Ethical Implications of Artificial Expression of Emotion by Social Robots in Assistive Contexts

Anouk van Maris

A thesis submitted in partial fulfilment of the requirements of the University of the West of England, Bristol, for the degree of Doctor of Philosophy

Faculty of Engineering and Technology

University of the West of England, Bristol

February 2021

Acknowledgements

These acknowledgements feel as important as the dissertation itself, as without these people I would never have been able to conduct this research and present it in this work. First of all, a huge thanks to all my participants for your willingness to take part in my experiments, without all of you this work would not have existed.

The next people to thank are my supervisors. Your guidance has helped me become the strong and independent researcher I am today. Praminda Caleb-Solly, Nancy Zook, Sanja Dogramadzi, Matt Studley and Alan Winfield, thank you all so much. Special thanks to Praminda and Nancy, who have gone above and beyond by not only guiding me and giving advice at a professional level, but also provided solid support through my personal growth.

A huge thank you goes to my examiners. Shannon Vallor, Maartje de Graaf and Manuel Giuliani, thank you for giving me the opportunity to defend and improve the quality of this work through your constructive criticisms and thoughtful questions.

Next, the SOCRATES team has to be mentioned. Our bi-annual meetings were always useful as I received feedback from people who were less involved in the research. Without this project I never would have had the opportunity to acquire my PhD in such an exciting, multidisciplinary and international setting. This has been an experience I will cherish for life. Special appreciation goes to Antonio, who encouraged me when things got difficult and was the biggest supporter of my successes.

Besides all my international colleagues, I also made some great friends in the lab who made my PhD experience and moving to a foreign country fun, memorable and so much easier than expected. Tom, Ben, Greg, Alex and Gabi (Albi), thank you.

There is also the ECHOS research group that deserves to be mentioned! You always provided me with a safe place to express doubts and concerns. Also, your input during brainstorm sessions was extremely valuable due to the diversity of knowledge and experience within the group.

I also want to thank Gordon Darling, the security guard of the Bristol Robotics Laboratory. You put a smile on my face every single day, and were extremely helpful and patient when I was running experiments and needed clearance for my participants to enter and leave the lab.

The next appreciation goes to some of the most important people in my life. Mum and Dad, thank you for always believing in me and supporting me in every decision that I ever made in life, even if you were not sure it was the right decision for me. Also, spoiling me to death every time I come home to visit is definitely appreciated. Mich, 'sis', you are the big sister that people always read about in stories. Never further than a phone call away, tackling my problems before I realise I have them, and showing me the bigger picture and that life is good. The fact that your wardrobe always ends up as my wardrobe is also treasured. Dion, 'Meus', I cannot say anything else but thank you for always being there for my sister, my family and also for me. You truly are the brother I never had.

And finally, I definitely saved the best for last. Luc, where would I ever be without you? You are my partner in so many ways. From loving me and making me feel safe, to providing dinner and patiently listening to my rants, brainstorming with me about my research and telling me to not be a drama queen, you do it all. Thank you does not cover everything you do, but thank you anyway. You are my rock.

Abstract

This research investigated whether Artificial Expression of Emotion (AEE) by a social robot can lead to emotional deception and emotional attachment. The use of AEE can be beneficial, as it may encourage engagement, and can help to build trust in the robot. However, it may also lead to misplaced trust and false expectations of the robot's abilities, which in turn could lead to mental or physical harm. Even though the literature has raised ethical issues in the form of emotional deception and emotional attachment, research regarding potential negative consequences of these concerns is limited. As such, the impact of AEE was considered in this research. Knowledge on potential negative consequences is essential, as social robots are likely to become increasing prevalent in supporting assistive tasks for people who are vulnerable.

The impact of AEE was investigated through surveys, lab-based experiments and longitudinal field studies. Participants' opinion of the robot, acceptance of the robot and attachment to the robot, and physiological responses to the robot were investigated. Findings indicate that emotional deception and emotional attachment may have occurred, although their impact on users was low. These findings contributed to the development of a framework, that could help developers and producers of future social robots design social robot behaviours, with a view to limit negative consequences where possible.

This work contributes to the area of social robot ethics, highlighting issues for socially assistive robots based on findings from a range of user studies. Furthermore, a novel approach for conducting research in determining people's attitudes and perspectives on ethical issues is presented, which allows people to make more informed decisions while completing surveys. In addition, this research provides rich insights in user experience of socially assistive robots, which contributes to our understanding of the impact that the use of AEE may have on society.

Contents

1 Introduction

Welcome to the era of social robotics research. There are many applications where social robots can be beneficial, such as tutoring, customer engagement and companionship. Current successes of such robots (Chen, Park, and Breazeal, 2020; Murphy, Gretzel, and Pesonen, 2019; Barrett et al., 2019) make it likely that they will become an aspect of life for a large population. Social robots are ultimately designed to provide social support. This support can be either in the form of physical assistance, social assistance or both. The goal of these robots is to improve people's quality of life (Clabaugh and Matarić, 2018). Existing literature indicates that the use of such robots can be beneficial, for example as a therapeutic device for children with autism spectrum disorder (Wood et al., 2019), or as a device that can improve the quality of life of people with dementia (Kang et al., 2019). Several definitions have been presented in the literature, which will be considered in more depth in the literature review (Chapter 2). In this research, a social robot is regarded as *a robot for which interactions are of high importance to achieve its fundamental and intended functionality, that is ultimately designed to meet the needs of the people it interacts with.*

1.1 Motivation

One trait of a social robot is that they may have the ability to express artificial emotions. This can be an asset to the successful human-robot interactions and increased use of the robot, which means users can fully benefit from the services that the robot provides (Wilson, 2017). The ability to express emotion is a key characteristics for social robot acceptance (De Graaf, Allouch, and Van Dijk, 2015). However, knowledge regarding the impact of social robots with the ability to display realistic emotions through a combination of different modalities, from now on defined as *artificial expression of emotion* (AEE), is limited and needs to be investigated in more detail to foster successful human-robot interactions.

1.1.1 Social Robots and Expectations

The novelty of social robots compared to other robotic technologies is the combination of social skills and an anthropomorphic physical embodiment, and it is argued they are a new technological genre due to the emotional component involved in interactions with social robots (De Graaf, 2016). The possibilities that arise with the emergence of social robots, and the way they are portrayed in popular media, result in user expectations that may not match with the actual abilities of these robots (Beer et al., 2011). After all, even though a social robot can be capable of artificial expression of emotion, this does not mean it knows what emotions are, let alone understand them. Nonetheless, artificial expression of emotion can be misleading to the point that its users assume that the robot is capable of understanding their emotions, especially as humans have a tendency to ascribe human characteristics to objects (Nass et al., 1995). It is essential that these expectations are guided to be as close to the robot's abilities as possible. Not doing so may lead to unsuccessful human-robot interactions, which may result in disuse of the robot (Hancock et al., 2011). Furthermore, misguided expectations may lead to incorrect understanding of the robot's abilities and can lead to potentially harmful outcomes (Sparrow and Sparrow, 2006). These outcomes can either be physically harmful (e.g. falling over the robot as it was expected to have situational awareness through the way it behaved) or mentally harmful (e.g. getting upset as the robot did not respond appropriately due to a lack of contextual awareness).

1.1.2 Ethical Concerns of Social Robots

Reflecting on these potential incidents, emotional deception and emotional attachment are concerns that have been raised in the literature and can be applied to the use of AEE (Sullins, 2012; Sharkey and Sharkey, 2012). However, even though they have been identified in the literature, knowledge regarding their effects in practice is limited. It is essential to determine whether or not these concerns occur, as this will allow us to fully understand their impact and to address them accordingly. Furthermore, identifying concerns in the literature may discourage poor practice, but can also constrain research and development of new technologies if the concern raised proves to be unwarranted. The rise of ethical issues not only applies to social robotics but to many technologies and is expected to increase

(Vallor and Bekey, 2017), as can be seen in the number of sets and principles around the design and implementation of robotics and Artificial Intelligence systems that have been published in recent years (Jobin, Ienca, and Vayena, 2019).

1.1.3 Social Robots and Older Adults

MSCA-ITN project SOCRATES (grant agreement no. 721619) investigates different aspects of social robots and their use in European society, particularly focusing on interaction quality between social robots and older adults. This target group can benefit greatly from the services that social robots can provide (Banks, Willoughby, and Banks, 2008; Pu et al., 2019). It is even more important interactions between robots and older adults are successful now that people live longer and experience an increasing amount of health issues (World Health Organization, 2015) and loneliness (Khaksar et al., 2016).

However, age-related impairments increase the likeliness of older adults experiencing negative consequences of human-robot interactions (Sharkey and Sharkey, 2012). It is stated that the primary risk of interacting with robots in assistive contexts is that the person can become physically hurt (Feil-Seifer and Mataric, 2011). However, in this research I argue that the psychological impact can provide severe risks as well, which may lead to decreased well-being. This is especially the case when the robot provides mental support, for example to address loneliness. Therefore, this research aims to acquire a rich understanding of older adults' experiences when being exposed to AEE, to ensure negative consequences (both physical and mental) can be limited.

1.2 Context and Scope of Research

The first aim of this research is to determine potential negative consequences of AEE by a social robot, with a focus on older adults. The second aim is to develop a framework that can help developers and producers of future social robots to design social robot behaviours with minimal negative consequences. The main argument is that reflective and critical thinking is essential for the development of successful social robots.

1.2.1 Main Research Questions

An overview of existing ethical concerns that have been raised regarding social robots and AEE is provided in Chapter 2. Based on this review, it was determined that emotional deception and emotional attachment can be consequences of AEE. However, little is known whether they are reflected in practice as well as being raised in the literature. Therefore, two main research questions were established that are addressed in this work. The first research question focuses on potential ethical concerns as a consequence of AEE:

RQa: *'Can artificial expression of emotion by a social robot lead to emotional deception and emotional attachment?'*

The answer to this question is essential to ensure human-robot interactions are successful, as will be explained in more detail in Chapter 2. Next, it needs to be determined whether negative consequences can occur following AEE. This is addressed through the question:

RQb: *'Can artificial expression of emotion by a social robot result in negative consequences?'*

Emotional deception and emotional attachment have been addressed as ethical concerns at a theoretical level, as will be discussed in Chapter 2. However, there is only limited knowledge on whether they are reflected in practice, and whether negative consequences occur following these concerns. Therefore, RQa and RQb are investigated in this research through user studies. To be able to answer these questions, several sub-questions were formulated.

1.2.2 Supporting Research Questions

A methodology to execute the implementation and evaluation of AEE is outlined in Chapter 3. The implementation of AEE is provided in Chapter 4, as well as the evaluation of these behaviours. The following sub-question is addressed in this chapter:

RQ1: *'Is artificial expression of emotion as implemented in this research perceived as designed?'*

This is essential to establish, as otherwise the impact of AEE cannot be investigated. Once the answer to this question is 'yes', the impact that AEE can have on older adults can be evaluated. This is investigated in Chapter 5, which concentrates on two sub-questions. The first one investigates whether AEE can lead to emotional deception and emotional attachment in older adults:

RQ2a: *'Does AEE impact older adults' human-robot interaction experience?'*

This question will be answered by investigating results of specific measurements that may indicate emotional deception and emotional attachment. Furthermore, the impact of time is determined through the sub-question:

RQ2b: *'Does the impact of AEE on older adults change over time?'*

Time is an essential factor that may impact the success of human-robot interactions (Pripfl et al., 2016), as will be discussed in more detail in Chapter 5. Therefore, it is essential to investigate how users' behaviours when exposed to AEE may change over time. The final sub-question of Chapter 5 is:

RQ3: *'Can physiological data, speech prosody data and behavioural data provide valuable insights when used in longitudinal field studies in addition to questionnaires?'*

Successful human-robot interactions are of the essence for users to benefit from the services that a social robot can provide, as the robot may not be used if interactions are not successful. Using additional measures next to questionnaires can provide richer insights in users' needs and expectations. These insights may be especially important when considering vulnerable populations. These measures are presented and investigated in Chapter 5.

Chapter 6 presents the framework that is based on the results of the user studies of this research. However, as human-robot interaction experiments often have limitations, it should be determined whether the findings are reflected in the opinion of the general public. Therefore, one sub-question that Chapter 6 addresses is:

RQ4a: *'How do the results found in this research compare to the opinion of the general public?'*

More specifically, it is of high importance that the findings of this research represent the opinion of a large population, in order for them to be suitable as input for a framework. Therefore, the final sub-question investigated in this work is:

RQ4b: *'Can the findings of this research be used for the development of a framework on ethical AEE?'*

Finally, Chapter 7 considers the answers to the sub-questions (RQ1 through RQ4), in order to address RQa and RQb.

1.3 Dissertation Outline

This research investigates whether concerns regarding social robot ethics that have been considered in the literature are reflected in practice. Furthermore, the importance of critical thinking for the development of social robots and social robot behaviours is highlighted. Finally, a framework is developed that can help developers and producers of future social robots to design social robot behaviours that will result in minimal negative consequences. The outline of this work is as follows:

Chapter 2: Literature Review - Provides an overview of existing ethical concerns regarding the use of social robots where particular emphasis is placed on emotional deception and emotional attachment as ethical concerns. This chapter supports the formulation of RQa and RQb.

Chapter 3: Research Methodology - Presents an overview of how artificial expression of emotion was investigated, and justifies what user studies were conducted and why.

Chapter 4: Designing Artificial Expression of Emotion - Describes the modalities that were used to implement artificial expression of emotion and describes how it was tested whether the implemented behaviours were perceived as intended. This chapter addresses RQ1.

Chapter 5: Evaluating Artificial Expression of Emotion with Older Adults - Presents the preparation and execution of a longitudinal field study with older adults, which is the main user study of this research, that investigates the impact of emotional

deception and emotional attachment through artificial expression of emotion. RQ2a, RQ2b and RQ3 are discussed in this chapter.

Chapter 6: A Framework for Ethical Design of Artificial Expression of Emotion - Reviews whether the findings from the user studies are reflected in the opinion of the general public through an online survey, and describes how a framework is developed based on these findings. The research questions addressed in this chapter are RQ4a and RQ4b.

Chapter 7: Discussion - Summarises the achievements accomplished in this research, and assesses the impact that the main findings of this research may have, highlighting strengths and weaknesses and providing possibilities for future work. The main research questions, RQa and RQb, are addressed in this chapter.

Appendices - A large amount of data was collected for each of the user studies, additional analyses and information not central to the dissertation can be found in the appendices.

1.4 Contributions

Ethical HRI Literature - This research contributes to the groundwork of ethical HRI, by highlighting the importance of ethical AEE to ensure successful human-robot interactions, where psychological and physiological effects that AEE may have on social robot users are considered.

Ethical HRI Framework - A framework is presented that can help developers and producers of future social robots to design social robot behaviours that have minimal negative consequences. Reflective questions and recommendations are provided that can assist in identifying ethical concerns of social robot behaviours.

Ethical HRI Assessment - A survey is introduced that assesses the opinion of the general public on the findings from this research. The high reliability of the survey makes it suitable for future use. The approach of the survey is novel, as it supports participants to make more informed decisions by providing real contextual scenarios.

Ethical HRI Evaluations - This research presents two online surveys that evaluate the opinion of the general public. Furthermore, three formal studies are presented that report on participants' experiences, including one longitudinal field study. These studies provide us with a richer understanding of the impact that AEE can have on user experience, and the effect that longitudinal field-studies can have on older adults over time. This experiment with older adults resulted in several methodology recommendations on conducting a longitudinal field studies with vulnerable populations.

1.5 Publications

The implementation and evaluation of AEE, that are discussed in Chapter 4, have been presented in:

Van Maris A., Zook N., Caleb-Solly P., Studley M., Winfield A., Dogramadzi S. (2018). Ethical considerations of (contextually) affective robot behaviour. In Bringsjord, S., Tokhi M. O., Ferreira M. I. A. and Govindarajulu N. S. (Eds.), *Hybrid Worlds: Societal and Ethical Challenges - Proceedings of the International Conference on Robot Ethics and Standards (ICRES 2018)* (pp. 13-19). CLAWAR Association LTd.

A pilot study was conducted in order to develop the longitudinal field-study with older adults, as presented in Chapter 5. This pilot study is described in:

Van Maris A. (2018). The effect of affective robot behaviour on the level of attachment after one interaction. *In Proceedings of the International PhD Conference on Safe and Social Robotics (SSR 2018).* (pp. 1-3).

The longitudinal field study with older adults (Chapter 5), that was conducted to determine the impact of AEE on older adults, is published in:

Van Maris A., Caleb-Solly P., Zook N., Studley M., Winfield A., Dogramadzi S. (2020). Designing ethical social robots - a longitudinal field study with older adults, *Frontiers in Robotics and AI.* Volume 7, Issue 1. (pp. 1-14).

An important incidental finding of the longitudinal field study with older adults, as also discussed in Chapter 5, is highlighted in:

> **Van Maris A.**, Dogramadzi S., Zook N., Studley M., Winfield A., Caleb-Solly P. (2020). Speech related accessibility issues in social robots. *In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction.* (pp. 505-507).

And finally, even though the framework that was developed in this work (Chapter 6) has not been published yet, initial recommendations were provided in:

> **Van Maris A.**, Zook N., Studley M., Dogramadzi S. (2019). The need for ethical principles and guidelines in social robots. In Tokhi M. O., Ferreira M. I. A., Govindarajuli N. s., Silva, M., Virk, G. S., Kadar, E. and Fletcher S. R. (Eds.), *Artificial Intelligence, Robots and Ethics - Proceedings of the Fourth International Conference on Robot Ethics and Standards (ICRES 2019)* (pp. 19-24). CLAWAR Association Ltd.

1.6 Related Publications

This research led to several collaborations with other researchers, that has not been discussed in detail in this dissertation.

The implemented behaviours to convey AEE by a social robot have been used in other work, where it was investigated how AEE, among other factors, impacted robot acceptance:

> Bishop L., **Van Maris A.**, Dogramadzi S., Zook N., (2019). Social robots: the influence of human and robot characteristics on acceptance. *Paladyn Journal of Behavioral Robotics,* Volume 10, Issue 1. (pp. 346–358).

A secondment at SoftBank Robotics resulted in a collaboration with a colleague of the SOCRATES project and one of the main researchers of the company, where AEE through speech was explored:

> **Van Maris A.**, Sutherland A., Mazel A., Dogramadzi S., Zook N., Studley M., Winfield A., Caleb-Solly P. (2020). The impact of affective verbal expressions in

social robots. *In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction.* (pp. 508-510).

Collaboration with a co-worker who focused on social robot persuasion resulted in an initial taxonomy for emotional deception by a social robot:

Winkle K. & **Van Maris, A.** (2019). Social influence and deception on socially assistive robotics. In Tokhi M. O., Ferreira M. I. A., Govindarajuli N. s., Silva, M., Virk, G. S., Kadar, E. and Fletcher S. R. (Eds.), *Artificial Intelligence, Robots and Ethics - Proceedings of the Fourth International Conference on Robot Ethics and Standards (ICRES 2019)* (pp. 45-46). CLAWAR Association Ltd.

Finally, a discussion with a colleague on critical reflection of new technologies led to a short paper on ethical concerns of embodied prosthetic devices:

Ward-Cherrier B. & **Van Maris, A.** (2019). Ethical implications of embodied prosthetic devices. In Tokhi M. O., Ferreira M. I. A., Govindarajuli N. s., Silva, M., Virk, G. S., Kadar, E. and Fletcher S. R. (Eds.), *Artificial Intelligence, Robots and Ethics - Proceedings of the Fourth International Conference on Robot Ethics and Standards (ICRES 2019)* (pp. 142-143). CLAWAR Association Ltd.

2 Literature Review

This chapter introduces different ethical approaches that have guided the development of principles in robot ethics. It follows with an overview of ethical concerns regarding the use of social robots that have been raised in the literature. After that, it is described how machines and social robots may express artificial emotion. This topic is investigated as emotions are an essential aspect of human-human interaction (Keltner and Haidt, 1999), and play an important role in our lives (Picard, 2000). This indicates that artificial expression of emotion by social robots can improve human-robot interaction quality (Breazeal and Scassellati, 1999). However, it is essential that not only the benefits but also the potential negative consequences are known, which is discussed next in this chapter. After that, it is highlighted which ethical concerns that have been raised in the literature are investigated in this work and why. It should be noted that this chapter provides a general overview of topics related to this research. More detailed overviews, for example on how AEE can be implemented, and what kind of ethical frameworks for robot development already exist, will be provided in the relevant chapters.

2.1 Ethics in Robotics

Asaro has argued that there are at least three forms of 'ethics in robotics', namely the ethics of people who design and use robots, ethical systems that are built into robots, and the ethics of how people treat robots (Asaro, 2006). The approach 'ethical systems that are built into robots' is later identified as 'machine ethics', where the other two approaches are defined as 'the ethical application of robots and AIs', also known as 'robot ethics' (Winfield, 2019). Ideally, research in robot ethics would focus on these three approaches at the same time as they are all addressing the fundamental issue of how human and robot behaviour should be regulated (Asaro, 2006). However, these distinctions highlight that there are different research areas within 'ethics in robotics' that require specific attention and they need to be investigated separately before they can be merged to ensure safe and successful deployment of robots into society.

Research in different areas of robotics such as machine ethics (e.g. Anderson and Anderson, 2007), and work on ethical perspectives in social robotics (e.g. Huber, Lammer, and Vincze, 2014), do often build on principles of biomedical ethics that are derived from moral norms (Beauchamp and Childress, 1979). They define four clusters of moral principles:

1. respect for autonomy - acknowledges one's right to hold views and make choices intentionally, with understanding and without influencing the action;
2. non-maleficence - harm to humans or humanity, either by action or inaction, should be avoided;
3. beneficence - do good, among other things by preventing and/or removing harm;
4. justice - ensure a fair distribution of goods in society and look at the role of entitlement.

Another work that has inspired research on robot ethics is the science fiction short story 'Runaround' from Isaac Asimov, where he introduces three laws of robotics (Asimov, 1941). Even though this is a fictional work, it has shaped expectations of how robots should act around humans for a long time, and also inspired other principles of robotics (e.g. Boden et al., 2017) and responsible robotics (e.g. Murphy and Woods, 2009). The three laws of robotics entail the following:

1. No human may come to harm either through action or inaction by a robot;
2. A robot must obey orders given by a human, unless this is in conflict with the first law;
3. A robot must protect its own existence, unless this is in conflict with the first or second law.

Both the biomedical principles and the three laws of robotics have influenced research in robot ethics, which can be expected as there are similarities. For example, the principle of 'non-maleficence' is similar to Asimov's first law, where the goal is to avoid harm to humans and humanity. They also both clarify that harm can follow not only from action but also from inaction. However, Asimov's second law is in contrast with the biomedical

principle ‘respect for autonomy’. The second law states that ‘A robot must obey orders given by a human, *unless this is in conflict with the first law*’. This second part of the law contradicts the principle of respect for autonomy, which highlights the rights to hold one’s own views and make choices intentionally, without their actions being influenced and while being understood. If a person would command the robot to harm them, this should be respected according to the biomedical principle ‘respect for autonomy’. However, it should no be adhered to following the second rule of Asimov, as that will result in harm to the person, which should be prevented. This indicates that even though principles and rules from other fields can guide development of but not pass as principles in robot ethics, and specific principles to address ethical concerns are required.

The influence of both the principles of biomedical ethics and the three laws of robotics can be found in the Principles of Robotics (Boden et al., 2017). These are five principles that are described to the general public as follows:

1. Robots should not be designed as weapons, except for national security reasons;
2. Robots should be designed and operated to comply with existing law, including privacy;
3. Robots are products: as with other products, they should be designed to be safe and secure;
4. Robots are manufactured artefacts: the illusion of emotions and intent should not be used to exploit vulnerable users;
5. It should be possible to find out who is responsible for any robot.

