the huge difference in mood, atmosphere, legibility and perception when going from the "Dance with me" scenario to the "Follow me" one, where the passing from loud music to silence impacted the participant in equal measure as the change in robot's behavior did, going from caretaker to companion. In the reversed case, the participant reaction was instead more positive. This could have led to some bias in the answering of the questionnaires and the interaction willingness.
Overall, the implementation of the feedback helped correct mistakes, misconceptions and bad practices and design, reduce the study duration, maintain the participant interested and focused, while also learning do's and don'ts, being a successful "Wizard-of-Oz" in tricking people to believe in robot autonomy, and getting accustomed with the equipment and driving Sphero, so as to be well prepared for the actual study experiments with the signed up participants.

3.4.6 Limitations

The paper, from a replication point of view, is unfortunately very scarce in information, with only superficial mentions of important methods, approaches or ideas, or simply lacking them at all. This left many aspects of the study replication to the decision of design, implementation and based on the best logical solution, given the circumstances. All the differences, weighing this study to the original one, collected during the unfolding of the study replication design, literature research and pilot trials are collectively presented here in the form of an enumeration list, as follows, whereas the main ideas and key words in each entry has been highlighted in bold:

1. **OS**: The study was conducted with **31 participants**, 20 males (65%) and 11 females (35%). The mean age of subjects was 22 years (range: 18-26). 90% of the participants were students and 10% student workers. Knowledge and experience was OK, according to Faria et al. [26] Of these, **only 17 interacted with Sphero**, 12 males and 5 females. Of the participants, 12 had previous knowledge of the robots and 4 interacted previously with them.
   RS: The study is conducted with **30 participants**, split evenly among males and females. The age of 20 of the subjects is in the range of: 18-27, with a mean of 25, the other group varying from 28 to 43. Participants are allowed to come from the university, but only around 20% do so. It is not possible to faithfully reproduce each and every result from the original paper, as the

participants only have, to some extent, the age in common with the prior ones; hence adding some volatility to the findings. The setting environment, tools and unfolding of the replication also differ, given the resources and space available. Moreover, looking at the big picture, it can be said with confidence that the generalized and summarized results are set to deviate slightly from the original ones nonetheless, given that culturally speaking USA's and Europe's, more precisely Austria's, inhabitants are not so different in their perspective and reaction towards robots.
The experience of the participants is also varied and unique; the criteria of them not having previous knowledge of, or at least interacted with, Sphero is straying away from the original study.

2. **OS**: The robot model Sphero 1.0 used in the study has a **blank, white exterior**.
RS: The next iteration of the robot model, namely Sphero SPRK+ has been provided by the supervisor on behalf of the university, with the version at hand disposing of a **transparent shell**. This might influence the participants, being more obvious for them that they are dealing with a machine by being able to see its components and inner workings. Otherwise, SPRK+ and 1.0 pose the same capabilities and features.

3. **OS**: The original study bases the outcome off of human behavior around **pets**.
RS: The robot behaviour is inspired by pets, but the **anthropomorphization** is left to the participant to decide for themselves. Especially the "Dance with me" scenario is a good example of behavior deviating from the source of inspiration being pets, moving more into innovating new behavior tailored to robots, based on their capabilities.

4. **OS**: **Participants interacted with one of the two robots**, namely Sphero or BB8, whereas the results discussion and presentation had been to some extent aggregated.
RS: **Only Sphero is regarded in the study replication**, with results being shown aggregated where meaningful, and split into the two participant age groups otherwise.

5. **OS**: All the participating subjects **signed** an informed consent **previous** to the experiment.
RS: The participant are handed an **informed consent and GDPR document** at the beginning of the experiment.

6. **OS**: The subject starts the experiment seated in a chair and answers a **pre-questionnaire**.
RS: The paper gives no information on the pre-questionnaire at all, assuming that this is related to collecting some background information on the participant, deducing it from the aggregated information provided on the participants. This is organized into **7 questions**, looking to anonymously gather non-sensitive information on the participant's age, gender, occupation or profession, study or expertise, and, if any, past contact with robots, knowledge

and experience with Sphero. The participant is as well seated with the researcher.

7. **OS**: The study was conducted with one participant at a time. The experiment was created using the **"Wizard-of-Oz"** technique, with the robots being controlled by **one** of the **researchers**, as well as using pre-programmed routines. **RS**: There is **only one person** to conduct, observe, drive and control the experiment. There will be no pre-programming, as the robot has to quickly adjust to unexpected and unforeseen interaction and reaction from the participant, make it look more natural and less predictable, but instead legible. Only "Wizard-of-Oz" will be employed, which assures an unfaithful replication even in one's own study, due to different reiteration with each participant.

8. **OS**: When the participant finishes filling in the papers, the **robot enters the room** from an open door located in front of the participant.
RS: In the study replication, the **robot is hidden behind the seat** of the participant, under some tables to cover it from sight.

9. **OS**: **One of the researchers is in the same room** as the participant during the experiment, who is told that the researcher is there only to monitor the experiment. The researcher remains the whole time at a considerable distance from the participant and the robot, providing privacy for their interaction. Moreover, the instructions provided to the participant are that this work aims at testing how people and robots interact together.
RS: **The only researcher is in the same room**, monitoring, observing and controlling Sphero. They are there to provide instructions and guide the participant through both the interactive and documents-filling part of the study. The disclosed scope remains the same, no further information is divulged until after the experiment had concluded, so as to not influence the participant in any way or give away the "Wizard-of-Oz" approach.

10. **OS**: The main goal of the user study was to test if the robots could **correctly communicate** with the participants and **convince** the participants to follow them. To test if the participants understood the robot's intentions, they would need to follow the robot from the **room** where the interaction was initiated and head towards a stand in another division of the lab. This stand had three bowls with candies on top of it and a congratulations message for participants.
RS: Given the current **test lab** environment at disposal, it is not possible to film with one camera in a 2+ room environment. The scenario has to be changed and adapted. Sofas or furniture must be hiding the reward, **improvising a wall**. The reward can fluctuate as long as it is something sweet, hence the Kinder chocolate bars. Whatever the difference in interior arrangement, there must be **enough space** for the robot to dance and move freely. The angle of the filming needs to cover everything. The **researcher** would have to sit as the experiment conductor, observer and "wizard" in the corner of the room. The Hawthorne effect can not be avoided, neither people not realizing that the "wizard" is controlling the robot.

11. **OS**: The study took place in a lab room, which resembles a **living room** of a home. Distinct paths and measures are given for the robot's actions.

RS: The interaction takes place mainly in the central part of a **library room** reserved at the institute where the thesis is hosted, so as the robot can be driven around easily and unhindered. As "homey" as the library room at the institute can look, there is no other option for me. Important to notice is, that after the experiment, the similarity between the picture from the test lab in Fara et al. paper [26] looked almost identical with the test lab of the replication study. Furthermore, especially worth mentioning is that the **camera placement** is crucial to capture the entire interaction in this wide space. The paper gives no details on the distinct paths and measure beyond the *Figure 3.6*, hence, although sizes and trajectories will differ, the unfolding will be similarly recreated.

12. **OS**: It is mentioned that the **Sphero Edu App** is used, whereas an own **Android application** for the experiments was created. The robot was controlled with a joystick, remediating some of the sporadic or accelerated amok movements caused by the touchscreen maneuvering.
RS: However, as no code, sources or further details are provided, and assuming that a "Wizard-of-Oz" approach allows for more spontaneous adjustments and solutions, these will be **omitted from the replication**. Only the official app is used.

13. **Original Study (OS)**: **The pre-programmed routines** were previously described as follows. *Flashing behavior*, when the robot flashes for 0.25s with random colors to attract attention. *Moving in and away from person*, like in *Figure 3.2*, where in each repetition the robot moves back and forth, employing a square-type movement to gain attention. Movements to and away from a person to convey the intention that the robot wants the person to follow it. Expressing two emotional behaviors through a combination of both lights and movement. *Expressing happiness* through a yellow light, colour which is generally associated with joy, and the robot's speed was increased. *Expressing sadness* through the displayed color of red was greatly dimmed and the speed was heavily decreased.
Replicated Study (RS): **The movement and the flashing patterns should not be programmed**, as just a complete simulation using with "Wizard-of-Oz" allows for better adaptability to the participant actions. The drawing in *Figure 3.6* does not convey an adaptable motion and a fluid interaction. At the same time however, it affects the exact reproducibility of the interaction, with the robot varying the flash duration, the movements and their evenness, as well as the reactions from participant to participant, based on what the "wizard" is observing. Moreover, this requires the "wizard" to be in the same room with the participant, and, as this anyway also takes up the roles of the mediator, observer, and guide, the Hawthorne effect is hard to omit, as well as the people not realizing that the robot is being controlled by the other person.

14. **OS**: The **experimental scenario** unfolds as follows: the robot and the participant meet; the latter steps outside the room to the corridor to follow the robot; where they find the stand with candies and a congratulations message,

saying the experiment was over.
RS: The replication study implements the **same script and steps**, with the exceptions given in divergences to the test lab, manpower and hiding the "Wizard-of-Oz" technique. The **participant is given no explanation** if asking questions regarding the robot, to support the perception of autonomy and animacy. Moreover, the replicated study expands with a **prologue**, a **second scenario**, as well as more questionnaires, adding to the value gained and data collected, as well as the influence on the participant.

15. **OS**: The robot does **three attempts** at most to convince the participant to follow it, given the scenario progression above. Each session approximately lasted the same time. As mentioned earlier, there is a **sign congratulating** the participant for having followed the robot and signaling the end of the study. Questionnaires are handed afterwards to be filled in.
RS: The "Follow me" and "Dance with me" **scenarios** will be presented **interchangeably**, for an equal number of times, with each session planned to last approximately the same. The questionnaires of Godspeed and on proximity will be handed after the first scenario, whichever it may be, including the first PANAS table sheet. The second PANAS table and the informal interview are to be presented only after the second scenario, to collect any changes in emotions, perception and feelings, and to wrap the whole experiment up, giving the participants the chance to speak up their mind. This approach also offers the chance of discussing differences between the PANAS tables on the spot and even integrating them in further guiding the informal interview start or unfolding. **Not generating bias** is the goal in this approach.

16. **OS**: The Godspeed questionnaire references only the ***Perceived Intelligence*** and ***Likability*** dimensions. The participants were also asked to evaluate the **proximity** they felt to the robot, with the intent to assess if people felt close to the robot and if they perceived it as a "social entity", resulting in a bar chart.
RS: The Godspeed questionnaire is extended by the dimension of ***Animacy***, especially due to its relevance to the "Dance with me" scenario. The proximity questionnaire is **recreated** from the original paper it is traced back to, where it was originally taken from. [59] Additionally, having the Godspeed and proximity questionnaire after the one or the other scenario yields **different perceptions and interpretations** of the robot accordingly, which are important to note when comparing the results to the original ones. The other two dimensions available in the original questionnaire, namely that on anthropomorphism and safety, were not considered relevant due to the obvious non-humanoid and ball-shape of Sphero.

17. **OS**: Further specific questions were asked to the participants to assess the **readability of robot intents**. The participants had to select one from **five options** in order to associate robot's behaviors to intents.
RS: The **questions and options are not presented directly** but can be extracted from the results discussion towards the end of the paper. These were "When the robot touched me and moved away was for:", with the expected answer "ask to follow it"; and "When the robot flashed was for:",

with the expected answer “ask for attention”. These are built into the **informal interview**, with a duplicate for the “Dance with me” scenario, adjusted to it naturally, and an additional question regarding Sphero dancing and its perceivability. Instead of presenting them in the form of a questionnaire, an informal interview was introduced, to more easily and naturally discuss these aspects in more depth and detail, as well as any other results from the other questionnaires, especially the PANAS one.

18. **OS**: The **behaviors** of the participants were also analyzed during the interactions and measured how long it took them from the moment when the first contact with the robot was established to the moment the participant decided to follow the robot. The results are described in terms of questionnaires and behavioral **analysis**. Different factors were considered for the behavioral analysis. The effectiveness of the robot in persuading the participants to follow it was analyzed. In order to test the significance of the results a Binomial Test was performed, which proved that the proportion of participants that followed the robot was higher than a random distribution with a p =0.001. Furthermore, a Pearson’s chi-square test was performed, to attest whether or not there is an association between the robot interacted with and the proximity felt by the participants towards it; as well as an ANOVA, to investigate whether the participants felt a significant difference between the robots, with regards to the Godspeed questionnaire.
    RS: The **video recordings** need to be timed, structured and classified when analysing them. The Binomial Test and other applied statistics need to be researched for repetition relevance. Otherwise, the results are left to interpret as close as possible, given the superficial information provided and assuming that some work previously carried out is still valid and redundant to recheck. The Pearson’s chi-square method, as well as the ANOVA, are not relevant, as there is only one robot used in the study replication. **Speech, physical and behavior** are the best values for the analysis of the video footage and the comparison with the original study. Differences in culture and language, interjections and expressions, considering USA and Europe are very interesting to note, if observed whatsoever. All other data and questions are to be used from the original paper as a reference.

Chapter 4

Findings

The study replication required 2 weeks for carrying out each of the experiments. These were stacked together in different days, depending on the availability of the participants, as well as of the test lab. In total 7 days were necessary to complete the study replication unfolding. Having two Sphero robots helped with planning consecutive experiments in a given day, with each robot lasting around 2 to 3 hours, ideally swapped with the other one, which was fully charged, after 2 hours and 30 minutes or 3 experiments on average. One experiment run was planned for about 20 minutes, with a buffer time of 10 added minutes, as well as an extra 10 minutes to reset the test lab and prepare it for welcoming the next participant. Summarised, 40 minutes session were held each evening of the 7 consecutive days, as follows:

- Day 1 took 3 hours for setup, 3 participants and cleanup.
- Day 2 took 3 hours and 30 minutes for setup, 4 participants and cleanup.
- Day 3 took 2 hours for setup, 2 participants and cleanup.
- Day 4 took 4 hours and 30 minutes for setup, 5 participants and cleanup.
- Day 5 took 5 hours and 30 minutes for setup, 8 participants and cleanup.
- Day 6 took 2 hours for setup, 2 participants and cleanup.
- Day 7 took 4 hours and 30 minutes for setup, 6 participants and cleanup.

Following the results in the original paper, the Godspeed, Proximity and Behaviour questionnaire followed, given in a somewhat lacking order. The results are cited, found under the sections *"IV.User Study"* and *"V.Discussion"*, from the paper as it can be seen in the following textbox:

"[...] We analyzed the effectiveness of the robot in persuading the participants to follow it. In the Sphero trial, 12 of 17 participants followed it. [...] In particular the average time that the participants took to follow the robots was 2m38s for Sphero [...]
The questions asked were: to "When the robot touched me and moved away was for:" the expected answer was "ask to follow it". However the most common answer, with 48% of answers, was "ask for attention" followed by "ask to follow it", with 32% of the answers.
To "When the robot flashed was for:" the expected answer was "ask for attention". In this case, most of the participants, 74%, answered as expected, followed by "displaying happiness", with 23% of the answers. [...]
The participants were also asked to evaluate the proximity they felt to the robot, with the intent to assess if people felt close to the robot and if they perceived it as a "social entity" [...] The results obtained also show that the participants feel some connection with the robots, grading the proximity, more frequently, as a 3 or a 4 in the scale. [...]
In terms of the Perceived Intelligence and Likability the results of the questionnaires show that the participants liked the robots they interacted with (Sphero: M = 4.40, SD = 0.50;[...] in a scale from 1-5), and thought that they were fairly intelligent (Sphero: M = 4.0, SD = 0.61). [...]". [26, p. 667–668]

There is a lot to unpack here, so a structure and re-organization is necessary for a clearer and easier understanding. The remaining of this chapter details the findings of the study replication, as well as the aggregated findings from the original paper, following the layout example provided in the original version of these in the citation above. As the original paper already has an order of presenting the study details and experiment findings, the red line here will follow the same principles and order of presentation. Information and data is added beyond the original paper, where available and considered relevant.

4.1 Study Participants

To start of with the key aspect of the study, the participant sample of the original study is described in chapter *"IV.User Study/A. Sample"* as follows:

"[...] The study was conducted with 31 participants, 20 males (65%) and 11 females (35%). The *mean age* of subjects was 22 years (range: 18-26). 90% of the participants were students and 10% student workers. 84% of the subject sample studies in the area of Science and Technology, 3% studies Economics, 2% in the Social Sciences field and 2% study Art.
Seventeen participants (12 males) were allocated to the condition where they interacted with Sphero 1.0. Of these seventeen participants, nine (53%) had previous experience with robots but only four (24%) knew Sphero 1.0. Also, three participants (18%) had already interacted with Sphero 1.0. The other fourteen participants (8 males) interacted with BB- 8. Ten of them (71%) had previous experience with robots, and seven of them (50%) knew BB-8 [...]". [26, p. 666]

Only the 17 people who interacted with Sphero 1.0 are of interest to this study analysis. In the study replication, the following can be summarised about the persons who took part in it:
The replicated study was conducted with 30 participants, 15 males and 15 females, namely 50% each. The *mean age* of subjects is 27.3 years, *the median* and *the mode* being equally 27, from the overall range of 18 to 43 years of age. These are further split in two age groups, as presented under *Methodological Approach*:

1. A younger age group, consisting of 20 participants equally split among males and females, with an age span closely tied to the original participants group, namely in the range of 18 to 27 years of age, with 25.2 being their mean, 26 their median, and 27 their mode. It was difficult recruiting enough young participants, so the higher boundary was incremented by one year so as to not stray too far away from the original participant's details. This participant group is also referred to as Age I from this point onwards.

2. An older age group, consisting of 10 participants equally split among males and females, representing the previous generation, namely in the range of 28 to 43 years of age, with 32.5 being their mean, 30 their median, and 30,8 their mode. This participant group is also referred to as Age II from this point onwards.

Of all participants 18 mentioned to be still studying, out of which only 4 (13.5%) considered themselves to be full-time students. Of those studying, approximately 71,5% (13) of the subjects study in the area of Science and Technology, 11,5% (2) study Human Medicine, 5,5% (1) study Architecture, 5,5% (1) study Economics, and 5,5% (1) in the Social Sciences.
Nonetheless, the 26 participants, except for those studying full-time, were working part-time or full-time in the following areas respectively: approximately 61% (16) of the work in the field of Science and Technology, 23% (6) in Business and Economics, 4% (1) in the Medical System, 4% (1) in the Social System, 4% (1) in Architecture, and 4% (1) in Arts.
When asked about having had any interaction, experiences or gathered knowledge of robots, only 17 (56%) responded affirmative. However, what each understood under robots varied from home, cleaning and cooking robots, to exposition exhibits, study courses material and voice chat AI. Only 1 participant, 3.33% more precisely, had previous knowledge of and theoretical experience with Sphero from an earlier university course, it's interaction, behavior and physical capabilities were though not experienced hands-on.
In each group, participants were split equally based on their gender, as well as by trying to mix the repetition order of the experiments, so as for it not to be the same, in order to follow either the "Follow me" (*abbr. F*) with "Dance with me" (*abbr. D*), shortly **F-D**, or the "Dance with me" with "Follow me", abbreviated **D-F** from now on.
A general overview of the data and information presented so far can be observed under *Table 4.1*.

Participant Age and Gender	Occupation, Profession	Study, Expertise	Experience Robots
24, Male	Student	MSc. Media and Human Centered Computing	Yes
26, Female	Market Manager	International Business	Yes, Home Robots

Participant Age and Gender	Occupation, Profession	Study, Expertise	Experience Robots
26, Female	Software Developer	MSc. Media and Human Centered Computing	Yes
27, Female	Controller	International Business Administration	Yes
27, Male	Software Developer	Computer Science	No
30, Male	Test Engineer	Electrical and Electronic Engineering	No
25, Female	Student	Medicine	No
38, Male	Software-Architect	Computer Science	Yes, GPT3 and Roomba
26, Female	Software Developer	Computer Science	Yes, Lisbon Museum Robot, Roomba
23, Female	UX/UI Designer	MSc. Media and Human Centered Computing	Yes, Thesis Research
28, Female	Student/Nurse	Medicine	Yes
28, Male	Software Engineer	MSc. Software Engineering	Yes, Smart Cleaning, Home Robots
27, Male	Tech Employee, Student	Environmental Engineering	No
25, Female	Associate Tax	Economic law	No
21, Male	Student	MSc. Media and Human Centered Computing	Yes, School Course
27, Male	Software Developer	Media informatics	Yes, Home Robots
18, Female	Student	Physics	Yes, Lego Mindstorms, Vacuum Cleaner
26, Female	Student, Dance teacher	MSc. Media and Human Centered Computing	No
27, Male	Employee, Student	Macroeconomics, Tiler	No
25, Female	Employee, Student	MSc. Architecture	No
30, Female	Software Tester	MSc. Media and Human Centered Computing	Yes, University Course and Sphero
28, Male	Student, part time job	Computer and Data Science	Yes, Vacuum, Home Robot
43, Female	Independent Financial Consultant	College for Chemical and Environmental Technology	No
27, Male	Corporate Structure Financial Analyst	Financial Science	No
38, Male	Test Engineer	MSc. Agriculture Sciences	Yes, Vacuum, Home Robot
27, Male	Architect	BSc. Architecture	Yes, Cleaning, Home Robot
25, Female	Social Worker	None	Yes, Exhibition on Artificial Intelligence
25, Male	Software Engineer	Computer Graphics	No
30, Female	Digital Marketing	None	Yes, Alexa, Bixie
32, Male	Test Automation Engineer	BSc. Computer Science	Yes, Cleaning Robots at the Airport

Table 4.1: Replication Study Participant Description

4.2 Study Structure and Duration

The experiment duration was planed with an average of 20 minutes per participant in mind, with each of its parts structured, designed and rehearsed to last appropriately, depending on the willingness of the participant to interact, communicate and answer, as well as on their shown interest in Sphero and the study.
In reality, the following timestamps were recorded as the minimum and maximum boundary for each part of the experiment:

- Introduction: min. 2 minutes and 20 seconds, max. 5 minutes and 20 seconds
- “Follow me”: min. 30 seconds, max. 4 minutes
- First Questionnaires: min. 4 minutes and 10 seconds, max. 11 minutes
- “Dance with me”: min. 1 minute and 19 seconds, max. 1 minute and 36 seconds
- Second Questionnaires: min. 2 minutes and 5 seconds, max. 7 minutes

With the exception of participants losing their interest in the experiment, when first presented with Sphero, resulting in a duration of only 12 minutes and 37 seconds, the shortest full run-through lasted 13 minutes and 6 seconds, while the longest took up to 31 minutes and 50 seconds.
Around 75% of the experiments lasted in the boundary of 15 to 22 minutes, with the average rounding up at 19 minutes and 40 seconds, a satisfying result for the foreseen planed unfolding of the study.

The duration of each participant’s interaction in each respective scenario, as well as the order in which they were presented with these, can be observed under *Table 4.2*. The duration values highlighted in bold represent the minimum and maximum duration for each given scenario considering the entire participant count.
As the dance scenario was designed to last around the same time for all participants, there is no major difference of interest in the time taken with the robot during it, although the values and means are also included in the table.
However, if the participant first interacted with Sphero in the sequence F-D or D-F, it was interesting to see how the new dancing scenario influenced the duration of the “Follow me” scenario, as well as how the following scenario influenced the interaction and behavior in the “Dance-with-me” scenario.
The *Table 4.2* is concluded with the representation of the means for each age group, scenario order and aggregations of groups and scenarios.
These are summarised as follows:

- The mean duration of **all participants** together, **irrelevant** of which **order** they had experienced the scenarios in, was *2 minutes and 4 seconds* for *“Follow me”*, and *1 minutes and 29 seconds* for *“Dance with me”*.
- The mean duration of **all participants** together, who went through the scenarios in the order **F-D**, was *higher* for *“Follow me”*, namely *2 minutes and 15 seconds*, and *equal*, *1 minutes and 29 seconds* for *“Dance with me”*.

- The mean duration of **all participants** together, who went through the scenarios in the order **D-F**, was *lower* for *"Follow me"*, namely *1 minutes and 51 seconds*, and almost *equal*, *1 minutes and 28 seconds* for *"Dance with me"*.

- The mean duration of **Age I participants** together, who went through the scenarios in the order **F-D**, was almost *equal* for the aggregated *"Follow me"*, namely *2 minutes and 16 seconds*, and almost *the same*, as in *1 minutes and 30 seconds*, for *"Dance with me"*.

- The mean duration of **Age I participants** together, who went through the scenarios in the order **D-F**, was the *lowest* of all for *"Follow me"*, namely *1 minutes and 50 seconds*, and a few seconds *shorter* the previous run order, as in *1 minutes and 27 seconds*, for *"Dance with me"*.

- The mean duration of **Age II participants** together, who went through the scenarios in the order **F-D**, was the *highest* of all for *"Follow me"*, namely *2 minutes and 26 seconds*, and comparably *equal* the other means, as in *1 minutes and 28 seconds*, for *"Dance with me"*.

- The mean duration of **Age II participants** together, who went through the scenarios in the order **D-F**, was *higher* than that of Age I for *"Follow me"*, namely *1 minutes and 66 seconds*, and still *equal* the run order, as in *1 minutes and 28 seconds*, for *"Dance with me"*.

- To sum up, the mean duration of **Age I participants**, **irrelevant** of which **order** they had experienced the scenarios in, was *2 minutes and 3 seconds* for *"Follow me"*, almost identical to the aggregated mean of all participants, and precisely *1 minutes and 29 seconds* for *"Dance with me"*; with that of **Age II participants** landing at *2 minutes and 11 seconds* for *"Follow me"*, *higher* to the aggregated mean of all participants, and 1 second slower, namely *1 minutes and 30 seconds*, for *"Dance with me"*.

Participant Age and Gender	**Duration F**	**Duration D**	**Scenario Order**
24, Male	2m	1m 20s	D-F
26, Female	2m 45s	1m 25s	F-D
26, Female	1m 21s	**1m 19s**	D-F
27, Female	2m 20s	1m 23s	F-D
27, Male	1m 20s	1m 22s	D-F
30, Male	1m 5s	1m 24s	D-F
25, Female	1m 30s	1m 26s	F-D
38, Male	**4m**	1m 28s	D-F
26, Female	1m 57s	1m 30s	F-D
23, Female	1m 10s	**1m 36s**	D-F
28, Female	2m 50s	1m 24s	F-D
28, Male	1m 5s	1m 32s	D-F
27, Male	2m 25s	1m 33s	F-D
25, Female	1m 29s	1m 25s	D-F

Participant Age and Gender	Duration F	Duration D	Scenario Order
21, Male	1m 45s	1m 33s	F-D
27, Male	2m	1m 29s	D-F
18, Female	1m 40s	1m 35s	F-D
26, Female	2m 7s	1m 31s	D-F
27, Male	2m 18s	1m 30s	F-D
25, Female	2m 30s	1m 33s	D-F
30, Female	1m 55s	1m 32s	F-D
28, Male	**30s**	1m 28s	D-F
43, Female	3m 30s	1m 22s	F-D
27, Male	2m 35s	1m 31s	F-D
38, Male	2m 12s	1m 29s	F-D
27, Male	2m	1m 31s	F-D
25, Female	2m 35s	1m 33s	D-F
25, Male	2m 40s	1m 32s	F-D
30, Female	2m 48s	1m 30s	D-F
32, Male	1m 45s	1m 35s	F-D
Mean Age I, II	**2m 4s**	**1m 29s**	**F-D, D-F**
Mean Age I, II	**2m 15s**	**1m 29s**	**F-D**
Mean Age I, II	**1m 51s**	**1m 28s**	**D-F**
Mean Age I	**2m 3s**	**1m 29s**	**F-D, D-F**
Mean Age I	**2m 16s**	**1m 30s**	**F-D**
Mean Age I	**1m 50s**	**1m 27s**	**D-F**
Mean Age II	**2m 11s**	**1m 30s**	**F-D, D-F**
Mean Age II	**2m 26s**	**1m 28s**	**F-D**
Mean Age II	**1m 56s**	**1m 28s**	**D-F**

Table 4.2: Scenario Duration in the Replication Study

Table 4.3 displays the results of whether or not the participants had successfully interacted with Sphero in each of the scenarios, basing the criteria on original conditions, as well as added ones. Originally, the study was interested in observing whether or not the persons understood that they were supposed to follow Sphero, resulting in a success rate of 71%. In the replicated study, around 57% of all participants, namely 17, understood what Sphero wanted to communicate them, with 43%, so 13 persons, not following it. By further splitting the findings based on age groups, it results that 65% of participants aged between 18 and 27 years old followed Sphero, while only 40% of those aged between 28 and 43 year old did likewise.
An additional aspect to keep track of arose during observations of the experiments unfolding; one related to the scenario's finishing reward and announcement sign. Exactly half of the participants missed it, due to them being to focused on Sphero to look up to the chair. Of these 15 participants, 8 participants did not understand to follow Sphero, hence did not find the reward; and 7 understood that Sphero wanted them to follow it, yet by having their eyes fixated on Sphero they missed the reward, despite it being in front of them. Another 5 participants did not understand that Sphero wanted them to follow it, yet by going after Sphero they found the reward

nonetheless. Age was not considered as being a relevant factor here, as the sample sizes of each result respectively are too small.
Moving over to the "Dance with me" scenario, the rate of participants understanding it's behavior as communication dancing is at 80% of all participants, with only 6 participants being either not sure of whether Sphero intended to dance or not, or interpreting it in completely other ways. The other way around, not as many participants were willing to actually dance with Sphero. Of these, 3 belong to Age I and 2 to Age II.
With disregard to Sphero's intention, around 60% actually danced to/with Sphero, a total of 18 persons. By combining the two columns and values, just 15 participants perceived Sphero as dancing and responded in a similar manner in their interaction. In more detail, only 5 persons belonging to Age II danced to Sphero, with the other 5 either not understanding it willing to dance and not dancing themselves, namely 2 persons, or understanding that it was dancing but not feeling like returning the behavior, namely 3 persons. Of those in the Age I group 7 did not dance, where one person did not understand that Sphero was dancing, or where not sure of it, and did not dance either, while 6 understood that it was dancing but did not want to dance as well.
To sum up the success rate for each scenario based on the two criteria respectively, 7 participants (23% rounded down) understood Sphero in the "Follow me" scenario, while 13 (43% rounded down) successfully passed the "Dance with me" one, by having a "Yes" in both columns of the according scenario.
Of all scenarios combined, only 4 participants (13% rounded down) comprehended Sphero accurately, offering the sought-after answers to their interpretation and perception of it, with 2 persons (7% rounded up) missing by it's purpose in all scenarios and questions.
The order of the scenarios seems to have impacted the Yes/No ratio , with only 5 of the D-F, versus 8 of the F-D scenario order not understanding the follow wish of Sphero, and reversed when analysing the dance wish of Sphero, with 1 of the F-D and 5 of the D-F not perceiving the intention communication and behavior of Sphero.
The presented results are written, with the percentages elaborated also for each scenario order, at the end of *Table 4.3*.
Of those marked with a "No", half danced to Sphero nonetheless, presumably the ones who were not sure of how to interpret the robot's behavior.

Participant Age and Gender	**Followed Sphero?**	**Reward Found?**	**Sphero Dancing?**	**Participant Danced?**	**Scenario Order**
24, Male	Yes	Yes	No	Yes	D-F
26, Female	Yes	Yes	Yes	Yes	F-D
26, Female	Yes	Yes	Yes	No	D-F
27, Female	Yes	Yes	Yes	Yes	F-D
27, Male	Yes	Yes	No	No	D-F
30, Male	Yes	Yes	Yes	No	D-F
25, Female	No	Yes	Yes	Yes	F-D
38, Male	No	No	No	No	D-F
26, Female	Yes	Yes	No	Yes	F-D

Participant Age and Gender	Followed Sphero?	Reward Found?	Sphero Dancing?	Participant Danced?	Scenario Order
23, Female	Yes	No	Yes	Yes	D-F
28, Female	No	No	Yes	No	F-D
28, Male	Yes	No	Yes	Yes	D-F
27, Male	No	No	Yes	No	F-D
25, Female	Yes	No	Yes	No	D-F
21, Male	No	Yes	Yes	No	F-D
27, Male	Yes	No	No	Yes	D-F
18, Female	Yes	Yes	Yes	Yes	F-D
26, Female	No	Yes	Yes	Yes	D-F
27, Male	No	Yes	Yes	No	F-D
25, Female	No	No	Yes	No	D-F
30, Female	Yes	Yes	Yes	Yes	F-D
28, Male	No	No	Yes	Yes	D-F
43, Female	No	No	Yes	No	F-D
27, Male	Yes	No	Yes	Yes	F-D
38, Male	No	Yes	Yes	Yes	F-D
27, Male	Yes	No	Yes	Yes	F-D
25, Female	Yes	No	Yes	Yes	D-F
25, Male	No	No	Yes	Yes	F-D
30, Female	No	No	No	No	D-F
32, Male	Yes	Yes	Yes	Yes	F-D
Yes % Age I, II	13%				**F-D, D-F**
Yes % Age I, II	23%		43%		**F-D, D-F**
Yes % Age I, II	38%		63%		**F-D**
Yes % Age I, II	21%		36%		**D-F**
Yes % Age I	35%		36%		**F-D, D-F**
Yes % Age I	30%	35%	50%	40%	**F-D**
Yes % Age I	35%	20%	30%	25%	**D-F**
Yes % Age II	30%		50%		**F-D, D-F**
Yes % Age II	20%	30%	50%	30%	**F-D**
Yes % Age II	20%	10%	30%	20%	**D-F**

Table 4.3: Scenario Goals and Behavior Expectations of the Replication Study

4.3 Godspeed Questionnaire

Continuing with the questionnaires following the first scenario, the first one to discuss is that of Godspeed on perception evaluation, how the robot appeared to be for the participant. All these tables, respectively found under *Table 4.4*, *Table 4.5*, *Table 4.7*, and *Table 4.8*, portrait the accumulated results for each dimension, with the accumulated number of participants who marked a specific score in the range from 1 to 5, for each of the presented attribute scales. The tables have been split based on the age group of the participants and the scenarios they had filled in the questionnaire after. The aggregated general analysis is deducted by analysing the union of the tables. Interpreting the results found in each of the tables, the minimum and maximum mean (M) and standard deviation (SD), rounded up or down to two decimals, define the boundaries of all scales within a particular dimension, showing the trend in the participant's answers.

"Follow me" with Age I (18 to 27)							**M**	**SD**	**Mode**
Likeability							**3,98**	**1**	**5**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dislike		1	1	4	5	Like	**4,18**	**0,94**	**5**
Unfriendly	1		3	4	3	Friendly	**3,72**	**1,14**	**4**
Unkind		1	3	3	4	Kind	**3,91**	**1**	**5**
Unpleasant			5	3	3	Pleasant	**3,82**	**0,83**	**3**
Awful			4		7	Nice	**4,27**	**0,96**	**5**
Perceived Intelligence							**3,69**	**0,95**	**3**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Incompetent		1	4	3	3	Competent	**3,73**	**0,96**	**3**
Ignorant		1	2	5	3	Knowledgeable	**3,91**	**0,9**	**4**
Irresponsible		1	6	2	2	Responsible	**3,46**	**0,89**	**3**
Unintelligent		1	2	5	3	Intelligent	**3,91**	**0,9**	**4**
Foolish		1	7		3	Sensible	**3,46**	**0,99**	**3**
Animacy							**3,58**	**1,22**	**4**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dead			3	3	5	Alive	**4,18**	**0,83**	**5**
Stagnant			1	3	7	Lively	**4,54**	**0,65**	**5**
Mechanical	4	4	2	1		Organic	**2**	**0,95**	**1; 2**
Artificial	1	3	2	5		Lifelike	**3**	**1**	**4**
Inert		1	2	5	3	Interactive	**3,91**	**0,9**	**4**
Apathetic		1	2	6	2	Responsive	**3,82**	**0,83**	**4**

Table 4.4: Godspeed Perception Score of **F** with *Age I*, where the rating scale ranges from *1 = totally the negative state* to *5 = totally the positive state* and the numeric values represent the count of the occurrence

The 11 participants, who first experienced the “Follow me” scenario, belonging to Age I, as presented in *Table 4.4*, perceived Sphero as being *quite* likeable, rounding up to the mean value of *4*, but also intelligent and animate likewise; whereas the latter two dimensions were rated half-way between the values *3* and *4*, namely *neutral* and *quite a bit*. The biggest deviation in scoring occurred when rating the animacy, while the lowest mode was observed at the perceived intelligence. The top 3 positions occupied by characteristics of Sphero being controlled to take up the role of a “leader”, asking to be followed, were lively (4,54), nice (4,27) and alive, as well as likeable (both rated at 4,18). Its weakest characteristics are grouped under animacy, with a somewhat mechanical presence (2), seeming to be neither artificial, nor lifelike (3), and being perceived rather neutral when considering it’s sense of responsibility and sensitivity (3,46), if rounded down.

“Dance with me” with Age I (18 to 27)							**M**	**SD**	**Mode**
Likeability							**4,44**	**0,72**	**5**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dislike				4	5	Like	**4,55**	**0,5**	**5**
Unfriendly			3		6	Friendly	**4,33**	**0,94**	**5**
Unkind			3	1	5	Kind	**4,22**	**0,92**	**5**
Unpleasant				4	5	Pleasant	**4,55**	**0,5**	**5**
Awful				4	5	Nice	**4,55**	**0,5**	**5**
Perceived Intelligence							**3,35**	**0,92**	**3**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Incompetent		2	2	5		Competent	**3,33**	**0,82**	**4**
Ignorant		1	4	3	1	Knowledgeable	**3,44**	**0,83**	**3**
Irresponsible		1	4	2	2	Responsible	**3,55**	**0,96**	**3**
Unintelligent		2	2	4	1	Intelligent	**3,44**	**0,96**	**4**
Foolish	1	1	4	3		Sensible	**3**	**0,94**	**3**
Animacy							**3,80**	**1,16**	**5**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dead		1	2	1	5	Alive	**4,11**	**1,1**	**5**
Stagnant				2	7	Lively	**4,77**	**0,42**	**5**
Mechanical	1	1	5	2		Organic	**2,88**	**0,87**	**3**
Artificial	1	3	3	1	1	Lifelike	**2,77**	**1,13**	**2; 3**
Inert		1		4	4	Interactive	**4,22**	**0,92**	**4; 5**
Apathetic		1		6	2	Responsive	**4**	**0,82**	**5**

Table 4.5: Godspeed Perception Score of **D** with *Age I*, where the rating scale ranges from *1 = totally the negative state* to *5 = totally the positive state* and the numeric values represent the count of the occurrence

The 9 participants, who first experienced the “Dance with me” scenario, belonging to the younger age group, namely Age I, as presented in *Table 4.5*. They perceived Sphero as being *more* likeable than otherwise, with a mean value of *4,44*, but less intelligent and animate than in the other scenario order; whereas the latter two dimensions were rated between the values *3* and *4*, namely *3,31* and *3,80*. The biggest deviation in scoring occurred again when rating the animacy, while the lowest mode observed at the perceived intelligence persisted. Regarding the mode, the animacy was also rated more often a 5 than before, presumably due to the nature of the scenario and its interaction. The top 3 positions occupied by characteristics of Sphero playing the role of an "entertainment robot", hierarchically lower than the human participant, were still lively (4,77), likeable, nice, as well as the new one pleasant (all three rated at 4,55), followed by friendly (with the rating 4,33). Its weakest characteristics are grouped under animacy and perceived intelligence, with a tendency towards mechanical presence (2,77) and artificial aspect (2,88), as well as with a neutral sensitivity perception (3), yet aiming to be a somewhat more than neutral when referring to its competence (3.33) and imposed intelligence (3,44).

"Follow me" with Age II (28 to 43)							**M**	**SD**	**Mode**
Likeability							**4,44**	**0,64**	**5**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dislike				2	3	Like	**4,6**	**0,49**	**5**
Unfriendly				2	3	Friendly	**4,6**	**0,49**	**5**
Unkind			1	2	2	Kind	**4,2**	**0,75**	**4; 5**
Unpleasant				2	3	Pleasant	**4,6**	**0,49**	**5**
Awful			1	2	2	Nice	**4,2**	**0,75**	**4; 5**
Perceived Intelligence							**3,6**	**0,75**	**3**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Incompetent			2	3		Competent	**3,6**	**0,49**	**4**
Ignorant			3	1	1	Knowledgeable	**3,6**	**0,8**	**3**
Irresponsible			2	1	2	Responsible	**4**	**0,89**	**3; 5**
Unintelligent			3	2		Intelligent	**3,4**	**0,49**	**3**
Foolish		1	1	3		Sensible	**3,4**	**0,8**	**4**
Animacy							**3,33**	**1,53**	**5**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dead				2	3	Alive	**4,6**	**0,49**	**5**
Stagnant				3	2	Lively	**4,4**	**0,49**	**4**
Mechanical	3	1	1			Organic	**1,6**	**0,8**	**1**
Artificial	4	1				Lifelike	**1,2**	**0,4**	**1**
Inert			1	2	2	Interactive	**4,2**	**0,75**	**4; 5**
Apathetic			2	1	2	Responsive	**4**	**0,89**	**3; 5**

Table 4.6: Godspeed Perception Score of **F** with *Age II*, where the rating scale ranges from *1 = totally the negative state* to *5 = totally the positive state* and the numeric values represent the count of the occurrence

The 5 participants, who first experienced the "Follow me" scenario, belonging to the older age group, namely Age II, as presented in *Table 4.7*. They perceived Sphero as being *as* likable *as the participants of Age I in scenario D*, with the same mean value of *4,44*; but also intelligent and animate to some extent; whereas the latter two dimensions were rated half-way between the values *3* and *4*, namely *neutral* and *quite a bit*. Although Animacy was rated the lowest of the three it offers the best mode score of 5 out of all tables presented here. The biggest deviation in scoring occurred when rating the animacy, while the lowest mode was observed at both the perceived intelligence and the animacy. The top 3 positions occupied by characteristics of Sphero playing the role of a "leader", hierarchically higher than the human participant, were alive, likeable and friendly (all with a rating of 4,6), followed by lively (4,4) and kind, nice and interactive (all placed at 4,2), the latter scoring a new high in comparison with the other age group. Its weakest characteristics are grouped under animacy, with the lowest score due to a lifelike aura (1,2), appearing very mechanical (1,6) and, referring to perceived intelligence, seeming a little intelligent and sensible (both rated at 3,4).

"Dance with me" with Age II (28 to 43)							**M**	**SD**	**Mode**
Likeability							**4,16**	**0,83**	**4; 5**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dislike				2	3	Like	**4,6**	**0,49**	**5**
Unfriendly			1	1	3	Friendly	**4,4**	**0,8**	**5**
Unkind			3	2		Kind	**3,4**	**0,89**	**3**
Unpleasant		1		2	2	Pleasant	**4**	**1,1**	**4; 5**
Awful				3	2	Nice	**4,3**	**0,49**	**4**
Perceived Intelligence							**3**	**1**	**4**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Incompetent		1	1	3		Competent	**3,4**	**0,8**	**4**
Ignorant	1	1	1	2		Knowledgeable	**2,8**	**1,17**	**4**
Irresponsible	1	1	2	1		Responsible	**2,6**	**1**	**3**
Unintelligent		1	1	3		Intelligent	**3,4**	**0,8**	**4**
Foolish	1		3	1		Sensible	**2,8**	**0,98**	**3**
Animacy							**3,46**	**1,28**	**4**
Negative State	**1**	**2**	**3**	**4**	**5**	**Positive State**	**-**	**-**	**-**
Dead			1	3	1	Alive	**4**	**0,63**	**4**
Stagnant				1	4	Lively	**4,8**	**0,4**	**5**
Mechanical	2	1	1	1		Organic	**2,2**	**1,17**	**1**
Artificial	2		1	2		Lifelike	**2,6**	**1,36**	**1; 4**
Inert		1	1	2	1	Interactive	**3,6**	**1**	**4**
Apathetic			3	1	1	Responsive	**3,6**	**0,8**	**3**

Table 4.7: Godspeed Perception Score of **D** with *Age II*, where the rating scale ranges from *1 = totally the negative state* to *5 = totally the positive state* and the numeric values represent the count of the occurrence

The 5 participants, who first experienced the "Dance with me" scenario, belonging to the older age group, namely Age II, are presented in *Table 4.8*. They perceived Sphero as being *less* likeable than the other 5, with a mean value of *4,16*. The score for intelligence is at its *lowest* out of all 4 tables (rated 3), with the animacy slightly *improved* over the Age II with F (rounded at 3.46). The biggest deviation in scoring occurred again when rating the animacy, with the perceived intelligence encompassing the lowest scores. Regarding the mode, these were rated similarly to the participants of F order, with likeability being tied between 4 and 5. The top 3 positions occupied by characteristics of Sphero playing the role of an "entertainment robot", hierarchically lower than the human participant, were still and improved lively (4,8), likeable (4,6) and friendly (4,4), with nice (4,3) and pleasant (4) following only after them, rated lower than the other 5 have. Its weakest characteristic is still the mechanical aspect (2,6), followed by the responsible and artificial aspect (2,6), as well as knowledge and sensitivity impression (2,8). Compared to the other 5 Age II participants who followed F-D, the negative aspects were rated even lower than otherwise.