The three laws of robotics are mentioned by the authors of these principles, where they highlight the fact that the work introducing the three laws is fictional and stress that not robots themselves, but humans are responsible for robot behaviours and actions. This becomes clear when looking at the principles in more detail. For example, the larger goal of the first principle can be interpreted as ensuring that no human comes to harm (countering the use of robots as weapons), which is similar to Asimov’s first law. However, the focus now lies on humans preventing harm through development and implementation of the robot, instead of the robot preventing harm itself.

Even though the authors do not mention the principles in biomedical ethics, they can be interpreted from the principles of robotics as well. For example, the final principle on responsibility for robots can be interpreted as the biomedical principle of justice, where identifying responsible parties will be an essential step to ensuring that robots are used in a safe and responsible manner and their use is fair for everyone.

The authors' aim of publishing these principles was to start a more open discussion on ethical issues in robotics (Boden et al., 2017). This was successful, as there are many documents addressing these principles. For example, it is argued that robots are more than tools (Prescott, 2017), that they never should be allowed to decide whether to kill a person or not (Sharkey, 2017), and that the impact of the principles in areas like social and legal conduct is limited as there are no clear definitions of how the authors define terms as 'robots', 'tools' and 'agents' (Voiculescu, 2017). As it is clear these principles have been discussed thoroughly I will not go into detail any further in this work. However, I would like to highlight one concern that not only applies to these principles but many works that raise potential ethical concerns of social robots (e.g. Sharkey and Sharkey, 2012). As can be seen in the fourth principle, there is a concern on deception and exploitation of vulnerable populations by robots. I agree that this is an important consideration, and that people that are more vulnerable to potential negative effects of robots should be protected at all cost. However, highlighting this concern for vulnerable populations only, undermines the importance to determine potential negative consequences of deception (and other psychological concerns) for *all populations*, as I argue that many, if not all, populations can be psychologically impacted by certain robot behaviours, especially in the case of social robots. For example, if a social robot displays a certain behaviour, any layperson can be deceived and believe the robot is capable of experiencing emotions. This does not only apply to vulnerable populations. This is a gap that I aim to address in the framework presented in Chapter 6, where I highlight the need to determine consequences for the intended target audience, regardless of whether this audience is categorised as vulnerable.

Finally, another ethical approach that has impacted ethics in robotics is philosophical ethics, and more specifically normative ethics, which investigates questions on how one should act morally. One approach to answer these questions which has also been used in robotics is virtue ethics (e.g. Sparrow, 2016). Virtue ethics considers, among other things, the role of emotions in our moral life and the fundamentally important questions of what

sorts of persons we should be and how we should live (Hursthouse and Pettigrove, 2020). In virtue ethics one aims to cultivate desirable character traits, while trying to get rid of undesirable character traits (Russell, 2013). It is different from other ethical theories as it focuses on the character of agents and not on their actions (Sparrow, 2020). De Graaf argues that people ascribe virtual virtue to robots as their function is to contribute to the good life of people (De Graaf, 2016). The authors of the principles in robotics did not mention virtue ethics in their work, but one could argue that they adopted a virtue ethics approach through their goal to address designers and developers to urge them to 'do the right thing' while developing robots. Virtue ethics is becoming increasingly popular in philosophy of technology (Coeckelbergh, 2020). This has led to the introduction of the term 'technomoral virtues', which are traits that must be cultivated more successfully for people to be able to live with emerging technologies (Vallor, 2016), and therefore should be pursued in the development and use of social robots.

It can be concluded that the overarching goal of robot ethics is to ensure that the potential negative consequences, both of the development and the use of robots, to humans and society are limited, if not prevented. The next section will present ethical concerns regarding the use of social robots that have been raised in the literature.

2.2 Social Robots

To be able to understand ethical concerns that apply to social robots specifically, I first had to establish what a social robot entails. As mentioned in the introduction, there is no universally agreed definition for a social robot. A common factor for several definitions that have been proposed, is that a social robot should be capable of social communication in a human-like manner (De Graaf, 2015). Sub-categories of social robots have been provided in the literature, such as socially intelligent robots, socially interactive robots and sociable robots (Dautenhahn, 2007b). A definition of a socially intelligent robot is that it is 'an embodied agent that is part of a society of robots and/or humans' (Dautenhahn and Billard, 1999). A definition for a socially interactive robot is that it is 'an agent for which interaction is important' (Fong, Nourbakhsh, and Dautenhahn, 2003). As the focus of this research is on older adults and how they may be impacted by robots providing support, the term socially assistive robots should be considered. They are defined as the intersection of

physically assistive robots and socially interactive robots (Bertel, 2011). A different type of assistive robots is care robots, which are introduced as 'robots that use interactions as a means to meet care needs' (Van Wynsberghe, 2012). This research focuses on the impact that interactive behaviours of social robots may have on people in an assistive context. During the experiments conducted in this research, the robot provides support in the form of cognitive stimulation and social engagement. This assistive functionality, taking into consideration that older adults are the intended target audience, resulted in a definition that has taken inspiration from definitions for 'socially assistive robots' and 'care robots'. Hence, a social robot is defined in this research as *a robot for which interactions are of high importance to achieve its fundamental and intended functionality, that is ultimately designed to meet the needs of the people it interacts with.*.

2.2.1 Ethical Concerns of Social Robots

Older adults are a popular target group when it comes to developing applications for social robots. This is due to the growing number of older people who need care, without the systemic capacity to meet these growing care needs (World Health Organization, 2015). This incapacity to provide care for older adults is one of the reasons that research into social and assistive robotics is so attractive (Sparrow and Sparrow, 2006). Some beneficial uses of social robots for older adults are, for example, that the global ageing society will require ongoing care and health education, and receiving help from a robot may sometimes be less stigmatising for the older adults (Breazeal, 2011). Looking at the use of AEE, this can be beneficial for people experiencing negative consequences of loneliness (e.g. a decrease in cardiovascular function; Cacioppo and Patrick, 2008). Furthermore, it can help with encouraging engagement, building empathy and engendering trust (Hung et al., 2019). However, due to the novelty of the use of AEE, knowledge regarding potential negative outcomes is limited.

Frennert and Östlund established several matters of concern regarding the use of social robots for older adults. One of these concerns entails the ethical implications of using such robots (Frennert and Östlund, 2014). Examples of ethical implications are reduced human contact, loss of control, loss of personal liberty, loss of privacy, matters regarding responsibility, infantilisation, the quality of the support provided, emotional deception and consequences of emotional attachment (e.g. Sharkey and Sharkey, 2012; Sullins, 2012;

Sparrow and Sparrow, 2006; Turkle, 2006; Coeckelbergh, 2010). Furthermore, it is argued that ethical concerns not only apply to the user, but also to others involved in meeting the needs of the user (Vallor, 2011). These concerns have been highlighted in the literature, but counterarguments are provided as well. For example, the use of social robots may reduce social interaction with other people. However, if human contact was already limited, the use of social robots may reduce loneliness and increase conversation opportunities, not only for human-robot interactions but also for human-human interactions (Sharkey and Sharkey, 2012).

Even though some counterarguments are provided, many concerns raised in the literature are not analysed further (Vandemeulebroucke, Casterlé, and Gastmans, 2018). Identifying concerns in the literature may discourage poor practice, but it can also constrain research and development of new technologies. Therefore, these concerns need to be analysed in more detail to determine whether they are warranted.

Other identified ethical implications are loss of control and loss of personal liberty. However, robots can also provide older people with the opportunity to self-manage their well-being and the ability to reduce risks (Callén et al., 2009).

Loss of privacy is an important ethical concern. Nonetheless, this issue arises for different types of new technologies such as artificial intelligence systems and is being investigated in these research fields as well (e.g. Price and Cohen, 2019). Therefore, this topic will not be discussed further in this report. This applies to matters regarding responsibility as well, which is being researched through e.g. autonomous cars (Baumann et al., 2019).

Encouraging people to interact with robots, that may sometimes look like toys, may give them the feeling of being infantilised. However, this can be addressed by taking into account the aesthetics of the robot, which has also been identified as a matter of concern (Frennert and Östlund, 2014). Also, sometimes it may be assumed that older adults can feel infantilised when encouraged to interact with a robot, where this may actually not be the case. This will depend on both the cognitive abilities of the older adults, as well as their subjective response and societal perception of the robot.

The final two concerns that were investigated in this research and will be introduced later in this chapter are emotional deception and emotional attachment. These two concerns

were chosen as there is a gap in the literature reviewed to date, regarding research investigating their potential negative consequences. Also, their potential negative consequences can be indirect, long-term effects of AEE, and these consequences will most likely occur at a psychological level where the other concerns introduced here are direct consequences of interactions with social robots. This makes it more difficult, but even more important, to investigate them and mitigate their negative consequences where possible.

2.3 Artificial Expression of Emotion

Even though research on emotions and affect can be traced back to the 19th century, it has not been linked to machines until the end of the 20th century (Tao and Tan, 2005). Affective computing is described as computing that arises from, relates to or influences emotions (Picard, 2000). More specifically, it entails assigning human-like capabilities to computers such as interpretation, observation and generation of artificial emotions (Tao and Tan, 2005).

It is expected that people interact with social robots in a natural and social way (Weijers, 2013), as they do this with computers as well (e.g. Nass, Steuer, and Tauber, 1994). Furthermore, as mentioned before human-like communication is important for successful human-robot interactions. Therefore, affective computing is an important aspect of social robotics (Paiva, Leite, and Ribeiro, 2014).

It has been stated that affective computing can increase human-computer interactions (Tao and Tan, 2005). The display of emotions has been deemed important for successful human-robot interactions (Dautenhahn and Billard, 1999), as this can improve a robot's ability to communicate with a person (Kirby, Forlizzi, and Simmons, 2010). Affective computing has been used to generate emotions and explore the effect it has on people interacting with social robots. The focus of this research is not on how to generate emotions through learning systems, but rather on the effect that displaying emotions may have on people. AEE can be displayed using several modalities, such as body language and behaviour (Mandler, 1980), dialogue (Aristoteles, 1960) and speech (Fong, Nourbakhsh, and Dautenhahn, 2003). This will be explored in more detail in Chapter 4.

The literature has indicated that there are many positive effects of AEE by social

robots. It has also shown that there are several ethical concerns regarding the use of social robots. Ethical concerns that can follow from AEE are emotional deception and emotional attachment, which are discussed in the following sections.

2.4 Emotional Deception

Deception occurs when false information is communicated to benefit the communicator (Arkin, Ulam, and Wagner, 2012). It implies that an agent acts accordingly to induce a false belief in the one being deceived (Hyman, 1989). This can also mean that no information is communicated at all (Dragan, Holladay, and Srinivasa, 2015). Deception occurs in many different practices, like forgery, psychic hoaxes, practical jokes and white lies (Hyman, 1989). Looking at the principles of biomedical ethics, one could argue that deception breaches the principle of 'respect for autonomy', as the freedom to hold views and make decisions is denied by not providing the correct or complete information available.

Danaher distinguishes between three forms of robotic deception: external state deception, superficial state deception and hidden state deception (Danaher, 2020). External state deception implies that the robot uses a deceptive signal from the external world to indicate a false belief. Danaher mentions the example of a medical diagnostics robot giving an incorrect impression of one's health to tempt one into a medical treatment that is not needed. Superficial state deception entails the robot using some deceptive signal to indicate it has a certain internal state that it lacks in reality. AEE is categorised in this form of robotic deception. Finally, hidden state deception indicates that a robot obscures the presence of its internal state through the use of a deceptive signal.

Robot deception differs from philosophical definitions of deception as the latter often relates to intentions, beliefs and desires of the receiver, and it is hugely debated whether robots are capable of having intentions, beliefs and desires (Danaher, 2020). Furthermore, many ethical theories from a philosophical origin involve solely intentional deception and do not mention that deception can occur unintentionally as well (Trivers, 2011; Sharkey and Sharkey, 2020), which will be discussed more next.

2.4.1 Intentional and Unintentional Deception

Just like robot ethics was inspired by philosophical and biomedical approaches, deception can be approached through these perspectives as well (Shim and Arkin, 2013). It is highlighted that they both discuss the division of deception into being either intentional or unintentional (Dragan, Holladay, and Srinivasa, 2015). Unintentional deception occurs when a certain feature of the (unintentional) deceiver causes expectations that the deceiver means to evoke. This is also known as physical deception. Intentional deception occurs when the deceiver is aware of the fact that a certain feature will raise false expectations. This is called behavioural deception, as it is often the behaviour from the deceived person that causes the formation of these expectations. For a robot to be able to successfully perform a deceptive action, it requires specific knowledge about the person that it intends to deceive (Wagner and Arkin, 2011). Also, the robot should convey its intentions and have a theory of mind for the person being deceived to be able to manipulate their beliefs (Dragan, Holladay, and Srinivasa, 2015). Some researchers claim that social robots do not intend to deceive, and therefore any intentional deception comes from developers or users (Matthias, 2015; Coeckelbergh, 2012). However, it should be considered that these actors can also evoke unintentional deception of the robot. This can be true for AEE, as it can be implemented with the goal to create more natural interactions and increase social robot acceptance, with emotional deception being an unintended consequence of AEE.

Some might argue that unintended negative ethical consequences of a design decision (either hardware or software) is not an ethical shortcoming but a design issue (Misselhorn, Pompe, and Stapleton, 2013). Nonetheless, a multidisciplinary approach is required for successful human-robot interactions (Lindblom and Andreasson, 2016), which should also include ethical design considerations.

2.4.2 Opinions Regarding the Use of Deception

Opinions regarding the question of whether (emotional) deception is right or wrong are divided. Some say it is unethical, as it encourages users in self-deception (e.g. Sparrow, 2002). Furthermore, even though there may be some health benefits, it is argued that overall well-being is not improved when using deception, as it disengages people from reality and creates moral failures (Sparrow and Sparrow, 2006). It is stated that deception may lead to

vulnerable people misunderstanding the robot's abilities (Sharkey, 2014). However, as I already argued before, deception, and especially unintentional deception through the use of AEE, can occur to any layperson and not only to vulnerable populations with cognitive limitations. Noel and Amanda Sharkey support thise argument of misunderstanding the robot's ability by focusing on the deceived and not the deceiver (Sharkey and Sharkey, 2020) and propose that in the absence of intention false beliefs can still be created.

However, others are of the opinion that deception is ethically correct if it increases benefits for the deceived (Shim and Arkin, 2013). An example where deception benefits the deceived is the use of the placebo-effect. This deception being beneficial is also called benevolent deception (Adar, Tan, and Teevan, 2013). Benevolent deception has always been part of medical care (Jackson, 1991), and may even be required to act morally (Arkin, Ulam, and Wagner, 2012). Others argue that a robot is (morally) allowed to be deceptive, as long as there is no betrayal of trust, along with other criteria (Matthias, 2015). If the deception is in the interest of the patient (thus benevolent deception), there is no betrayal or breach of trust. Trust is an important factor to consider, as a breach of trust may result in a different outcome of the interaction (Hancock et al., 2011).

One of the issues regarding deception is that (often unintentional) deception is not always regarded as deception, especially if there are no clear harmful consequences. Some argue these deceptions are 'harmless fun' (Sharkey and Sharkey, 2020), where others claim that it should not be regarded as deception (Danaher, 2020). I personally experienced that there is a negative association with the word 'deception', and therefore stakeholders may be unwilling to use it, especially in cases where deception is harmless. However, as social robotics technology is expanding and many consequences are not yet known, they should be considered with extreme care and additional consideration through the use of the word 'deception' is desirable. As long as the impact of social robots on individuals and society is unknown, it is essential to tread with care, as seemingly innocent features may have (unexpected) substantial negative consequences.

2.4.3 Deception and AEE

Artificial expression of emotion by a robot may also be described as the appearance of human-robot affection, as real affection requires the experience of emotions, which is

difficult (if possible) to implement in robots (Weijers, 2013). This appearance of affection may be regarded as emotional deception, as the robot shows affection where it does not feel affection, and thus provides inaccurate and misleading information about its internal state (Fulmer, Barry, and Long, 2009). It is stated that people are very vulnerable to emotional manipulation by AI and robotic systems (Scheutz, 2011). Deception is created when robots are used in assistive settings (Sharkey and Sharkey, 2011), since the robot's social behaviour often does not correspond with its actual capabilities. AEE is found ethically dubious, as it can lead to people believing their relationship with the robot is mutual (Turkle, 2007).

It should be noted that emotional deception may often occur unintentionally. As emotional deception can be a consequence of artificial expression of emotion and has both advantages and disadvantages, it is important to investigate whether the benefits outweigh the risks. Therefore, it is important to investigate and understand people's responses to AEE, and determine whether there is a cause for ethical concerns as emotional deception.

2.4.4 Measuring Deception

The goal of AEE is not to deceive people but to provide one or more of the benefits presented earlier in this chapter. Therefore, if deception occurs as a consequence of AEE, this should be classified as unintentional deception. This type of deception is difficult to measure directly so should be measured indirectly through other phenomena. One approach can be to investigate anthropomorphism - the attribution of human traits to non-human entities - and the level to which this occurs during interactions. As stated by Danaher, misrepresentation by a robot can be in the form of speech, behaviour or physical appearance (Danaher, 2020). Anthropomorphic cues can be used to conceal non-anthropomorphic abilities in a robot (Kaminski et al., 2016), and a taxonomy has been developed on anthropomorphic cues that can give misleading impressions of a robot's goals (Leong and Selinger, 2019).

Another means to measure the occurrence of deception is through the investigation of social presence - whether the robot is perceived as a social entity (Heerink et al., 2010). A connection was found between a robot's social abilities and participants' sense of presence of the robot (Heerink et al., 2008), making it possible to measure deception through social

presence.

2.5 Emotional Attachment

Attachment can be defined as a lasting psychological connection between human beings (Bowlby, 1969), also known as Attachment Theory. This theory was based on research that determined the importance of an infant's relationship with their mother for social, cognitive and emotional development (Bowlby, 1958). This theory can be extended to other agents as well, as besides a connection between humans, attachment can also entail a relationship between people and domestic pets (Coeckelbergh, 2011). Furthermore, it is possible to become attached to an object (Keefer et al., 2012; Turkle, 2011), which can be defined as an emotional bond between a person and the technology they use (Suh, Kim, and Suh, 2011). This is possible as people have a tendency to respond socially to computers, which is known as The Media Equation (Reeves and Nass, 1996).

When interacting with a robot, people can form emotional attachments to this robot (Borenstein and Arkin, 2019). For example, reports have shown that attachment to military bomb disposal robots already occurs (Michel and Carpenter, 2013). It is likely that attachment to social robots becomes more intense when the sophistication of these robot increases (Borenstein and Arkin, 2019), for example through AEE. It was found that affective modelling can induce feelings of attachment (Moshkina et al., 2011), increasing the likelihood of attachment when AEE occurs. This is supported by the argument that social interactions that are more engaging (e.g. through AEE) can lead to attachments (Sharkey, 2014).

If AEE can lead to emotional attachment and people's needs following this attachment are met by the robot's abilities, then it can be argued that the use of AEE and the consequent attachment are virtuous, as it enhances the good life. This would improve quality of life, thus 'doing good' and addressing the biomedical principle on beneficence. However, if attachment grows following AEE and the robot is not capable of meeting the needs that come with this attachment, then both the biomedical principle on non-maleficence and the first law of robotics are breached as this may lead to harm, which is discussed more in the next section.

2.5.1 The Impact of Emotional Attachment to Social Robots

Before social robots can successfully be integrated in older adults' lives, the effects of interacting with such a robot should be known first. It can be positive to become attached to a robot, as high attachment can lead to increased use of the technology (Li, Browne, and Chau, 2006), where the person can optimally benefit from the services the robot can provide. Examples of these services are alleviating loneliness and improving a person's well-being (Hutson et al., 2011).

However, there are also disadvantages to becoming emotionally attached to a robot. For example, it may increase a person's level of dependence on the robot (Sharkey and Sharkey, 2012). Furthermore, increased use of the robot may lead to social isolation (Feil-Seifer and Mataric, 2011; Sharkey, 2014). Also, vulnerable users may not understand why the robot is taken away (e.g. when it breaks), which can cause distress and loss of the benefits the robot can provide (Feil-Seifer and Mataric, 2011). Furthermore, forming bonds with social robots may create a moral obligation toward them, which may not be in the best interests of human well-being (Bryson, 2018). Therefore, the social and psychological impact of human-robot relations should be analysed (Reis, Collins, and Berscheid, 2000; Ostrowski et al., 2019). Finally, it is argued that attachment can be a potential harm of unintentional deception (Sharkey and Sharkey, 2020), for example through AEE.

2.5.2 Attachment and AEE

AEE may influence one's level of attachment to a robot, since affective behaviour can result in a more natural interaction with it (Kirby, Forlizzi, and Simmons, 2010). However, as discussed in the previous section, the user may be deceived by AEE, as this may imply some level of autonomy, and raise false expectations regarding the robot's abilities (Sparrow and Sparrow, 2006; De Graaf, 2016). Therefore, it is important to determine whether AEE elicits emotional attachment, and if the advantages of becoming emotionally attached to a robot outweigh the potential disadvantages.

2.5.3 Measuring Attachment

No specific means to measure attachment to robots have been established so far. In theory, any of the types of attachment mentioned before (human, pet, object) can be explored. Depending on the robot's appearance and functionality it can differ whether it is perceived as more similar to a human or a machine (Weijers, 2013), and therefore the form of attachment may differ per robot. This also applies to pet robots where either animal attachment or object attachment can be explored.

The difficulty lies in determining what approach is best for certain situations, due to the contrast in psychological and ontological status of social robots (Coeckelbergh, 2011; Prescott, 2017). Social robots may be categorised as machines or tools, but it appears that humans behave differently toward them, like they have psychological capacities (Prescott and Robillard, 2020). This ambiguity has led to the suggestion that social robots belong to a new ontological class (De Graaf, 2016; Kahn Jr and Shen, 2017).

Social robots can access forms of social support that are less natural for people such as sensors and databases, and it can be expected that human-robot relationships will be different from human-human relationships (Prescott and Robillard, 2020) and robot capability does not match human intelligence yet. Therefore, it can be concluded that human-attachment questionnaires are less suitable for human-robot interaction research. However, as suggested by the Media Equation, it is still possible that people become attached to robots, also if their relationships will differ from human-human relationships.

Pet attachment questionnaires would be suitable only for social robots with a functionality similar to that of pets, and even then it is uncertain this approach is useful as earlier work indicated that participants' attachment to robots was not as typically defined in human-pet or human-human relations (Huang, Varnado, and Gillan, 2013). Also, current social robots are perhaps not advanced enough yet to imitate reciprocal affection. As mentioned before, this is not a requirement for object attachment. For these reasons, it would be most useful to consider existing object-attachment questionnaires to measure human-robot attachment.

As mentioned before, attachment already occurred with military robots, where it was preferred a broken robot was repaired instead of replace (Michel and Carpenter, 2013).

This is another reason to measure robot attachment as object attachment, as people are more likely to repair an object when it breaks down (Schifferstein and Zwartkruis-Pelgrim, 2008). Furthermore, product personalisation leads to increased attachment (Mugge, 2004), and robot personalisation is an active research area in HRI (e.g. Gordon et al., 2016). Therefore, object attachment seems the most suitable alternative currently available to measure human-robot attachment.