The original study recorded having analyzed the following outcome:

- Likeability: the participants liked the robots they interacted with, M = 4,4, SD = 0,5
- Perceived Intelligence: the participants thought that the robots were fairly intelligent, M = 4, SD = 0,61

So far the results presented were classified according to the age of the participant and the scenario order. Thus, unifying these by one of the two criteria and both is necessary for a valid and sustainable comparison with the original values. The process and results step-by-step for each combination is described in the next paragraphs.

First, grouping the results of the questionnaire by age for the group Age I, encompassing 20 of the 30 participants. The union of *Table 4.4* and *Table 4.5* yield the dimension of Likeability with M = 4,19, SD = 0,91 and Mode = 5; the one of Perceived Intelligence with M = 3,54, SD = 0,95 and Mode = 3; and the Animacy respective one with M = 3,7, SD = 1,2 and Mode = 4. This lowers overall the score of the Animacy from a 5, as observed during D, to the 4 captured during F. Compared to the original results, these are much smaller, with a mean smaller by 0,21 for Likeability and 0,46 for Perceived Intelligence, while the answers are also more varied and far stretched placed, as the standard deviation is almost double in these replication results. However, this approach is the most faithful one to the original participant's group age, deviating from this only through the addition of the 27th year of age.

Second, grouping the results of the questionnaire by age for the group Age II, encompassing the other 10 of the 30 participants. The union of *Table 4.7* and *Table 4.8* yield the dimension of Likeability with M = 4,3, SD = 0,75 and Mode = 5; the one of Perceived Intelligence with M = 3,3, SD = 0,94 and Mode = 4; and the Animacy respective one with M = 3,39, SD = 1,4 and Mode = 4. This union also sees the Perceived Intelligence being upped to a 4, instead of a 3 as given in scenario

F. Compared to the original results, these are not that lower, considering the group Age 1 , with a mean smaller by 0,1 for Likeability, however by 0,7 for Perceived Intelligence, with the standard deviation is almost tripled for the latter. The Animacy also scored much lower than than for Age I, showing that the older generation is trusting the capabilities and intelligence of the robot less than the younger one. Nonetheless, this group is a completely different one than the original participant's, so a correlation can not be meaningfully established.

Third, grouping the results of the questionnaire by the scenario "Follow me", aggregating 16 of the 30 participants. The union of *Table 4.4* and *Table 4.7* yield the dimension of Likeability with M = 4,1, SD = 0,93 and Mode = 5; the one of Perceived Intelligence with M = 3,66, SD = 0,89 and Mode = 3; and the Animacy respective one with M = 3,5, SD = 1,33 and Mode = 4. Compared to the original results, these are further away than the previous tables, with a mean smaller by 0,3 for Likeability, the biggest difference yet, however by 0,44 for Perceived Intelligence, which is controversially the lowest difference to this point. The standard deviation was though almost doubled in every dimension. The Animacy scored in-between the Age I and Age II tables, as expected. This mix of participants, thus ages, lacks a faithful replication of values, as seen best in group Age I, additionally lacking 1 participant from the original minimum of 17. Yet, it is a more promising result the the results provided by Age II and Age I in some aspects, such as Perceived Intelligence and SD.

Fourth, grouping the results of the questionnaire by the scenario "Dance with me", aggregating 14 of the 30 participants. The union of *Table 4.4* and *Table 4.8* yield the dimension of Likeability with M = 4,34, SD = 0,77 and Mode = 5; the one of Perceived Intelligence with M = 3,23, SD = 0,97 and Mode = 4; and the Animacy respective one with M = 3,68, SD = 1,22 and Mode = 4 and 5. Compared to the original results, the Likeability is the closest so far to the original results, with a difference of only 0,06 in rating. The Perceived Intelligence however is at its furthest, with a difference in rating of 0,77. The standard deviations recorded are the lowest ones, compared to the other three combinations so far. Furthermore, the Animacy score is almost not distinguishable from the one proposed in the first combination, with a difference of only 0,02 in rating and standard deviation. The mode for it hereby ties between 4 and 5. Unfortunately, the same issues presented under the previous combination apply here as well, with an even lower count of participants, missing 3 from the original one. Moreover, the most important aspect, is that the scenario is a completely different one. It is nonetheless very interesting to observe such a similarity in Likeability and a lower score than that of Age I in Animacy combined from F and D; posing good questions to be elaborated in chapter *Discussion*.

Finally, by combining all age groups and scenarios together, the most consistent and numerous database is generated, covering 30 participants and 480 answers. The union of *Table 4.4*, *Table 4.5*, *Table 4.7* and *Table 4.8* yield the dimension of Likeability with M = 4,27, SD = 0,86 and Mode = 5; the one of Perceived Intelligence with M = 3,46, SD = 0,96 and Mode = 3; and the Animacy respective one with M = 3,58, SD = 1,28 and Mode = 4. As the thesis aimed at replicating the original study to confirm or disprove the presence of the replication crisis in its case, also expanding the study scenarios and participants, and removing bias, this combination is supposed to be the most real, trustworthy and suitable choice for comparing and discussing the different outcomes. However, the values are still not as close to the original ones, showing bigger differences than the previous combination of scenario D. The Likeability dimension is still the closest to the original rating, with only a 0,13 difference, but the Perceived Intelligence is off by 0,54 and the standard deviations are also higher by at least 50%. The Animacy worsened, as it was expected due to the union, proving again that the scenario incorporating dancing scored better Animacy through it's design, interaction and unfolding than the follow scenario. The results will be further discussed in the next chapter.

4.4 Proximity Questionnaire

In the next questionnaire, the participants were asked to evaluate the proximity they felt to the robot, assuming that they had perceived it as a "social entity", using the "Inclusion of the Other in the Self Scale" of the original paper. [59] The original result was a rather neutral one, with the participants placing the robot on a scale from 1 to 7 just on 3, respectively 4, alternating the values frequently, as it can be seen under *Figure 4.1*.
This supports the claim, that Sphero was not viewed as belonging to the intimate space, although also not to the social one, despite the robot and the participant meeting for the first time, but instead as a social entity in the personal radius. [6]
Important to mention is also, that the results offered in the original paper combine the results from both the participants having interacted with Sphero and of those having interacted with BB8. Hence, it is unclear how big the standard deviation is and which of the robots increased the average, except for the case that they were perceived equally. This assumption is also supported by Faria et al. [26], following their claims after the Godspeed questionnaire analysis, namely that "[...] A One-way ANOVA was performed and showed that the participants did not notice a significant difference between both robots, with the perceived intelligence being $F(1,29) = 0.401$, $p = 0.53$ and the likability $F(1,29) = 0.009$, $p = 0.92$. [...]". Whichever of the robots scored higher remains with the researchers.
Figure 4.1 The bar charts under *Figure 4.2*, *Figure 4.3*, *Figure 4.4*, and *Figure 4.5*, display the aggregated results for each proximity level, showcasing a similar design inspired by the original chart, for the replication's sake. In spite of the number of participants being too low, especially those belonging to *Age II*, for a meaningful quantitative analysis and display. However. to compare faithfully to the original study, this is shown split up based on the age group of the participants and the scenarios they had filled in the questionnaire after. Aggregated general interpretations are deducted by analysing the union of the results in the charts, and presented in the paragraphs below.

The average of the answers given by the 16 participants belonging to Age I, who checked the questionnaire after the "Follow me" scenario, is showcased in percentages under *Figure 4.2*. As this age group and scenario are a close replica of the original ones, it is of high importance to acknowledge both the similarities and the differences. Beginning with the latter, where the original proximity question saw some answers for the supposedly intimate space, namely ratings 6 and 7, the one

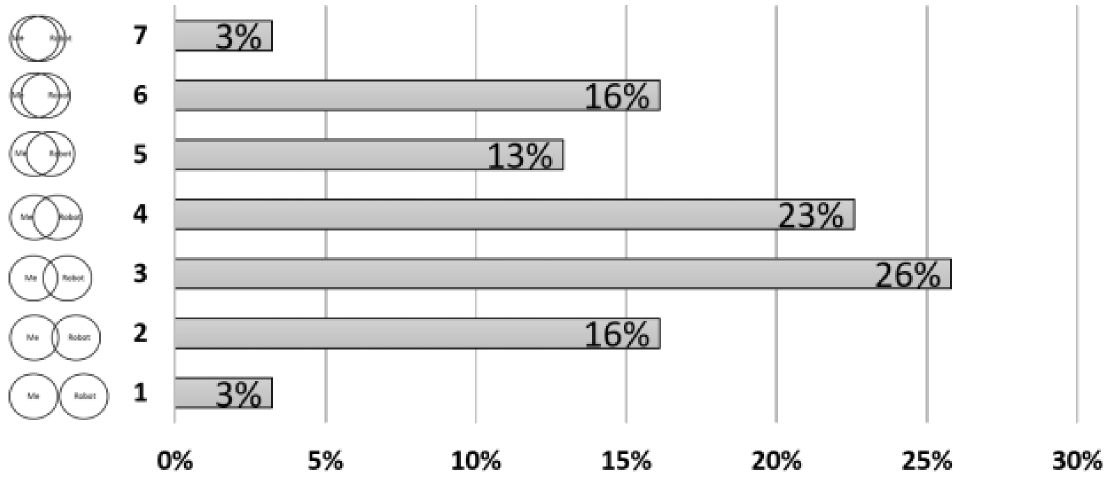


Figure 4.1: Original Results of the Proximity felt by the Participants with the Robots, using the Inclusion of the Other in the Self Scale [26]

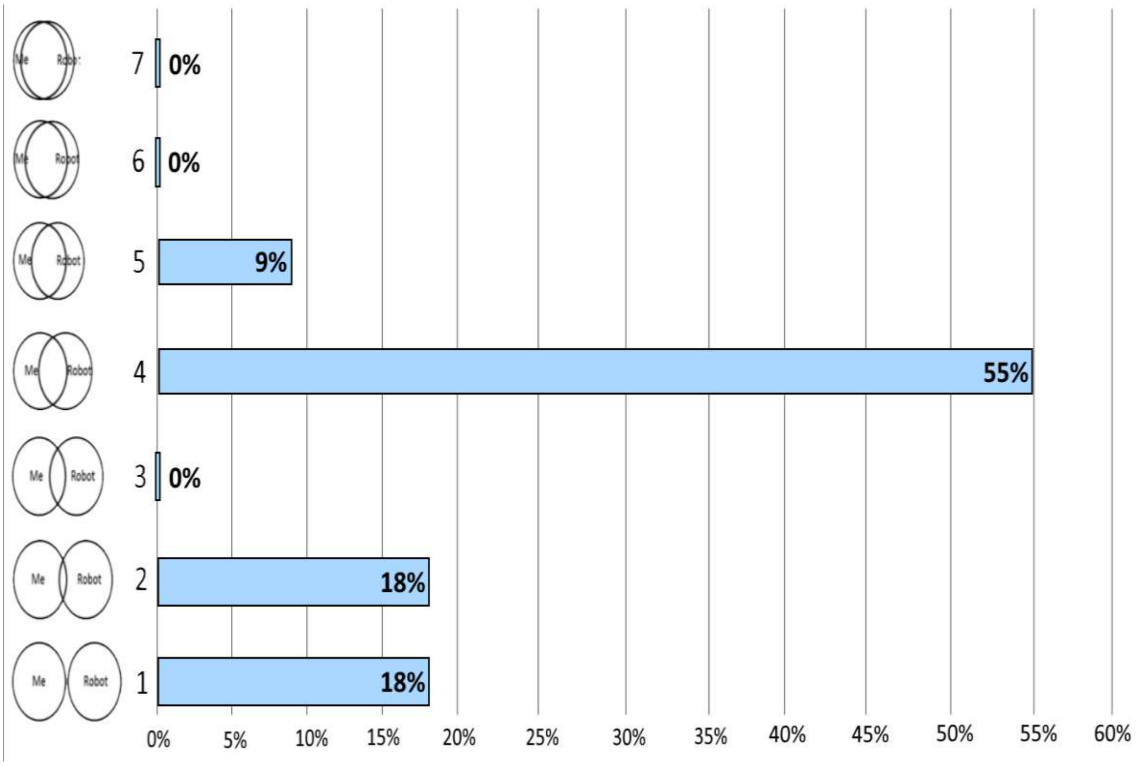


Figure 4.2: Proximity felt by *Age I* in **F**, where 1 = totally social space and 7 = totally intimate space

presented under the aforementioned figure saw none of the participants cross those options. Neither was the personal space rating 3 crossed at all. Despite this result though, the 55% observed under the rating 4 is higher than the original percentages for both 3 and 4 combines (48%). More participants considered the social distance adequate, by crossing 18% ratings 1 and 2, which is almost double the percentage of the original 1 (2%) and 2 (16%) combined. Option 5 saw a somewhat similar amount of markings (9%), with the replicated study having one person less (around 4%) not crossing it than in the original study (13%).

Another possible explanation is that, under the novelty and Hawthorne effect, the participants might have been more inclined to choose the neutral option, as in the 4 in the middle, than the other ones. However, to argue against such an assumption, as it can be seen under the *Appendix*, the options were lined out, hence by switching rows and having more time to offer, the participants would be less inclined to just mark the option in the middle and move on to the next questionnaire. If that would have been the case, the option would have then been either 2 or 3 on the first row, or 5 on the second. Instead, 55% chose the options furthest to the right, thus supposedly also the one most appropriate to describe their perception of Sphero. This assumption is valid for every other figure displaying results of the proximity questionnaire.

The following figures for Age I and Age II might seem redundant, due to their

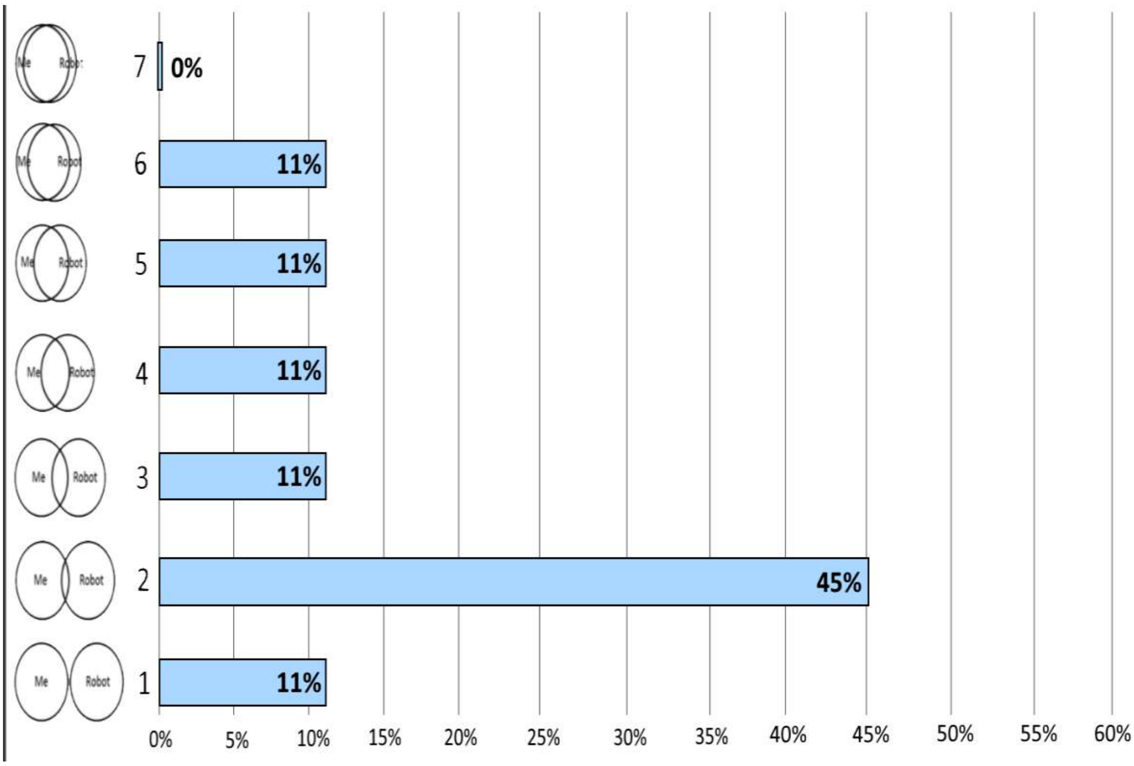


Figure 4.3: Proximity felt by *Age I* in **D**, where 1 = totally social space and 7 = totally intimate space

low sample number. These are displayed nonetheless, for a complete and detailed overview of the experiments and their outcome.
The average of the answers given by the 9 participants belonging to Age I, who checked the questionnaire after the “Dance with me” scenario, is showcased in percentages under *Figure 4.4*. As this scenario is a new addition to the study, it is of high importance for the following *Discussion* chapter. In summary, despite Sphero

portraying an entertainment, dancing robot, the distance that the participants felt towards it increased. Most participants felt estranged from the robot, as they could not connect to it during the experiment, as they did in the "Follow me" scenario, hence 45% of them marked a 2 on the scale. Another reason could be the inversion of the paradigm, from a companion to a care-taker one [39] , thus resulting in participants involuntarily feeling superior and implicitly treating the robot differently, more carefree and breezy, seeing it being tasked to dance and not communicating and interacting with it. The robot appears more as a machine. Additionally, one participant each marked the other ratings except for 7, displaying that for some a dancing robot was something completely out of their comfort, interest and social zone, while others felt even more bonded to it, hence the higher 5 and 6 scores than in the "Follow me" scenario.

The average of the answers given by the 5 participants belonging to Age II, who

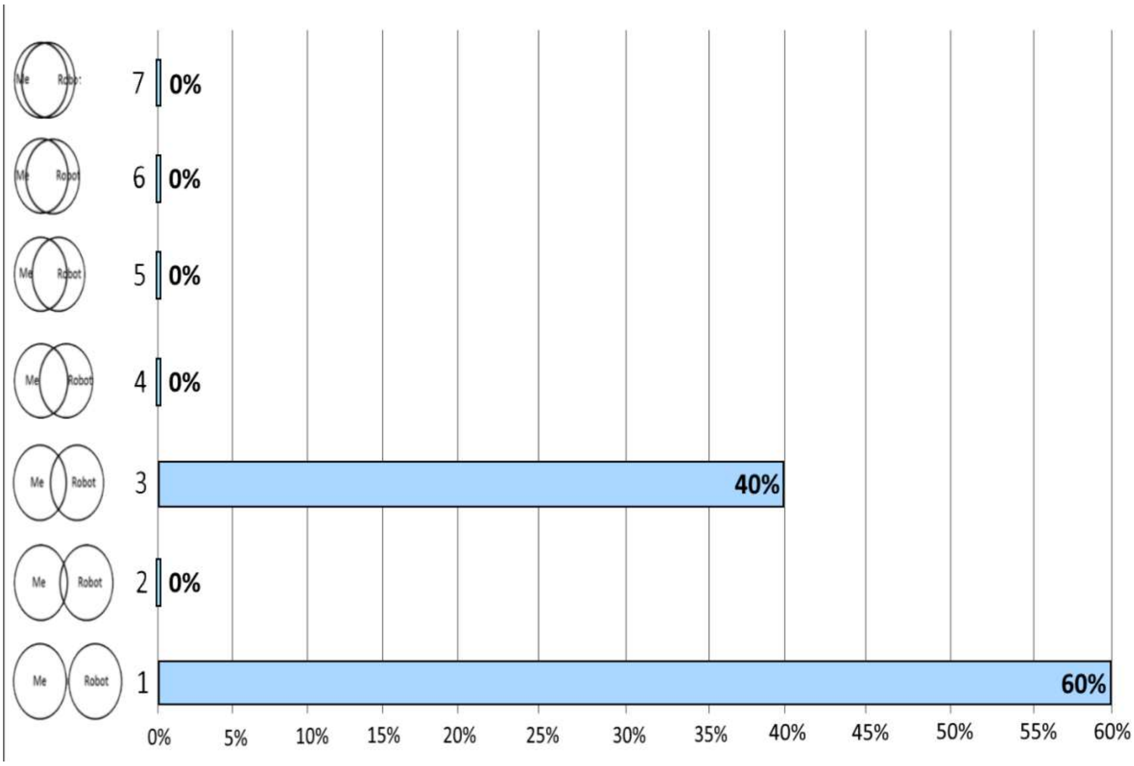


Figure 4.4: Proximity felt by *Age II* in **F**, where 1 = totally social space and 7 = totally intimate space

checked the questionnaire after the "Follow me" scenario, is showcased in percentages under *Figure 4.5*. As this age group is a newly added one to expand the originals study, their perception is of high importance for the following discussion. With regards to the information presented under *Related Work/HRI*, more precisely that the elder people, hence in this case generations, will feel more estranged to robots and technology, while also being more impacted by its novelty, is supported by these results. Compared to *Figure 4.2*, 40% of the participants rated Sphero highest on the social level of 3, with the other 60% viewing it as a complete stranger, namely a 1.

Figure 4.5 continues the analysis of the age group, but also in relation to the new scenario. Building upon the findings of the previous figure, it can be clearly deducted, that an entertaining robot was much better accepted by the older group than as presented in the “Follow me” scenario. Thus, only 20%, namely 40%, rated it as being a 3, respectively a 1, whereas the remaining 40% felt immediately more connected to the robot and rated it at 5.

The proximity results overall speak for themselves, showing that an entertaining

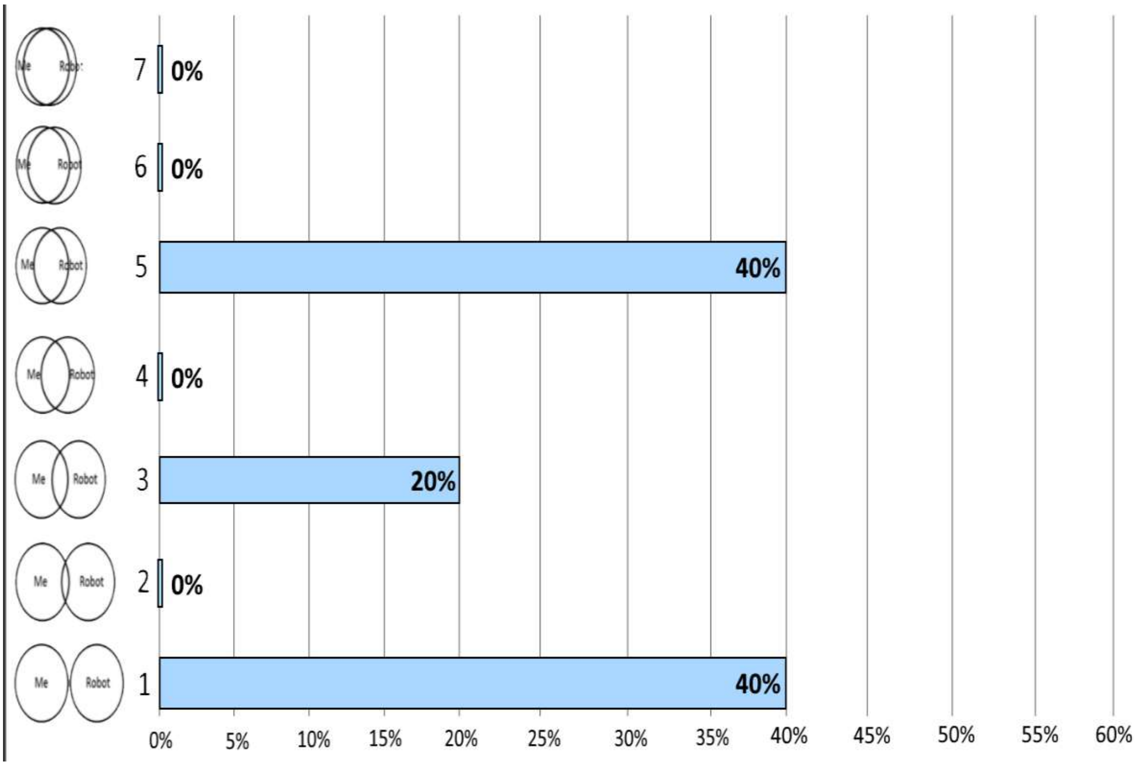


Figure 4.5: Proximity felt by *Age II* in **D**, where 1 = totally social space and 7 = totally intimate space

robot is more readily welcomed than an commanding, task-oriented one, considering the perceived distances towards it. This affirmation will be taken up in the discussion as well.
Following the findings portrayal structure implemented so far, the following list is intended to showcase the outcome by combining the results based on the age and the scenario order.

- Age I in F and D, the merge of *Figure 4.2* and *Figure 4.3* - In total 20 participants divided as follows: 0% rated Sphero being a 7 on the scale, 5% rated Sphero being a 6 on the proximity scale, 10% rated Sphero being a 5 on the proximity scale, 35% rated Sphero being a 4 on the scale, 5% rated Sphero being a 3 on the scale, 30% rated Sphero being a 2 on the scale, and 15% rated Sphero being a 1 on the scale. The human peer felt mostly that Sphero belonged either to the social or to the personal space in equal measure. Compared to the original result of mostly 3 and 4 ratings, only half the reproduced result coincides, the other common rating being a 2 instead of 3.

- Age II in F and D, the merge of *Figure 4.4* and *Figure 4.5* - In total 10 participants divided as follows: 0% rated Sphero being a 7, 6, 4 or 2 on the scale, 20% rated Sphero being a 5 on the proximity scale, 30% rated Sphero being a 3 on the proximity scale, and 50% rated Sphero being a 1 on the scale. The feeling of Sphero being a stranger belonging to the social space were much higher than with Age I, in spite of the positive effect that the dancing scenario has had on the participants, insufficient to make a significant change. Next to the original result of mostly 3 and 4, this goes to show that a quarter chose the same rating, with another quarter having chosen 2 and a half 1.

- Age I and II in F, the merge of *Figure 4.2* and *Figure 4.4* - In total 16 participants of mixed ages, divided as follows: 0% rated Sphero being a 7 or 6 on the scale, 6,25% rated Sphero being a 5 on the proximity scale, 37,5% rated Sphero being a 4 on the scale, 12,5% rated Sphero being a 3, as well as a 2, on the scale, and 31,25% rated Sphero being a 1 on the scale. By combining the two Ages both the predominant perception of personal space attribution and that of the social space attribution carry over to the results, displaying a sample undecided in which direction to classify Sphero. Next to the original result of mostly 3 and 4, this goes to show that the majority did indeed choose, with 1 taking the second place as a preference, followed 3 being as often chosen as 2.

- Age I and II in D, the merge of *Figure 4.3* and *Figure 4.5* - In total 14 participants of mixed ages, divided as follows: 0% rated Sphero being a 7, 7,1% rated Sphero being a 6, 21,5% rated Sphero being a 5 on the proximity scale, 7,1% rated Sphero being a 4 on the scale, 14,2% rated Sphero being a 3, 28,6% rated Sphero being a 2, and 21,5% rated Sphero being a 1 on the scale. Compared to Age I and II in F, the distribution was more even in the case of the D scenario. Due to the distant approach perceived by the Age II group, the proximity hereby is much more socially oriented than personal, compared to the previous combination, with approximately 64,5 percent rating Sphero between a 2, a 1 and a 5 in order. Next the original result of mostly 3 and 4, this goes to show that 4 and 3 were only the 4th and the 5th option, the latter sharing the frequency spot with the rating 6. Hence, they are not the most frequent and preferred options at all.

- Age I and II in F and D, the merge of *Figure 4.2*, *Figure 4.3*, *Figure 4.4* and *Figure 4.5* - In total 30 participants of mixed ages and scenario order, the most substantial database usable, divided as follows: 0% rated Sphero being a 7 on the scale, 3,3% rated Sphero being a 6 on the proximity scale, 13,3% rated Sphero being a 5 on the proximity scale, 23,3% rated Sphero being a 4 on the scale, 13,3% rated Sphero being a 3 on the scale, 20% rated Sphero being a 2 on the scale, and 26,8% rated Sphero being a 1 on the scale. This combination is comparable with the original chart found under *Figure 4.1* in overall scope and aim of the study. Next the original result of mostly 3 and 4, this goes to show that the preferred choice is the option 1 followed closely by the 4th and 2nd, with the 3rd sharing the 4th place with 5 and the 6th rating being the last one. This means that 1 and 4 are the most frequent options, in the order given, as opposed to 4 and 3.

The results from the Proximity Questionnaire support the Age I group as being closest to reaching the same results as the original study. However, these are yet far from similar, especially with the addition of the second participant group and interaction scenario, and will be discussed accordingly in the *Discussion*.

4.5 HRI and Behavior Analysis

Figure 4.6 displays the original observations recorded during the initial study. The original participants tested Sphero's physical condition, by hitting, stopping it in place, kicking it or moving it around, while attempting to play different games with it or reacting to it as to an animal. The speech examples presented rather seem like paraphrases of pbserved emotional reactions, as the statements leave a "polished" impression; however, in the original paper they were presented as if they were think aloud statements. These results also encompass the BB8 interaction so it is unclear how representative they are for Sphero.

Speech	**Physical**	**Behavior**
"Are you human?"	Pick the robot up and examine it.	Play hide-and-seek.
"Hi, I'm André. Nice to meet you!"		
"I'm going to teach you a game."		Play football.
"Are you running away from me, BB-8?"	Push the robot away with the feet.	
"Are you afraid of me?"		Play catch.
"Do you like your creators?"		
"What is your purpose? Just to exist?"	Grab the robot on the floor and stop it.	Wave goodbye to the robot.
"How can we play if you do not stop?"		
"Come here! Come here! Why are you running away?"	Put the robot on the table.	Call the robot like it was a dog.

Figure 4.6: Original Example of Interactions between the Participants and the Robots [26]

The information presented in the lists below is derived from interview and observational data from the replication experiments; similarities and patterns were identified, independent of age group and scenario order. Hence age and scenario order variances are excluded, for a better overview of how participants reacted to their scenarios and Sphero.
The informal interview of each participant, following each scenario individually, was aimed at replacing the questionnaire on behavior interpretation from the original

study, by communicating directly with the people, thus going more in depth with the topic. The original 2 questions were replicated for the "Dance with me" scenario, with an added 3rd one formulated more out of the desire to understand how legible and perceivable the robot dancing was, as well as what it meant for the participants. Each of the answers is presented in the form of a list, detailing all the gathered answers and their occurrence, with the first two belonging to scenario F, and the three afterward to scenario D. These lists are found below, approximated and stacked together in occurrence if interpreted as highly similar, linked to the participant's respective age group. Important to note is that some participants had more opinions and answers than others, hence the numbers do not count up to the total participant sample. The questions and the answers are listed as follows:

F: When the robot touched me and moved away was for:

- 13 of Age I (65%) and 3 of Age II (30%) responded as "ask to follow it".
- 1 of Age I (5%) responded as "to test my reactions".
- 6 of Age I (30%) and 2 of Age II (20%) responded as "to play with me, being playful".
- 3 of Age I (15%) responded as "being confused, disoriented or silly".
- 3 of Age I (15%) responded as "searching for something".
- 1 of Age I (5%) and 1 of Age II (10%) responded as "to fill the room in the background".
- 2 of Age I (10%) responded as "no practical usage, not understood at all".
- 2 of Age (10%) I responded as "following me".
- 1 of Age I (5%) responded as "still to dance to me without music".
- 1 of Age II (10%) responded as "running away from me".
- 1 of Age II (10%) responded as "to ask me to pick it up".
- 1 of Age II (10%) responded as "to provoke emotions".
- 1 of Age II (10%) responded as "to analyze me".

F: When the robot flashed was for:

- 7 of Age I (35%) and 5 of Age II (50%) responded as "asking for attention".
- 6 of Age I (30%) and 1 of Age II (10%) responded as "to communicate".
- 5 of Age I (25%) and 2 of Age II (20%) responded as "did not pay attention to it".
- 1 of Age I (5%) responded as "to mimick me".
- 3 of Age I (15%) responded as "to mark when it had reached it's goal".

- 1 of Age I (5%) responded as “to show it’s closeness to a person”.
- 1 of Age I (5%) responded as “signaling the distance towards the person”.
- 1 of Age I (5%) and 2 of Age II (20%) responded as “to display emotions”.
- 3 of Age I (15%) and 6 of Age II (60%) responded as “to show it’s current status, activity”.
- 1 of Age II (10%) responded as “to show that it can not carry out with it’s task, that it is stuck”.
- 1 of Age II (10%) responded as “being playful”.

D: When the robot touched me while dancing was for:

- 12 of Age I (60%) and 6 of Age II (60%) responded as “dancing with me”.
- 1 of Age I (5%) and 1 of Age II (10%) responded as “not perceiving me, being aware of me”.
- 2 of Age I (10%) responded as “playing with me”.
- 1 of Age I (5%)responded as “to befriend me”.
- 2 of Age I (10%)responded as “to interact with me”.
- 1 of Age I (5%)responded as “to gather information about the participant”.
- 2 of Age I (10%) responded as “to stay close me”.
- 1 of Age I (5%) responded as “to take up contact with me”.
- 1 of Age I (5%) responded as “to ask me to pick it up”.
- 1 of Age I (5%) responded as “following me”.
- 1 of Age II (10%) responded as “asking for attention”.
- 2 of Age II (20%) responded as “moving randomly”.

D: When the robot flashed (while dancing) was for:

- 16 of Age I (80%) and 5 of Age II (50%) responded as “reacting to music”.
- 2 of Age I (10%) and 1 of Age II (10%) responded as “changing the color based on the interaction”.
- 1 of Age I (5%) and 2 of Age II (20%) responded as “asking for attention”.
- 1 of Age I (5%) and 1 of Age II (10%) responded as “to simulate being alive”.
- 5 of Age I (25%) and 2 of Age II (20%) responded as “simulating a disco globe”.

- 2 of Age I (10%) and 1 of Age II (10%) responded as “communicating emotions”.
- 1 of Age I (5%) responded as “did not recognize the intent”.
- 2 of Age I (10%) responded as “reacting to me dancing”.
- 1 of Age I (5%) responded as “signaling (possible) collisions”.
- 1 of Age II (10%) responded as “being playful”.
- 1 of Age II (10%) responded as “communicating”.
- 1 of Age II (10%) responded as “displaying it’s current status”.
- 1 of Age II (10%) responded as “did not pay attention to it”.

D: How do you interpret it? Was the robot dancing, and if so what/why for?

- 12 of Age I (60%) and 6 of Age II (60%) responded as “I felt the robot was dancing”.
- 2 of Age I (10%) and 2 of Age II (20%) responded as “I felt the robot was entertaining me in a cheer-up, playful manner”.
- 3 of Age I (15%) and 1 of Age II (10%) responded as “I felt the robot was seeking to interact positively with me”.
- 3 of Age I (15%) and 2 of Age II (20%) responded as “I felt the robot was moving randomly”.
- 2 of Age I (10%) 2 of Age II (20%) and responded as “Good for animating, parties, elderly, lonely people or as a companion dancer at a club/bar”.
- 4 of Age I (20%) responded as “I felt the robot was reacting to music”.
- 1 of Age I (5%) responded as “I felt the robot was trying to make me follow/mimick it”.
- 2 of Age I (10%) responded as “I felt the robot was following me”.
- 1 of Age II (10%) responded as “asking for attention”.
- 1 of Age II (10%) responded as “not understood”.

The difference in results and the lack of other options being raised by the participants of the replicated study, can be also traced back to the fact that the participants were not presented with a questionnaire, containing a fix set of possible answers to choose from for each question, but instead were interviewed and invited to speak openly and freely. Moreover, by collecting the thoughts of every participant, some might have proposed more than one option, which would have not been possible with a questionnaire, having the “one-answer-only” constraint. The influence this might have had on the study is crucial to keep in mind when reading through these

findings.
The questions asked for the "Follow me" scenario were expecting the same results as initially designed for. However, despite them being replicated for the "Dance with me" scenario, the answers were changed accordingly: to "When the robot touched me while dancing was for:" the expected answer was "to simulate dancing with me OR invite me to dance"; and to "When the robot flashed (while dancing) was for:" the expected answer was "to point out that it is dancing OR reacting to music". These answers were broader accepting, with two possible outcomes, as the goal and aim of the scenario was not designed so strictly and could be thus left for interpretation, given that it is in the right, indicated and desired direction. For the last open question about the participants interpreting the robot, there was no expected result. The questions served more as an introduction for the other two, to help the participants start thinking about the dynamic and musically experience that they just went through.

Originally, the most common answer to the question "When the robot touched me and moved away was for:", with 48% of answers, was "ask for attention" followed by "ask to follow it", with 32% of the answers. In the replicated study, 53% of all participants, 81% belonging to Age I and 29% to Age II, answered correctly, namely "ask to follow it". Following this, the most popular answer with 27% was "to play with me, being playful". The rate of answers with the expected outcome is higher than in the original study, approximately 1.7x greater in fact. Despite the other options not being the same, the original expectation was replicated approximately correctly, with a difference of 5%, considering also the broader age span. Regarding the latter, it is very interesting to notice, that the younger age group had a doubled percentage of understanding the robot than the elderly did, linking it to the differences presented in the literature review to be discussed in the *Discussion*. Gender did not play a role, as there was no pattern observed to affect the outcome.
Moving on, originally to "When the robot flashed was for:", most of the participants, 74%, answered as expected, followed by "displaying happiness", with 23% of the answers. In the replicated study, 40% of all participants, 35% belonging to Age I and 50% to Age II, answered correctly, namely "ask for attention". Following this, the most popular answers were "to show its current status or activity", with 30% response rate; and "to communicate" and "did not pay attention to it", each with 23% occurrence percentage. The obtained results are almost half than the original ones, with "displaying happiness" occurring only from 10% of the participants. Only one of the participants did not touch Sphero at all during a given F scenario and could thus not see the colour change and communication meaning. Here is the form of questioning to blame, having an interview instead of a questionnaire, leaving the interviewee's more space to express their thoughts and formulate them freely. Interesting enough, the elderly achieved a much higher percentage of the expected answer than the younger group. Despite the other's answers being sprung out of the thought of asking for attention, Sphero was not perceived as such, hence the replication failed in this case.
When asked to interpret the "Dance with me" scenario, 60% of the participants understood that it was dancing, with 13% each either considering it to be just entertaining, as in not necessarily having a fitting movement pattern enough to be dancing, or just seeking to interact, or indeed just moving randomly. A sixth of

them did however praise it's use for animating parties, animals, children or elder persons.
The same percentage, namely 60%, also understood that Sphero was bumping into them to make them dance, invite them to it or simulate pair dancing. However, particular individuals still considered it playing, looking to interact or signalise its presence, as well as analyse the participant in their reactions or moving randomly.
Furthermore, 70% of the participants thought Sphero was reacting to music. 23% found it stimulating like an entertaining and inviting disco globe, while 10% still found it to be seeking attention, as it did with the previous scenario and questionnaire.
Overall, the answers for the "Dance with me" scenario had a much higher appropriate response rate than both the original and replicated results for the following scenario. This supports Sphero's legibility and natural perception as an entertainment robot, labeling this part of the experiment as being a success.
Further interpretation of the scenario order effect on the participants, as well as the observed reoccurring behaviors, patterns and trends, are presented bundled up with the PANAS results of their feelings, namely under the name-like following section.

There are a few recurring aspects to detail and present, both in the scenario design and experiment unfolding of the original paper, as well as of the replicated study hereby.
The most predominant of them is the lights feature of Sphero, used solely as its singular means to communicate with the participants. Originally intended for blue to catch attention and signalise happiness, for red to signalise dislike, upset, anger and with white/light blue as being the neutral colour, all the other colours being used only in the disco globe, such as green, yellow and pink, the success rate of changing the colour and always hitting the right RGB lay with the researcher. Thus, under the pressure of being the "wizard" and controlling the robot seamlessly and unnoticed, the same colour was not obtained 100% during each of the experiment runs in the exact same values, timestamp or setting. Hence, the following list of results is an subjective interpretation out of the perspective of the participant, regarding the colour palette and message that they experienced. There is no clear majority in interpreting the colours, however it can be said that most often the participant understood red as meaning dislike or error, blue as being default or movement and light blue as meaning that Sphero is following. Except for the red colour, the rest of the palette used was not understood as intended, at least not with a satisfactory number of participants. The following enumeration displays how often a given colour was observed, both in F and D scenarios, as well as all the various meanings it had been attributed, with an additional occurrence number, if available. Not all participants were paying attention to the colours and some missed them completely, thus there is not one with an occurrence level of 30. The occurrence rate and that of each interpretation do not sum up, as some participants offered more answers than one. Another interesting aspect is that of the really low recognition of the yellow colour. Considering that it was used only in D, where colours were flashing, the assumption that it would be easily mistaken with white or gray due to the brightness levels, as it happened similarly with light blue, means that using it for communication might have indeed not proven efficient at all. The following list displays all recorded interpretations for each reportedly observed colour and behavior, while

the *Table 4.6* offers a summarised overview of it, the intention of communication through colour and lights, how often each colour or pattern was observed and how often these were correctly interpreted. As all colours had no intention beyond that of simulating a disco globe in scenario D, the table is relatable only to the scenario F, unless specifically stated otherwise.

- **Red**, observed by 28 of the participants: 6 regarded is a warm colour, associated with blushing, loving, affectionate or being friendly; 2 perceived it as meaning being worried; 1 suggested that it meant it showed interest, while also being controlled by me; 10 understood that it disliked being touched; 1 thought that knew it was being held; 2 said it inspired calmness; 2 saw it as expressing happiness; 2 found it entertaining and chat-inviting; 4 realised it when they picked Sphero up; 1 thought it signalised collision, danger, being spun; 4 a problem; 2 interpreted it as being lost, confused and/or complaining; 3 suggested it meant that Sphero was less inclined to interact; 1 interpreted it as communicating something;

- **Blue**, observed by 27 of the participants: 1 found it to mean not interacting; 1 associated it with rolling around and flashing; 5 suggested it signaled being active; 5 saw it as a neutral color; 5 felt it was a cold color, 1 suggested it was deactivated or blinking; 2 felt it was thinking meanwhile; 1 thought about phones and suggested it displayed full battery; 1 reported it as being happy; 1 said that it was interesting, fitting the transparent build, being the most artificial color, neon lamp like; 1 thought it communicated like an alright, OK; 6 associated it with it being on the floor; 2 associated it with spinning; 3 suggested it inspired them calmness; 4 said it was asking them to follow it; 2 said it was asking for attention; 1 suggested it looking for closeness; 2 said that it was signaling mimicking them; 10 thought of it as being the default state; 1 interpreted it as being operable; 1 interpreted it as communicating something; 1 associated it with the music; 1 found it unfriendly; 6 however regarded it as being friendly; 1 associated it with signaling an error;

- **Pink**, observed by 15 of the participants: 8 regarded is a warm colour, associated with being friendly, seeking closeness or greeting; 2 perceived it as meaning being worried; 3 meant that it was signaling it not being on the floor; 1 thought of it as meaning it is searching; 3 understood that it disliked being touched; 2 suggested it being affectionate; 1 thought that knew it was being held; 2 said it inspired calmness; 2 saw it as expressing happiness; 2 found it entertaining and chat-inviting; 2 realised it when they picked Sphero up; 1 interpreted it as being lost, confused and/or complaining;

- **Green**, observed by 13 of the participants: 1 reported it as being happy; 1 meant it induced less stress; 1 thought it was more active and blinking; 1 thought it communicated like an alright, OK; 1 associated it with it being on the floor; 1 associated it with spinning; 2 said that it was signaling following them; 1 said that it was signaling mimicking them; 1 meant that Sphero was listening; 1 associated it with an error; 1 said that it was marking being not centered; 1 associated it with the music; 1 found it unfriendly;

- **White**, observed by 6 of the participants: 1 interpreted it as communicating something; 1 found it unfriendly; 1 thought of it showing that it is being held; 3 found it a cold, distant color;

- **Purple**, observed by 5 of the participants: for 1 it signaled to stop; 1 interpreted it as communicating something; 2 saw it as displaying dislike; 1 thought that it was due to being picked up;
- **Yellow**, observed by 4 of the participants: 1 understood that it disliked being touched; 1 meant it induced less stress; 1 found it as following; 1 suggested it supposed it hinted at an error;
- **Light Blue**, observed by 4 of the participants: 2 said that it was signaling following them; 1 suggested that it was signaling mimicking them; 1 meant it was cold;
- **Gray**, observed by 2 of the participants: 1 found it cold; 1 found it unfriendly;
- **Orange**, observed by 2 of the participants: 1 understood that it disliked being touched; 1 realised it when they picked Sphero up;
- **Cyan**, observed by 1 of the participants: 1 saw it when having found the chocolate and reached the end of scenario F.
- Then there is ***blinking*** as a feature, reported by 18 participants: 5 associated it with picking Sphero up; 4 saw it as a form of communication; 3 thought it meant being active; 3 suggested it showed happiness; 3 remarked it as asking for attention; 2 said they observed it when synchronised with the music.