2.6 Summary

Several ethical concerns have been raised in the literature regarding the use of social robots, which have been presented in this chapter. Furthermore, it was found that the use of AEE may result in some of these concerns, namely emotional deception and emotional attachment. Potential negative consequences of these concerns are provided in Table 2.1. As this table indicates, scientific literature has indicated that the use of social robots and AEE may result in negative consequences. However, analysis whether these concerns are reflected in practice is limited, but required to shape perspectives on these concerns (De Graaf, 2016). Little is known on how AEE may impact the robot users. It is important that user studies are conducted to investigate this impact, to gain understanding in user experience and guide the development of AEE to meet their needs. Potential means to measure emotional deception and emotional attachment have been presented in this chapter. Following these insights, the mains research questions of this research are defined as follows:

> **RQa:** *'Can artificial expression of emotion by a social robot lead to emotional deception and emotional attachment?'*
>
> **RQb:** *'Can artificial expression of emotion by a social robot result in negative consequences?'*

It was expected that AEE can indeed lead to emotional deception and emotional attachment, as this has been stated in the literature before (e.g. Fulmer, Barry, and Long, 2009; Borenstein and Arkin, 2019). Several negative outcomes have been addressed in the literature as a consequence of emotional deception or emotional attachment, of which some are summarised in Table 2.1. Negative consequences are often paired with vulnerable

users with cognitive impairments. The participants in this research were vulnerable to the extent that they lived in a semi-independent retirement village as they were not capable to live fully independently, but were all cognitively healthy. Furthermore, the level of AEE used in this research was designed to be low, to determine whether a baseline use of AEE has implications for human-robot interactions. Therefore, it was not expected that negative consequences of AEE would occur in this research.

Before providing details on how AEE was implemented in a social robot, and how emotional deception and emotional attachment were evaluated, Chapter 3 will first provide an overview of the approach taken in this research to investigate these research questions.

Ethical concern	**Positive impact**	**Negative impact**
Emotional deception	Health benefits	Disengages people from reality
	Act morally	Misplaced trust in the robot
		Misunderstanding of the robot's abilities
Emotional attachment	Alleviate loneliness	Increase level of dependence on robot
	Improve well-being	Social isolation
		Feelings of distress when separated from robot

TABLE 2.1: Example positive and negative consequences of emotional deception and emotional attachment, which may occur through AEE.

3 Research Methodology

The two main questions explored in this research are:

RQa: *'Can artificial expression of emotion by a social robot lead to emotional deception and emotional attachment?'*

RQb: *'What is the impact of emotional deception and emotional attachment on user experience?'*

There were several challenges in addressing these questions concerning the validity, reliability and reproducibility of this research. First, existing measures for emotional deception and emotional attachment are limited. Therefore, new and suitable measures had to be considered, which may impact the validity and reliability of the research. Second, it was challenging to recruit participants that are representative of the goals of this research. To gain as much information as possible, it was decided to adopt a mixed method approach. Furthermore, the decision to run a field study instead of a lab-based study may have impacted the reproducibility of the work.

Several user studies were conducted to address the research questions. An overview of what experiments were performed in order to develop a framework on AEE is provided in more detail in Section 3.1, along with the rationale for the sequence in which the studies were conducted. The materials and measures used in this research are outlined next. Finally, an explanation is provided on how data was processed and analysed.

3.1 Surveys and User Studies Conducted in this Research

This research investigated how older adults react to emotive robot behaviours, and whether potential negative ethical consequences occurred. A combination of online surveys, lab-based experiments and field studies was used to accomplish this. Surveys were used to gather the view of the general opinion as this allowed me to reach a larger target audience compared to physical experiments. However, as people may respond differently to a robot

displayed on a screen compared to a robot physically present in the same room (Deng, Mutlu, and Mataric, 2019), it was essential to test and evaluate people's responses through physical experiments as well. However, lab-based studies are likely to evoke different responses from participants compared to when these studies are conducted 'in the wild' (Jung and Hinds, 2018). Therefore, a field study with older adults was conducted as well, as this would result in more natural responses and give truer insights in their user experience.

As mentioned before, there are several challenges in accomplishing the goal of this research. The number of participants representative of the target audience was expected to be limited, as this target audience may be hesitant to interact with technologies unknown to them (Wu et al., 2014). Furthermore, the number of people with the baseline level of vulnerability required for this project is limited. This research aimed to recruit participants that needed support and were not able to live independently. However, to ensure a baseline of vulnerability was represented, the amount of support they needed had to be minimal, and they could have any cognitive impairments. Therefore, it was decided to run a longitudinal field study with this participant group to address the research questions. Before this field study was conducted, pre-studies were run to ensure that AEE was perceived as designed, in order to get the most out of the data from the field study. A between-within mixed method approach was taken for this field study, to be able to address the research questions as completely as possible. The next sections will summarise what type of experiments were conducted and with what purpose. The studies are presented and were performed in chronological order, where the output of one study would provide input for the design of the following study.

3.1.1 Testing Artificial Expression of Emotion

The first step of the research was to implement AEE. This implementation is presented in Chapter 4. Next, it had to be determined whether AEE was perceived as designed. This was investigated through an online survey, where participants saw short video-recordings of the different robot behaviours. They had to rate the behaviours on a scale from sad to happy. It was decided to use a survey for this test, as this allowed for recruitment of a large participant group, which would increase the reliability of the results. As this was the first step of the research, it was essential that reliability of the results of this survey was high. More details on the survey and its results are provided in Chapter 4.

As the survey established that AEE was perceived as designed, it was tested whether this remained true during physical human-robot interactions as well. This study is described in Chapter 4 as well.

3.1.2 Investigating the Effect of Artificial Expression of Emotion

Once testing of the behaviours was completed, a longitudinal study with older adults was prepared that would investigate the impact of emotive robot behaviours on a vulnerable population. This study was designed as a field study, as this would help participants feel at ease as they were in their home environment instead of a laboratory setting. Also, as mentioned before older adults can greatly benefit from social robots. Therefore, it is essential to include them in user studies and the development of social robots, to ensure their needs are met. Furthermore, the field study allowed recruitment of participants with lesser mobility, which resulted in a more diverse and realistic participant group than when the study would have been conducted at a laboratory. Finally, the study was conducted over a longer period of time to account for the novelty effect and to investigate whether participants' behaviour changed over time. The study is presented in Chapter 5. Keeping in mind that a framework would be developed on AEE based on these findings, it was essential to ensure this study was engaging for participants. Therefore, a short pilot study was conducted first, which is also presented in Chapter 5.

3.1.3 Validating the Findings regarding Artificial Expression of Emotion

Several challenges had to be addressed in this research. The number of participants of the longitudinal field study with older adults was small, making it difficult to generalise the findings and use them for the development of a framework. Therefore, one final online survey was conducted to investigate whether the findings from the user studies were reflected in the opinion of the general public (also described in Chapter 6). Initially, the goal was to distribute this survey online to reach a large and diverse target audience, and distribute physical copies in retirement villages to engage users that are perhaps less likely to be reached through online surveys. However, due to COVID-19 distribution of physical copies was no longer possible, and participants were recruited online only.

The findings from the field study with older adults were transformed into statements, and participants were asked to rate to what extent they agreed with these statements. The results of all user studies and this final survey, together with the observations made while designing, performing and analysing these experiments, were used to update the framework. This final framework can help future researchers and developers to realise whether they are, likely unintentionally, misleading intended users with the artificial expressivity they are implementing in social robots. This can lead to users being emotionally deceived or becoming emotionally attached to the robot, leading to misplaced trust and potentially harmful outcomes. Figure 3.1 provides a roadmap of the steps that were taken in this research to develop the final framework.

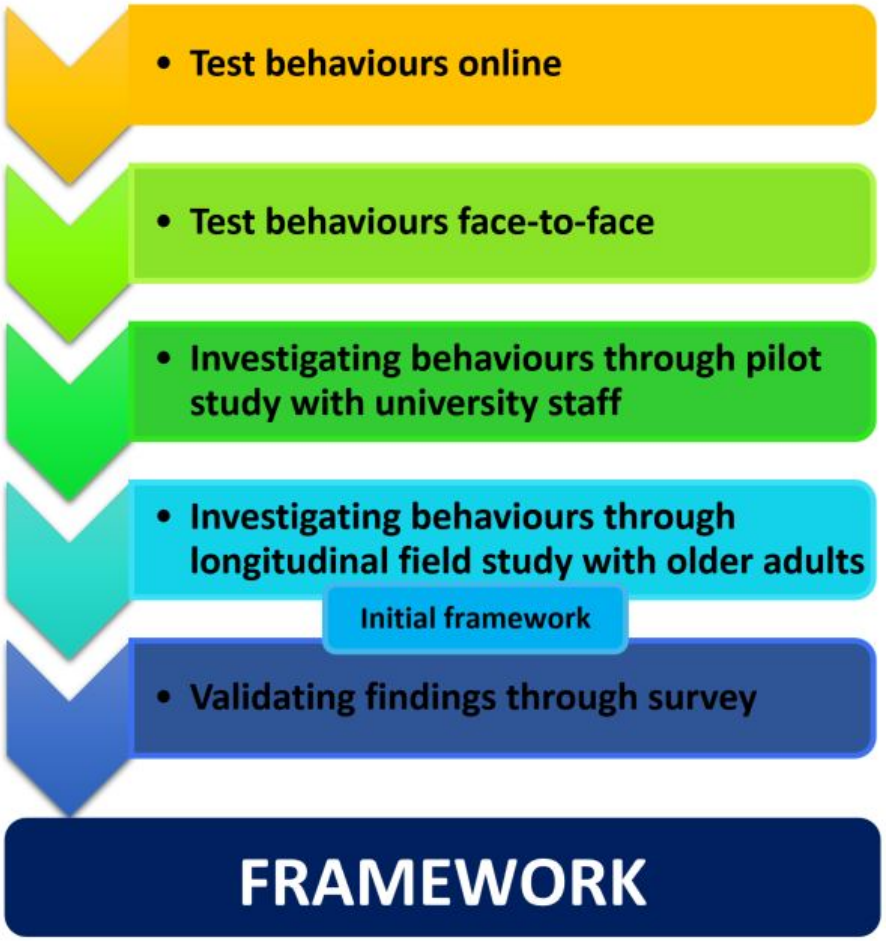


FIGURE 3.1: Roadmap to develop the framework.

3.2 Procedure

A Wizard-of-Oz strategy was used for the face-to-face experiments, which means the robot was manually operated during the human-robot interactions. This approach is well established within HRI research and measures human responses to robot actions (Steinfeld, Jenkins, and Scassellati, 2009). The use of this strategy was essential for this research, as interactions had to be as natural as possible in order to investigate participants' responses to AEE, and therefore should not impacted by a potential system failure.

The behaviours that would be displayed in this research were completely pre-programmed, and were manually prompted when the robot should continue with the interaction, waiting for participants' responses without needing to use speech recognition. It was ensured that additional prompts like 'yes', 'no', 'I don't know' or 'I do not understand, let us continue with the interaction' were implemented as well in case participants provided unexpected responses. It was decided to pre-program the interactions and manually operate the robot, as at the present time the robot's speech recognition is not always reliable, which may result in a frustrating experience for the participants when they are not accurately understood and need to repeat themselves. As this research focuses on whether and how participants respond to artificial expression of emotion by the robot, it was essential that as many external factors as possible that could potentially influence the participants' responses were eliminated. By manually prompting the robot to continue the interaction the accuracy of the conversation from the robot's side could be manually prompted for a more natural interaction.

3.3 Materials

The robot used for the implementation of the behaviours and used for all studies in this research is Pepper from SoftBank Robotics (Pandey and Gelin, 2018). Pepper is a humanoid robot of approximately 1.5m tall. An image of Pepper is provided in Figure 3.2.

There are several reasons why this robot was used. First of all, participants were meant to focus on the behaviour that the robot was displaying, and not be distracted by their mental model and expectations of how a robot should behave. This meant that the use of pet robots was not optimal, as speech was included in this research. This was

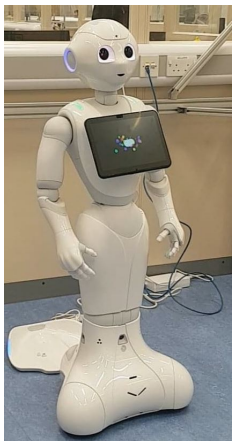


FIGURE 3.2: Social robot Pepper.

decided, as the main study of this work would be a longitudinal user study with older adults, and without speech it may be difficult to keep participants engaged without the sessions becoming too repetitive. Therefore, a humanoid robot was chosen for this research. Large humanoid robots like Baxter were not a suitable option, as their size may intimidate participants. However, the robot should not be regarded as a toy either, excluding NAO as a suitable platform. Therefore, it was decided to use Pepper for this research, as it meets the qualifications of not being perceived as a toy and not be intimidating due to its size. Also, Pepper is currently used for many studies with older adults (Carros et al., 2020; Unbehaun et al., 2019; Bechade et al., 2019). Furthermore, it is being introduced to the general public outside of laboratory settings such as museums and shopping malls (Allegra et al., 2018; Niemelä et al., 2019). This may results in Pepper being an inspiration for the design of future social robots, making it easier to apply the results of this work to research with future robots. Finally, it is cost-effective to use this robot as it is not very expensive for a robot and provides all modalities required for this research.

To be able to implement different emotive robot behaviours the software Choregraphe (Version 1.5.5.5) was used, which is provided by SoftBank Robotics as well (Pot et al., 2009). How this software was used is described in more detail in Section 4.2. Surveys were designed using the software Qualtrics (Qualtrics, 2013) and distributed via email using Outlook 365 (Murray, 2011), and through the online platform Prolific (Palan and Schitter, 2018). Gathered data was pre-processed in Microsoft Excel 2016 (Carlberg, 2017), and analysed using IBM SPSS Statistics for Windows, version 25 (SPSS, IBM, 2017).

3.4 Measures

Both subjective and objective measures were used in this research. Questionnaires are often used measurements in HRI studies and provide useful initial insights. Unfortunately, there are also disadvantages to the use of questionnaires. For example, they are always conducted after the interaction, so participants have to think back to the interaction and how they felt about it to give a response. Their answers in hindsight may not always be similar to the feeling they experienced during the interaction. Also, participants may, either consciously or unconsciously, try to give the answer they think is expected. This indicates that only the use of questionnaires in HRI studies is not sufficient to get a complete insight in participants' experiences.

Therefore, data from additional measurements like video-recordings and physiological recordings was gathered as well. These types of data, all with their own advantages and disadvantages, may provide additional insights to the findings from the questionnaires.

3.4.1 Questionnaires

The following questionnaires were used in this research to validate, investigate and evaluate AEE. Examples of these questionnaires are provided in Appendix A.

3.4.1.1 Demographics

Age, gender and level of education were collected for all experiments. Additionally, participants were asked to rate their familiarity with social robots and robotic technologies on a scale from 1 (not familiar at all) to 5 (very familiar). Furthermore, participants of the longitudinal field study were asked to provide how often they used technologies such as a smartphone, tablet and laptop/desktop.

3.4.1.2 Questions on Displaying Emotions

While testing the implementation of AEE, a 5-point sad-to-happy scale was used where participants were asked to rate the emotional state of the robot, as shown in Figure 3.3. More details are provided in Section 4.3. However, when evaluating AEE in the field study, this scale could no longer be used, as the robot now showed both sad and happy

behaviours during the interactions (if any). Therefore, participants had to rate to what extent they agreed on a 5-point scale from 1 (fully disagree) to 5 (fully agree) with the statements 'Based on my interaction with Pepper, I feel that it is capable of communicating an emotion of happiness' and 'Based on my interaction with Pepper, I feel that it is capable of communicating an emotion of sadness'.

FIGURE 3.3: Five-point scale from sad (1) to happy (5) which was presented to participants to rate the robot's emotional state.

3.4.1.3 Godspeed Questionnaire

All constructs of the Godspeed questionnaire (Bartneck et al., 2009), except for the construct 'animacy', were used to test and evaluate AEE. The construct 'animacy' focuses on how lifelike a robot appears to be, which was not the main interest of this research. Therefore, as several questionnaires were required for this research and it was aimed to exhaust participants as little as possible, it was decided to not include the construct in this research. Questionnaire items were presented on a 5-point semantic differential scale presenting two opposing words for each item (e.g. unintelligent - intelligent). Participants were asked to use this scale to indicate to what extent the item applied to the robot's displayed behaviour.

3.4.1.4 Almere Model of Acceptance

Most constructs of the Almere Model (Heerink et al., 2010) were used to test and evaluate AEE. Excluded constructs were facilitating conditions, intention to use and perceived adaptability, as these constructs were not representative for the experiment conditions. For example, participants were not provided with the opportunity to use the robot outside of experiment sessions. Therefore, these constructs were not included in this research. One may argue that excluding constructs from this questionnaire is a mistake as they form a model to predict usage of a technology. However, as the focus of this research was to evaluate AEE and not predict usage of the technology, it was deemed acceptable to exclude

these constructs. Questionnaire items were presented on a 5-point scale from 'strongly disagree' to 'strongly agree'. The original model uses a 7-point scale, but to preserve consistency it was decided to present all questionnaires with 5-point scales.

3.4.1.5 PANAS: Positive And Negative Affect Schedule

Participants' level of affect was measured through the Positive And Negative Affect Schedule (PANAS; Watson, Clark, and Tellegen, 1988), and has been used in HRI research before (e.g. Rosenthal-von der Pütten et al., 2013). This self-report scale measures independent levels of positive affect and negative affect, and consists of 10 positive and 10 negative emotion words. Participants rated to what extent each word described their emotions during the past week on a 5-point scale from not at all (1) to very much (5). The minimum score for both positive and negative explicit affect was 10, the maximum score 50. High numbers indicate a high level of affect (either positive or negative). This questionnaire was used to evaluate whether participants' mood was impacted by AEE, considering that an increase in negative affect or decrease of positive affect is an unwanted outcome of human-robot interactions.

3.4.1.6 IPANAT: Implicit Positive and Negative Affect Test

As the PANAS can be susceptible to social desirability, the Implicit Positive and Negative Affect Test (IPANAT) was used to determine implicit affect (Quirin, Kazén, and Kuhl, 2009). This questionnaire presented participants with six non-existing words (e.g. SUKOV, TALEP) for which participants had to rate whether the words sounded happy, helpless, energetic, tense, cheerful and/or inhibited to them. The minimum score for implicit affect was 1, the maximum score 4. This was the only questionnaire where a 4-point scale was used instead of a 5-point scale. It was important that participants had to pick a positive or negative option as the ambiguous nature of the questionnaire would likely have resulted in participants always choosing the middle option.

3.4.1.7 MOCA: Montreal Cognitive Assessment

If older adults are recruited as participants, there is always a possibility that the cognitive functioning of the participant has deteriorated, which may influence their responses to the robot. As this research aims to evaluate a 'baseline' impact of AEE, it was important

that participants of the longitudinal field study were cognitively healthy. Therefore, the Montreal Cognitive Assessment was used to measure performance in executive functioning, memory, language, attention and visuo-spatial perceptual skills (Nasreddine et al., 2005). This assessment was only used for some of the participants, as a locksmith was available for the other participants, who could determine participants' cognitive functioning and the use of MOCA was not needed. Tasks included a visuospatial and attention assessment among other things, as presented in Appendix A7. If participants scored less than 26 points, they were excluded from the study.

3.4.1.8 Attachment Style

Attachment takes into account whether people are positive or negative about themselves, and about others (Bowlby, 1958). This leads to a 2x2 (avoidance x anxiety) attachment model that addresses four possible attachment styles (Bartholomew and Horowitz, 1991). These styles are secure, preoccupied, dismissive and fearful. People with a fearful attachment style are unwilling to allow themselves to be vulnerable to their partner, and often worry about the attention they receive (Brennan, Clark, and Shaver, 1998). These people often tend to become attached to objects (Neave et al., 2016). Participants' attachment style was gathered to investigate whether this influenced participants' level of attachment to the robot. In order to assess attachment types, participants of the pilot study that took place before the field study were asked to fill in the Adult Attachment Scale (AAS; Collins and Read, 1990). However, it was found that this scale provides only three different levels of attachment, where the literature distinguishes between four attachment styles (Bartholomew and Horowitz, 1991). Therefore, the Experiences in Close Relationships Inventory was used for the field study with older adults (Brennan, Clark, and Shaver, 1998). Statements involving '(romantic) partners' were adapted to a more general variation of 'people that are dear to me' to prevent potentially upsetting participants if a (romantic) partner was no longer with them. The four attachment styles are secure, fearful, preoccupied and dismissive, of which participants with a fearful attachment style may potentially be more prone to becoming attached to or dependent on the robot(Brennan, Clark, and Shaver, 1998). The questionnaire had statements for which participants had to indicate to what extent it applied to them, on a 5-point scale from 'strongly disagree' (1) to 'strongly agree' (5).

3.4.1.9 Object Attachment

The object attachment questionnaire used in this research was adapted from an existing consumer attachment questionnaire (Schifferstein and Zwartkruis-Pelgrim, 2008), which was recommended by a researcher in consumer psychology. Adaptations entailed changing 'this product' from the original questionnaire to 'Pepper'. For example the statement 'I have a bond with this product' was adapted to 'I have a bond with Pepper'. Participants were asked to rate to what extent certain statements applied to them on a 5-point scale from not at all (1) to very much (5). A question on the intention to use the robot if they had the opportunity was added to this questionnaire, as this can be an indicator of attachment to the object (Wilson, 2017).

3.4.1.10 Interview

At the debrief session, after all questionnaires were filled in once more, participants were debriefed and asked several questions in a face-to-face interview. Notes were taken of the participants' responses to the questions asked during this interview. This interview is described in more detail in Section 5.3.3.1.

An overview of when each questionnaire was used is provided in Table 3.1.

	Test AEE online	**Test AEE face-to-face**	**Pilot for field study**	**Longitudinal field study**	**Online evaluation of findings**
Demographics	x	x	x	x	x
Emotion scale	x	x			
Emotion questions			x	x	
Godspeed		x		x	
Almere		x		x	
PANAS		x		x	
IPANAT		x		x	
Attachment style			x	x	
Attachment			x	x	
MOCA				x	
Interview				x	

TABLE 3.1: Overview of questionnaires used in this research, and indication of when they were used.

3.4.2 Other Measures

Additional measures were used for the longitudinal field study with older adults to evaluate the impact of AEE. This was deemed essential as questionnaires are less appropriate

for investigating the impact of new innovations, and it can be challenging to measure user experiences and potential consequences (De Graaf, 2016). Data gathered include physiological data, speech prosody data, and behavioural data. Physiological data has shown an increase in arousal when participants were confronted with a situation where a robot was tortured compared to when they were confronted with a pleasant situation (Rosenthal-von der Pütten et al., 2013). Speech prosody data has been used as an indicator for increased arousal as well, where it was found that both pleasant and unpleasant conditions resulted in increased arousal compared to neutral conditions (Cohen et al., 2009). Research has indicated that nonverbal behaviour communicates our impressions more clearly than words do (Bente et al., 2008). Therefore, participants' behaviour was recorded to investigate whether they showed levels of discomfort. As they were only used for the experiment that investigated the impact of AEE on participants in the longitudinal field study with older adults, these measures are explained in more detail in Chapter 5, in Section 5.3.2.

3.5 Data Processing and Analyses

Data from questionnaires and physiological data was pre-processed in Excel and then transferred to SPSS for analyses. It should be noted that although questionnaires were analysed on a 5-point scale, participants were given the options 'strongly disagree' to 'strongly agree', and not the numbers. This was decided to ensure participants would pick the option they felt most comfortable with, and not be biased by the number representing an option. Physiological data was pre-processed in Excel and analysed using Kubios (Tarvainen et al., 2014). Speech prosody data was analysed using the software Praat (Boersma, 2001). Video-recordings were coded into behaviour schemes in Excel. During face-to-face interviews with older adults, the answers were written down by the experimenter to invite participants to elaborate more in their answers. These replies were pre-processed in Excel and analysed using SPSS.

Deciding on what analysis to use was not always easy. Many questions used Likert scales, and there is a debate in the literature as to whether parametric or non-parametric tests should be used for analysis, as there are arguments for both (e.g. Mircioiu and Atkinson, 2017 for parametric, Schrum et al., 2020 for non-parametric). Therefore, both parametric

and non-parametric tests were conducted for the online evaluation of the implemented behaviours. As there were no critical differences between the results of parametric and non-parametric analyses, it was decided to use parametric analyses for this research. Also, a senior statistician was consulted for the final survey that investigated the opinion of the general public who also suggested the use of parametric tests (Derrick and White, 2017). The results of the non-parametric analyses for the online evaluation can be found in Appendix C.

3.6 Ethical Approval

Ethical approval was gathered for all studies conducted during this research. The initial survey testing the behaviours and its accompanying face-to-face study were both approved under reference number UREC.17.09.05. The field study with older adults was approved under reference number FET.18.02.030, and the pilot study with university staff was approved as an amendment to this field study. The final survey reviewing the findings with the general public was approved under reference number FET.19.12.023.

Informed consent was gathered before each experiment and participants were debriefed on the aim of the whole research and how the study they participated in fit in the whole research at the end of each experiment.

3.7 Summary

This chapter provided an overview of the studies conducted in this research, and explained how these user studies addressed certain challenges of this work. Table 3.2 provides an overview of the studies conducted in this research. The next chapters will present the surveys and user studies and their findings in more detail, followed by the development of the framework. First, Chapter 4 will describe in more detail how AEE was implemented and tested. Next, Chapter 5 presents the studies that were conducted to evaluate whether AEE could lead to emotional deception and attachment. Chapter 6 then presents the framework that was developed based on the findings from the user studies described in Chapters 4 and 5. It also describes the survey that was conducted to evaluate whether the findings were reflected in the opinion of the general public, and how the framework

was updated accordingly. Finally, Chapter 7 addresses the main research questions of this research, as well as its limitations and potential future work.

Study purpose	**Type of study**	**N**	**Duration**	**Measures**
Test AEE online	Online survey	161	10 minutes	Questionnaires
Test AEE during physical interactions	Lab-based study	48	20 minutes	Questionnaires
Prepare evaluation of AEE	Lab-based study	17	20 minutes	Questionnaires Physiological data
Evaluate AEE through longitudinal field study with older adults	Field study	14	8 x 20 minutes & 2 x 30 minutes	Questionnaires Interview Physiological daata Speech prosody data Behavioural data
Validate findings from previous studies	Online survey	239	10 minutes	Questionnaires

TABLE 3.2: Overview of the studies conducted in this research to develop a framework on AEE.

4 Designing Artificial Expression of Emotion

This research aimed to investigate whether artificial expression of emotion by a social robot raises ethical concerns, focusing on the impact that AEE may have on older adults. This was investigated, as it is important to understand ethical concerns prior to using social robots, especially when they interact with vulnerable populations. First, it was explored how robot emotions can be designed and implemented. Next, it was tested whether these behaviours were perceived as intended, which is the research question investigated in this chapter: 'Is artificial expression of emotion as implemented in this research perceived as designed?'. The chapter starts with a description of existing work on how robots can artificially express emotions. Once implemented, the behaviours were tested through an online survey to ensure a large participant group with varied demographics was reached. However, as the literature has shown that participants can perceive virtual agents differently from embodied robots (e.g. Vasco et al., 2019, the implemented robot behaviours were tested through physical face-to-face interactions as well. This chapter provides a more detailed description of the work that was presented at the International Conference on Robot Ethics and Standards in Troy, NY in August 2018 (Van Maris et al., 2018).

4.1 Existing Work on Implementing Expression of Artificial Emotions in Social Robots

There are several modalities that can be used for artificial expression of emotion by a robot. Factors used by several researchers to implement artificial expression of emotion include emotional valence and arousal (Beck et al., 2013; Thimmesch-Gill, Harder, and Koutstaal, 2017; Saerbeck and Bartneck, 2010). Emotional valence specifies whether the quality of a stimulus is positive or negative. Emotional arousal indicates the level of energy of the emotional expression. Example modalities to differentiate between emotions

include the robot's body language, head position and speech prosody. Several researchers (Erden, 2013; Häring, Bee, and André, 2011) that investigated artificial expression of emotion through body language, based their behaviours on Coulson's work (Coulson, 2004). In this work, descriptions of expression of emotion through body posture were used on computer-generated mannequin figures to investigate the attribution of emotion to static body postures. Besides static body postures, emotions can also be expressed through movement (Beck et al., 2013; Saerbeck and Bartneck, 2010; Häring, Bee, and André, 2011). It was found that movement speed and curvature influenced participants' perception of emotion in the robot. Acceleration of movement affected arousal, where acceleration positively correlated with perceived arousal. Furthermore, acceleration and curvature of the movements impacted valence, both on the positive and negative dimensions (Saerbeck and Bartneck, 2010).

Proxemics is a third factor that can be used for artificial expression of emotion (Beck et al., 2013). However, changing the distance between the robot and participant during an interaction will likely influence participants' focus on the robot's behaviour. Therefore, it was decided that proxemics would not change during interactions and the distance between the participant and the robot remained the same.

Changing eye-colour to artificially display emotion was investigated as well, where the LEDs around the eyes of a NAO robot would change colour according to the emotion it was displaying (Häring, Bee, and André, 2011). This was found to be an unreliable feature, as lighting conditions impacted the brightness of the colours used, resulting in emotions being perceived differently than intended. Furthermore, it was found that human-like emotive non-verbal behaviour by a robot had greater impact on its participants than robot-specific emotive non-verbal behaviour (Rosenthal-von der Pütten, Krämer, and Herrmann, 2018), and changing eye-colour through LEDs is a robot-specific feature that is not human-like. Therefore, it was decided to not use eye colour as a means to express artificial emotions.

Facial expressions can be used to display artificial emotions as well. However, some robots have set facial expressions that cannot be changed. Also, it was found that facial expressions may be difficult to perceive (Thimmesch-Gill, Harder, and Koutstaal, 2017). The combination of these two factors resulted in facial expressions not being used to display emotions in this research.

Head position was found to influence arousal and valence (Beck et al., 2013). Tilting the head upwards led to an increase in arousal and valence, where tilting the head downwards resulted in a decrease of arousal and valence. In other research where artificial expression of emotion was correctly recognised, the head was tilted upwards for positive emotions and downwards for negative emotions (e.g. Erden, 2013; McColl and Nejat, 2014).

Furthermore, vocal prosody can be used to convey emotion in social robots as well (Crumpton and Bethel, 2016). Pitch, duration and intensity are features that are found to often correlation with emotion recognition during human-human interactions (Vinciarelli et al., 2008). These features have also been used in speech synthesis of technological devices with positive outcomes such as paying more attention to the road while driving (Nass et al., 2005) or increased learning gains (D'mello and Graesser, 2013). Vocal prosody research in social robots entails both research regarding utterances (Read and Belpaeme, 2010) and speech prosody (Niculescu et al., 2013).

4.2 Implementing Emotive Behaviours

Different robot modalities were used to implement happy, sad and neutral robot behaviour, based on existing literature as previously described in Section 4.1. It was decided to categorise positive emotions as 'happy' and negative emotions as 'sad'. It was deemed acceptable to do this, as the emotive behaviours in the studies were not intended to elicit strong emotional responses, but rather indicators of a change in arousal. Also, the context of the interaction could potentially help participants determine whether the emotions were positive or negative. If this proves to be sufficient, it would save time for social robot researchers and developers when creating different robot behaviours. The approach of displaying low levels of arousal was taken as it is important to understand the impact of 'base levels' of AEE, as we need to be sensitive to the population that is interacting with the robot and their reactions to low and high levels of arousal. An overview of all modalities that were used to implement the different emotive behaviours can be found in Table 4.1.

One modality that differed for each emotive behaviour is the robot's head position. The chin would point upwards for happy behaviour, point forwards for neutral behaviour and downwards for sad behaviour. This characteristic was chosen to display different emotions as this had been determined as a useful characteristic before (e.g. Johnson and Cuijpers,

	Head position	Arm position key posture	Trunk position key posture	Movement dynamics	Voice pitch	Talking speed
Happy	Slightly upwards	One/both arms far outwards	Slightly tilted backwards	High	High	High
Neutral	Forwards	One/both arms slightly outwards	Original position	Normal	Normal	Normal
Sad	Slightly downwards	One/both arms close to body	Slightly tilted forwards	Low	Low	Low

TABLE 4.1: Modalities that were adjusted for the implementation of the different emotive behaviours.

2019). The pitch of the robot's voice was different for each behaviour; it was higher for happy behaviour and lower for sad behaviour with the pitch for neutral behaviour being the average of the pitches for happy and sad behaviour. The same strategy was used for the robot's talking speed, where it would talk faster when showing happy behaviour and slower when showing sad behaviour, neutral behaviour once again being the average of the speeds for happy and sad behaviour. The pitch and volume were manipulated using Choregraphe (introduced in Section 5.3.2). The software provides the option to make the robot say anything using so-called 'say'-boxes as shown in Figure 4.1. There are also parameters that shape the robot's voice as well as speed of talking. I used these parameters to distinguish between happy, sad and neutral speech characteristics. The default settings of such say-box are shown in Figure 4.2.

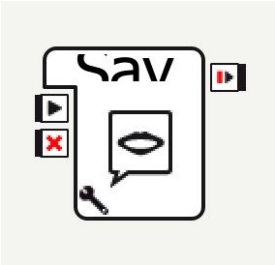


FIGURE 4.1: 'Say'-box in Choregraphe.

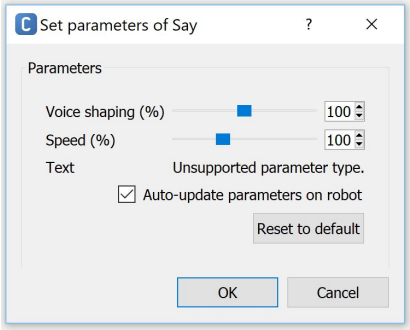


FIGURE 4.2: Default settings of 'Say'-box in Choregraphe.

The range for voice shaping runs from 50 to 150 and the range for speed runs from 50 to 200 in %, with the default value being 100 for both parameters. Taking into account that older adults would be interacting with the robot in a later study who may have hearing aids or become overwhelmed by the experience, I decided to set the parameters for the 'neutral' behaviour a bit lower than the default setting provided by Choregraphe so the robot's voice pitch would be a bit lower and the speed of talking a bit slower, giving the robot a calmer appearance. In the end, after testing the parameters with a dozen of lay users, I decided on the parameter settings shown in Table 4.2.

	Happy	**Neutral**	**Sad**
Voice shaping (%)	120	90	65
Speed (%)	80	75	65

TABLE 4.2: Parameter settings of the robot's voice for each behaviour.

Lastly, the robot's body motions (moving arms and changing head position) were more elaborate for happy behaviour and more compressed for sad behaviour, with gestures for neutral behaviour being the average of the other two. Choregraphe provides a library with pre-determined body motions. This library was explored for existing behaviours that fit the requirements for sad, neutral and happy behaviour as described above. It has been used to display emotions in Pepper before (Marmpena, Lim, and Dahl, 2018). Besides these predefined movements, Choregraphe's library also provides animated 'say'-boxes. These boxes combine a movement with text, where the text parameters can be manipulated similarly to a normal 'say'-box. They were also used to implement AEE in this research. An overview of the predefined behaviours used to implement AEE can be found in Appendix B. Snippets of happy, neutral and sad behaviour can be found in Figure 4.3.

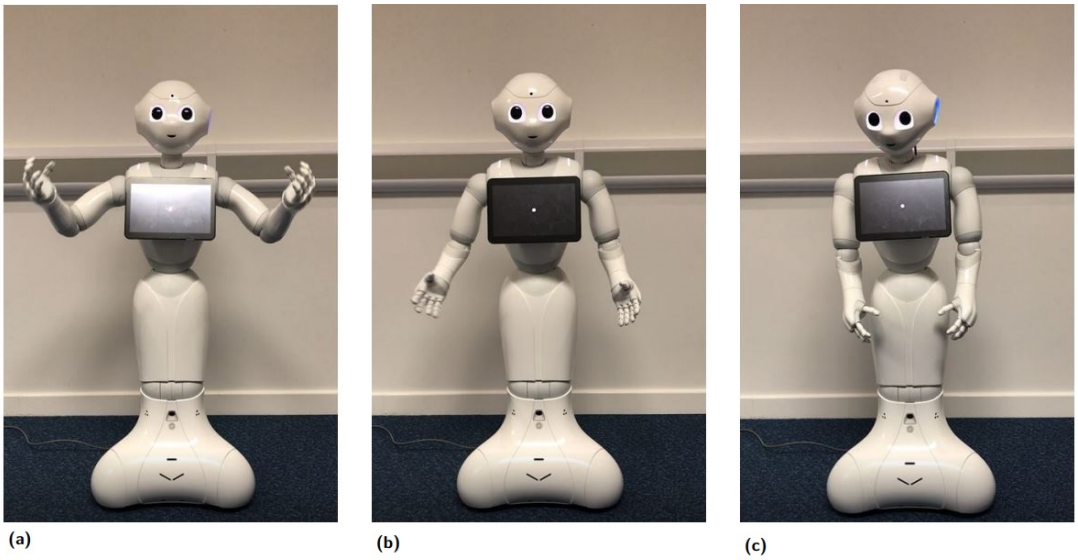


FIGURE 4.3: Snippets of happy (a), neutral (b) and sad (c) robot behaviours.

4.3 Online Evaluation of Emotive Behaviours

The first step of developing a framework addressing AEE by social robots contained online evaluation of the implemented robot behaviours, as can be seen in Figure 4.4. An online survey was conducted to test whether the different robot behaviours were perceived as intended. Participants were asked to rate the robot's behaviours on a scale from happy to sad, and to ensure they were rating the robot's behaviour and not the content of what it was saying, I searched for a sentence the robot could say to accompany the behaviour with little emotional value to it. I decided that facts about nature or the robot itself could provide suitable sentences for this survey as many of these facts are objective and therefore their content may have a lower chance of triggering an emotional response from the participant. The following options were considered more thoroughly:

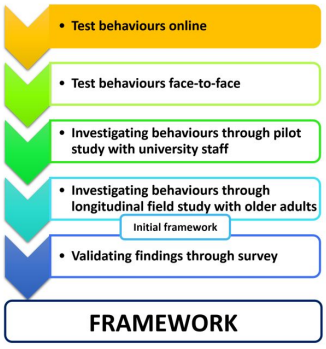


FIGURE 4.4: Current stage of framework development.

- A bear has 42 teeth.
- The country Brazil is named after a tree.
- I am 1.2m tall.
- I weigh 28 kg.

Eventually I decided to use the sentence *'The country Brazil is named after a tree'* for the survey, as hearing about a bear's teeth may evoke an emotional response since bears are predators, and facts about the robot itself may indicate some level of consciousness in the robot. This was undesirable as this may lead to participants potentially being deceived which is the main focus of this research and therefore any unintentional deception from the studies should be prevented where possible.

The three behaviours were displayed in the survey as illustrated in Table 4.1: for happy behaviour, the robot's head was tilted upwards, the robot would stretch its arms sideways and would speak in a fast speed and a high pitch saying *'The country Brazil is named after a tree'*. For neutral behaviour, the head was pointed forwards, the arms would stretch sideways but not as far as for the happy behaviour, and speed of speech and pitch of voice

were slightly less fast and high. Finally for sad behaviour, the robot's head was tilted slightly downwards, there would be a small sideways stretch of the arms and the speed of speech and pitch of the voice were slow and low. These behaviours are shown in Figure 4.3.

4.3.1 Participants

Participants were gathered through distributing the link to the survey to colleagues and family via email, asking them to participate and distribute the survey further to other colleagues and family, and ensuring them to do this was completely voluntary. In total 161 participants completed the experiment and consented to their data being used for analysis. Out of 161 participants, 98 identified as male, 61 identified as female and 1 participant preferred not to say. Their age ranged from 18 to 74 years old. Due to the question of age being presented as different age ranges where the participant had to select what range applied to them, it was not possible to acquire mean and standard deviations for age. The age distribution per age range is displayed in Table 4.3. Familiarity with social robots was low (M = 1.88, SD = 1.09 on a 5-point scale from not familiar at all to extremely familiar). Lastly, the cultural background of participants was not diverse enough to investigate further, with 78% of participants being either Dutch or British.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**Total**
Number of participants in age range	18	43	20	28	45	7	161
Percentage of all participants per age range	11	27	12	17	28	7	100

TABLE 4.3: Distribution of participants that completed the online survey based on age range.

4.3.2 Materials

The software used for implementing the emotive robot behaviours is Choregraphe, as discussed in Section 4.2. The survey was produced using the online software Qualtrics. A link to this survey was distributed to colleagues, friends and family via email with the request to fill in the survey and distribute it further.

4.3.3 Measures

Participants were asked to provide demographic information (age, gender, cultural background and familiarity with social robots). Their perception of the robot's emotional state was measured through a scale from sad to happy (See Figure 3.3). This scale was inspired by the 'affective slider' that was introduced as measurement of human emotions (Betella and Verschure, 2016) and used in other HRI research as well (Marmpena, Lim, and Dahl, 2018). No other measurements were taken during this survey.

FIGURE 4.5: Five-point scale from sad (1) to happy (5) which was presented to participants to rate the robot's emotional state.

4.3.4 Procedure

At the start of the survey, participants were asked to give consent to use their data. It was explained to them that their data would be anonymised immediately, and therefore it would not be possible to withdraw their data from the study once they completed it. To account for this, participants were asked once more at the end of the survey whether they consented to their data being used for analysis.

Participants were shown nine video fragments that each lasted for 3 to 4 seconds. All three behaviours (happy/neutral/sad) were shown three times to each participant. For each fragment, they had to rate the robot's emotive behaviour on a 5-point scale from sad to happy as shown in Figure 4.5 (1 = sad and 5 = happy). Showing behaviours less than three times could result in participants rating the intended emotion by accident rather than through recognising the behaviour, and showing fragments more often could lead to participants learning the differences between the behaviours, and basing their ratings on these differences instead of how they perceived the behaviour. The first video-fragment always showed neutral behaviour; the behaviour of the other eight fragments was randomly ordered.

At the end of the survey, participants were asked once more whether they were happy

to share their data, as the data was anonymised immediately and participants could not withdraw later on. If they selected 'no', their data would be deleted.

4.3.5 Results

As mentioned in Chapter 3, gathered data of this survey was analysed using both parametric and non-parametric tests, before deciding parametric tests would be used for the remainder of this research. The non-parametric tests results of this survey can be found in Appendix C.

First, the average rating for each behaviour (happy, sad, neutral) was calculated, where a low rating would indicate the behaviour was perceived as sad and a high rating would indicate the behaviour was perceived as happy. As this study used a within-subject design, a repeated measures ANOVA was conducted to determine whether each behaviour was rated differently. As sphericity was breached ($\chi^2(2) = 33.03$, $p < 0.01$), a repeated measures ANOVA with Greenhouse-Geisser correction showed that the emotional state of the robot was rated significantly different for each implemented behaviour ($F(1.68, 269.46) = 287.47$, $p < 0.01$, $\eta_p^2 = 0.64$). Post hoc tests showed that happy behaviour was rated significantly higher than neutral and sad behaviour ($p < 0.01$ for both), and neutral behaviour was rated significantly higher than sad behaviour ($p < 0.01$). More details can be found in Figure 4.6.

4.3.5.1 Age

The effect of age was investigated as this research aims to investigate the impact of AEE on older adults. A Pearson product-moment correlation was conducted to determine the relationship between participants' age and their rating of the implemented robot behaviours. A strong negative relationship was found between age and ratings for both happy and neutral behaviour ($r(161) = -0.38$, $p < 0.01$ and $r(161) = -0.33$, $p < 0.01$ respectively). Furthermore, a positive correlation between age and rating for the sad behaviour was found ($r(161) = 0.17$, $p = 0.03$). A one-way MANOVA was conducted to investigate the impact of age on participants' ratings of the emotive robot behaviours, which showed that age significantly impacted participants' ratings ($F(5, 155) = 2.83$, $p = 0.02$, $\eta_p^2 = 0.08$). Also, a weak interaction effect was found between participants' ratings of the implemented behaviours and their age ($F(8.65, 268.01) = 1.58$, $p < 0.01$, $\eta_p^2 = 0.08$). Post hoc pairwise

FIGURE 4.6: **p $<$ 0.01; Average ratings of happy, neutral and sad robot behaviour from sad (1) to happy (5).

comparisons were performed to determine the impact of each age group. Significant differences between age groups can be found in Figure 4.7. Exact values can be found in Appendix C. This figure shows that different age groups perceived happy and neutral behaviour significantly differently, but there was no significant difference in rating for sad behaviour based on age.

4.3.5.2 Gender

Independent samples t-tests were conducted to investigate whether participants' gender impacted their rating of the different emotive robot behaviours. Gender did not significantly impact the rating of the happy behaviour ($t(157)$ = -0.93, p = 0.35), nor did it impact the rating of neutral behaviour ($t(157)$ = -1.00, p = 0.32) or sad behaviour ($t(157)$ = -0.29, p = 0.77).

4.3.5.3 Familiarity with social robots

One-way MANOVA was conducted to determine whether participants' familiarity with social robots impacted their ratings for the implemented behaviours. The results showed that participants' perception of happy and sad robot behaviours significantly differed depending on their familiarity with social robots ($F(4, 156) = 4.06$, $p < 0.01$, $\eta_p^2 = 0.09$

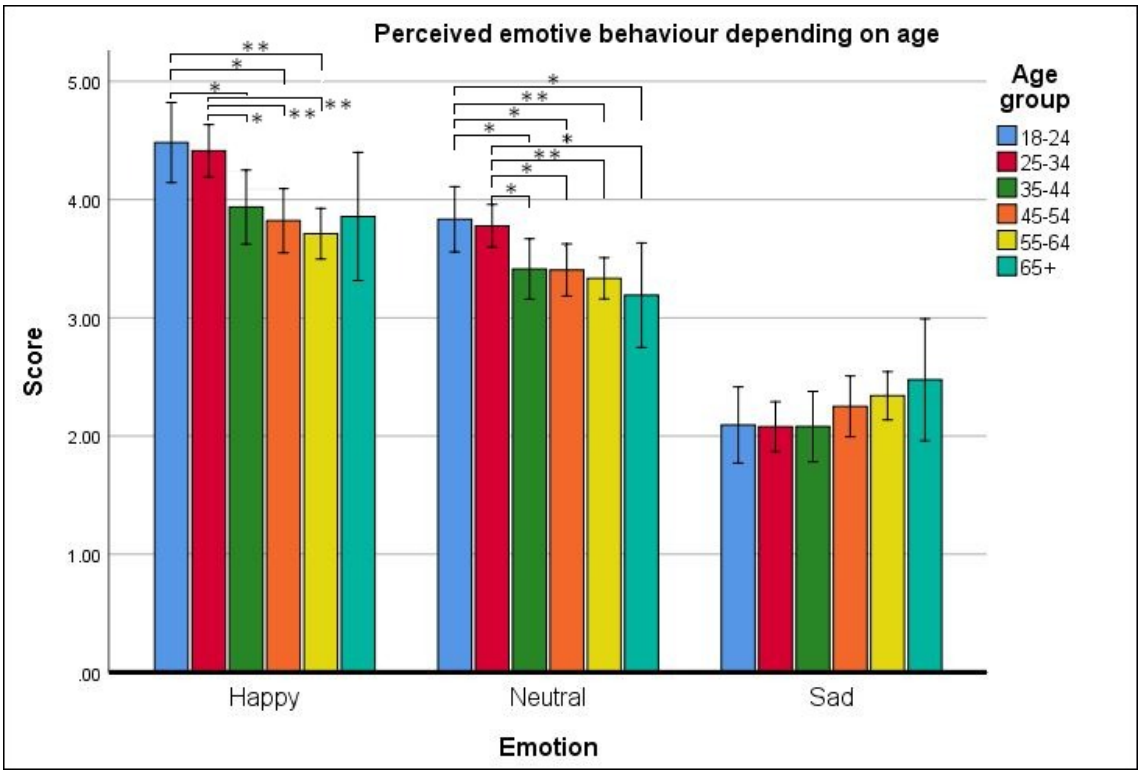


FIGURE 4.7: ** $p < 0.01$, *$p < 0.05$; Average ratings of happy, neutral and sad robot behaviour for each participant age group on a scale from sad (1) to happy (5).

and $F(4, 156) = 2.67$, $p = 0.04$, $\eta_p^2 = 0.06$ respectively). This difference was not significant for neutral robot behaviour ($F(4, 156) = 2.16$, $p = 0.08$, $\eta_p^2 = 0.05$). Post hoc pairwise comparisons were conducted to investigate where the differences occurred, which are presented in Figure 4.8. Exact values are provided in Appendix C.