Light Colour	**Communication Intention**	**Observed Rate (OR)**	**Correct Interpretation Rate (% from OR)**
Red	Dislike, sad, and angry	93%	39%
Blue	Follow me, happy, friendly, attention asking	90%	48%
White	Neutral state colour	20%	none
Gray	Alternative interpretation for white	7%	none
Light Blue	No intention, mistaken for white or blue due to Sphero's brightness and contrast levels	13%	50%
Pink	Unintentional occurrence when manually adjusting Sphero's light from blue to red and vice versa, same meaning as red	50%	40%
Purple	Unintentional occurrence when manually adjusting Sphero's light from red to blue and vice versa, same meaning as red	16%	60%

Light Colour	Communication Intention	Observed Rate (OR)	Correct Interpretation Rate (% from OR)
Green	Just a disco globe colour to music and dancing	43%	15%
Yellow	Just a disco globe colour to music and dancing	13%	none
Orange	Just a disco globe colour to music and dancing	7%	none, mistaken for red
Cyan	Just a disco globe colour to music and dancing	3%	none, mistaken for reward signal
Blinking Light	Mimic communication with a given colour, attract attention, simulate wording and dancing to music	60%	100%

Table 4.8: Sphero's Light Communication Overview

Another important aspect and association in the original study is that of Sphero with an animal, rather a pet. During the study replication, 27% of the participants, age-independent, associated Sphero with various pets, the most common ones phrased being twice a cat or a small dog, once a hamster or otherwise any pet or an animal in general. Whether this is enough to support a claim will be elaborated in the discussion. The participants who thought about Sphero in this way were not all owners of pets themselves, more precisely only half of them, with a quarter never having owned and the other quarter having owned at some point in their lifetime. The robot reminded them of a specific animal due to its behaviour, interaction and communication. According to them, it tried to analyse the participant, check the environment, seeming curious, watching and investigative, but at the same time interested in them. Half of the participants suggested it being an ideal toy for pets, kids or even a pet substitute for the elderly or just to animate by dancing in clubs and bars. Other animal-like behaviour mentioned was Sphero following them and looking for ways to approach them. going from being proactive, trying to impress and come off as playful, mostly during "Dance with me", to calmer and holding more back, depending on the reaction of the participant, mostly during "Follow me".
Other participants, namely 20% of them, stated that it first reminded them of a home robot, especially a cleaning robot, such as a vacuum cleaner. Task-oriented and just interacting with the closest object or subject in the room, also because of bumping into them, they felt like it as a lifeless entity which they were hindering in its pathfinding with their presence. Others thought of it completely out of a pointless, senseless, mechanical perspective, with random movement and decision making. However, half of them mentioned after a while, or an entire scenario, that it had taken them a while to notice it actually looking to interact with them and not just being a lifeless robot carrying out tasks. Furthermore, 3 participants felt

stressed for throwing Sphero around (similar to a bowling ball), when they realized that it actually had a purpose. Feeling guilty was not out of financial reasons, as it would be with damaging an expensive home-robot, but rather out of an compassionate and social entity perspective.

Continuing on some of the feelings displayed, which are presented in detail in the following section *PANAS*, 3 to 5 participants interpreted Sphero as running away from them, especially those in the D-F scenario order. They could not understand how it communicates and to what it reacts, especially being presented with D, where Sphero was an entertainment robot first. It was observed, that when Sphero started as being a companion, on the same hierarchical level with the participant, that the interaction was more seamless and quicker understood than otherwise. Once people fixated themselves an idea of the robot it was harder to change it, thus the impression of it running away from them, being confused, mysterious, cold, unfriendly and distant.

Regarding the scenarios' remarks: 3 participants found the dance not necessarily fitting to the music, as having no pattern, and also having no idea if the lights fitted to the music, however they were influenced by the music and over-stimulated; 3 noticed that the pathfinding was lackluster, 4 had a problem with the speed being either too slow or too fast; 4 meant that it had no meaning; 2 supposed that Sphero was unaware of them, not perceiving their movements and refusing to follow them; 2 were worried not to step it; and almost all of the participants reported having a good mood during the "Dance with me" scenario. Without the music, in the order D-F, they were confused, as mentioned earlier, and thought about Sphero being disoriented without the music, thus colliding more often into objects or them. Not dancing had also made them disappointed. Regarding the closeness of the robot, one participant even mentioned, that they had felt the robot much closer as a social entity, than the robot had felt towards them. For other two participants, the robot failed: one mid "Follow me" scenario, due to it running into an error and spinning uncontrollably until reseted and reconnected, fortunately the participant was seated and not interested in the robot and did not realize the "wizard's influence" behind it; and another one between scenario, due to low battery. These incidents however did not impact the opinion of the participants in any negative way, nor positive.

Nonetheless, more than half of the participants expressed the wish to test the robot more, while a quarter mentioned it being good for sale, as well as one of them in particular questioning how the music impacted the robot behavior and suggested different genres and approaches for future work, all to be discussed in the remaining chapters of this thesis.

Following, even though the scope of the replicated study was to remove bias, this might have found a way in through the specific experiences created by the different scenario order. Example given, the huge difference in mood, atmosphere, legibility and perception when going from the "Dance with me" scenario to the "Follow me" one was observed only after a few experiment runs, where the passing from loud music to silence impacted the participant in equal measure as the change in robot's behavior did, going from caretaker to companion. In the reversed case, the participant reaction was instead more positive. This could have led to some bias in the answering of the questionnaires and the interaction willingness. Moreover, more than 2/3 of the participants anthropomorphized Sphero, by often referring to

it using the pronouns he/him and never as a she/her. This even led to the researcher using the wrong pronouns when referring to the robot, subject to the influence of the other peer. It is an interesting observation, aligned with the literature review findings [39] [2] [5] [4], however the discussion and future investigation remains to unveil why Sphero was viewed and addressed to as a male rather than a female. Considering that the participants divided equally in German and English speakers, in German-speaking countries it is most likely a grammatical pronoun aspect influencing the male gender attribution. For the English speaking persons however, it remains a question, of whether the Sphero's appearance design and its blue colour influenced participants' observed gender attribution in any way.
The next tables, found in *Figure 4.7* and *Figure 4.8* combined, display the observation made during the replication study. The physical results can be highlighted as being highly similar, with the participants testing the robot, constraining, hitting or moving it repeatedly in many ways. There was only one participant who did not touch Sphero and around five who did not block, hit or hold it. The behaviour is very different compared to the original one. While around 6 participants did have the impression of Sphero impersonating various pets, all the other games, attempted by the original participants, were not reproduced. Instead, the participants focused more on the novelty effect of Sphero, on its surprise effect approaching them from behind, as on the "wizard" in the room. At least one third of participants asked the researcher, whether they were controlling the robot, with 4 of them realising it after further intentional tests of Sphero's capabilities and the researcher's coordination. Although the researcher did not reveal that they controlled the robot, participants were still asked to imagine it as being so, so as to not influence the recorded answers. The participants tried mainly to understand Sphero, followed it, analysed it from afar and close-up, hindered and hurting it in different ways with their members, danced with or to it, or just sat on the chair, laid on the ground or stood unmoved to understand how it works. Although Sphero was often perceived as playful, they did not play with it in the given scenario duration. Moreover, the verbal communication was much more restrained. This resumed itself to asking the researcher questions about the robot or expressing joy, surprises or confusion, as they felt it in the moment. Elaborate sentences, as presented in the original table under *Figure 4.6*, almost philosophical questions, were not posed. Despite this, one can see the similarities in what the participants felt and how they reacted, all classified under the influence of novelty, curiosity, confusion, joy and being startled. Only one in ten participants actually spoke thinking out loud, following the suggestion of the researcher to feel free to interact, do, or speak however and whatever they want. It is possible, that, through the initial encouragement of the researcher, the participant actually became reserved, finding it odd to speak alone. However, it was proven beneficial, as shown during the pilot test trials, to help the participants get up and interact with it more comfortably and naturally.
In the scope of the replication, the physical column is the one best reproduced, with the other only qualitatively analysed and approximately joined together.

Speech	Physical	Behavior
Ah, it's red now.	Crouch, use hands, arms and feet, move around it.	Touch, grab it. Pick it up. Throw it around. Hold it between arms in place or in the palm. Block it with arms or feet. Kick it. Hop, skip around it. Shove it with the feet. Place it on a table or chair.
Ah, it gets red, he is furious, he does not like it, hurting him.		
It's again red.		
He follows me! Does he see me?		Moving away from Sphero.
He comes again to me, or? It is not moving anymore.		
Can I touch it? I do not want to break it. Can I pick him up? He is so small.		Reach out to Sphero to grab it or offer it an open hand.
Oh my god I expected something but not this, OK, I am trying to understand it. Do I have to move?		
Wow, hi! Can I touch it?	Turn towards Sphero and crouch on the seat. Touching it whenever possible.	Analysing it, focussing on it closely, looking for a pattern. Seems curious.
Hello! What, he is so sweet.		
I wonder what he is good for, can it be that he does it randomly?		
This is not a game, is it safe?		
Oh, how cute you are, hello.		
Totally the surprise, cool!		

Figure 4.7: Interaction Speech, Physical and Behavior Observations

Speech	Physical	Behavior
Is it supposed to follow me or something?	Standing in place, laying on the ground, sitting next to Sphero or keeping distance towards it. Switching positions and postures.	Testing the robot's capabilities, analyzing it, looking it.
Does he react to my voice? Does he react to humans?		
Does he have a scanner, a camera, a sensor? How does he see me?		
Ouch! Don't hurt yourself, I am afraid it is hurting itself		Sphero bangs into table or chair foot.
-Laughter- Oh, it keeps spinning!	Moving around to dance with it. Crossing legs with Sphero spiralling between them. Trying to copy or come up with a choregraphy.	Investigative, stiff or freeform dancing.
He is really funny, I must admit.		
Where is the robot? Aaah.	Seated turning around.	Sphero approaching from behind.
I do not understand how it is supposed to work.		
It follows me around!		Looking scared of it.
I do not know what to do.	Dance around it, step and hop.	Try to imitate or understand the choreography, dance. Laugh, cheer, exclamate joy, surprise and confusion.
You are moving it, damn. Ask the researcher to pick up their hands.	Look distrusting.	Testing the wizard of control over Sphero.
Are you controlling it? Do you move it?	Picking Sphero up next to the researcher.	Look distrusting at Sphero and the researcher.
Without music I can focus more on it.		Following Sphero.
Do you want to follow or should I follow you?	Reaching out to touch it seated.Stand and look at it.	Mesmerized by Sphero, also relaxed to just observe it.
		Disinterested in the interaction, not motivated to stand up.

Figure 4.8: Interaction Speech, Physical and Behavior Observations Continuation

4.6 PANAS Questionnaire

After each scenario, the participants were asked to fill in a PANAS table sheet, in addition to the original questionnaires found in the paper. Containing 20 standardized emotions and 4 possible choices, participants can express how the interaction with Sphero made them feel, in the given context. The results, grouped by age group and scenario order, are presented in the following tables, from *4.9* to *4.12*. A blank field means that the emotion in the respective intensity was not felt by any of the participants. A blank row translates to the emotion being marked as "not at all", not having occurred during any of the experiments.
The PANAS questionnaire was used not only to collect how people thought and felt about the respective scenario, but also to facilitate the start of the informal interview, detailed under the previous section on HRI and behavior. Moreover, the goal was to observe and simultaneously record any changes in the participant's interpretation, going through two scenario, thus also becoming loosened from the grip of novelty and the Hawthorne effect. Employing PANAS proved also suitable for analysing the impact of Sphero in a long-term interaction study, considering at least that the replicated study was longer than the original one.

Scenario:	**"Follow me" with Age I**				**"Dance with me" with Age I**			
Emotion/ Intensity	**A little**	**Moderately**	**Quite a bit**	**Extremely**	**A little**	**Moderately**	**Quite a bit**	**Extremely**
Interested		1	6	4			4	7
Distressed	3	2			3			
Excited	3		4	4	1	2	2	6
Upset		1			1			
Strong		1	1			2		
Guilty		2			1	1		
Scared	2							
Hostile	1	1	2	5	1	5	2	2
Enthusiastic	1	1	2	5	1	5	2	2
Proud	2		1	1			1	1
Irritable	1				1			
Alert	1	1	4	1	1	1	3	2
Ashamed								
Inspired		1	2	4	1	1	3	3
Nervous	2		1		1			
Determined						2		2
Attentive		2	2	3		1		5
Jittery	3				3			
Active	4	1	1	4	2	2	1	6
Afraid		1			1			

Table 4.9: PANAS Results of **Age I** after *F*, where the rating scale ranges from *blank = not at all* to *extremely = totally experienced* and the numeric values represent the count of the occurrence

Starting with the *Table 4.9*, representing the group Age I in the scenario order F-D, the feeling of novelty can be observed through the aggregation of different emotions. Side by side with the F scenario, participants felt 10% less hostile and enthusiastic, 7% less distressed, nervous and scared. This shows that after the surprise of meeting Sphero and the novelty effect blurred out, and they supposedly got to understand how Sphero functioned and that it was not a threat, the negative feelings, as well as the positive enthusiastic ones, dropped in intensity. Moreover, with an increased rating of being interested, excited and inspired, it can be deduced that Sphero was viewed more positively and friendly by dancing than by trying to make people follow it, in a cold, unfriendly and commanding way.

The increase in positive feelings in the scenario order F-D, as well as the decrease in the other way around, of D-F, is a trend observed throughout the entire study replication, repeated in the other tables and results as well. The fading of the novelty effect is also an established and recurring phenomena. Additionally, due to the shift in paradigm of Sphero's behavior, communication and purpose, as well as of the widely different scenario design and goal, most participants were confused to observe, that what they had initially learned about the robot was now not applying anymore, e.g: first flashing to communicate then randomly to music, and first flashing randomly for entertainment and then actually trying to impersonate a social entity vice-versa. The increase in the determination and active feelings is also a repetitive and expected happening, due to the nature of the D being more active and dynamic than F. Also, most of the participants were more determined to understand Sphero further, following the scenario order D-F.

Scenario:	**"Dance with me" with Age I**				**"Follow me" with Age I**			
Emotion/ Intensity	**A little**	**Moderately**	**Quite a bit**	**Extremely**	**A little**	**Moderately**	**Quite a bit**	**Extremely**
Interested	1		2	5		1	3	5
Distressed		1			1			
Excited		3	2	4	2	1	4	1
Upset		1			1			
Strong		1	1		3		1	
Guilty					3			
Scared	6							
Hostile					1	1		
Enthusiastic	1	1	2	4		3	3	2
Proud		3			2		1	1
Irritable			1		1	1		
Alert	5	1	1	1	1	3	2	
Ashamed					1			
Inspired	2	4		1	2	2	2	
Nervous	2				1	1		
Determined	1	3				2	1	
Attentive	2	2	2	1		3	4	
Jittery	2			1	1		1	
Active	1	2	3	2		4	4	
Afraid								

Table 4.10: PANAS Results of **Age I** after *D*, where the rating scale ranges from *blank = not at all* to *extremely = totally experienced* and the numeric values represent the count of the occurrence

Continuing with *Table 4.10*, representing the group Age I in the scenario order D-F, the recurring drop in novelty can be observed through the decrease of the intensity of feelings such as distressed, upset, scared, excited, enthusiastic and attentive. The occurrence and intensity of interested and determined went up when switching from F to D, as the participants were first now confronted with Sphero acting as a conversation and leading partner, rather than just a dancing robot. Feelings such as strong, hostile, proud and determined also arose or increased due to this shift in paradigm. The participants finally had the chance to pick it up, kick it and study it without being distracted by its chaotic movement and music, hence felt that they had power over it and were content with their accomplished understanding, while wishing to find out more. The physical trial of Sphero also led to them feeling strong but guilty or ashamed, for having supposedly harmed an innocent robot, as described by them. The shift was also often described as leading to the robot seeming much more distant, cold, unfriendly and distant towards the participants, as opposed to the F-D order, where Sphero had opened up in the second scenario. The active levels dropping were expected, as the scenario was more a cognitive than physical one, with nervous' increase in intensity being a surprise, as some participants did not yet comprehend the shift and change in behavior. The trend of participant excitement and interest dropping in the scenario order D-F, as opposed to F-D, is observed throughout the entire study and is a valid perspective for the remaining tables as well.

Scenario:	**"Follow me" with Age II**				**"Dance with me" with Age II**			
Emotion/ Intensity	**A little**	**Moderately**	**Quite a bit**	**Extremely**	**A little**	**Moderately**	**Quite a bit**	**Extremely**
Interested			3	2			1	4
Distressed	1		1				1	
Excited			3	2		1	3	1
Upset	1	1						
Strong		3	1			1		
Guilty			1					
Scared	1		1					
Hostile								
Enthusiastic		1	2	2		1	2	1
Proud	2		1		1			
Irritable	1							
Alert	1	1	3		1		2	1
Ashamed	1							
Inspired		2	1	1	1		1	2
Nervous	1	1			1			
Determined	2	2				2		
Attentive		1	1	2		1	1	2
Jittery		1	1		1			
Active		2	2	1		1	3	1
Afraid								

Table 4.11: PANAS Results of **Age II** after *F*, where the rating scale ranges from *blank = not at all* to *extremely = totally experienced* and the numeric values represent the count of the occurrence

Switching to the group Age II, and starting with the scenario order F-D as presented under *Table 4.11*, the same conclusions can be drawn as for the Age I. Interest spiked in intensity, following the scenario F with D, and saw a doubling of the highest intensity for being inspired, as did for the feeling of being alert, due to the new behavior expressed by Sphero. By having gotten accustomed to Sphero and with the first interaction round behind them, with the pause facilitating a moment to think and meditate, the decrease in the feelings of distressed, excitement, scared, nervous and jittery is logically deduced and as expected. The lowering in occurrence of strong, guilty, proud, irritable and determined, by 50% to 75% respectively, is also normal due to the change in scenario: the participants did not have the chance to grab and physically interact with Sphero during D as much as they had during F, due to it being constantly on the move and less inviting to be touched, kicked, picked up or having its path blocked. Likewise also the increase in activity is an expected outcome. An interesting finding here nonetheless is that the participant's attention and feeling inspired increased in intensity by 7%, respectively stayed the same, showing that the Age II was more interested in Sphero as an entertainment robot and not dropping due to the novelty effect dissipating.

Moving on to the last table, namely *Table 4.11*, a comparison between the group Age II in scenario order F-D and D-F can be sketched. The downwards trend of D-F positive emotions is obvious, with these participants noting a greater decrease in occurrence and intensity for feelings such as interested, excited, enthusiastic, irritable, alert, inspired, active and afraid, ranging from a drop by -20% up to -100%. Besides the expected outcomes presented so far, it was interesting to observe, that for these participants the attention stayed the same, with the inspiration following closely behind. This goes to show a trend for Age II being much more invested in new technology than Age I, supported by claims made under the chapter *Related Work/HRI*. Pride was also noted by 20% of the participants, due to them understanding how Sphero functions during their second interaction, as well as being enthusiastic, despite lower in intensity, being marked by a further participant than during D. Some had felt overwhelmed by the dancing robot and were welcoming a moment of peace and analysis, having the opportunity to get to understand and know Sphero better. The equality in occurrence and intensity of strong is a special case, as the participants interacting with Sphero grabbed it in both scenarios. Others felt upset by the switch from an entertainment to a "dull" robot, hence the increase in feeling upset by 20% and distressed by 10% in their lowest intensity available.

Scenario:	**"Dance with me" with Age II**				**"Follow me" with Age II**			
Emotion/ Intensity	**A little**	**Moderately**	**Quite a bit**	**Extremely**	**A little**	**Moderately**	**Quite a bit**	**Extremely**
Interested	1		1	3			2	2
Distressed	1				2			
Excited				4	1		2	1
Upset					2			
Strong			1				1	
Guilty	1				1	1		
Scared	1							
Hostile								
Enthusiastic			2	2		1	2	1
Proud					2			
Irritable	1							
Alert	2						1	
Ashamed					1			
Inspired	1		1	1		2		1
Nervous	2				1			
Determined		1					3	
Attentive		1	2	1		1	2	1
Jittery	1							
Active		1	3			1	1	1
Afraid	1							

Table 4.12: PANAS Results of **Age II** after *D*, where the rating scale ranges from *blank = not at all* to *extremely = totally experienced* and the numeric values represent the count of the occurrence

Overall, it can be deduced that what was assumed about the younger and older persons in the literature review is true, concluding the findings in the tables. Age I graded on average twice as more often the robot as being exciting, interesting and them themselves enthusiastic about it, curios of what its capabilities are. However, they also lost their interest much faster in it, than Age II did. Due to the number of participants in Age II being half of those in Age I assumptions can only be made half-based, however a trend in being more attentive and inspired, but also more negative, unfriendly towards Sphero during the novelty effect, stand out from the table, as well as the explanations for different feelings and beyond them behaviors portrayed by the participants. The findings presented here, as well as the emotional progression trends, recurring patterns, emotions and behaviors observed during the experiments in relation to the PANAS questionnaire are further detailed and discussed in the following chapter.

Chapter 5

Discussion

Considering that the findings of the replicated study have been thoroughly laid on paper and compared to the original study, the discussion serves to offer an answer for the research questions, but also a deeper look into the outcome of the replication study as it is, as well as to judge the claims made by the original paper in its own *Findings, Discussion* and *Conclusion* chapters, especially considering that these were marked by Faria et al. [26] as being not conclusive, yet however enough to base affirmations, conclusions and learnings on them. All the citations that follow are extracted from their scientific paper. Knowledge gained from the *Related Work* is applied in relation to the findings of the replication study to offer a trustworthy, valid and founded discussion perspective, affirmations and claims.