4.3.6 Discussion

The results from this online survey show that the implemented emotive behaviours are perceived by the general public as intended. It was found that people from different age groups perceive the robot differently. Negative correlations were found for happy and neutral behaviour and participants' age. Furthermore, a positive correlation was found for sad behaviour and participants' age. These findings are strengthened by the finding that age significantly impacted participants' ratings of happy and neutral behaviour. This indicates that participants rated the emotive behaviours less extreme as their age increased, as illustrated in Figure 4.7. This raises the question why this finding occurred. It is possible that younger participants are more familiar with new technologies which reduces the novelty of interacting with a machine, which may potentially lead to them perceiving the

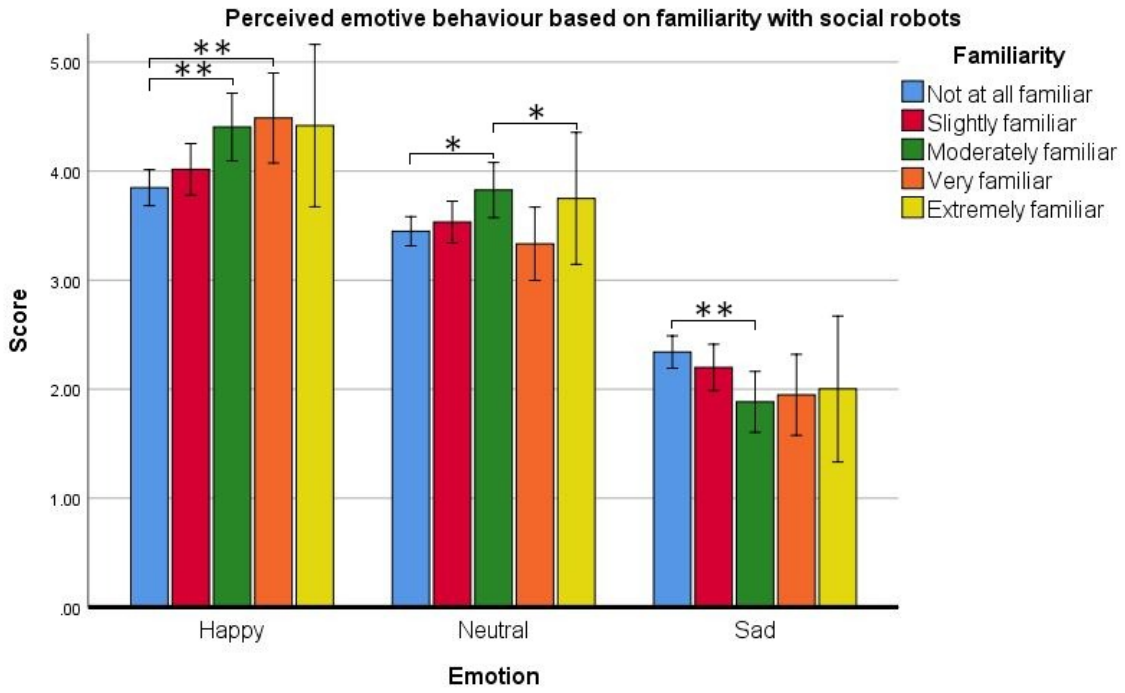


FIGURE 4.8: $^{**}p < 0.01$, $^{*}p < 0.05$; Average ratings of happy, neutral and sad robot behaviour based on participants' familiarity with social robots on a scale from sad (1) to happy (5).

robot as more emotive where older participants may perceive the robot more as a machine still. Nonetheless, as the intended target audience for the longitudinal field study is older adults, this is something to take into account during further research.

Even though the findings support that implemented behaviours are perceived as intended, it needs to be established whether the emotive behaviours are still perceived as intended during face-to-face interactions, as several studies have shown that participants perceive virtual agents on a screen differently from physically embodied robots that are in the same room as them. Following this, it can be investigated whether emotional deception and/or emotional attachment occur when a robot displays AEE during face-to-face human-robot interactions. Therefore, the next study describes a study where it is determined whether the perception of the implemented behaviours is perceived as intended during face-to-face interactions.

4.4 Face-to-face Evaluation of AEE

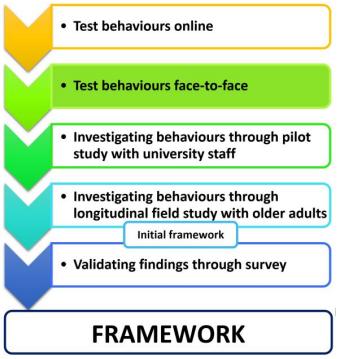


FIGURE 4.9: Current stage of framework development.

It was expected that recruiting older adults would be challenging, as older adults may be less keen to try new technologies (Wu et al., 2014). Furthermore, care facilities needed to be approached and be willing to provide means to conduct an experiment. Therefore, it was decided to only recruit older adults for the longitudinal field study that would be conducted at a later stage, and recruit participants of a different age group for this physical study where the perception of the implemented emotive robot behaviours was tested in more detail. Even though the results from the online survey showed differences between age groups, the overall result showed that the emotive behaviours were rated significantly differently for each age group. Therefore, I deemed it acceptable to use participants of a different age group for this study and keep interested older adults for the longitudinal field study that would be conducted later in the research. The goal of this face-to-face study was to investigate whether the emotive behaviours are still perceived differently when the robot is physically present in the room instead of virtually displayed on a screen. To investigate this, participants would interact with a robot that initially showed neutral behaviour so participants had something to build their mental model of the robot on, followed by a story that the robot told either displaying happy, neutral or sad behaviour.

4.4.1 Participants

In total 48 Psychology students (32 female, 16 male; age $M = 21$, $SD = 3$) from the university participated in this study. They were recruited through the university's Psychology Participant Pool and received one Psychology participation point for taking part in the experiment. Familiarity with social robots was low ($M = 1.19$, $SD = 0.45$ on a 5-point scale from not at all familiar to extremely familiar). There were four conditions, with 12 participants (8 female, 4 male) for each condition. The robot displayed either happy behaviour while telling a cheerful story, neutral behaviour while telling a cheerful story, neutral behaviour while telling a somber story or sad behaviour while telling a somber

story.

4.4.2 Materials

As mentioned before the robot that was used for this whole research is Pepper. Qualtrics was used for the questionnaires from this study, which participants had to fill in using a tablet that was given to them. SPSS version 24 was used to analyse the data gathered. The cheerful and somber polar bear stories were taken from earlier work with an undergraduate student where we used the behaviours I implemented to investigate what user characteristics influenced user acceptance (Bishop et al., 2019).

4.4.3 Measures

Demographics data (age, gender, familiarity with social robots) was gathered at the start of the experiment. Participants' mood was measured at the start and end of the interaction to investigate whether their mood was influenced by the interaction and potentially the different robot behaviours. Mood was measured as this may influence participants' perception and acceptance of a robot (Baisch, Kolling, and Knopf, 2017), and earlier work that I had done had indicated that participant mood may influence robot acceptance (Bishop et al., 2019). Also, if mood was negatively impacted due to interactions with the robot, this would be a negative ethical consequence of such interactions. Mood was measured both explicitly (PANAS, see Section 3.4, Watson, Clark, and Tellegen, 1988) and implicitly (IPANAT, Quirin, Kazén, and Kuhl, 2009) to account for participants potentially providing answers they felt were socially acceptable with respect to how they were feeling. Constructs from the Godspeed questionnaire (Bartneck et al., 2009) and Almere questionnaire (Heerink et al., 2010) were used to determine whether deception potentially occurred. These questionnaires were given after the interaction. Finally, the question from the online survey regarding the robot's emotional state (see Figure 3.3) was asked after the interaction as well. More detailed descriptions of these questionnaires are provided in Section 3.4. Participants would initially see the robot introduce itself showing neutral behaviour, after which it would show either happy, neutral behaviour. As this allowed participants to see a change in the robot's behaviour (if there was any) and have a reference point to 'neutral' behaviour, I believed participants would be able to rate the robot emotions as intended, even though the study design was now between-subject and

they would not see all behaviours. By prolonging the introduction of the robot, where it would no only introduce itself but also provide some background information on where it was built, I aimed to ensure participants got used to the robot and would see a change in its behaviour if there was any.

4.4.4 Experimental Setup and Procedure

Participants were located opposite the robot with a table between them and the robot that was used to fill in questionnaires. The robot was located in front of a blue screen, behind which I was located with a laptop to manually operate the robot during the interactions without the participants being able to observe this.

Participants interacted with a robot that displayed either happy, neutral or sad behaviour while telling a story about polar bears. Taking into account the longitudinal field study that would be conducted later on, it was deemed impossible to have several interactions without any emotional context in them, as even topic discussed in a didactic setting may lead to participants responding to the context. Therefore, context was added to this experiment. The robot's behaviour would either be neutral or appropriate for the story context (happy behaviour for cheerful context, sad for somber context).

Before the start of the interaction, participants were asked to fill in questionnaires and to say to the robot 'I am ready' once they completed the questionnaires. This allowed me to know when to manually prompt the robot to start the interaction. Participants were unable to see the robot was operated manually. The robot would first introduce itself and give some information about itself showing neutral behaviour. Once the introduction was finished, it would tell a story about polar bear cubs while showing either happy, sad or neutral behaviour. After the interaction participants were given more questionnaires to fill in, after which they were debriefed about the goal of the study.

4.4.5 Results

First, it was investigated whether the three implemented behaviours were perceived differently. A one-way ANOVA was conducted (happy x neutral x sad) which showed a significant difference in how the implemented behaviours were rated ($F(2, 45) = 3.82$, $p = 0.03$, $\eta_p^2 = 0.15$). Post hoc pairwise comparisons showed that happy behaviour was rated significantly higher than sad behaviour ($p = 0.02$) and neutral behaviour was rated

significantly higher than sad behaviour ($p = 0.02$). However, happy behaviour was not rated significantly higher than neutral behaviour ($p = 0.56$). This can also be found in Figure 4.10.

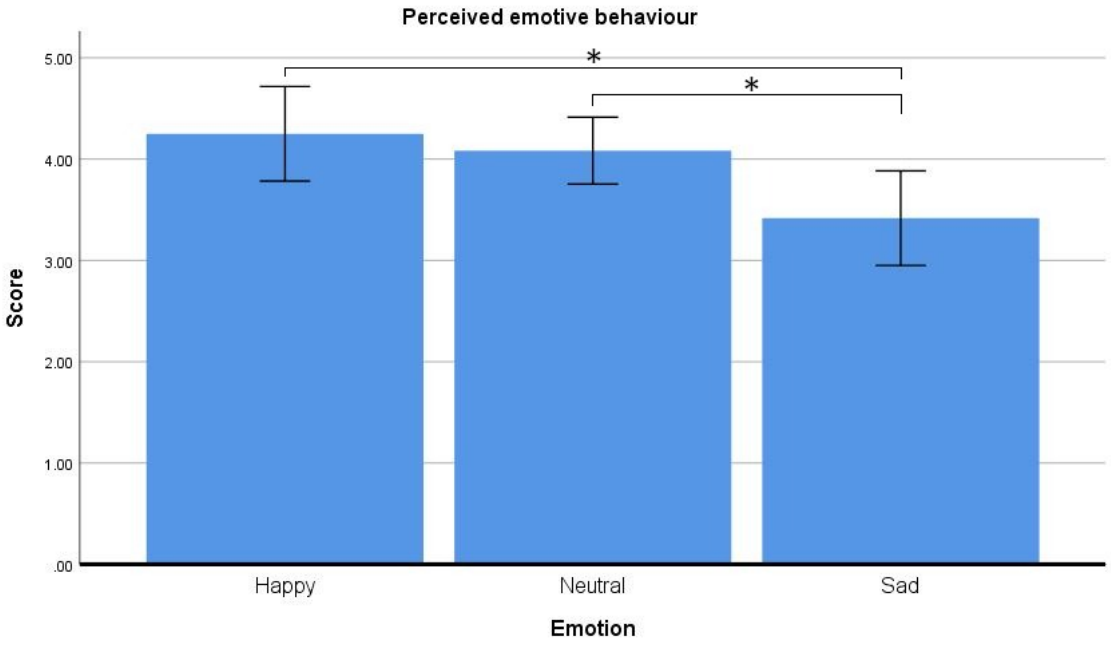


FIGURE 4.10: *$p < 0.05$; Average ratings of happy, neutral and sad robot behaviour on a scale from sad (1) to happy (5).

4.4.5.1 Age

As participants were students of similar age ($M = 21$, $SD = 3$), no analysis was conducted to investigate the impact of participant's age on their perception of the implemented behaviours.

4.4.5.2 Gender

An independent samples t-test was conducted to investigate the impact of gender on the perception of the implemented behaviours. No significant impact of gender was found ($t(46) = 0.12$, $p = 0.91$).

4.4.5.3 Familiarity with Social Robots

As most participants indicated they were not at all familiar with social robots, and only few indicated otherwise, no analysis was conducted to determine the impact of participants'

familiarity on their perception of the implemented behaviours.

4.4.5.4 Story Context

One-way ANOVA was conducted to determine whether story context (cheerful x somber) impacted participants' perception of the robot's emotional state. Results showed a significant impact ($F(1, 44) = 4.58$, $p = 0.04$, $\eta_p^2 = 0.09$) where participants scored the robot's emotional state more positive for the cheerful story ($M = 4.21$, $SD = 0.72$) compared to the somber story ($M = 3.71$, $SD = 0.91$).

Looking at the questionnaires given, and more specifically looking at the questionnaire constructs that may indicate potential occurrence of emotional deception by the robot (anthropomorphism, social presence), one-way ANOVA indicated that these constructs were not significantly impacted by the displayed emotive robot behaviour ($F(2, 45) = 0.57$, $p = 0.57$, $\eta_p^2 = 0.03$ for anthropomorphism, $F(2, 45) = 0.21$, $p = 0.81$, $\eta_p^2 = 0.01$ for social presence), nor did participants' gender impact results of these constructs ($F(1, 46) = 0.08$, $p = 0.77$, $\eta_p^2 = 0.01$ for anthropomorphism, $F(1, 46) = 1.05$, $p = 0.31$, $\eta_p^2 = 0.02$ for social presence). Even though none of the differences are significant, ratings for the sad robot or somber story are always higher for the constructs investigated for the potential occurrence of emotional deception, as can be found in Figure 4.11.

4.4.5.5 Mood

Participants' mood was measured before and after the interaction to investigate whether the interaction and potentially the emotive robot behaviour had an impact on participants' mood. Paired samples t-tests showed that explicit negative mood was significantly lower ($t(47) = 7.01$, $p < 0.001$) after interacting with the robot ($M = 15.27$ out of 50, $SD = 6.82$) than it was at the start of the experiment ($M = 19.44$ out of 50, $SD = 7.67$), as can be seen in Figure 4.12. Although there was a nearly significant decrease in implicit negative effect ($t(47) = 2.00$, $p = 0.052$), no other mood measures significantly changed over time ($t(47) = 0.49$, $p = 0.63$ for explicit positive mood, $t(47) = -1.38$, $p = 0.17$ for implicit positive mood).

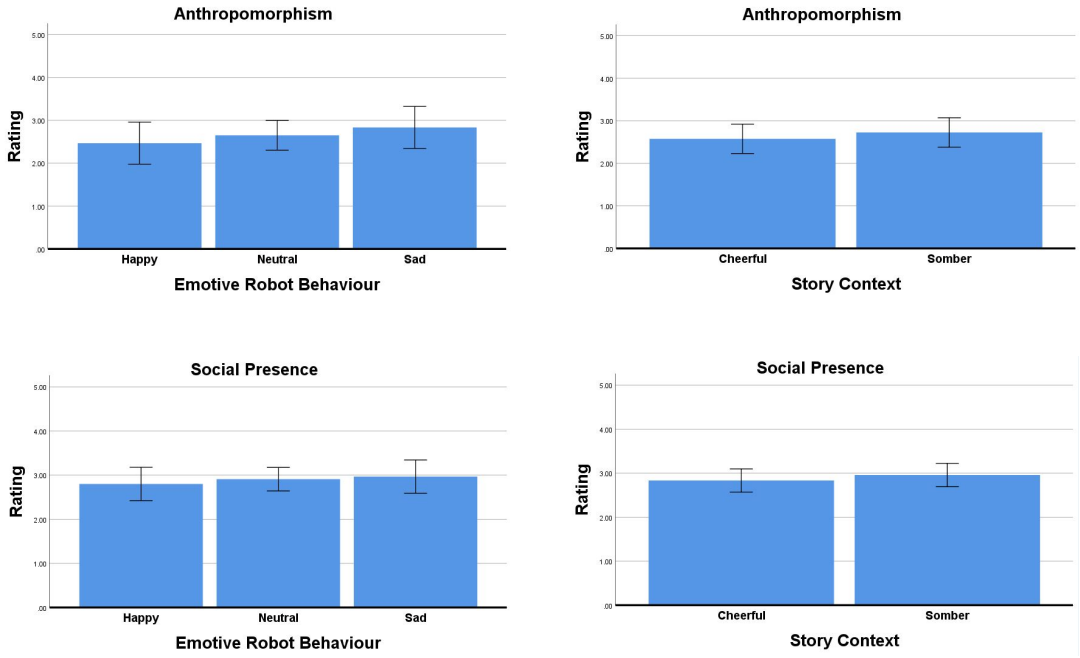


FIGURE 4.11: Average ratings for anthropomorphism and social presence depending on the robot's displayed behaviour (happy, neutral, sad) and story context (cheerful, somber) on a scale from not at all (1) to very much (5).

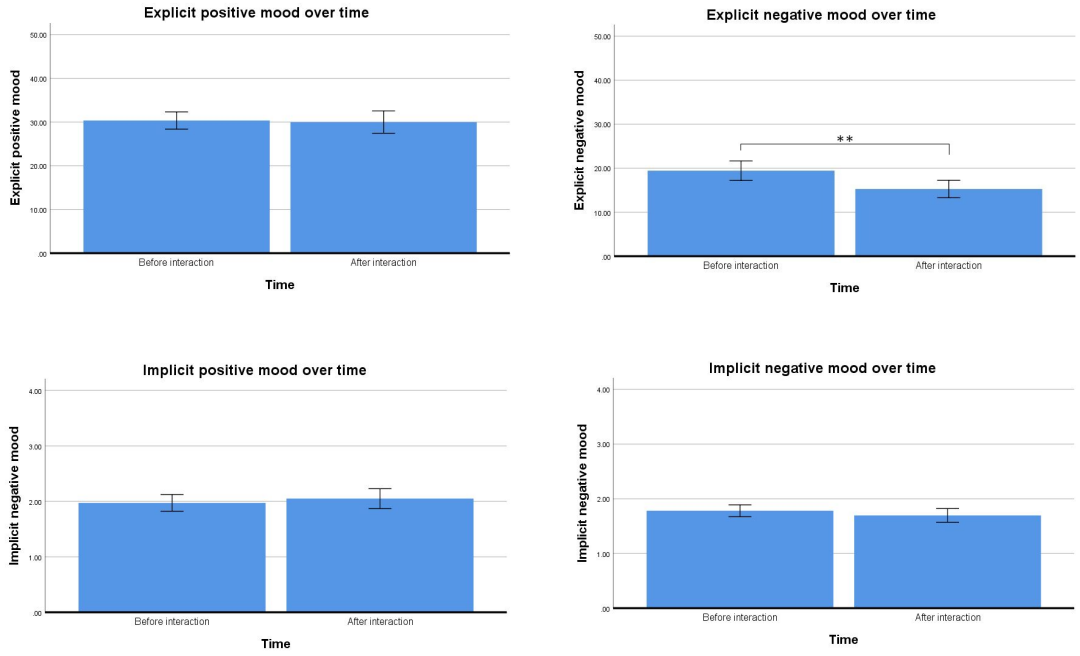


FIGURE 4.12: $*p < 0.05$; Participants' average scores for explicit (max. 50) and implicit (max. 4), positive and negative mood before and after the interaction.

4.4.6 Discussion

The goal of this study was to investigate whether the implemented emotive robot behaviours were perceived as designed during face-to-face interactions. Even though the happy behaviour was still rated highest and therefore most happy and the sad behaviour lowest and therefore most sad, the difference of the ratings between behaviours was not always significant. More specifically, happy behaviour was not rated significantly happier than neutral behaviour. Looking at the averages for these ratings, it appears that participants perceived the neutral behaviour as relatively happy and therefore happier than intended. It should also be noted that even though sad behaviour was rated significantly lower than happy and neutral behaviour, the overall rating for neutral behaviour was higher than intended. However, it should be taken into account that the participants in this user study were all aged between 18 and 24 years old. If we look back to Figure 4.7 from the online survey, it can be found that participants of this age group rated the neutral behaviour as relatively high. Therefore, the narrow age range of the participants in this study may be a reason why happy and neutral behaviour are not rated significantly different. The low number of participants may also have an impact on the findings, and it is also possible that participants did not expect the robot to display sad behaviour. Furthermore, it is possible that participants interpreted the neutral behaviour as happy until they saw the happy behaviour. The aesthetics of the robot may also have an impact as Pepper's mouth is shaped like a smile which may have been more clearly visible during the face-to-face study. Finally, as familiarity was low participants may have been distracted by the novelty of seeing a robot for the first time and therefore not have been able to pay attention to the behaviour the robot was displaying. Nonetheless, the goal was to determine whether the emotive behaviours were interpreted as designed. As participants were able to recognise happy and sad behaviour, it can be concluded that this was true.

4.4.6.1 Story Context

As described in Section 4.4.4, story context was added to the interaction, to investigate whether that influenced participants' perception of the emotional state of the robot. In hindsight I probably would have approached this matter differently. I still believe it was justified to introduce context in this study as it is difficult to find a topic that does not

evoke an emotional response and remain interesting for several sessions. However, more conditions were required for me to be able to properly analyse the impact of story context. I investigated whether context had an impact on how emotions were perceived but did not take into account what the impact of displayed emotion on story context could be. I should have added two more conditions where the robot would tell a neutral story while displaying either happy or sad behaviour. Perhaps the weak effect of story context on participants' perception of the emotional state of the robot where the robot's emotional state was rated happier for the cheerful story than for the sad story was the reason why happy behaviour was no longer rated significantly happier than neutral behaviour. However, as the effect size of story context was quite small ($\eta_p^2 = 0.09$) and the differences between the three emotive behaviours were the same when taking into account story context I deemed it safe to accept that the results were still sufficient for me to continue with the next study.

4.4.6.2 Emotional Deception

Looking at the findings from the questionnaires it appears that there was limited evidence of emotional deception due to the robot's displayed behaviour. However, it is interesting that for all constructs that may indicate occurrence of potential deception ratings for happy behaviour were lowest and sad behaviour were highest as can be seen in Figure 4.11. This may indicate that artificial expression of sad emotion is more likely to result in potential emotional deception than artificial expression of happy emotion. Perhaps this occurs as participants expected the robot to show happy behaviour where sad behaviour was less expected, which may have resulted in them updating their mental model of the robot's abilities. Furthermore, it is possible that sad behaviour evoked a stronger empathetic response than happy behaviour. It is also possible that participants expected the robot required a certain awareness of context for it to be able to respond appropriately to a sad context, which again may lead to potential emotional deception. It is important to consider the impact of this difference in the context of social robots interacting with vulnerable populations. The person interacting with the robot may be less aware of being deceived or respond stronger to the displayed behaviour, which may result in the person trying to console the robot and putting themselves in uncomfortable positions and potentially at risk of harm.

4.4.6.3 Participant Mood

The finding that both implicit and explicit positive mood did not decrease during the interaction with the robot is reassuring, as this indicates that there is no negative impact of the robot's behaviour on participants' mood. It would be problematic if this were the case as that would indicate the behaviour had a negative impact on how participants were feeling, which is not a desirable outcome of any human-robot interaction. The finding that negative mood significantly decreased over time is not completely unexpected as negative mood implies feelings of being nervous and/or tense, which could occur at the start of the experiment, especially as familiarity with robots was low. It is likely these feelings decreased once participants became more familiar with the robot, its behaviour and what was expected of them during the experiment.

As the findings of this study were similar to the findings for participants of the same age group from the online survey where the emotive robot behaviours were perceived as intended I was able to continue with the preparation of the longitudinal field study with older adults which will be presented next in Chapter 5.

4.5 Summary

The goal of this chapter was to implement different emotive robot behaviours and validate whether they were perceived as intended. The research question investigated in this chapter was: 'Is artificial expression of emotion as implemented in this research perceived as designed?'. First, it was determined what robot modalities could be used to implement AEE. Next, they were tested on social robot Pepper. This was first evaluated through an online survey where a more diverse target audience could be reached. As the results of this survey indicated the implemented behaviours were perceived as intended, a physical user study was performed next to investigate whether this also applied to face-to-face interactions compared to participants observing a virtual agent. Even though neutral and happy behaviour were not perceived significantly differently by participants, their age range was limited and findings were similar to findings from the survey for participants of the same age range for happy and neutral behaviour. Therefore, it can be concluded that emotive behaviours were perceived as designed, and the research can continue with the preparation and execution of a longitudinal field study with older adults. This field

study can possibly provide more insights regarding ethical concerns of AEE for potentially vulnerable populations, which will be described next in Chapter 5.

5 Evaluating Artificial Expression of Emotion with Older Adults

This research addressed the question whether AEE can lead to ethical concerns in the form of emotional deception and emotional attachment. The previous chapter described the implementation and validation of AEE, which were the first steps in determining whether AEE can lead to negative consequences for older adults. No negative impact of the emotive behaviours on the students was found, deeming it reasonable to continue and investigate the impact of the behaviours on older adults. This chapter describes the preparation and execution of a longitudinal field study with older adults, which was conducted to investigate the impact of AEE on vulnerable populations. First, an overview is provided of existing longitudinal user studies with social robots and older adults. This is followed by the description of a small pilot study that was performed with university staff aged over 50, to establish whether the chosen interaction topic was suitable for the field study. Some of this work was presented at the International PhD Conference on Safe and Social Robotics that was held in Madrid in October 2018 (Van Maris, 2018). Next, the longitudinal field study with older adults is presented. In this study, several measures were included to investigate whether they can contribute to this type of study that may involve vulnerable populations. The field study and results from the questionnaires used in this study have been published in a topical issue of Frontiers in Robotics and AI (Van Maris et al., 2020b).

In summary, three research questions will be investigated in this chapter through a longitudinal field study with older adults. The research questions addressed in this chapter are:

1. 'Does AEE impact older adults' human-robot interaction experience?'
2. 'Does the impact of AEE on older adults change over time?'
3. 'Can physiological data, speech prosody data and behavioural data provide valuable insights when used in longitudinal field studies in addition to questionnaires?'

5.1 Longitudinal HRI Experiments

Kerstin Dautenhahn highlighted the importance of conducting longitudinal experiments in 2007 (Dautenhahn, 2007a). Even though such experiments bring additional challenges compared to experiments that consist of a single sessions, it provides insights on user experience and how their responses to the robot may change over time. This is important to address, as otherwise the robot may not be used or used less once the novelty effect has worn off (e.g. Kanda et al., 2004; Fernaeus et al., 2010).

Many longitudinal studies are conducted either with children or with older adults. This is understandable, as these are two target audiences that can benefit from services that social robots can provide and therefore their experiences should be evaluated. Studies involving children investigated whether perceived social presence would change over time (Leite et al., 2009), whether a peer-like social robot could enhance children's language growth (Kory-Westlund and Breazeal, 2019), and improving social skills in children with autism spectrum disorder (Scassellati et al., 2018). Studies with older adults focus for example on acceptance of social robots depending on cognitive health (Wu et al., 2014), suitability of robot platforms in care homes (Carros et al., 2020), whether they can improve well-being (Wada and Shibata, 2007) and acceptance over time (De Graaf, Allouch, and Dijk, 2016).

Where possible, recommendations from the literature were used in the development of the longitudinal field study conducted in this research. One study found that the robot's perceived social presence decreased over time, and suggested that memories could sustain this (Leite et al., 2009). Therefore, the interactions in the field study were implemented such that the robot would summarise the topics that were discussed during their previous meeting.

5.2 Preparing a Longitudinal Field Study with Older Adults

It was decided that the focus would be on older adults when investigating the impact of AEE, as they are a target group that can benefit from social robots a lot. However, they are also more likely to be vulnerable for negative consequences through age-related impairments. To try and evoke natural responses, it was decided to conduct this study in their home-environment. This would help them feel more comfortable during the experiment. Furthermore, this allowed for recruitment of a user group that is more representative of the target population than participants that are able to come to a laboratory for several sessions, as participants with mobility issues could now participate as well.

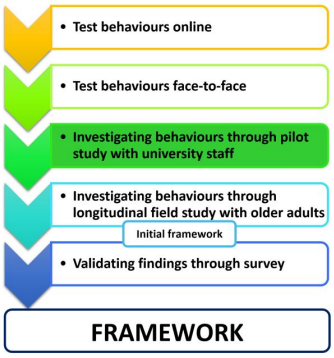


FIGURE 5.1: Current stage of framework development.

To be able to understand the impact of AEE, it was essential that this field study lasted over a longer period of time. Existing research has indicated that participants' behaviours towards the robot change once the novelty effect has worn off (e.g. Leite, Martinho, and Paiva, 2013). Running a longitudinal field study provided insights on participants' responses to the robot and how they changed over time. Furthermore, it provided more realistic feedback as social robots are likely intended to provide support over a longer period of time, and not just for a single session.

It was decided that the interactions were going to be didactic, to ensure a baseline level of AEE would be provided and interactions would not become too personal for participants. This topic had to stay relevant and interesting for several sessions. Eventually, the Seven Wonders of the World (both of the Modern and the Ancient World, from now on called 'Wonders') were selected for this study. Before the field study was designed, a pilot study with university staff aged over 50 was conducted to investigate whether this was a suitable topic for the study with older adults. Participants would interact either with the emotive or non-emotive robot. It was decided to recruit university staff aged over 50, as it was aimed to ensure as many older adults as possible were recruited for the longitudinal field study, and staff over 50 were closest to represent the intended target group.

Attachment had not been investigated in the studies where AEE was tested, as the aim of these studies was to ensure that the implemented behaviours were suitable for this research. To ensure suitable measures for attachment were used in the longitudinal field study, this pilot study used different questionnaires to measure attachment.

Finally, the goal of the longitudinal field study was to investigate the suitability of other measures besides questionnaires. One of these measures was the use of physiological measures, and the Empatica E4 wristband was used in the pilot study to determine whether this was a suitable device. This device was used, as it was available for use and was recommended by colleagues.

5.2.1 Participants

In total 17 participants (9 male, 8 female) completed the pilot study (*Min. age* = 53, *Max. age* = 71, $M = 60$, $SD = 4.9$). Participants were randomly assigned to the emotive (8 participants, 4 female) or non-emotive condition (9 participants, 4 female). Participants were invited through an email that was distributed to university faculty staff. Participants were slightly familiar with social robots ($M = 1.94$, $SD = 1.20$ on a 5-point scale from not familiar at all to extremely familiar).

5.2.2 Materials

The robot Pepper was used for the interactions with the participants. Physiological data was recorded using Empatica E4 wristband[1] (see Figure 5.2). The Empatica E4 provides the option to measure both heart rate variability and electrodermal activity, which are the two physiological measures of interest for this research, as they indicate changes in arousal. The sampling rate for HRV is set at 1 Hz, for EDA this is 4 Hz. The sensor is very unobtrusive for participants to wear. It has a PPG sensor which is required to analyse HRV and an EDA sensor as well. The device was started before the interaction was started and only taken off after completing post-study questionnaires, to gather a baseline level for HRV and EDA.

[1] https://www.empatica.com/research/e4/

FIGURE 5.2: Empatica E4 wristband which was used to gather EDA and HRV. The picture shows the placement of the wristband (left) and the location of the PPG sensor (right).

5.2.3 Measures

At the start of the experiment, participants were asked to provide demographic information (age, gender, familiarity with social robots on a scale from 1 - not familiar at all to 5 - very familiar). Even though attachment was expected to be low after a single interaction, it was measured in this study to test whether the measurements found for attachment were suitable for the longitudinal field study with older adults. First, participants were given a questionnaire to determine their attachment style (Collins and Read, 1990). Furthermore, they were given an attachment questionnaire when the interaction was finished (Schifferstein and Zwartkruis-Pelgrim, 2008). More details about these questionnaires were provided in Section 3.4. Finally, participants were asked their opinion on the interaction, and whether they would be keen to hear more on the topic in future sessions.

5.2.4 Experimental Setup and Procedure

Participants were located opposite the robot and were placed such that they were unable to see the robot was manually operated, using a laptop. As described in Section 4.4, the robot was manually operated during each user studies, to ensure the interaction flow was natural and speech-recognition software was not needed. There was a small table located between the robot and the participant that was used for questionnaires. The experiment started with initial questionnaires, followed by placement of the Empatica E4. After that, the interaction between the participant and robot began. The interaction started with the

robot asking the participant whether they could name any of the Wonders. If participants could list some of the Wonders, the robot would say 'These are indeed Wonders of the World. The full list consists of...'. Otherwise, the robot would just list the wonders. Next, the robot would ask the participant to choose what Wonder they wanted to be informed about in more detail. All Wonders were shown on the robot's tablet in case participants did not know them all. The robot would provide information on the Wonder that was chosen, and then ask what wonder the participant would like to discuss next. This process was repeated until all Wonders were discussed. First all ancient Wonders were discussed, followed by all modern Wonders. After the interaction, participants were asked to fill in the attachment questionnaires followed by a debriefing on the goal of the experiment and the research.

During the interaction, the robot would show either emotive or non-emotive behaviours. In the non-emotive condition, the robot would show neutral behaviours during the interaction. In the emotive condition, the robot would display contextually suitable emotions, for example by saying with a sad voice that a monument got destroyed in a fire showing sad behaviour or saying with a happy voice that a monument was still mostly intact showing happy behaviour. The same characteristics for happy, neutral and sad behaviours as described in Chapter 4 were used.

5.2.5 Results

The main goal of this study was to determine whether Seven Wonders of the World is a suitable topic for a longitudinal study, and to determine the usefulness of human attachment and object attachment questionnaires for human-robot interaction experiments.

5.2.5.1 Suitability of the Topic

All participants reported being interested in learning more about the Wonders and indicated they would have liked to receive more information on the Wonders during the interaction.

5.2.5.2 Attachment to the Robot

As it was found that the attachment style questionnaire used in this study only covered three styles, where four styles are presented in the literature (Bartholomew, 1990), this

measure was not analysed.

Overall, attachment to the robot was low to medium (M = 2.44, SD = 0.69 on a 5-point scale). Averages and standard deviations per condition can be found in Figure 5.3.

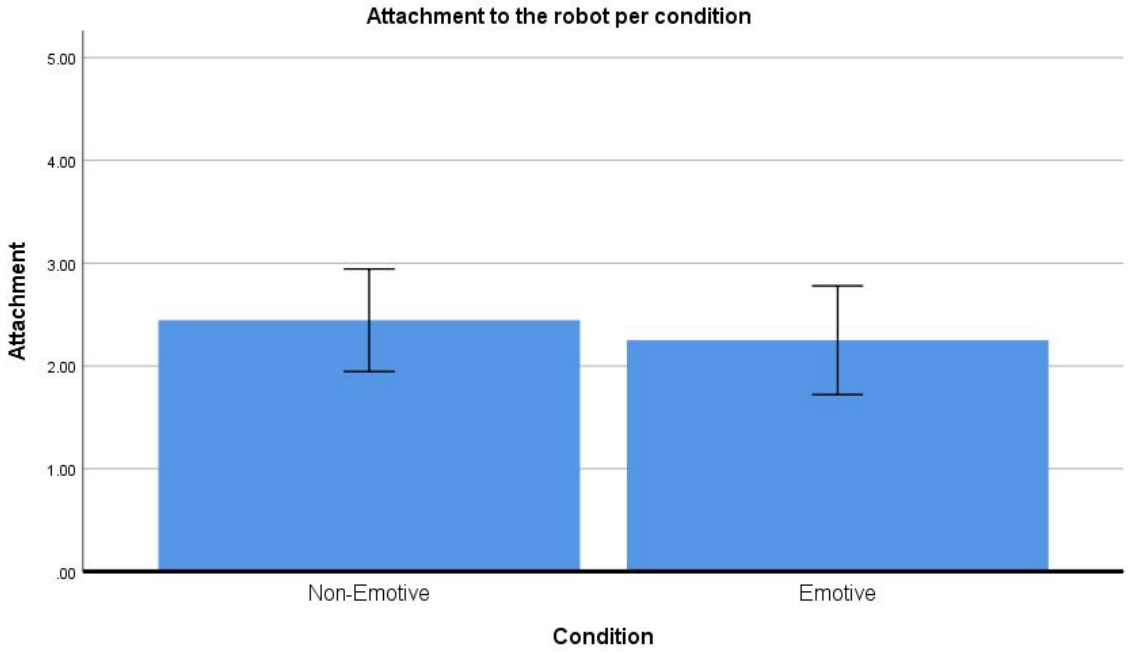


FIGURE 5.3: Average attachment to the robot per condition (non-emotive x emotive), with the number indicating the level of attachment (1 = low attachment, 5 = high attachment).

Reliability of the questionnaire was acceptable (Cronbach α = 0.72). Independent samples t-tests were conducted to investigate whether attachment to the robot was significantly influenced by the robot's behaviour (emotive x non-emotive). No significant impact was found ($t(15) = 0.36$, $p = 0.72$).

Age As the age range was narrow, no analyses were performed to investigate whether age had an impact on participants' level of attachment to the robot.

Gender An independent samples t-test was conducted to investigate the impact of gender on participants' level of attachment to the robot. No significant impact was found ($t(15)$ = -0.68, $p = 0.51$).

5.2.5.3 Familiarity with Social Robots

As most participants were not familiar with social robots, no analyses were conducted to investigate the impact of participants' familiarity with social robots on their level of

attachment to the robot.

5.2.5.4 Physiological Data

The Empatica E4 was used to measure heart rate variability and electrodermal activity. However, it was found that the sampling rate used to gather HRV (1 Hz) is not sufficient to gather reliable data (Laborde, Mosley, and Thayer, 2017). Furthermore, the EDA values recorded by Empatica E4 were extremely low and therefore deemed unreliable. Therefore, the data gathered through this device has not been analysed.

5.2.6 Discussion

The aim of this study was to investigate whether the topic on Wonders of the World was suitable for the longitudinal field study with older adults. All participants reported they found the topic interesting and would be interested in more sessions where the robot would inform them in more detail on the Wonders.

Looking at attachment, it was found that attachment to the robot was low to medium, nor was this impacted by the behaviour the robot displayed. This was expected, as the participants interacted with the robot for a single session and the displayed arousal by the robot was designed to be low.

Brief analysis of the data the Empatica E4 provided showed that a better sensor was required for the field study with older adults, as the sampling rate was too low or the data gathered provided unexpected values that did not seem reliable. Therefore, it was decided to add the GSR+ sensor from Shimmer Sensing that allows for a higher sampling rate (users can adjust the sampling rate up to 2048 Hz compared to 4 Hz for Empatica E4) and thus a more precise measure of heart rate variability and electrodermal activity.

Other observations from this pilot study were that the table blocked the lower half the robot from view. Also, it was observed that people would lean forward on the table, making behaviour analysis more difficult. Therefore, it was decided that a small table that could easily be removed would be used for the field study with older adults. This resulted in a table being present to fill in questionnaires but not blocking the view of the robot.

Finally, it was observed that participants focused on the robot's tablet that displayed the list with Wonders of the World and not the robot's behaviour. This was reported by some participants as well. Therefore, it was decided that the tablet would not be used during the field study with older adults.

In summary, the topic was deemed suitable for the longitudinal field study. Furthermore, it was found that the Empatica E4 wristband did not provide reliable data and an alternative with more flexible sampling rates had to be considered. This was also found for the attachment style questionnaire. The alternatives are described in the next section, where the longitudinal field study is presented.

5.3 Conducting a Longitudinal Field Study with Older Adults

This field study was conducted to investigate whether participants can be emotionally deceived by AEE. It also investigated whether participants can become attached to a robot over time and whether attachment is higher when AEE occurs. This field study was conducted twice, one time in two retirement villages each. Residents in these villages have their own apartments and live semi-independently, but are not able to live fully independently and are provided with assistance if needed. In the remainder of this dissertation, when a distinction needs to be made between the two villages, the village where the study was conducted first will be called Village A, and the other village will be called Village B. It was decided to conduct this experiment twice as there were only four participants with usable data from Village A, which is not enough for reliable data analysis. The decision to decide what study design was ideal and what was practical was difficult. A between-subject design where responses to AEE were compared with responses to non-emotive behaviour would be ideal to determine the impact of AEE on user experience. However, as mentioned before it was unlikely that the participant group would be big enough for this design to provide reliable data. Another option was

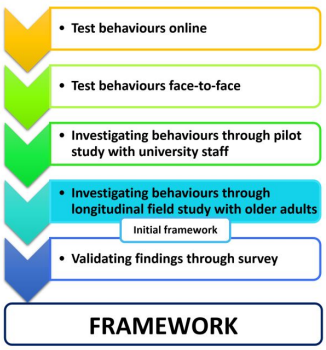


FIGURE 5.4: Current stage of framework development.

a within-subject design. However, as AEE was designed to be subtle, analyses may not be able to identify a potential impact. Therefore, a between-within mixed design was chosen for this experiment. As it was recognised early on that the number of participants for Village A was going to be low, it was decided to treat them as control group which meant they interacted with a robot that displayed non-emotive behaviour only. Village B provided ten more participants that would form the test group and interacted with a robot that displayed both emotive and non-emotive behaviour.

5.3.1 Participants

In total 17 older adults participated in this field study. Seven participants were recruited through Village A (control group) and ten participants were recruited through Village B (test group). As this study was directed toward typical ageing, prior to scheduling sessions, participants were asked to self-report health issues/diagnoses (e.g. dementia) that could affect their ability to complete measures or limit their capacity to consent. No participants were excluded based on this criteria. In addition, participants from Village B had their capacity for informed consent monitored by a locksmith (an individual who monitors residents). As part of the procedure, participants in Village A were administered using the Montreal Cognitive Assessment test (MOCA; Nasreddine et al., 2005) for overall cognitive function. Based on these scores, data from two participants was excluded as they scored below 15 (out of 30) where all other participants scored between 26 and 30, which indicated high cognitive function. One participant from Village A was unable to complete the experiment; all participants from Village B completed the experiment. As such, data from 14 participants were included in the analyses. Table 5.1 provides an overview of the participant demographics per village and overall. Table 5.2 provides demographics information per participant (randomly ordered per village). This table also provides attachment style and level of attachment, which will be discussed in more detail later.

5.3.2 Materials

Once again Pepper was used for the interactions. To be able to gather as much knowledge on the participants' experiences with the robot, other forms of data besides questionnaires were gathered as well. As mentioned in the pilot study, one of these measures was

		Age		Gender	
	N	*M*	*SD*	*M*	*F*
Village A	4	77	11	3	1
Village B	10	76	8	6	4
Total	14	76	9	9	5

TABLE 5.1: Distribution of participants' age and gender per retirement village and for the whole experiment.

Participant	Group	Gender	Familiarity	Attachment Style	Level of Attachment
1	Control	M	Somewhat	Fearful	Medium
2	Control	M	Low	Secure	Low
3	Control	M	Low	Secure	Low
4	Control	F	Low	Dismissive	Medium
5	Test	M	Somewhat	Secure	Low
6	Test	M	Low	Secure	Medium
7	Test	F	Low	Fearful	Low
8	Test	F	Low	Dismissive	Medium
9	Test	M	Low	Dismissive	Medium
10	Test	M	Low	Fearful	High
11	Test	M	Low	Secure	Medium
12	Test	F	Low	Dismissive	Low
13	Test	M	Low	Fearful	Medium
14	Test	F	Low	Secure	High

TABLE 5.2: Case characteristics of the user trials.

physiological data (HRV and EDA). A Shimmer GSR+ sensor[2] was used (see Figure 5.5), as the pilot study with university staff indicated the Empatica E4 wristband did not provide reliable data. However, Empatica E4 was used for gathering data as well, in case GSR+ would provide faulty recordings. The GSR+ was chosen as it provides the option to measure both physiology features at the same time, provides a higher sampling rate and places sensors at different locations that may be more precise than the Empatica E4 wristband. On the GSR+, the PPG sensor was placed on participants' earlobe instead of their wrist while the EDA sensor was placed on their fingers. This allowed for putting both devices on participants at the same time without the devices disrupting each others' recordings. The sampling rate used on the GSR+ is 128 Hz which is recommended in the software provided by the company to analyse HRV. The Empatica E4 was still used as back-up in case recordings with the GSR+ sensor would be faulty. As will be discussed in the following section, speech prosody data and behavioural data were recorded as well. Video-recordings were taken with a Nikon D3100 to be able to collect this data.

5.3.3 Measures

As this research investigates topics that may not be best measured using solely questionnaires, this field study combined several different measurements to investigate potential occurrence of deception and attachment. This resulted in insights on what measurements can be useful additions to the use of questionnaires in HRI research to retrieve richer and more insightful data.

5.3.3.1 Questionnaires

All questionnaires used in the user face-to-face study with Psychology students were used in this field study as well. The questionnaires used in this study are the following:

- **Demographic information:** Besides the questions used in previous studies (age, gender, familiarity), participants were asked how often they used technologies such as smartphone, tablet and laptop/desktop.
- **Adapted Godspeed questionnaire:** The adaptation is the same as the one used in previous studies, that can be found in Appendix A1.