5.1 Discussion Comparison

> "[...] The participants that interacted with the robots in this experiment understood what the robot meant to communicate, albeit the fact that most of the participants did not associate the behavior of touching and moving away as a request to follow. The fact that said association was not made in most cases is interesting, because most of the participants understood that the robot wanted them to follow it. As such, this may indicate that a request, such as asking to follow, needs more than a simple action of moving in, touching and moving away from a person. [...] interactions were very different from one participant to another [...]". [26, p. 668]

The original study claim was supported by 32% of the 17 participants, who answered as expected to the behavior questionnaire.
Considering that the replicated study achieved 53% of the participants answering correctly, it poses a special case. Logically speaking, the affirmation is clearly negated by the replication study not managing to replicate its value, surpassing it instead by an increase or 66% in valid answers. From a findings perspective however, the results speak for themselves and strongly justify the conclusion drawn. It is also true, that participants followed Sphero willingly or not. Even of those who did not understand it as such, 13 of 30 participants, only 5 actually did not get up from their seat to interact with Sphero, resulting in 83% of participants actually being led to the finish chair with chocolate. This again is also a much higher percentage than the original 71% success rate. Not to be overlooked of course, only 57% indeed

understood that they had to follow Sphero. Still, 25 people did complete the scenario nonetheless. This all aligns with the suggestion of expanding on the "Follow me" order communication, beyond just bumping into the person and lighting up. In this regard, the original claim is upheld, although the values are not the same. To this extent, it is also important to mention the different socio-demographics of the sample. Values might never be the same unless for one the participants really are as similar as possible amongst studies, and for the other the communication is much improved, both in its perceivability as well as legibility.

> "[...] The majority of the participants tended to perceive the robot as a "social entity". In fact, some of the participants engaged in conversation with the robot while others tried to communicate with it through gestures. [...]". [26, p. 668]

The original values were both of personal level, namely rating 4 and 3 obtained by the questionnaire. The replicated study showed, that only a 50% match could be achieved. Age I group, closest to the original one, reported most often 4 and 2 following the "Follow me" scenario, thus varying between personal and social space. The "Dance with me" scored even lower, with half of the answers being 2 and 1. Age II reported 2, 5 and 1 during D, and 1 and 3 during F. Even combined, the results clearly point towards 4, then 1 being the preferred answers. Thus, despite the Age II group lowering the score, it can be said with certitude that participants were feeling so close to Sphero. The "Dance with me" scenario was also reported more distant, whereas the robot portrayed a perceived-as-mindless entertainment robot, not necessarily interacting but just dancing around. Assuming that the original participants were more inclined towards informatics, the majority of them having that background and being also much younger than the participants in the replication study, that the results could have been representative for the sample, that they have indeed felt more personal than social towards Sphero during the replication the original claim is not strongly supported and evidenced.

> "[...] The results show that most of the participants understood what the robot meant to communicate and both robots conclude that the use of behaviors inspired in pets is useful when developing communication between humans and robots that rely only in simple ways to communicate. However, not all of the participants understood what the robots' meant to communicate. Observing the data recorded, no conclusion about the reason for this fact could be extracted, since the results of the questionnaires are similar for both cases where the robots' were followed and when they were not. [...]" [26, p. 668–669]

Beginning with the first claim, this was based on a 48% correct answer rate to the original questions, assuming however that the other participants also understood Sphero although not as following. How this is possible is inexplicable, yet the replicated results were much higher than the original ones, sitting at 65% for Age I and just 30% for Age II. As the first group is more relevant for the original replication, the original supposition is indeed confirmed, being supported by higher numbers

than before. However, looking at just the original numbers and deductions these are not trustworthy alone.
The pet behaviour was observed by less than 40% of the participants. Half of those who observed it failed to correctly understand Sphero in its intent and purpose, associating it too much with a curious small cat or small dog, or simply with a home cleaning robot, not worth interacting or any attention. Borrowing from pets was indeed useful, but it needs to be further expanded upon with the unique features and capabilities of the robot in itself, as Keunwook stated. [8].
The results in questionnaires were indeed similar, both for participants who interacted with Sphero and of those who did not. The novelty effect, high in excitement, curiosity, hostility, unease and enthusiasm could be observed no matter the participant, dropping after the second scenario in intensity. However, the drop of positive emotions, attentions, activeness and determination was much lower and slower than with those participants who did not interact with Sphero at all. No parallel can be drawn upon gender, background and interaction level. However, as it was also stated in the literature review, a difference between the older and the younger age group could be remarked. The latter one had higher expectations of Sphero, were less negative impacted by it initially and quicker to jump into the interaction, forming ideas and drawing conclusions. They were also the ones to lose faster the intensity of emotions displayed, as they got to assume that they know what Sphero does, proving not worth further time and attention. The other age group however, although being a bit more reluctant towards Sphero, showed a much more emotional and timely investment in it, their interest not fading with the progression of the experiment but the other way around, being more interested in how it works, proposing ideas of and for it, as well as discussing its practical usage, as well as technical capabilities, in much more detail. A closer future look into age differences in interaction, with a bigger sample is important to determine further conclusions and analysis.

> "[...] Both the perceived intelligence and likability had similar scores. [...]". [26, p. 669]

As shown under *Findings/Godspeed Questionnaire*, this is not completely replicated either. While the likeability results, both for the groups individually and combined, tended to be almost as high as the original results, the perceived intelligence was rated by around half a point lower. Hence, not only were they not equal, but the original claim on its intelligence, perceivability and legibility in communication, is further undermined. Sphero is indeed a likeable robot, but it is not perceived as intelligent due to it's embodiment and functionality. Moreover, having applied the "Wizard-of-Oz", with a few participants pointing it out, it might have definitely lowered their standards in perception of Sphero. The original claim is undoubtedly false, assuming that the bias in the participant sample led to its demise.

"[...] When talking with some of the participants that did not follow the robots, they reported that they did not understand that they could leave the room and because of that they did not follow it. Besides this fact, one other reason, that seems to us as the motive for people not following the robots, is that perhaps humans do not associate only the physical movement to the intention of the movement and so some participants did not associate the intention to the movement. [...]". [26, p. 669]

This original assumption can only be compared to, as the experiments unfolded hereby differently. There were no two rooms for the participants to feel intimidated by, not knowing their boundaries. Moreover, they were specifically instructed to interact as they wish, the rules and boundaries of the extent of what is permitted was clearly communicated to them. Thus, around 80% did stand up and many of them followed Sphero, knowingly or unknowingly, as presented in the previous chapter. The reason why some participants did not do so could have been the disassociation between the movement and incentive, but most of them did not display interest in it and were not eager to get up at all. Watching was enough, as it often is with pets or home robots as well, coming from them. Sphero was very dynamic and graphic, not only in D but also in F, so the movement in itself was most likely not the reason.

"[...] The results about the proximity, similar in both the followed and not followed cases, are interesting, since they show that these simple robots, can not only communicate but also trigger a bond with people, which is a helpful feature not only for people to understand robots more easily, but to open a new spectrum of applications for robots of this kind. [...]". [26, p. 669]

Having already discussed on proximity earlier, the observed bond with people is not something the results of this thesis are supporting. While indeed Sphero was often anthropomorphized and rated a 4, it was also rated a 2, with the older participants placing it more often than not in a social spectrum. For such a bond to be successful in future application and implementation, it needs to be rated at least a 4, if not a 5, to cross the intimate space of the participant's perception. A cleaning robot breaking down and being replaced by a newer model, versus a house pet dying and being mourned over, are not the same thing emotionally speaking.The lack of anthropomorphism seems to negatively impact the bonding, even if the robot is perceived as interactive, smart and friendly the robot is. The robot can however attract attention and raise through its novelty but not beyond a technological curiosity level, and it most likely can not maintain it, an issue observed by other researchers as well and mentioned in the literature review. [27]
On a different note, the results on proximity would lead one to speculate that the caretaker paradigm is not as impactful as the companion one. Having a companion robot, a conflict due to a living and non-living being standing at the same hierarchical level [53], having the same social status in commanding each other, as it was with Sphero during "Follow me", definitely raised the perception on it as a social entity, as otherwise in the entertaining "Dance with me" scenario, where participants rather perceived it as a simple machine and looked down upon it.
Nonetheless, more data is required for the previous claims made here, presenting an

ideal opportunity for future studies.

> "[...] Thus, this study showed that using implicit communication based on behaviors and actions associated to other beings with similar communication capabilities (e.g. pets), lets people understand what a robot wants to communicate, even when the robot lacks the usual means of communication. This fact is of great importance because sometimes it is not viable to use a robot capable of talking or moving arms, to communicate and interact with people, and with a model of communication like this it is possible to use a simpler robot to achieve similar communication outcomes. [...]". [26, p. 669]

While true that the Sphero behavior was inspired by pets, most clearly seen in its pattern to attract attention during F, also employed partially during D, however by replacing the light communication with a disco globe effect, not many participants associated it with a pet. It could be that people subconsciously understood its intention due to it, considering those who did, although it was not associated with an animal. When a robot lacks the usual means of communication it is important to not try and mimick only humans and animals, as confirmed by researching it under *Chapter 2*, but also try to bring in its own approach, tailored to fit the capabilities while not obstructing it's function. [28][15] [14] [7] The most important thing for a robot to achieve in its interaction is to provoke strong emotions, as observed during the experiments and suggested by literature sources as well. [7] [30] A simpler robot could as a following be used conditionally, whereas people would require training and learning to understand it, e.g. a robot based only on lights or movement, similar to a street light, or its design and capabilities would have to be greater improved, refined and extended. A robot with arms on the other hand, would actually be better at communicating due to it being able to express postures and gestures [47] [16] [11]

Lastly, the proposed future work was due to be analyzed hereby as well, with regard to the overall findings and the present comparison discussion.

> "[...] In the future, we intend to extend the approach developed in this work - using pet like behaviors - with other non-anthropomorphic robots like aerial drones or Sphero Ollie (the other robot from Sphero), since these robots have different constraints and offer new interaction possibilities. [...] It is important to study how these drones can communicate with people in a fluid and natural way, without having to resort to external devices like smartphones. [...]". [26, p. 669]

This is a very interesting claim, as drones are even less anthropomorphic and more machine-like than most non-humanoid robots. Furthermore, it is unclear how these are expected to communicate, as there are only viewed through the caretaker paradigm prism, having a function to fulfill and nothing beyond it. Lacking any anthropomorphic similarity, it is highly likely that without a voice a communication between and humans would be impossible to facilitate. The other proposed robot

Ollie is also discontinued and information, except for an archived FAQ, can not be found of it on the internet. [65] At the time the original study was undertaken it might have been an option, as was the hype with drone delivery systems making a name for itself in the field of informatics, but nowadays neither of these seem an important future research to undertake. Future work should instead focus more on the human-side of the HRI, as well as polishing, refining and further innovating in non-verbal communication possibilities using Sphero. There is still a lot unknown or not systematically enough explored about the different persona background and characteristics and their influence the interaction with the robot.

"[...] Another aspect that is interesting to study related with how people react to this type of robots in an open space when they are not expecting to be approached by a robot. [...]" [26, p. 669]

This citation was highlighted at the end to discuss the approach from behind. Both the original study and the replicated study made use of it, despite it going against social norms and indications, as mentioned in the review of the literature. [35] [34] It is unknown why the original research made use of it, however it did help the singular researcher in replicating the study to hide, most of the times at least, the employment of the "Wizard-of-Oz" technique. Yet, for a robot to be even more acceptable, thus also obtain higher proximity, perception and behavior scores with the aforementioned questionnaires, it should definitely be aware of human social and politeness norms, as well as possess more anticipatory capabilities, life-form-like behaviour and cognitive power, as confirmed by the literature research. [16] [19] [40] [39] [13] To this extent future work can be undertaken, with a more technical approach than faking the robot's movement, whereas the suggestion on proximity, approach, navigation, social intelligence and emotional impact are indeed regarded in the study. Then the real value of the interaction will be truly observed and analyzed.

5.2 Study Replication Crisis

The research questions need to be addressed and answered appropriately, to conclude this thesis research.

How accurate are the results of a reproduced study in comparison with the original one?

This question has already been addressed indirectly in many places over the past two chapters. These are summarised as follows:

- **"Follow me" scenario**: the success rate was higher than originally presented with, thus actually confirming the previous unfounded claim, that most of the participants understood what Sphero tried to communicate. Considering both age groups altogether, this replication has failed in obtaining the same results but managed to achieve the same conclusion.

- **Godspeed questionnaire**: the results deviation was high, especially considering the Perceived Intelligence dimension. However, factors such as researcher, context, participant background and the "Wizard-of-Oz", as well as the Hawthorne effect have to be taken into consideration. Likeability almost obtained the same values for all groups, both original and current, yet it is not enough to support the full replication of the study.
- **Proximity questionnaire**: here again, the results could be half-way replicated, leading to a more negative outcome than originally presented. Thus, the replication crisis also becomes apparent here.
- **Behaviour analysis**: although the new participants had a much wider array of speech, physical and behavioral outings, as well as much broader and personalized answers for the questions on Sphero's behavior, there could be seen a similarity in physical-behavioral reactions and popular answers. However, basing this conclusion only on the vast data available, and the somewhat unfounded claims made by the original researchers, is not enough to rule out the replication crisis. The results were after all not the same. Moreover, this qualitative data gathering and analysis is very much subject to each individual participant, hence also favoring the replication crisis apparition. Nonetheless, this analysis is the closest one to the original out of all aspects of the study.
- **Robot behaviour**: the navigational and lightening communication of Sphero was reproduced as originally intended, however it was not understood by everybody, as the paper claimed that most did. Most did understand it communicating indeed, but only a bit over half actually understood the desired intention and incentive. Just because people interpreted the behaviour of the robot in a for research favourable condition, does not mean that the original results were replicated. Hereby as well is worth considering all the study unfolding differences. The lights were moreover not perceived as blue and red, but as a multitude of colours, also due to the "Dance with me" scenario, with many participants not even paying attention to them, not distinguishing the colours or not associating any meaning to them. Thus, it is reasonable to conclude that the exact same outcome could not be replicated. As stated earlier, the effect and direction of the study might have been reproduced at a higher success rate, but the data and analysis of it are not identical.
- **Study duration**: Although too subjective to take into consideration, it is worth mentioning that overall the participants spent less time in each of the scenarios than the original ones did, yet another deviation from the original study.

To conclude this question, to faithfully replicate a study is dependent on so many factors, that it was almost impossible. Despite the numerous guidelines presented in the *Related Work* [22] [21] [25], and the many limitations and deviations from the original study in the *Methodological approach* and *Findings*, it is still very challenging to accurately pinpoint the exact ratios in equations between the two studies, considering all values, factors and settings. Thus, as with every iteration of the study, it is not feasible to obtain the same results in any way, 100%, even if a time-travel would be possible, not only due to the researcher, materials and procedure, but also because of the participant as a not static being themselves.

How valid and trustworthy is the original HRI study considering a comparison with its replication?

Building up on the previous question, it can be concluded that the original paper is not grounded on much supportive evidence and data, making it less trustworthy. Lacking information, presumably presenting assumptions made and conclusions drawn on unexplained or insufficient data, as well as overall missing substrate to be successfully replicated leads to my judgement of the original paper as poorly written up scientific work. The study replication is meant to salvage and extract everything possible out of it, fill in the blanks, add diversity to the study and expand it for more valid results and less bias flowing into them, all while aiming at creating a baseline for future work. The attempt of this thesis suggests that the comparison indeed raises questions on the validity of the original paper. However, it is crucial to further replicate this in the future and improve the claims based on results, continuously refining the approach, the materials and the scope of the study so as to at some point indeed reach a universal truth, unchangeable if replicated or varying just by tiny margins, including all kinds of individuals and robots and a focus on human-sided robot interaction.
By wrapping these two research questions together this thesis would suggest, that the replication crisis phenomena is indeed affecting the scientific field of HRI, requiring further research, conclusions and measures to correct it.

How well can a different non-verbal scenario with the Sphero robot be interpreted, replicating a previous study?

Having expanded the study by adding the "Dance with me" scenario, it is important to discuss this at last in detail. Answering the question, the scenario was received very positively, achieving overall higher scores in both perception; proximity, mostly for Age I; interpretation and behavior. It could be clearly observed how much more positive, free and open participants felt the moment they had catalogued Sphero as being a mindless entertainment machine, a caretaker robot, as opposed to the companion robot, challenging them cognitively and socially at the same time. Observing the results, going from F to D and D to F pointed out major differences, resulting in people ending the experiment with D having a much more positive impression of Sphero, presumably having learned how to correctly interpret it during F, than otherwise. When confronted with F participants were confused, annoyed and somewhat unprepared, as the robot had already been analyzed and classified in their minds, with it now turning the tables and forcing them to start from 0 with its totally new behavior. The aspect of anthropomorphization observed during the study deserves its own future research into gender and roles attribution of robots based on their features and characteristics. Sphero in itself was very suitable for the dancing scenario, with it being a ball, especially able to spin in place, interact with the participant and force them to move, so that it does not constantly bump into them. The genre and design of the scenario and dance choreography in itself favored a positive reaction, with the question remaining, as posed by one participant during the interview, how the robot would react and be perceived, where the music of a different genre, rhythm, style and nature. Basing it also on the teaching from the "Follow me" scenario, this is the only research question which gathered a positive outcome, supporting the creation, innovation and addition of as many non-verbal scenarios as possible in future replications and work, to better test Sphero and other non-humanoid robots for an extended duration of time, in a prolonged interaction session and to collect more varied results, which combined offer the big picture on the human-robot-interaction to this regard.

Chapter 6

Conclusion

In this work, the goal was to replicate an original scientific paper, aiming to discover whether or not it was possible for a non-humanoid and non-verbal robot, lacking means of communication like speech, gestures and gaze, to communicate its intentions in an understandable way; the robot being one Sphero SPRK+ and a BB8.
The replicated study used all possible information from the original paper, given its scarcity in detail, lack of crucial information and unfounded claims made following its output. Using only a Sphero and a room available at the university, the study was designed and carried out as similar to the original as possible, taking all the materials and limitations presented under the 3rd chapter into consideration. Furthermore, it was extended by an additional age group, added variety in unfolding and a new scenario, to further collect more data, remove bias and explore the capabilities of the Sphero robot in a non-verbal communication context.
Concluding the research questions, the results supported the existence of the replication crisis, with the original study falling into its sphere of influence, while at the same time tried to offer an improved version of it, a 2.0 of "Follow me", with added guidelines and elaborated details on settings for future replication. With each iteration of the replication, the results are expected to become closer and closer to the previous ones, until a replication issues free variant can be obtained.
Learning from the experience collected in this study, it is crucial to have more researchers available, to ease the work during preparation, unfolding and grading of the study. Moreover, the interaction should be automated and continuously improved, to avoid the "Wizard-of-Oz" and Hawthorne influences. Furthermore, it is crucial to obtain the exactly same materials, samples, procedures and outcomes when considering the replication crisis, which will be never fully manageable. Even if everything is carried out identically, and Sphero has the exact same movements and lighting duration, also challenged by the unforeseeable participants behavior, given that these are also identical in background, knowledge and character, from one study to the next, it is impossible to remove the fluctuating aspect of the human invariance from any given HRI study.
Otherwise, the spectrum of non-verbal communicative, non-humanoid robots is in dire need for innovation and design of interaction and behavior, not extracted, influenced or plainly mimicking human or pet ones. Robots require their own interaction perspective, principles and rules, while also following existing norms the more intelligent they become. And humans need to learn about them as being robots, not mere machines, as well.

Future research should be centered around replicating the current study, using it as a baseline to start from, and not only. With regard to the replication study, the first step would be the improvement of the materials, test lab and robot consciousness, navigation and behavior, especially by automating it. Once a test lab has been arranged for and materials refined, to a point where the similarity with the original study is higher than the 99 percentile, and repeatedly reproducible in a static level, the focus needs to change towards the participants. Hereby, it is impossible to achieve the same participants reactions, even using the same ones at two different points in time, be it even today and tomorrow. Instead, future replication should aim at diversifying the data they collect, hence also the participants, continuously added on previous finds until an extensive database can be established, which can predict the outcome of any given person with a fairly high success rate and ever-increasing tiny margin for error. The robot itself should be be reused until no new results can be gathered, at which points other non-humanoid and non-verbal communicative robots should be swapped for with. The scenarios themselves need to be refined and diversified as much as possible, to further enhance the quality and variety in data collected. Once almost every combination possible will have been tested, recorded and aggregated, one could really start to question the validity of the replication crisis and find solution for it, at least in HRI. Then indeed the study will be replicated trustworthy every single iteration, obtaining the most meaningful results for future research to be built upon it as some sort of a universal truth.
The focus on participant, scenario and robot aside, other derivative future work starting with this thesis can be in the direction of dancing robots, their benefits for the elderly, inactive or demented people, or the design of non-verbal interaction, communication and behaviour at large, as too few research has been carried out in this area, with most of it focusing mainly on humanoid robots.

List of Figures

List of Tables

Appendix

Informal Interview Questions for me to ask

1. **"When the robot touched me and moved away was for:"**
"Als der Roboter mich berührte und sich entfernte, war das für:"

   (*the expected answer was* "ask to follow it" / "bitten, ihm zu folgen").

2. **"When the robot flashed was for:"**
"Als der Roboter blinkte, war für:"

   (*the expected answer was* "ask for attention" / "um Aufmerksamkeit bitten").

3. **"Was the robot dancing? If so, what/why for? How do you interpret it:"**
"Hat der Roboter getanzt? Wenn ja, warum und wozu? Wie interpretieren Sie ihn:"

   (*no expected answer, just a check-up question due to the scenario's novelty*).

4. **"When the robot touched me while dancing was for:"**
"Als der Roboter mich beim Tanzen berührte, war das für:"

   (*the expected answer was* "to simulate dancing with me - invite me to dance" / "zu simulieren, dass er mit mir tanzt - mich zum Tanzen auffordert").

5. **"When the robot flashed was for:"**
"Als der Roboter blinkte, war für:"

   (*the expected answer was* "point out that it is dancing - reacting to music" / "darauf hinweisen, dass es sich um einen Tan - eine Reaktion auf Musik handelt").

Personal Background Information

In order for the experiment to be analyzed thoroughly and for the outcome to be interpreted correctly, the following information disclosure is required.
Yes/No and one-word answers are sufficient but any further details are of *great* significance to the study.

- Age:
- Gender:
- Occupation or Profession:
- Study or Expertise Background:
- Past Contact and Interaction with Robots:
  (*if any)*
- Knowledge of Sphero:
  (*if any)*
- Experience with Sphero:
  (*if any)*

The information provided will be anonymised, aggregated, safely stored and used with discretion, there will be no identification of or tracing back to your own persona. For more information on this please request the *Data Privacy Consent*.

Informationen zum persönlichen Hintergrund

Damit das Experiment gründlich analysiert und die Ergebnisse richtig interpretiert werden können, ist die Offenlegung der folgenden Informationen erforderlich. Ja/Nein-Antworten und Ein-Wort-Antworten sind ausreichend, aber alle weiteren Details sind für die Studie von großer Bedeutung.