[2]http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

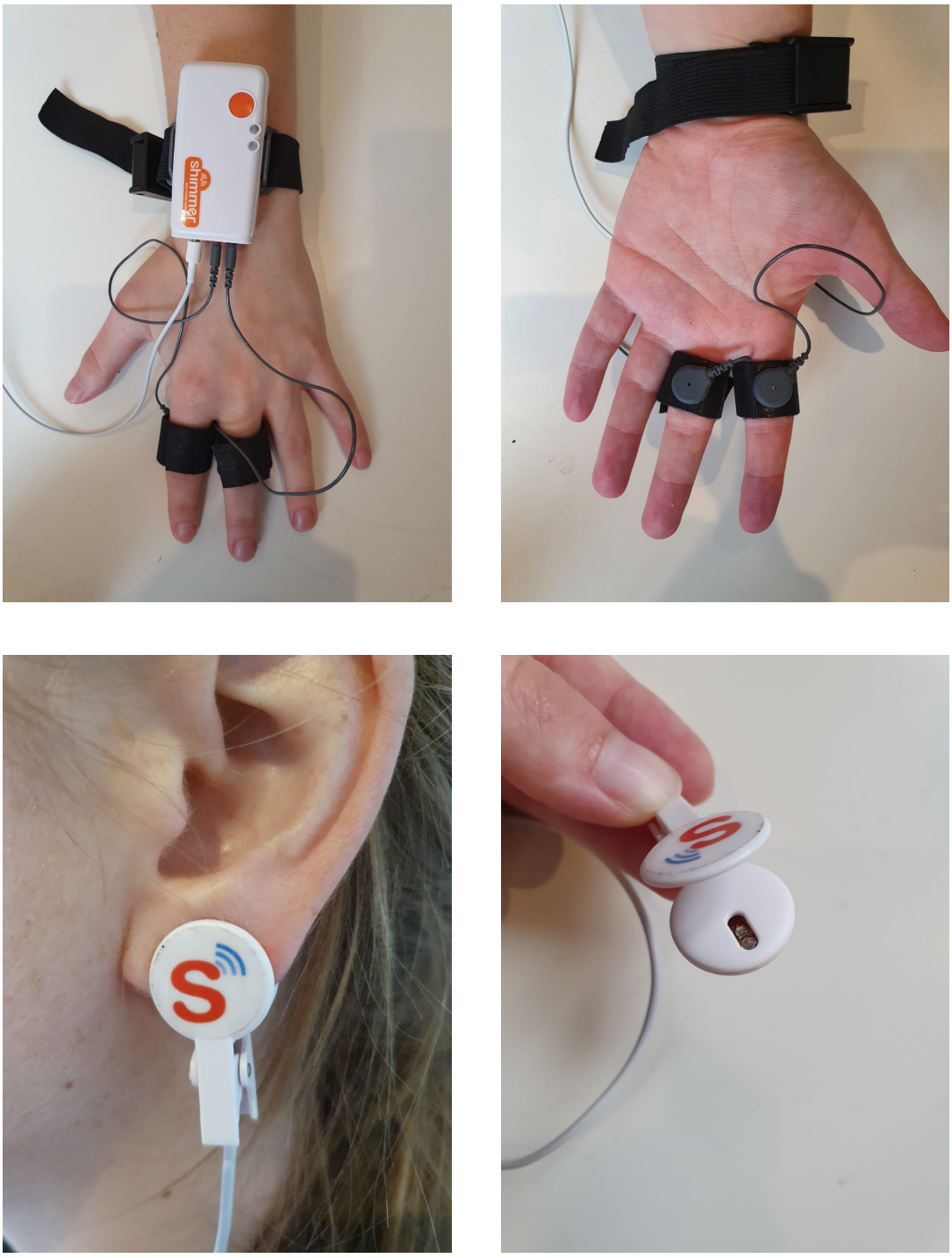


FIGURE 5.5: GSR+ sensor used to gather EDA and HRV. EDA is measured through sensors placed on the base of the fingers (top) and HRV is measured through a sensor placed on the earlobe (bottom).

- **Adapted Almere Model:** The same adaptation as used in previous studies. The adapted questionnaire is provided in Appendix A2.

- **PANAS:** Measurement of explicit mood, used in previous studies as well and presented in Appendix A3.

- **IPANAT:** Measures implicit mood to account for responses impacted by social desirability. This questionnaire is provided in Appendix A4.

- **MOCA:** Montreal Cognitive Assessment - used to measure general cognitive performance including executive functioning, memory, language, attention and visuospatial perceptual skills (Nasreddine et al., 2005). This assessment was only used for participants from Village A, as there was a locksmith available who determined participants' cognitive functioning for Village B. This assessment is presented in Appendix A5.

- **Attachment style:** The Experiences in Close Relationships Inventory was adapted for this field study (Brennan, Clark, and Shaver, 1998). Statements involving '(romantic) partners' were adapted to a more general variation of 'people that are dear to me' to prevent potentially upsetting participants if a (romantic) partner was no longer with them. This questionnaire is provided in Appendix A6.

- **Attachment:** The same attachment questionnaire was used as the one provided in the pilot study and can be found in Appendix A7.

- **Interview:** Once participants were debriefed, they asked several questions in a face-to-face interview. Notes were taken of the participants' responses to the questions asked during this interview to allow them to provide richer answers. This interview is provided in Appendix A8.

Demographics questionnaires were always administered first, followed by MOCA and the attachment style questionnaire. Items in the attachment style questionnaire were randomised. All other questionnaires and their items were randomly ordered as well, with the exception of the face-to-face interview that was always conducted last. An overview is provided in Table 5.3.

	T1 (Introduction)	T2 (Condition 1 complete)	T3 (Condition 2 complete)	T4 (Debrief)
Godspeed	x	x	x	x
Almere		x	x	x
Attachment		x	x	x
PANAS	x	x	x	x
IPANAT	x	x	x	x
MOCA (Village A only)	x			
Demographics	x			
Attachment Style	x			
Interview				x

TABLE 5.3: Overview of when questionnaires were administered during the experiment. Godspeed measured participants' perception of the interactions, Almere measured their acceptance of the robot. PANAS and IPANAT measured participants' mood, and MOCA measured participants' cognitive performance.

5.3.3.2 Physiological Data

Physiological data was gathered during the field study with older adults, to investigate whether they can be useful measures to indicate distress and therefore potential negative consequences of AEE. More specifically, heart rate variability (HRV) and electrodermal activity (EDA) were recorded. These were gathered as they may provide indications of stress and/or arousal (Nikolić et al., 2016). HRV has been used as a stress measure during human-robot interaction before, where HRV indicated that participants were less stressed when a robot approached them using a trapezoidal trajectory compared to a linear trajectory (Kühnlenz et al., 2018). Furthermore, it was found that EDA significantly differed for specific robot motions (Dehais et al., 2011). Therefore, these measures appeared useful to be included in this research. The difficulty with these measurements is that it is hard to determine whether the physiological change is due to a positive or a negative event. However, it is still valuable to determine whether a change occurs and combining these findings with findings from other types of data gathered may provide useful insights in participants' responses to the robot and its behaviour. Heart rate variability was measured through a PPG signal, that would provide details on the time between heart beats (inter beat interval; IBI), which was then used to determine heart rate variability. Electrodermal activity was recorded by gathering the galvanic skin response. More details on these two measures can be found in Section 5.3.6.

5.3.3.3 Speech Prosody

Besides physiological data, speech prosody can be an indicator of stress as well (Galaz et al., 2016). There are several prosodic features that can be analysed, such as pitch related features, energy related features and pause features (Ang et al., 2002). However, due to the didactic nature of the interaction, natural input from participants was limited, resulting in pre-determined pause features. Therefore, they were not investigated in this study. As pitch features were deemed more reliable than energy related features (Schuller, Rigoll, and Lang, 2003), pitch related features were analysed in this research. Pitch, duration and intensity are the three most used features of speech prosody (Vinciarelli et al., 2008). However, duration of the speech fragments was manipulated due to the nature of the interaction. Therefore only pitch and intensity were investigated in this study. All fragments where participants were talking to the robot were cut from video-recordings of the interactions. These fragments were turned into audio-files with the use of FFmpeg, a software used to handle multimedia files[3]. These audio-files were analysed through the software Praat (Boersma, 2001), a program often used in speech pathology studies (Cohen et al., 2009). After noise reduction, the minimum, maximum, average and standard deviation of both pitch and intensity were calculated for each audio-file. These features were compared when participants were interacting with the emotive or non-emotive robot. A difference could indicate stress, which may have been induced by AEE of the robot. More details on this analysis can be found in Section 5.3.7.

5.3.3.4 Behaviour Analysis

As mentioned before, physiological data can provide additional insights, as they gather real-life data and it can be analysed what behaviour the robot was displaying when an increase in arousal occurred in the participant. However, this data does not provide insights in why this arousal occurred - whether this was a positive or negative event. This is where behaviour analysis can provide useful insights. To be able to understand user experience, several observations are required over an extended period of time (De Graaf, 2016). Interactions between older adults and the robot were video-recorded to analyse participants' behaviour during the interactions. Comparing these recordings with the physiological data can provide insights whether the arousal that was found in the physiological data is positive or

[3]https://ffmpeg.org/

negative. Traditional research on non-verbal behaviour consists of three coding principles: generic coding, restrictive coding and evaluative coding (Frey and Pool, 1976; Bente et al., 2008). Generic coding entails general descriptions of body positions (e.g. upright trunk), where restrictive coding focuses more on details (e.g. hand gestures). These two coding systems consist of behavioural descriptions. Evaluative coding on the other hand provides subjective impressions of observers, where the observed behaviour is categorised immediately (e.g. scoring the level of friendliness of a behaviour). In this research, a combination of generic and restrictive coding was used, as none of the coders are experts in behavioural science and evaluative coding would provide very subjective findings. Generic and restrictive coding have disadvantages as well, as observers may be looking for specific behaviours and therefore miss other subtle behaviours. Therefore, coded recordings were independently coded by two observers, so their findings could be compared. If the inter-rater reliability value was at least substantial ($\kappa \geq 0.61$), the findings would be used for analysis. Video-recordings were coded by Psychology students who were conducting research as part of their work experience. They signed a non-disclosure agreement and were only given access to the recordings on my device when I was present. Each video-recording was coded by two students. An example table of coded behaviours is provided in Appendix F. More details on how behaviour was analysed in this study are provided in Section 5.3.8.

Physiological data, speech prosody data and behavioural data were gathered during each interaction, so data was gathered for eight sessions per participant.

5.3.4 Experimental Setup and Procedure

Participants interacted with the robot for eight sessions: two interactions per week for four weeks. Interactions lasted between five and eight minutes, and entailed the Ancient and Modern Wonders of the World. An example interaction can be found below. There would be preferably two or three but at least one day between sessions. Besides these eight interactions with the robot, there was an introductory session before the first interaction and a debrief session after the last interaction with the experimenter only. The robot was present during the introductory session and would introduce itself briefly to participants so they knew what to expect from the robot feel more comfortable when the first interaction started. The robot was not present in the room during the debrief session. For the participants

from Village B, the robot would display both emotive and non-emotive behaviour. It would either show non-emotive behaviour during the first four interactions and emotive behaviour during the last four interactions, or vice versa. The order of robot behaviors was counterbalanced between participants. The participants from Village A only saw the robot displaying non-emotive behaviour.

An example part of an interaction between the robot and a participant about the Statue of Zeus at Olympia; *R* = robot, *P* = Participant:

- **R:** 'As mentioned before the statue depicts Zeus sitting on a wooden throne. However, did you know that the whole statue and not only the throne was made of wood?'
  - *If **P** said 'no':* **R:** 'Yes, the whole statue was sculpted in wood. After that, Zeus was covered with ivory and gold plates.'
  - *If **P** said 'yes':* **R:** 'Indeed, the whole statue was sculpted in wood. After that, Zeus was covered with ivory and gold plates.'
- **R:**'Have you ever been to Olympia, or other places in Greece?'
  - *If **P** said 'no':* **R:** 'Now let us continue with...'
  - *If **P** said 'yes':* **R:** 'Would you like to tell me about it?'
    - *If **P** says 'no':* **R:** 'Ok, now let us continue with....'
    - *If **P** talks about positive experience:* **R:** 'That sounds nice. Now let us continue with....'
    - *If **P** talks about negative experience:* **R:** 'Sorry to hear that. Now let us continue with....'

Participants were seated opposite the robot. The distance between the chair and the robot was approximately 1.5m, which falls within the social space of Hall's proxemics categories (Hall et al., 1968), but approaches the personal zone as well as the threshold between these two zones is at 1.2 metres. The social space represents the distance between two strangers having a conversation, where the personal space represents the distance where two friends have a conversation. It should be noted that there were slight differences in distances between participants and the robot as participants could either lean forward or lean back in the chair, and there were some instances where a participant would slightly move the chair during the interaction.

FIGURE 5.6: Experimental set-up Village A.

FIGURE 5.7: Experimental set-up Village B.

Two sensors were used to gather physiological data. The sensor used to gather heart rate variability and electrodermal activity data is the GSR+ sensor from Shimmer Sensing. A PPG sensor would be attached to participants' ear lobe to gather HRV data. Two GSR sensors were used to gather electrodermal activity data - one would be placed around the base of the index finger of the non-dominant hand, and the other sensor would be placed around the base of the middle finger of the same hand. As discussed before, this sensor was chosen over Empatica E4 as this wristband provided unreliable data. However, it was used in this study as back-up as experience with GSR+ was limited. The Empatica E4 wristband was placed around the wrist of the dominant hand.

As mentioned in Chapter 4, Wizard-of-Oz strategy was used for all user experiments in this research to ensure the interaction would be as natural as possible.

It was taken into account that participants might deviate from the planned script. Therefore, additional responses like 'yes', 'no', 'I do not understand what you are saying' and 'Let us get back to the topic we were discussing' were pre-programmed as well.

In Village A only one room was available to conduct the experiment. Therefore, the experimenter was located behind the participants to ensure participants would not be distracted by the experimenter during interactions, and make sure they would not realise the robot was operated by the experimenter (see Figure 5.6). In Village B the experimenter was located in an adjacent room. The experiment room in which the interactions took place can be seen in Figure 5.7. The doors of both the experiment room and the adjacent room were open as to ensure the participants they could call for the experimenter if they felt uncomfortable. Participants were given a photograph of themselves with the robot and a £20 gift card for their participation during the debrief session. Contact details of the

experimenter were provided in case they wanted to see the robot again. This exit strategy was essential, as it would have been unethical to investigate their emotional attachment to the robot over a longer period of time and not provide them with the possibility to see the robot again.

5.3.5 Results from Questionnaires

5.3.5.1 Emotional Deception

As mentioned before, potential occurrence of emotional deception was measured through the construct anthropomorphism of the Godspeed questionnaire, the construct social presence of the Almere questionnaire and the statements regarding the robot being able to communicate an emotion of happiness or sadness. Independent samples t-tests were performed to investigate the influence of the robot's behaviour (non-emotive x emotive) on the factors mentioned above. Mixed between-within subjects analyses of variance were conducted to investigate the impact of group (control x test) and time (T1, T2, T3, T4 for anthropomorphism, T2, T3, T4 for social presence and communication of happiness or sadness). As shown in Table 5.4, anthropomorphism was not significantly influenced by the behaviour displayed by the robot (non-emotive x emotive), the group that interacted with the robot (control x test) or time (T1, T2, T3, T4), nor was there an interaction between group and time. Figure 5.8 shows that anthropomorphism was higher for participants of the test group for T2, T3 and T4, which means it was higher for all sessions except for when they had not interacted with the robot yet (T1).

Looking at social presence, a main effect was found of group (control x test) on the perceived social presence of the robot ($F(1,12) = 4.93$, $p = 0.046$, $\eta_p^2 = 0.29$). Participants in the test group that interacted with both the emotive and non-emotive robot perceiving the robot as more of a social entity than participants in the control group that only interacted with the non-emotive robot, and this remained the same over time. This can be seen in Figure 5.9. No other effects on the perceived social presence of the robot were found.

Investigating the statements *'Based on my interaction with Pepper, I feel that it is capable of communicating an emotion of happiness'* and *'Based on my interaction with Pepper, I feel that it is capable of communicating an emotion of sadness'*, no significant impact was found of either the robot's behaviour (non-emotive x emotive), group (control

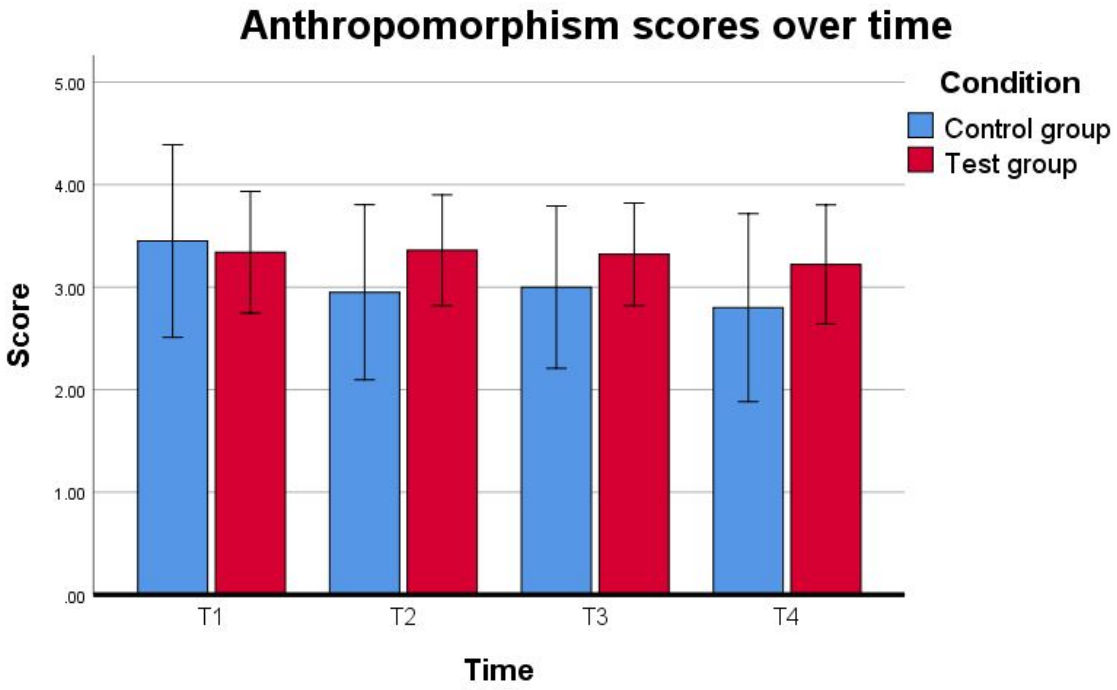


FIGURE 5.8: Average values of anthropomorphism for each experiment group over time from not at all (1) to very much (5). A distinction was made between participants that saw non-emotive robot behaviour only (control group) and participants that saw both emotive and non-emotive behaviour (test group).

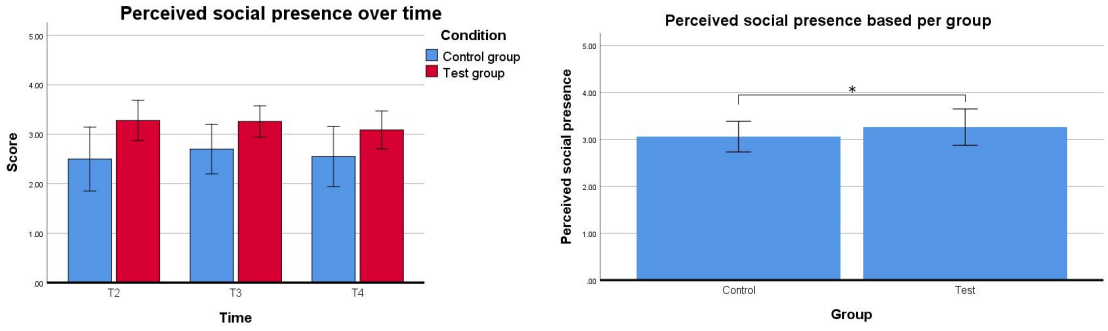


FIGURE 5.9: $*p < 0.05$; Average values to what extent participants perceived the robot as a social presence over time from not at all (1) to very much (5). A distinction was made between participants that saw non-emotive robot behaviour only (control group) and participants that saw both emotive and non-emotive behaviour (test group).

x test) or time (T2, T3, T4), nor was there an interaction effect between Group and Time (see Table 5.4). Even though no significant impact was found, Figure 5.10 indicates that participants of the test group perceived the robot being more able to communicate emotions than participants of the control group after they interacted with the robot (T2, T3). However, when participants were asked the same question during the debrief session (T4), participants of the control group perceived the robot as more capable of communicating emotions.

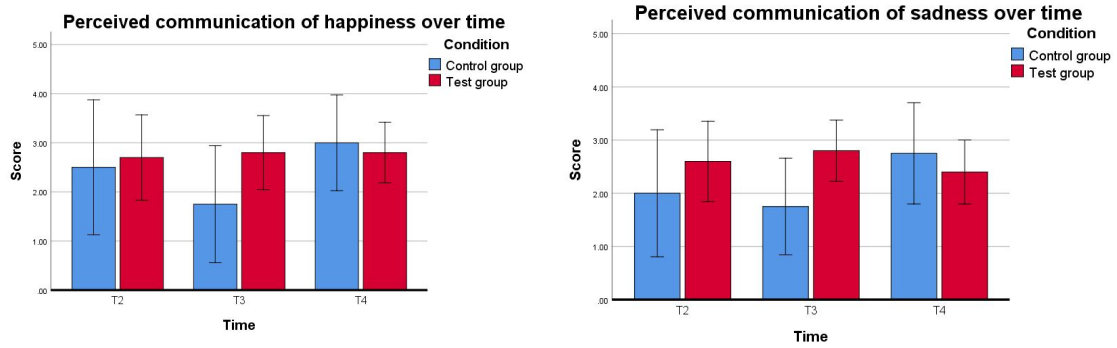


FIGURE 5.10: Average values to what extent participants perceived the robot was able to communicate an emotion of happiness (left) and sadness (right) over time from not at all (1) to very much (5). A distinction was made between participants that saw non-emotive robot behaviour only (control group) and participants that saw both emotive and non-emotive behaviour (test group).

	Behaviour (non-emo x emo)		**Group (control x test)**			**Time ([T1], T2, T3, T4)**			**Time x Group**		
	t	p	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Anthropomorphism	0.58	0.57	0.38	0.55	0.03	1.57	0.23	0.12	0.98	0.38	0.08
Social presence	-0.83	0.42	4.93	**0.05***	0.29	0.78	0.47	0.06	0.54	0.59	0.04
Happiness	0.03	0.98	0.45	0.52	0.04	1.82	0.19	0.13	1.89	0.17	0.14
Sadness	-0.92	0.37	1.06	0.32	0.08	0.56	0.58	0.04	2.56	0.10	0.18

*p $<$ 0.05

TABLE 5.4: Test results indicating whether several factors significantly impacted participants' perceived anthropomorphism of the robot, it being perceived as a social entity, and it being able to communicate emotions of happiness and sadness. 'Behaviour' distinguishes between the robot displaying either emotive or non-emotive behaviour, 'Group' distinguishes between participants that either saw only non-emotive behaviour (control) and participants that saw both emotive and non-emotive behaviour (test). Time indicates the times when questionnaires were taken (T1 / T4 for anthropomorphism, T2 / T4 for the other measures), and Time x Group represents a potential interaction effect between Group and Time. Significant findings are highlighted.