- Alter:
- Geschlecht:
- Beschäftigung oder Beruf:
- Studium oder Fachwissen Hintergrund:
- Kontakt und Interaktion mit Robotern in der Vergangenheit:
  (*falls überhaupt)*
- Kenntnisse über Sphero:
  (*falls vorhanden)*
- Erfahrung mit Sphero:
  (*falls vorhanden)*

Die zur Verfügung gestellten Informationen werden anonymisiert, aggregiert, sicher gespeichert und mit Diskretion verwendet, es findet keine Identifizierung oder Rückverfolgung zu Ihrer eigenen Person statt. Weitere Informationen hierzu finden Sie auf Anfrage im *Data Privacy Consent*.

How do you perceive the robot?

Please rate your impression of the robot on the following scales:

Likeability

Dislike	*1*	*2*	*3*	*4*	*5*	**Like**
Unfriendly	*1*	*2*	*3*	*4*	*5*	**Friendly**
Unkind	*1*	*2*	*3*	*4*	*5*	**Kind**
Unpleasant	*1*	*2*	*3*	*4*	*5*	**Pleasant**
Awful	*1*	*2*	*3*	*4*	*5*	**Nice**

Perceived Intelligence

Incompetent	*1*	*2*	*3*	*4*	*5*	**Competent**
Ignorant	*1*	*2*	*3*	*4*	*5*	**Knowledgeable**
Irresponsible	*1*	*2*	*3*	*4*	*5*	**Responsible**
Unintelligent	*1*	*2*	*3*	*4*	*5*	**Intelligent**
Foolish	*1*	*2*	*3*	*4*	*5*	**Sensible**

Animacy

Dead	*1*	*2*	*3*	*4*	*5*	**Alive**
Stagnant	*1*	*2*	*3*	*4*	*5*	**Lively**
Mechanical	*1*	*2*	*3*	*4*	*5*	**Organic**
Artificial	*1*	*2*	*3*	*4*	*5*	**Lifelike**
Inert	*1*	*2*	*3*	*4*	*5*	**Interactive**
Apathetic	*1*	*2*	*3*	*4*	*5*	**Responsive**

Wie nehmen Sie den Roboter wahr?

Bitte bewerten Sie Ihren Eindruck vom Roboter auf folgenden Skalen:

Sympathie

Abneigung	*1*	*2*	*3*	*4*	*5*	**Gefallen**
Unfreundlich	*1*	*2*	*3*	*4*	*5*	**Freundlich**
Gemein	*1*	*2*	*3*	*4*	*5*	**Nett**
Unangenehm	*1*	*2*	*3*	*4*	*5*	**Angenehm**
Schrecklich	*1*	*2*	*3*	*4*	*5*	**Lieb**

Wahrgenommene Intelligenz

Inkompetent	*1*	*2*	*3*	*4*	*5*	**Kompetent**
Ignorant	*1*	*2*	*3*	*4*	*5*	**Kenntnisreich**
Verantwortungslos	*1*	*2*	*3*	*4*	*5*	**Verantwortungsvoll**
Ungeschickt	*1*	*2*	*3*	*4*	*5*	**Intelligent**
Töricht	*1*	*2*	*3*	*4*	*5*	**Vernünftig**

Lebendigkeit

Tot	*1*	*2*	*3*	*4*	*5*	**Lebendig**
Stillstehend	*1*	*2*	*3*	*4*	*5*	**Lebhaft**
Mechanisch	*1*	*2*	*3*	*4*	*5*	**Organisch**
Künstlich	*1*	*2*	*3*	*4*	*5*	**Lebensecht**
Unbeweglich	*1*	*2*	*3*	*4*	*5*	**Interaktiv**
Apathisch	*1*	*2*	*3*	*4*	*5*	**Reagierend**

How did the interaction with the robot feel?

Please rate your perception of and feelings towards the robot by crossing the respective feelings in their perceived intensity:

Feeling/Rating	**A little**	**Moderately**	**Quite a bit**	**Extremely**
Interested				
Distressed				
Excited				
Upset				
Strong				
Guilty				
Scared				
Hostile				
Enthusiastic				
Proud				
Irritable				
Alert				
Ashamed				
Inspired				
Nervous				
Determined				
Attentive				
Jittery				
Active				
Afraid				

Scenario D

Wie hat sich die Interaktion mit dem Roboter angefühlt?

Bitte bewerten Sie Ihre Wahrnehmung des Roboters und Ihre Gefühle demgegenüber, indem Sie die jeweiligen Gefühle in ihrer wahrgenommenen Intensität ankreuzen:

Gefühl?Intensität	**Ein bisschen**	**Einigermaßen**	**Erheblich**	**Äußerst**
Interessiert				
Bekümmert				
Freudig erregt				
Verärgert				
Stark				
Schuldig				
Erschrocken				
Feindselig				
Begeistert				
Stolz				
Gereizt				
Wach				
Beschämt				
Angeregt				
Nervös				
Entschlossen				
Aufmerksam				
Durcheinander				
Aktiv				
Ängstlich				

Scenario D

How did the interaction with the robot feel?

Please rate your perception of and feelings towards the robot by crossing the respective feelings in their perceived intensity:

Feeling/Rating	**A little**	**Moderately**	**Quite a bit**	**Extremely**
Interested				
Distressed				
Excited				
Upset				
Strong				
Guilty				
Scared				
Hostile				
Enthusiastic				
Proud				
Irritable				
Alert				
Ashamed				
Inspired				
Nervous				
Determined				
Attentive				
Jittery				
Active				
Afraid				

Scenario F

Wie hat sich die Interaktion mit dem Roboter angefühlt?

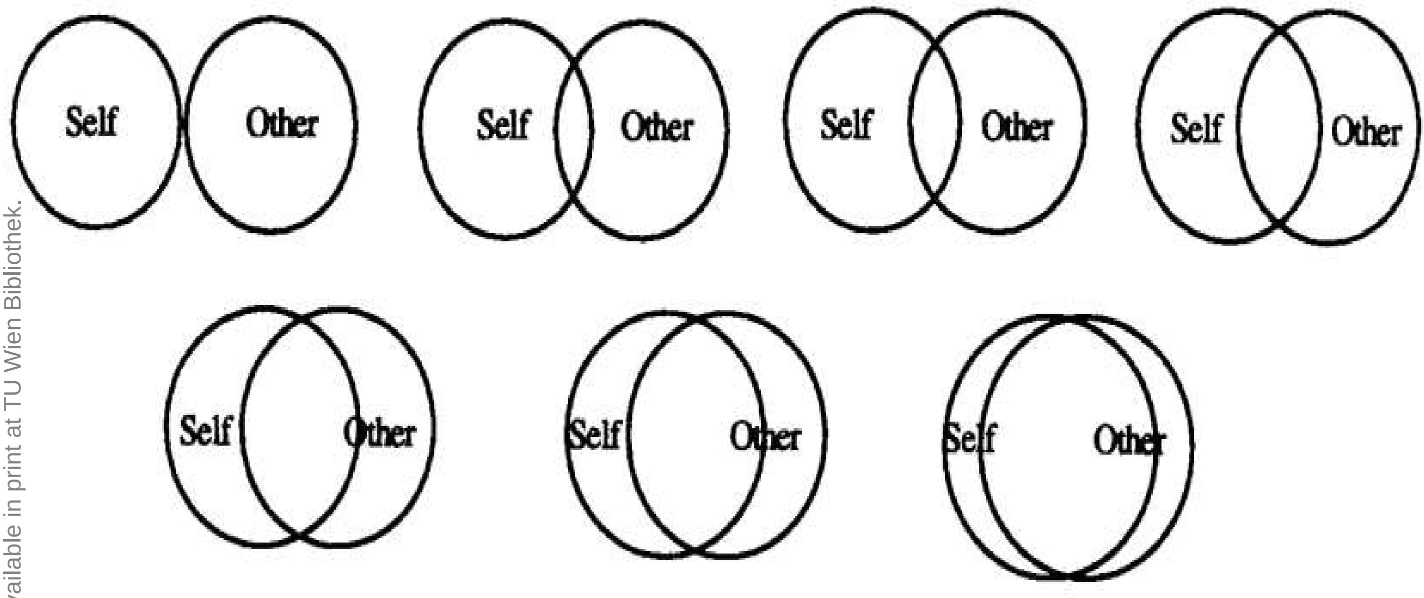
Bitte bewerten Sie Ihre Wahrnehmung des Roboters und Ihre Gefühle dem gegenüber, indem Sie die jeweiligen Gefühle in ihrer wahrgenommenen Intensität ankreuzen:

Gefühl?Intensitä	**Ein bisschen**	**Einigermaßen**	**Erheblich**	**Äußerst**
Interessiert				
Bekümmert				
Freudig erregt				
Verärgert				
Stark				
Schuldig				
Erschrocken				
Feindselig				
Begeistert				
Stolz				
Gereizt				
Wach				
Beschämt				
Angeregt				
Nervös				
Entschlossen				
Aufmerksam				
Durcheinander				
Aktiv				
Ängstlich				

Scenario F

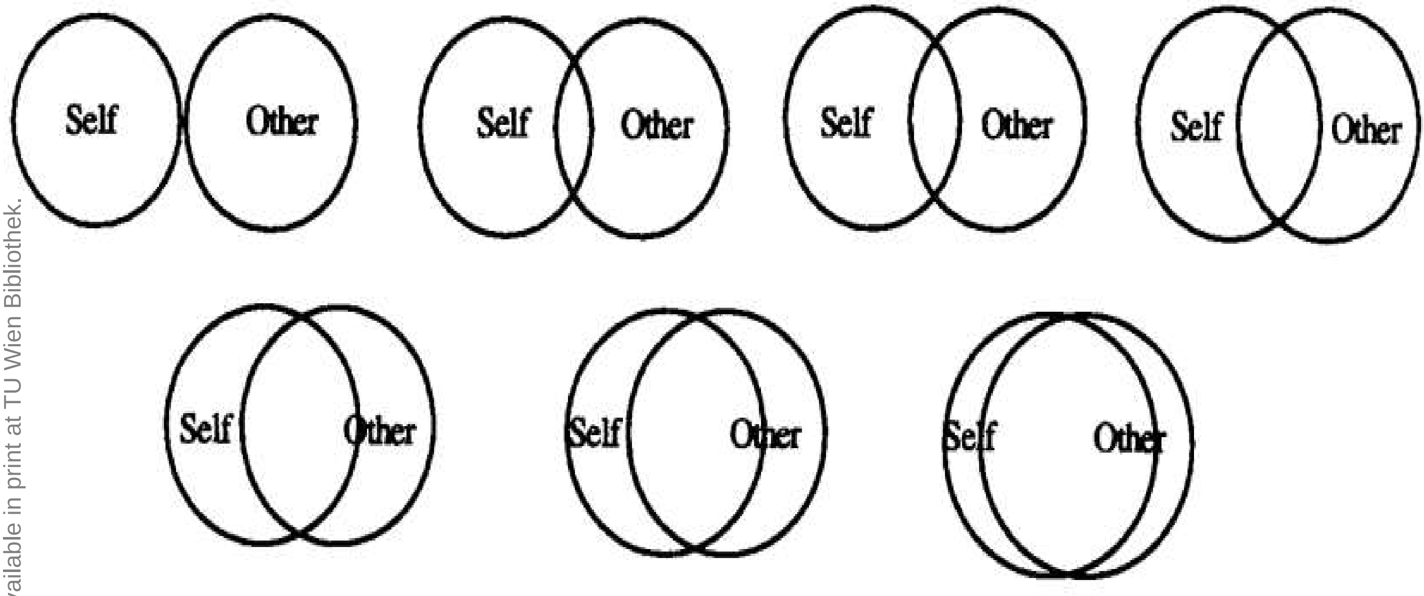
How close did you feel towards the robot?

Please circle the picture that best describes your relationship with and perception of the robot as a social unit:



Wie nahe haben Sie sich dem Roboter gefühlt?

Bitte kreuzen Sie das Bild an, das Ihre Beziehung mit dem und die Wahrnehmung dessen als sozialer Einheit am besten beschreibt:



TU Informatics | TU Informatics | TU Informatics

Human Computer Interaction Group

Research Study on the Interaction Between Humans and Robots

Background

The purpose of this study is to research, test and observe the interaction between humans and robots, and its unfolding. The information gathered, as well as its capturing process, will be used in the master thesis and accompanying seminar to help answer the research questions and write an academic paper. The research, supervised by Dr. Astrid Weiss, forms part of a Masters program at the Vienna University of Technology, namely "Human-Centered Computing".

The Study

The aim is to gain a better understanding of the extent of interaction between humans and robots. The information gathered will be used only to analyze the resulting transcriptions, observations and recordings, and summarize these findings and discussions in the final thesis. In this stage of the research study we would like you to agree to a live experiment meeting at the designated laboratory environment in the invitation, contained inside the technical university rooms. The session will take up to a maximum of 30 minutes in total. Your time and knowledge will be of utter importance for the undertaking and will yield in return gratefulness, as well as sweets, for your attendance.

Data Recording

You will be asked at the beginning of the experiment for an anonymous disclosure of your age, gender, profession, expertise and prior robot interaction. The experiments will be audio and video recorded with a camera. During the unfolding of the interaction, the researcher will be taking analog and digital notes and remarks using their laptop. The audio will be transcribed for further analysis of the findings. Together with the notes and interpretation of the video-recorded interaction, these will constitute the material of the final paper's practical work. Use of recorded material will be strictly limited in accordance with the consent given below.

The Research Group

The researchers who will be involved primarily with this study are the master student Alin Munteanu-Calen and his supervisor Dr. Astrid Weiss.

Privacy and Anonymity

Your name and other identifying sensible information will not be recorded nor reported in any documentation generated from this research study. All collected data will be held anonymous and stored safely in accordance with the Data Protection Act, and will not be used for any purpose for which it is not expressly permitted. If you would like to review the GDPR then request this at the beginning of the study.

Your Participation

Please note that you are free to participate in this study or not as you wish. You can withdraw your participation at any time without penalty. You can ask questions about the conduct of the experiment at any time. You can decline to answer any questions that you do not wish to answer. You can ask for any video, audio or notes to be deleted or not to be used in any way. You can request to review the material at any given time, as well as receive a copy of the published thesis.

By signing below, you have understood the information regarding this study and consent to taking part consent to video footage and audio samples of the experiment to be recorded for the purpose of the master thesis paper, and to recorded data, answers and interaction behaviors of me to be used, analyzed and presented in the academic master thesis, as well as presented in the seminar class and teaching department at the Vienna University of Technology.

Questions of Concerns

Dr. Astrid Weiss is the supervisor for this study and the master thesis work in general. If you have any questions, concerns or complaints about the study or the way it is being conducted, please feel free to contact her via email on astrid.weiss@tuwien.ac.at.

Furthermore, the student researcher will be happy to answer any questions you may have: Alin Munteanu-Calen, e1528098@student.tuwien.ac.at

I consent to taking part in this study **Yes / No**

I wish to receive a copy of this thesis after being finalized **Yes / No**
Email Address:_________________________

Participant's Full Name: ____

Signature:____Date:_________________________

Researcher's Full Name: Alin Munteanu-Calen

Signature:____Date:_________________________

Forschungsstudie über die Interaktion zwischen Menschen und Robotern

Hintergrund

Das Ziel dieser Studie ist es, die Interaktion zwischen Menschen und Robotern zu erforschen, zu testen und zu beobachten, wie sie sich entwickelt. Die gesammelten Informationen sowie der Erfassungsprozess werden in der Masterarbeit und dem begleitenden Seminar verwendet, um die Forschungsfragen zu beantworten und eine wissenschaftliche Arbeit zu verfassen. Die von Dr. Astrid Weiss betreute Forschungsarbeit ist Teil eines Masterstudiums an der Technischen Universität Wien, und zwar "Human-Centered Computing".

Die Studie

Ziel ist es, ein besseres Verständnis für das Ausmaß der Interaktion zwischen Mensch und Roboter zu gewinnen. Die gesammelten Informationen werden nur dazu verwendet, die daraus resultierenden Transkriptionen, Beobachtungen und Aufzeichnungen zu analysieren und diese Erkenntnisse und Diskussionen in der Abschlussarbeit zusammenzufassen. In dieser Phase der Forschungsstudie möchten wir Sie bitten, einer Live-Experiment-Sitzung in der dafür vorgesehenen Laborumgebung zuzustimmen, die sich in den Räumen der Technischen Universität befindet. Die Sitzung wird insgesamt maximal 30 Minuten dauern. Ihre Zeit und Ihr Wissen sind von größter Bedeutung für das Unternehmen und werden mit Dankbarkeit und Süßigkeiten für Ihre Teilnahme belohnt.

Datenaufzeichnung

Sie werden zu Beginn des Experiments um eine anonyme Auskunft über Ihr Alter, Ihr Geschlecht, Ihren Beruf, Ihr Fachwissen und Ihre bisherigen Erfahrungen mit Robotern gebeten. Die Experimente werden mit einer Audio- und Videokamera aufgezeichnet. Während der Interaktion macht der Forscher mit seinem Laptop analoge und digitale Notizen und Anmerkungen. Die Audioaufnahmen werden für die weitere Analyse der Ergebnisse transkribiert. Zusammen mit den Notizen und der Interpretation der auf Video aufgezeichneten Interaktion bilden diese das Material für die praktische Arbeit in der Abschlussarbeit. Die Verwendung des aufgezeichneten Materials wird gemäß der nachstehend erteilten Zustimmung streng begrenzt.

Die Forschungsgruppe

Die Forscher, die in erster Linie an dieser Studie beteiligt sein werden, sind der Masterstudent Alin Munteanu-Calen und seine Betreuerin Dr. Astrid Weiss.

Privatsphäre und Anonymität

Ihr Name und andere sensible Informationen, die zur Identifizierung dienen, werden weder aufgezeichnet noch in den aus dieser Studie hervorgehenden Unterlagen angegeben. Alle gesammelten Daten werden anonymisiert und in Übereinstimmung mit dem Datenschutzgesetz sicher aufbewahrt und nicht für Zwecke verwendet, für die sie nicht ausdrücklich zugelassen sind. Wenn Sie die GDPR einsehen möchten, können Sie dies zu Beginn der Studie beantragen.

Ihre Teilnahme

Bitte beachten Sie, dass es Ihnen freisteht, an dieser Studie teilzunehmen oder nicht. Sie können Ihre Teilnahme jederzeit widerrufen, ohne eine Strafe zahlen zu müssen. Sie können jederzeit Fragen über die Durchführung des Experiments stellen. Sie können die Beantwortung von Fragen verweigern, die Sie nicht beantworten möchten. Sie können verlangen, dass Video- und Tonaufnahmen oder Notizen gelöscht oder in keiner Weise verwendet werden. Sie können jederzeit Einsicht in das Material verlangen und eine Kopie der veröffentlichten Arbeit erhalten.

Mit meiner Unterschrift bestätige ich, dass ich die Informationen zu dieser Studie verstanden habe und mit der Teilnahme einverstanden bin, dass Video- und Audioaufnahmen des Experiments für die Masterarbeit aufgezeichnet werden und dass die aufgezeichneten Daten, Antworten und das Interaktionsverhalten von mir in der wissenschaftlichen Masterarbeit verwendet, analysiert und präsentiert werden, sowie in der Seminarklasse und in der Lehre an der Technischen Universität Wien vorgestellt werden.

Fragen und Feedback

Dr. Astrid Weiss ist die Betreuerin dieser Studie und der Masterarbeit im Allgemeinen. Wenn Sie Fragen, Bedenken oder Beschwerden über die Studie oder die Art und Weise ihrer Durchführung haben, können Sie sie gerne per E-Mail kontaktieren unter astrid.weiss@tuwien.ac.at.

Darüber hinaus beantwortet der studentische Forscher gerne alle Fragen, die Sie haben:
Alin Munteanu-Calen, e1528098@student.tuwien.ac.at

Ich bin damit einverstanden, an dieser Studie teilzunehmen **Ja / Nein**

Ich möchte ein Exemplar dieser Arbeit nach ihrer Fertigstellung erhalten **Ja / Nein**
E-Mail Adresse:________________________

Vollständiger Name des Teilnehmers: ____

Unterschrift:____Datum:_______________________

Vollständiger Name des Forschers: Alin Munteanu-Calen

Unterschrift:____Datum:_______________________

Privacy Policy for the Research Study on the Interaction Between Robots and Humans

Data protection and its safeguarding are important concerns of the Vienna University of Technology. The processing of personal data takes place in strict compliance with the principles and requirements set out in the GDPR1 and the Austrian GDPR2. The TU Vienna only processes the data that is necessary to achieve the intended purposes and always endeavors to ensure the security and accuracy of the data.

Project Overview

The purpose of this research study is to test how people and robots interact together given different cases, and is essential for carrying out the practical work in the frame of the master thesis. The thesis, under the supervision of Dr. Astrid Weiss, is required for the completion of a Masters program at the Vienna University of Technology, in the area of "Human-Centered Computing".

Data collection and processing

The research will use a qualitative observation, interview and questionnaire method to collect appropriate meaningful data. The declaration of consent is being made on a voluntary basis. Furthermore, the participant has the right to withdraw their consent at any time and stop the experiment without any consequences.

What data will be collected during the study?

The study will consist of a practical interaction experiment, which will be held in a test lab environment at the university and will be documented with audio and video recordings. The participant is not informed when the researcher starts recording, as this should in any way not influence their behavior.

The video and audio recording is partially transcribed and stored in a digital video and text format. In the course of this process, the participant will be anonymized, as identification by name is of no relevance. Instead, the participant will be assigned a code (E#). The collected information regarding job area, background, age, past experiences with robots and sex are only relevant in the aggregated analysis of the data. This prevents the easy identification of specific persons and the assignment of personal details in the statements to them. In case of other publications, no information will be disclosed, unless explicitly agreed upon with the participant.

The questionnaire and interview handed and carried out during the experiment will be stored digitally in text format, the answers being anonymously collected and interpreted to the benefit of gaining knowledge, being labeled only with the identification code. Other observations recorded in written format during the experiment will follow the same principle.

How will the data be used?

All written records and the recording collected during the experiment will be used primarily for this qualitative study, hence for the master thesis outcome and its analysis. In addition, the right to use the anonymized data for further research as appropriate is reserved. The data will be analyzed exclusively by Alin Munteanu-Calen, and reviewed by the master thesis group students, the students taking the seminar course on the master thesis, Dr. Astrid Weiss and Prof. Hilda Tellioglu.

All raw data material is treated and handled confidentially by the researcher. This means that all data will be stored in a password-protected environment on the TU Vienna's own network service (installed on in-house server) and backup copies of this material will be made at regular intervals. A unified naming system for the files will be used, namely <E#Code_Experiment_Date_VersionNumber.FileFormat> for experiment video recordings with participants and <E#_Questions_Date_VersionNumber.FileFormat> for their questionnaire and interview answers, and observations taken during the experiment.

Access to the raw data is limited to the researcher and their supervisor. The participant may view material pertaining to them upon request. The data carriers are encrypted according to the current state of the art. After the end of the project, the recordings will be archived in encrypted form at the TU Vienna.

Responsible:

Rectorate of Vienna University of Technology, Karlsplatz 13, 1040 Vienna

Data Protection Officer:

Mag. Christina Thirsfeld

Vienna University of Technology, Karlsplatz 13/018, 1040 Vienna

datenschutz@tuwien.ac.at

The following categories of data are processed during this data processing operation:

The following data will be captured and analyzed:

- Audio and video recording of the individual experiment.
- Handwritten and digital notes of the individual questionnaires, interview and observations.

Purpose of data processing

Data processing has the purpose of scientific evaluation and analysis.

Transmission

All evaluations and presentations of results that are published or passed on to third parties are carried out in anonymized and aggregated (summarized) form if explicitly agreed.

Legal basis for data processing

The processing and use of the participant's personal data is based on Art 6 para 1 e) DSGVO and Art 89 DSGVO iVm §3 UG and is limited to the above purposes. The processing of personal data is based on the principles and requirements set forth in the General Data Protection Regulation (GDPR), the Austrian Data Protection Act (DSG) and the Research Organization Act FOG.

Storage period / deletion period

The data will be stored as long as required by the legal retention periods or as required by the purpose.

Remedies

In connection with the processing of the participant's personal data, you have the rights of access, rectification, deletion, restriction of processing and objection, provided that the exercise of these rights is not likely to render impossible or seriously impair the achievement of the research purpose (Section 2d (6) FOG).

For this purpose, please contact: Dr. Astrid Weiss, astrid.weiss@tuwien.ac.at, or Prof. Hilda Tellioglu, hilda.tellioglu@tuwien.ac.at.

If the participant believes that the processing of their data violates data protection law or that their data protection rights have otherwise been violated in some way, the participant can complain to the competent supervisory authority: Austrian Data Protection Authority (DPA), Barichgasse 40- 42, 1030 Vienna, Austria.

Contact Information

In case of any further questions or concerns regarding the processing of the data, please contact:

Leading contact person at TU Vienna
Dr. Astrid Weiss
TU Vienna, Research area Human Computer Interaction
Argentinierstrasse 8, 2. Stock E193/5, 1040 Vienna
Tel: +43 1 58801 18735 , e-mail: astrid.weiss@tuwien.ac.at

General information on data protection can be found at the Austrian data protection authority at: https://www.dsb.gv.at/.

Bibliography

[1] Sung-En Chien et al. "Age Difference in Perceived Ease of Use, Curiosity, and Implicit Negative Attitude toward Robots". In: *ACM Transactions on Human-Robot Interaction* 8.2 (June 2019), 9:1–9:19. DOI: 10.1145/3311788. URL: https://doi.org/10.1145/3311788 (visited on 09/14/2022).

[2] Elizabeth Broadbent. "Interactions With Robots: The Truths We Reveal About Ourselves". In: *Annual Review of Psychology* 68, 2017 (Jan. 2017), pp. 627–652. URL: https://doi.org/10.1146/annurev-psych-010416-043958.

[3] R. Fawcett Nicholas and Eleonora Gullone. "Cute and Cuddly and a Whole Lot More? A Call for Empirical Investigation into the Therapeutic Benefits of Human–Animal Interaction for Children". en. In: *Behaviour Change* 18.2 (June 2001), pp. 124–133. ISSN: 2049-7768, 0813-4839. DOI: 10.1375/bech.18.2.124. URL: https://www.cambridge.org/core/journals/behaviour-change/article/abs/cute-and-cuddly-and-a-whole-lot-more-a-call-for-empirical-investigation-into-the-therapeutic-benefits-of-humananimal-interaction-for-children/798747FD2527E6D6E13C5B4DD4CEF311 (visited on 09/14/2022).

[4] Benedikt Leichtmann and Verena Nitsch. "Is the Social Desirability Effect in Human–Robot Interaction overestimated? A Conceptual Replication Study Indicates Less Robust Effects". en. In: *International Journal of Social Robotics* 13.5 (Aug. 2021), pp. 1013–1031. ISSN: 1875-4805. DOI: 10.1007/s12369-020-00688-z. URL: https://doi.org/10.1007/s12369-020-00688-z (visited on 09/14/2022).

[5] W.G. Volante et al. "Specifying Influences that Mediate Trust in Human-Robot Interaction". In: *HFES* 60.1 (Sept. 2016). URL: https://journals.sagepub.com/doi/10.1177/1541931213601402.

[6] Cindy L. Bethel. *Robots without faces: non-verbal social human-robot interaction - ProQuest*. en. URL: https://www.proquest.com/openview/b634c85ad4f80754a100c8 1?pq-origsite=gscholar&cbl=18750 (visited on 09/14/2022).

[7] Gabriella Lakatos et al. "Emotion Attribution to a Non-Humanoid Robot in Different Social Situations". en. In: *PLOS ONE* 9.12 (Dec. 2014). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0114207. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114207 (visited on 09/14/2022).

[8] Keunwook Kim. "Post-plant3: The Third Type of a Series of Non-humanoid Robots with an Embedded Physical Interaction: The development of non-verbal human-robot interaction framework and interactive robot prototypes". In: Oct. 2021, pp. 140–141. ISBN: 9781450386555. URL: https://kimkeunwook.com/Post-plant3 (visited on 09/15/2022).

[9] Shane Saunderson and Goldie Nejat. "How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction". en. In: *International Journal of Social Robotics* 11.4 (Aug. 2019), pp. 575–608. ISSN: 1875-4805. DOI: 10.1007/s12369-019-00523-0. URL: https://doi.org/10.1007/s12369-019-00523-0 (visited on 09/14/2022).

[10] K. Kosuge and Y. Hirata. "Human-Robot Interaction". In: *2004 IEEE International Conference on Robotics and Biomimetics*. Aug. 2004, pp. 8–11. DOI: 10.1109/ROBIO.2004.1521743.

[11] Elizabeth Cha, Maja Matarić, and Terrence Fong. "Nonverbal signaling for non-humanoid robots during human-robot collaboration". In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ISSN: 2167-2148. Mar. 2016, pp. 601–602. DOI: 10.1109/HRI.2016.7451876.

[12] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. "Legibility and predictability of robot motion". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ISSN: 2167-2148. Mar. 2013, pp. 301–308. DOI: 10.1109/HRI.2013.6483603.

[13] Guy Hoffman. "Anticipation in Human-Robot Interaction". en. In: *2010 AAAI Spring Symposium Series*. Mar. 2010. URL: https://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1192 (visited on 09/14/2022).

[14] Aaron G. Cass, Kristina Striegnitz, and Nick Webb. "A farewell to arms: Non-verbal communication for non-humanoid robots". In: *Proceedings of the Workshop on NLG for Human–Robot Interaction*. Tilburg, The Netherlands: Association for Computational Linguistics, Nov. 2018, pp. 22–26. DOI: 10.18653/v1/W18-6905. URL: https://aclanthology.org/W18-6905 (visited on 09/14/2022).

[15] Emily C. Collins, Tony J. Prescott, and Ben Mitchinson. "Saying It with Light: A Pilot Study of Affective Communication Using the MIRO Robot". en. In: *Biomimetic and Biohybrid Systems*. Ed. by Stuart P. Wilson et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 243–255. ISBN: 9783319229799. DOI: 10.1007/978-3-319-22979-9_25.

[16] Y. Kakazu, M. Wada, and T. Sato. *Intelligent Autonomous Systems: IAS-5*. en. Google-Books-ID: v0MzaDRFnzUC. IOS Press, 1998, pp. 352–359. ISBN: 9789051993981.

[17] Pauline L Hall and Zoey Malpus. "Pets as therapy: effects on social interaction in long-stay psychiatry". In: *British Journal of Nursing* 9.21 (Nov. 2000), pp. 2220–2225. ISSN: 0966-0461. DOI: 10.12968/bjon.2000.9.21.5425. URL: https://www.magonlinelibrary.com/doi/abs/10.12968/bjon.2000.9.21.5425 (visited on 09/14/2022).

[18] Aryel Beck et al. "Interpretation of emotional body language displayed by robots". In: *Proceedings of the 3rd international workshop on Affective interaction in natural environments*. AFFINE '10. New York, NY, USA: Association for Computing Machinery, Oct. 2010, pp. 37–42. ISBN: 9781450301701. DOI: 10.1145/1877826.1877837. URL: https://doi.org/10.1145/1877826.1877837 (visited on 09/14/2022).

[19] Jekaterina Novikova and Leon Watts. "A design model of emotional body expressions in non-humanoid robots". In: *Proceedings of the second international conference on Human-agent interaction*. HAI '14. New York, NY, USA: Association for Computing Machinery, Oct. 2014, pp. 353–360. ISBN: 9781450330350. DOI: 10.1145/2658861.2658892. URL: https://doi.org/10.1145/2658861.2658892 (visited on 09/14/2022).

[20] Astrid M. Rosenthal-von der Pütten, Nicole C. Krämer, and Jonathan Herrmann. "The Effects of Humanlike and Robot-Specific Affective Nonverbal Behavior on Perception, Emotion, and Behavior". en. In: *International Journal of Social Robotics* 10.5 (Nov. 2018), pp. 569–582. ISSN: 1875-4805. DOI: 10.1007/s12369-018-0466-7. URL: https://doi.org/10.1007/s12369-018-0466-7 (visited on 09/14/2022).

[21] Tony Belpaeme. "Advice to New Human-Robot Interaction Researchers". en. In: *Human-Robot Interaction: Evaluation Methods and Their Standardization*. Ed. by Céline Jost et al. Springer Series on Bio- and Neurosystems. Cham: Springer International Publishing, 2020, pp. 355–369. ISBN: 9783030423070. DOI: 10.1007/978-3-030-42307-0_14. URL: https://doi.org/10.1007/978-3-030-42307-0_14 (visited on 09/14/2022).

[22] Astrid Weiss. *One Cannot Know Everything - On The Need Of Epistemiological Diversity In Human-Centered HRI Research*. en. 2021.

[23] Benedikt Leichtmann and Verena Nitsch. "How much distance do humans keep toward robots? Literature review, meta-analysis, and theoretical considerations on personal space in human-robot interaction". en. In: *Journal of Environmental Psychology* 68 (Apr. 2020), p. 101386. ISSN: 0272-4944. DOI: 10.1016/j.jenvp.2019.101386. URL: https://www.sciencedirect.com/science/article/pii/S0272494419303846 (visited on 09/14/2022).

[24] Bahar Irfan et al. "Social Psychology and Human-Robot Interaction: An Uneasy Marriage". In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '18. New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 13–20. ISBN: 9781450356152. DOI: 10.1145/3173386.3173389. URL: https://doi.org/10.1145/3173386.3173389 (visited on 09/14/2022).

[25] Daniel Ullman, Salomi Aladia, and Bertram F. Malle. "Challenges and Opportunities for Replication Science in HRI: A Case Study in Human-Robot Trust". In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '21. New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 110–118. ISBN: 9781450382892. DOI: 10.1145/3434073.3444652. URL: https://doi.org/10.1145/3434073.3444652 (visited on 09/14/2022).

[26] Miguel Faria et al. "Follow me: Communicating Intentions with a Spherical Robot". en. In: (Aug. 2016). ISSN: 1944-9437. DOI: 10.1109/ROMAN.2016.7745189. URL: https://ieeexplore.ieee.org/document/7745189 (visited on 09/22/2022).

[27] A.V. Libin and E.V. Libin. "Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach". In: *Proceedings of the IEEE* 92.11 (Nov. 2004), pp. 1789–1803. ISSN: 1558-2256. DOI: 10.1109/JPROC.2004.835366.

[28] Luis Yoichi Morales Saiki et al. "How do people walk side-by-side? using a computational model of human behavior for a social robot". In: *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. HRI '12. New York, NY, USA: Association for Computing Machinery, Mar. 2012, pp. 301–308. ISBN: 9781450310635. DOI: 10.1145/2157689.2157799. URL: https://doi.org/10.1145/2157689.2157799 (visited on 09/14/2022).

[29] Cynthia Breazeal et al. "Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork". In: MIT Media Lab, USA: MIT Media Lab, Sept. 2005. URL: http://alumni.media.mit.edu/~guy/publications/BreazealIROS05.pdf (visited on 09/15/2022).

[30] Oren Zuckerman et al. "Companionship Is Not a Function: The Effect of a Novel Robotic Object on Healthy Older Adults' Feelings of "Being-Seen"". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–14. ISBN: 9781450367080. DOI: 10.1145/3313831.3376411. URL: https://doi.org/10.1145/3313831.3376411 (visited on 09/14/2022).

[31] António Santos et al. "Multi-sensor exercise-based interactive games for fall prevention and rehabilitation". In: *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. ISSN: 2153-1641. May 2015, pp. 65–71. DOI: 10.4108/icst.pervasivehealth.2015.259115.

[32] Laura Boccanfuso et al. "Emotional robot to examine different play patterns and affective responses of children with and without ASD". In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ISSN: 2167-2148. Mar. 2016, pp. 19–26. DOI: 10.1109/HRI.2016.7451729.

[33] David V. Lu and William D. Smart. "Towards more efficient navigation for robots and humans". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. ISSN: 2153-0866. Nov. 2013, pp. 1707–1713. DOI: 10.1109/IROS.2013.6696579.

[34] Satoru Satake et al. "How to approach humans? strategies for social robots to initiate interaction". In: *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. HRI '09. New York, NY, USA: Association for Computing Machinery, Mar. 2009, pp. 109–116. ISBN: 9781605584041. DOI: 10.1145/1514095.1514117. URL: https://doi.org/10.1145/1514095.1514117 (visited on 09/14/2022).

[35] Eleanor Avrunin and Reid Simmons. "Socially-appropriate approach paths using human data". In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. ISSN: 1944-9437. Aug. 2014, pp. 1037–1042. DOI: 10.1109/ROMAN.2014.6926389.

[36] Theresa Law, Josh de Leeuw, and John H. Long. "How Movements of a Non-Humanoid Robot Affect Emotional Perceptions and Trust". en. In: *International Journal of Social Robotics* 13.8 (Dec. 2021), pp. 1967–1978. ISSN: 1875-4805. DOI: 10.1007/s12369-020-00711-3. URL: https://doi.org/10.1007/s12369-020-00711-3 (visited on 09/14/2022).

[37] Christoph Bartneck, Kulic Dana, and Elizabeth A. Croft. "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". en. In: (Jan. 2008). URL: https://www.researchgate.net/publication/200508162_Measuring_the_anthropomorphism_animacy_likeability_perceived_intelligence_and_perceived_safety_of_robots (visited on 10/03/2022).

[38] Verena Wagner-Hartl et al. "Who Would Let a Robot Take Care of Them? - Gender and Age Differences". en. In: *HCI International 2020 - Posters*. Ed. by Constantine Stephanidis and Margherita Antona. Communications in Computer and Information Science. Cham: Springer International Publishing, 2020, pp. 196–202. ISBN: 9783030507268. DOI: 10.1007/978-3-030-50726-8_26.

[39] Kerstin Dautenhahn. "Socially intelligent robots: dimensions of human–robot interaction". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1480 (Apr. 2007), pp. 679–704. DOI: 10.1098/rstb.2006.2004. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2006.2004 (visited on 09/14/2022).

[40] Shikhar Kumar et al. "Politeness in Human–Robot Interaction: A Multi-Experiment Study with Non-Humanoid Robots". en. In: *International Journal of Social Robotics* (Aug. 2022). ISSN: 1875-4805. DOI: 10.1007/s12369-022-00911-z. URL: https://doi.org/10.1007/s12369-022-00911-z (visited on 09/14/2022).

[41] Viva Sarah Press and Hadas Erel. "Designing Non-Verbal Humorous Gestures for a Non-Humanoid Robot". In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI EA '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–7. ISBN: 9781450391566. DOI: 10.1145/3491101.3519924. URL: https://doi.org/10.1145/3491101.3519924 (visited on 09/14/2022).

[42] Lizzy Sinnema and Maryam Alimardani. "The Attitude of Elderly and Young Adults Towards a Humanoid Robot as a Facilitator for Social Interaction". In: *Social Robotics: 11th International Conference, ICSR 2019, Madrid, Spain, November 26–29, 2019, Proceedings*. Berlin, Heidelberg: Springer-Verlag, Nov. 2019, pp. 24–33. ISBN: 9783030358877. DOI: 10.1007/978-3-030-35888-4_3. URL: https://doi.org/10.1007/978-3-030-35888-4_3 (visited on 09/14/2022).

[43] B. Robins et al. "Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?" en. In: *Universal Access in the Information Society* 4.2 (Dec. 2005), pp. 105–120. ISSN: 1615-5297. DOI: 10.1007/s10209-005-0116-3. URL: https://doi.org/10.1007/s10209-005-0116-3 (visited on 09/14/2022).

[44] Jean-Loup Rault. "Pets in the Digital Age: Live, Robot, or Virtual?" In: *Frontiers in Veterinary Science* 2 (2015). ISSN: 2297-1769. URL: https://www.frontiersin.org/articles/10.3389/fvets.2015.00011 (visited on 09/14/2022).

[45] Julia Riedel et al. "The early ontogeny of human–dog communication". en. In: *Animal Behaviour* 75.3 (Mar. 2008), pp. 1003–1014. ISSN: 0003-3472. DOI: 10.1016/j.anbehav.2007.08.010. URL: https://www.sciencedirect.com/science/article/pii/S0003347207004861 (visited on 09/14/2022).

[46] Melissa Kont and Maryam Alimardani. "Engagement and Mind Perception Within Human-Robot Interaction: A Comparison Between Elderly and Young Adults". In: *Social Robotics: 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings*. Berlin, Heidelberg: Springer-Verlag, Nov. 2020, pp. 344–356. ISBN: 9783030620554. DOI: 10.1007/978-3-030-62056-1_29. URL: https://doi.org/10.1007/978-3-030-62056-1_29 (visited on 09/14/2022).

[47] Maha Salem et al. "A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction". In: *2011 RO-MAN*. ISSN: 1944-9437. July 2011, pp. 247–252. DOI: 10.1109/ROMAN.2011.6005285.

[48] Assem Kroma and Ali Mazalek. "Interfacing & Embodiment: "Baby Tango" Dancing Robot Attempts to Communicate". In: June 2021. URL: https://www.assemkroma.com/copy-of-samples-of-my-filmmaking-st-2?pgid=kj80s9ow-7b287577-3ef5-4987-9c18-069c03c114fd (visited on 09/15/2022).

[49] F. Tanaka and H. Suzuki. "Dance interaction with QRIO: a case study for non-boring interaction by using an entrainment ensemble model". In: *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*. Sept. 2004, pp. 419–424. DOI: 10.1109/ROMAN.2004.1374797.

[50] Marek P. Michalowski, Selma Sabanovic, and Hideki Kozima. "A dancing robot for rhythmic social interaction". In: *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. HRI '07. New York, NY, USA: Association for Computing Machinery, Mar. 2007, pp. 89–96. ISBN: 9781595936172. DOI: 10.1145/1228716.1228729. URL: https://doi.org/10.1145/1228716.1228729 (visited on 09/14/2022).

[51] *Programmable Robot Ball, Sphero SPRK+*. en. URL: https://sphero.com/products/sphero-sprk-plus.

[52] Maria Luce Lupetti. "Designing playful HRI: Acceptability of robots in everyday life through play". In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ISSN: 2167-2148. Mar. 2016, pp. 631–632. DOI: 10.1109/HRI.2016.7451891.

[53] Tom Sorell and Heather Draper. "Robot carers, ethics, and older people". en. In: *Ethics and Information Technology* 16.3 (Sept. 2014), pp. 183–195. ISSN: 1572-8439. DOI: 10.1007/s10676-014-9344-7. URL: https://doi.org/10.1007/s10676-014-9344-7 (visited on 09/14/2022).

[54] Tiffany L. Chen et al. "Older adults' acceptance of a robot for partner dance-based exercise". en. In: *PLOS ONE* 12.10 (Oct. 2017), e0182736. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0182736. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0182736 (visited on 09/14/2022).

[55] Wei Quan, Zhang haibin, and Naoyuki Kubota. "Elderly Health Promotion using Multiple Ball-robots based on Evolutionary Robot". In: *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)*. June 2020, pp. 000481–000486. DOI: 10.1109/SoSE50414.2020.9130424.

[56] Hua Peng et al. "Robotic Dance in Social Robotics—A Taxonomy". In: *IEEE Transactions on Human-Machine Systems* 45.3 (June 2015), pp. 281–293. ISSN: 2168-2305. DOI: 10.1109/THMS.2015.2393558.

[57] Akshay Potnuru, Mohsen Jafarzadeh, and Yonas Tadesse. "3D printed dancing humanoid robot "Buddy" for homecare". In: *2016 IEEE International Conference on Automation Science and Engineering (CASE)*. ISSN: 2161-8089. Aug. 2016, pp. 733–738. DOI: 10.1109/COASE.2016.7743475.

[58] Kazue Sawami. "Cognitive ability and psychological effectiveness of-brain training dance robot therapy for elderly people". en. In: URL: https://www.researchgate.net/publication/333906295_Cognitive_ability_and_psychological_effectiveness_of_brain_training_dance_robot_therapy_for_elderly_people (visited on 09/14/2022).

[59] Arthur Aron, Elaine.N Aron, and Danny Smollan. "Inclusion of Other in the Self Scale and the structure of interpersonal closeness". en. In: *Journal of Personality and Social Psychology* 63 (1992), p. 596. ISSN: 4. DOI: 10.1037/0022-3514.63.4.596. URL: https://ur.booksc.me/book/30414832/fbb407 (visited on 10/03/2022).

[60] *Star Wars Robots, Sphero BB8*. en. URL: https://sphero.com/pages/legacy-products.

[61] *Defining generations: Where Millennials end and Generation Z begins*. en. URL: https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/.

[62] *Sphero Edu App*. en. URL: https://edu.sphero.com.

[63] *Particle House - Moon and Tide*. en. URL: https://www.youtube.com/watch?v=P6jwzq2L3_o.

[64] *PANAS German Version*. de. URL: https://www.researchgate.net/publication/309210530_Deutsche_Version_der_Positive_and_Negative_Affect_Schedule_PANAS_GESIS_Panel (visited on 09/14/2022).

[65] *Sphero Ollie Robot Discontinued*. en. URL: https://www.theguardian.com/technology/2014/dec/04/sphero-ollie-review-the-remote-control-car-reimagined.