Investigating the effect of gender, a significant interaction effect was found between time and gender for the robot being able to communicating an emotions of sadness ($F(2,24) = 4.44$, $p = 0.02$, $\eta_p^2 = 0.27$). This was not significant for the robot being able to communicate an emotion of happiness ($F(2,24) = 0.14$, $p = 0.87$, $\eta_p^2 = 0.01$), nor was there an effect of time ($F(4,9) = 1.21$, $p = 0.37$, $\eta_p^2 = 0.35$) or gender ($F(2,11) = 0.47$, $p = 0.64$, $\eta_p^2 = 0.08$). Female participants agreed more with the statements that Pepper was capable of communicating an emotion of happiness and sadness than male participants, as can be seen in Table 5.5. Overall, participants' interpretation of the robot being capable of communicating emotions of happiness and sadness was low to medium ($M = 2.62$, $SD = 0.95$ on a 5-point scale).

	Happiness		**Sadness**	
	M	*SD*	*M*	*SD*
Male	2.52	1.17	2.30	1.00
Female	2.93	0.91	2.73	0.71
Overall	2.73	1.04	2.51	0.85

TABLE 5.5: Mean and standard deviation on a 5-point scale from strongly disagree (1) to strongly agree (5) for the statements *'Based on my interaction with Pepper, I feel that it is capable of communicating an emotion of happiness'* and *'Based on my interaction with Pepper, I feel that it is capable of communicating an emotion of sadness'*.

5.3.5.2 Emotional Attachment

Emotional attachment was directly measured through the attachment questionnaire and indirectly through the question 'Will you miss Pepper?' asked at the interview during the debrief session. Attachment to the robot fell in the low to medium range ($M = 2.68$, $SD = 0.64$). However, two participants (one male, one female) scored high on attachment ($M = 4.06$, $SD = 0.24$). These participants belonged to the test group and interacted with the robot displaying both emotive and non-emotive behaviour during the experiment.

Similar as when investigating emotional deception, independent samples t-tests were performed to investigate the influence of the robot's behaviour (non-emotive x emotive) on attachment. Mixed between-within subjects analyses of variance were conducted to investigate the impact of group (control x test) and time (T2, T3, T4) on attachment.

As can be seen in Table 5.6 there was no significant influence of either the robot's displayed behaviour, the experiment group or time on attachment, nor was there an interaction effect between time and group.

	Behaviour (non-emo x emo)		**Group (control x test)**			**Time (T2, T3, T4)**			**Time x Group**		
	t	p	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Attachment	-0.77	0.45	1.00	0.34	0.08	0.23	0.79	0.02	0.32	0.73	0.03
Intention to use	0.62	0.54	0.38	0.55	0.03	0.71	0.93	0.01	0.21	0.82	0.02

TABLE 5.6: Test results indicating whether participants' level of attachment towards the robot and intention to use the robot were impacted by the robot's behaviour, the participant group, time, and a potential interaction effect between group and time.

One-way ANOVA was conducted to determine whether participants' attachment style impacted their level of attachment to the robot. No significant impact was found ($F(2,11) = 0.35$, $p = 0.71$, $\eta_p^2 = 0.06$). As mentioned in Section 3.4, participants with a fearful attachment style may become more easily attached to the robot. Four participants had a fearful attachment style, with three of them having a medium or high level of attachment to the robot. Six participants had a secure attachment style and three of them had a medium or high level of attachment to the robot. Finally, four participants had a dismissive attachment style with three of them scoring medium on level of attachment to the robot. Therefore, even though participants with a fearful attachment style may become more easily attached to the robot, the number of participants in this study was too small to see a difference in participants' level of attachment to the robot based on their attachment style.

During the final interview, participants were asked whether they would miss Pepper, which may be an indicator of attachment. Four participants (three female, one male) reported that they indeed would miss Pepper ('Yes, I guess I will', 'Oh yes, definitely!'). Three of these participants were from the test group and one was from the control group. Two of them (both female) were categorised as having a low level of attachment to the robot.

Both in the attachment questionnaire and the final interview participants were asked whether they would use the robot if it were available to them at any time. During the interview, nine participants (six male, three female) indicated they would want to use the

robot on a daily basis in the future; these participants also reported they did not think they would get bored of the robot over time. Eight of these participants were from the test group and interacted with a robot that displayed both emotive and non-emotive behaviour. Four participants (two male, two female) declared they would want to use it on a weekly basis. One participant (male) would not want to use the robot at all. This participant stated that, although he liked interacting with the robot, he found that it was not sufficient for his needs as he preferred a robot that was capable of physical assistance. Results from the questionnaire showed that intention to use the robot was high and participants were consistent in their replies as the results were the same as for the interview. Taking into account the robot's behaviour, participant group and time, no significant effect of these factors was found on participants' intention to use the robot if they could, nor was there an interaction effect between time and group.

5.3.5.3 Emotional Deception and Emotional Attachment

Pearson correlation analyses were conducted to determine the relationship between participants' level of attachment to the robot, and the factors used to investigate emotional deception (anthropomorphism, social presence, communicating happiness and sadness). These analyses showed that there was a positive correlation between participants' level of attachment, and the extent to which they anthropomorphise the robot ($r(14) = 0.66$, $p < 0.01$), as well as their perception of the robot as a social entity ($r(14) = 0.42$, $p = 0.04$). Both anthropomorphism and the perception of the robot as a social entity increased when participants' level of attachment increased. No significant correlations were found between participants' level of attachment and their perception of the robot's ability to communicate happiness ($r(14) = 0.32$, $p = 0.12$) and sadness ($r(14) = 0.39$, $p = 0.06$).

Participants were categorized in one of three groups based on their level of attachment to the robot: low, (scores of 2 and below), medium (score of 3 at T2, T3 and/or T4) or high attachment (scores of 4 and above). A one-way ANOVA was conducted to compare participants' attachment categories with the factors used to investigate emotional deception. There was a significant difference between participants' attachment category and the robot's perceived social presence ($F(2,21) = 7.82$, $p < 0.01$, $\eta_p^2 = 0.43$), where participants with a low level of attachment perceived the robot significantly less as a social entity than participants with medium ($p = 0.04$) and high ($p < 0.01$) levels of attachment, and

participants with medium levels of attachment perceived the robot significantly less as a social entity than participants with high levels of attachment (p = 0.02). This can be seen in Figure 5.11. Participants' level of attachment did not significantly impact anthropomorphism ($F(2,21) < 0.01$, $p = 0.88$, $\eta_p^2 < 0.01$), communication of happiness ($F(2,21) = 2.23$, $p = 0.13$, $\eta_p^2 = 0.18$) or communication of sadness ($F(2,21) = 0.53$, $p = 0.60$, $\eta_p^2 = 0.05$).

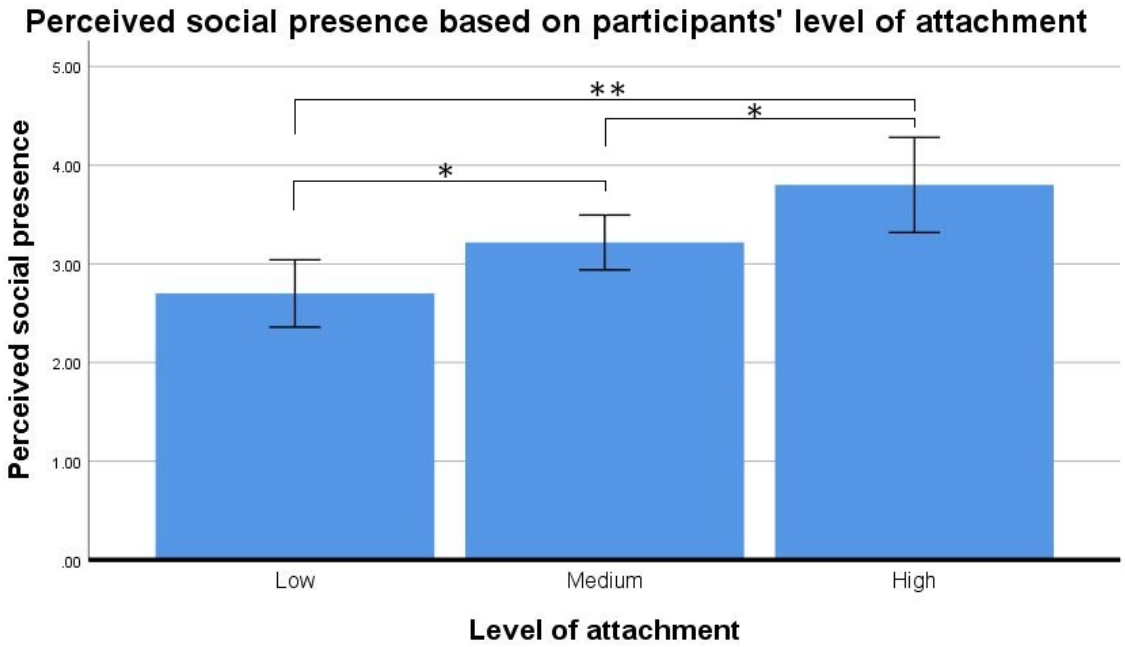


FIGURE 5.11: *$p < 0.05$; Participants' perception of the social robot as a social entity from strongly disagree (1) to strongly agree (5) based on their level of attachment (low, medium, high) to the robot.

Comparing participants' intention to use the robot with the factors used to investigate emotional deception, there was a positive correlation between participants' intention to use the robot and anthropomorphism ($r(14) = 0.47$, $p = 0.02$). There were no significant correlations between intention to use and social presence ($r(14) = 0.24$, $p = 0.27$), and the perceived ability of the robot to communicate happiness ($r(14) = 0.35$, $p = 0.09$) and sadness ($r(14) = 0.36$, $p = 0.09$).

A one-way ANOVA showed no significant influences of intention to use on anthropomorphism ($F(3,20) = 2.50$, $p = 0.09$, $\eta_p^2 = 0.27$), social presence ($F(3,20) = 0.55$, $p = 0.66$, $\eta_p^2 = 0.08$) and perceived ability of the robot to communicate happiness ($F(3,20) = 1.76$, $p = 0.19$, $\eta_p^2 = 0.21$) or sadness ($F(3,20) = 0.20$, $p = 0.89$, $\eta_p^2 = 0.03$).

Overall, these findings indicate that there may be an interaction between participants' level of attachment to the robot and the possibility of them being emotionally deceived by the robot.

5.3.5.4 Other Findings from Questionnaires

Differences between Emotive and Non-emotive Behaviour The effects of robot behaviour (non-emotive x emotive) on acceptance of the robot, perception of the interaction and participant mood was investigated using independent samples t-tests. Robot behaviour did not significantly impact any constructs of the Godspeed and Almere questionnaires. There were no significant differences in mood, both explicit and implicit, between the two groups.

Changes over Time The robot's perceived ease of use, a construct of the Almere questionnaire, significantly increased over time ($F(2,24) = 4.22$, $p = 0.03$, $\eta_p^2 = 0.26$). Time significantly decreased explicit negative affect as measured by PANAS ($F(2.09,27.14) = 5.10$, $p = 0.01$, $\eta_p^2 = 0.28$). No other effects of time were found, nor any interaction effect between time and group.

Observations from the Interview During the final interview, participants were informed on emotional deception and the importance to investigate this concern. After asking them whether they thought the robot had been deceptive during their interactions, some participants reported they found the robot was indeed deceptive ('I guess it was deceptive, as it showed some form of emotions'). These participants interacted with the emotive robot and either scored medium or high on attachment. They reported they thought this deception was acceptable, as otherwise the robot would have appeared too machine-like and not pleasant to interact with. The other participants did not find the robot deceptive, mainly because they thought of it as a machine and/or tool ('I take it for what it is: a distraction for when you are lonely', 'I realize it is a machine, therefore I do not find it deceptive').

Participants were asked whether they would miss the robot as a means to measure attachment. Four participants (one male, three female) indicated that they would miss the robot ('Yes, I guess I will', 'Oh yes, definitely!'). One of these participants was part of the

control group and only interacted with the non-emotive robot. The other three participants were part of the test group and saw both emotive and non-emotive robot behaviours. Two of these participants (both female) were highly attached to the robot according to the attachment questionnaire.

When asked whether they wanted to add anything else at the end of the interview, two participants said they enjoyed the interactions with Pepper - these participants scored high on attachment. One participant commented that Pepper is a likeable robot but preferred for it to have legs ('Please give it some proper feet so it can walk around').

Three participants could not imagine the robot ever playing a role in their life ('do not need a companion', 'happy talking to myself when I feel the need to'). Four participants would like to have a robot as a companion, and eight participants thought it would be useful as a helper. This could either be in the sense of helping with tasks, providing useful information or monitoring people's health.

5.3.6 Physiological Data

A GSR+ sensor was used for this field study to measure physiological data; both heart rate variability and electrodermal activity. Physiological data was gathered for two reasons. First, it was investigated whether wearable devices could be used for this type of human-robot interaction research in an older adult population. Second, it was gathered to investigate whether participants' responses to the robot changed over time. Especially when working with older adults these types of data may be useful. First of all, interacting with new technologies may have a large impact on older adults, who are perhaps less experienced in using new technologies. Real-time data gathering through these devices will provide insights for experimenters who can then address this, in case participants do not feel comfortable sharing this themselves.

5.3.6.1 Heart Rate Variability

The GSR+ allows users to use high sampling rates that can go up as high as 2048 Hz compared to 4 Hz for the Empatica E4. The sampling rate used for this research, 128 Hz, was recommended by the GSR+ manual for collecting HRV data. Furthermore, the GSR+ provides an earlobe sensor to measure HRV, which is more robust than hand-based HRV

sensors due to sensitivity to movement (Vescio et al., 2018). Both these sensors measure HRV through PPG. Sensors that use ECG are less sensitive to movement and might retrieve better data. However, there are several disadvantages to these types of sensors. For example, the sensors need to be placed on the torso which requires participants to temporarily remove some clothing which they may not appreciate. Also, conductive gels are required which may lead to allergic reactions. Finally, these sensors require expert placement, thus limiting their use for non-expert users. Therefore, PPG sensors placed on the earlobe may be the best alternative to potentially use in HRI research. More details can be found in a report that was written for a module. This report can be found in Appendix G.

There are several measurements that can be used to describe HRV. These are either time-domain, frequency-domain or non-linear measurements. For this study, RMSSD (root mean square of successive differences), which reflects the difference between heart beats, was analysed to determine HRV. This is a time-domain measurement taking into account the time between normal heart beats. This measurement was chosen as it is successful when analysing short-term data (Baek et al., 2015; Esco and Flatt, 2014), such as the data gathered in this experiment. Furthermore, it reflects the influence of the parasympathetic system, which becomes active when stress decreases and is measured in milliseconds. The sympathetic system becomes active when stress increases, but effects are slower (measured in seconds) and therefore not suitable for HRV analysis. More detailed arguments for the decision to use RMSSD and how it was calculated can be found in Appendix G. Table 5.7 provides the RMSSD values per participant per session, and Figure 5.12 shows normal RMSSD values as provided by the Kubios user guide (Tarvainen et al., 2014).

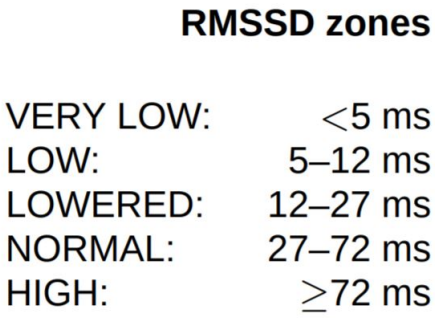


FIGURE 5.12: RMSSD zones provided by Kubios User Guide. Low zone indicates high levels of arousal and vice versa.

Participant	Interaction							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	24	24	26	22	28	31	22	18
2	18	26	22	23	26	23	24	28
3	19	17	18	13	19	15	18	12
4	73	70	54	45	35	53	62	59
5	79	129	132	154	74		118	
6	9	43	20		26	17	25	23
7		16	41	43	34	25	35	17
8		45	47	55		32	92	35
9	21	51	54	55	30	43		
10	30	55	43	100	68	59	81	69

TABLE 5.7: RMSSD values for each interaction for all participants. Empty slots indicate faulty readings by the GSR sensor.

The Kubios software that was used to analyse the data provides a stress index based on the input. This stress index is based on Baevsky's work who introduced stress index as a measure of heart rate variability (Baevsky and Berseneva, 2008). Overall, participants' HRV values felt primarily within the typical, non-aroused range, as can be seen in Table 5.8. Even though Figure 5.13 shows that these values are slightly elevated, Baevsky argumented in the original work that elevated stress levels are observed by adults in rest. As the values provided by Kubios as derived from Baevsky's stress index, it can be assumed that therefore elevated stress levels observed in older adults in rest indicate they are calm and not emotionally aroused.

		Stress zones
	($\sqrt{SI}$)	(Baevsky's SI)
VERY HIGH:	$\geq$30	($\geq$900)
HIGH:	22.4–30	(500–900)
ELEVATED:	12.2–22.4	(150–500)
NORMAL:	7.1–12.2	(50-150)
LOW:	$<$7.1	($<$50)

FIGURE 5.13: Stress levels provided by the Kubios User Guide and original stress zones provided by Baevsky.

Participant	Interaction							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	13	12	14	15	8	10	16	15
2	14	12	14	14	14	16	15	12
3	14	16	16	15	18	17	17	22
4	8	8	9	10	12	12	9	9
5	8	3	5	5	8		5	
6		21	13	16	14	19	14	20
7		20	12	12	17	24	12	24
8		10	9	10		13	4	15
9	12	9	9	8	11	10		
10	10	7	8	7	7	7	6	6

TABLE 5.8: Stress level for each interaction for all participants. Empty slots indicate faulty recordings from the GSR+ sensor.

5.3.6.2 Electrodermal Activity

As mentioned in Section 5.2.6, the EDA data provided by Empatica E4 was unreliable. This data provided values ranging from 0.006 to 0.04 where average EDA values range from 1 to 20 μS. Although the GSR+ sensor provided reliable HRV data, the EDA data was still very low, ranging from 0.02 to 0.7. After consulting an expert colleague on EDA and the Shimmer Sensing user support it was confirmed that these are indeed not normal values, not even for older adults. After testing both devices on myself and colleagues, it was found that Empatica E4 still provided unreliable values, but GSR+ values now were within the normal range. Perhaps the galvanic skin response is lower for older adults (Psychophysiological Research Ad Hoc Committee on Electrodermal Measures et al., 2012). Nevertheless, as the unexpected values could not be explained, it was decided to not analyse this data further.

5.3.7 Speech Prosody

As mentioned in Section 5.3.3.4 fragments where participants were talking to the robot were extracted from video-recordings and turned into audio-files to analyse using the software

Praat Version 6.1.03 (Boersma, 2001). For each audio-file the minimum, maximum, mean and standard deviation of the pitch and intensity were calculated.

These files could only be compared within participants as pitch and intensity levels differ per person. As it was the intention to investigate whether the robot's behaviour influenced participants' responses, data was only analysed for the participants that saw both the emotive and non-emotive behaviours of the robot. As this whole process was quite time-demanding, it was decided to first analyse half of the interactions for each participants (two out of four interactions with the non-emotive robot and two out of four interactions with the emotive robot).

Paired samples t-tests were performed to investigate whether any of the features investigated for pitch and intensity significantly differed when interacting with the robot displaying non-emotive behaviour compared to the robot displaying emotive behaviour. The exact *t*- and *p*-values of pitch and intensity for each participant can be found in Appendix D.

Results showed that for two participants, pitch-features significantly differed depending on the displayed robot behaviour. For one participant, the mean pitch was significantly higher ($t(19) = 6.75$, $p < 0.01$) when interacting with the emotive robot ($M = 447.79$) compared to the non-emotive robot ($M = 221.62$). For the other participant, the standard deviation of the pitch was significantly lower ($t(12) = 2.37$, $p = 0.04$) when interacting with the emotive robot ($M = 44.08$), compared to the non-emotive robot ($M = 85.01$).

Looking at intensity-features, results show that for two participants the minimum intensity was significantly lower ($t(19) = -2.37$, $p = 0.03$ and $t(12) = -2.93$, $p = 0.01$) when interaction with the non-emotive robot ($M = 11.46$ and $M = 8.27$) compared to the emotive robot ($M = 19.81$ and $M = 21.00$). For one participant, the maximum intensity was significantly higher ($t(18) = 3.41$, $p < 0.01$) when interacting with the non-emotive robot ($M = 65.95$) compared to the emotive robot ($M = 59.82$). Finally, for one participant the mean intensity was significantly higher ($t(20) = 2.15$, $p = 0.04$) when interacting with the non-emotive robot ($M = 43.09$) compared to the emotive robot ($M = 38.70$).

For all these participants, only the features mentioned were significantly impacted by AEE, and none of the other features that were investigated. This indicates that AEE did not

appear to systematically impact participants' speech prosody. Therefore, it was decided to not analyse the remaining interactions.

As mentioned before, it was decided to not analyse duration of the speech fragments, as they were limited due to the nature of the interaction. However, the amount of words used for each interaction was analysed, to investigate whether different robot behaviours evoked less or more elaborate responses from participants. Initial transcriptions of these recordings show that participants asked the robot more often to repeat itself when no subtitles were provided, as it was only asked once when subtitles were provided as compared to 15 times when subtitles were not provided. All four participants in the first setting asked the robot to repeat itself, indicating that it was not due to one participant being hard of hearing. This strengthens the observation that participants had trouble understanding what the robot was saying. It was also noticed that participants would blame themselves for not understanding the robot, claiming that they either needed a hearing aid or that the hearing aid they were wearing must be the problem. This again shows the importance of a social robot being accessible to all its users, as interactions to benefit the user should never result in the user feeling bad about themselves.

Not providing subtitles resulted in relatively richer interactions, with participants commenting more often on what the robot was saying without it asking them a question when subtitles were not provided, as can be seen in Table 5.9. However, the number of words that participants use during the interaction appear to be fewer for participants that had no subtitles, as also shown in Table 5.9. Perhaps this is due to the fact that participants with subtitles have a reminder of the question or comment from the robot and therefore are able to provide more elaborate answers.

	Comments		Words	
	M	*SD*	*M*	*SD*
Control (no subtitles)	20.00	5.57	40.00	15.88
Test (subtitles)	12.13	12.01	45.00	10.85

TABLE 5.9: Mean and standard deviation for total amount of comments and words spoken per interaction for each participant group (test x control). 'Comments' are statements or utterances from the participant without the robot asking them a question. 'Words' is the total number of words used by participants during an interaction.

5.3.8 Behavioural Analysis

As mentioned before, behaviour analysis could provide additional insights on why specific findings occurred. It was used in this research to determine whether behavioural analysis is a feasible measure for longitudinal field studies with participants that may have accessibility issues that prevents them from filling in questionnaires. For example, if physiological data showed increased levels of arousal, behavioural data could provide richer insights on whether this arousal was positive or negative. Therefore, interactions were video-recorded, so participants' behaviour could be analysed. However, once the first video-recordings were coded, it was found that participants responded on the content of the interaction itself (e.g. laughing when the robot asked them whether they knew how long the Great Wall of China was and they guessed wrong) rather than solely the robot's behaviour. Therefore, although it may still be a useful metric for HRI that can provide many useful insights, this data was not analysed further in this research. An example of the coding system that was used to code the video-recordings can be found in Appendix F.

5.3.9 Incidental Findings

As mentioned in Section 5.2.6 I initially decided to not use the tablet for the field study with older adults as this could distract them from the behaviour the robot was displaying. However, I soon found that participants had issues with understanding what the robot was saying. This was not caused by the robot's volume as participants reported the volume was loud enough and I did not need to speak louder for them to be able to hear me. This made me realise that there may be an issue with the robot's pronunciation which may cause issues when overall hearing deteriorates. As it may have been the case that participants had to get used to the robot's way of speaking I waited for another session. However, as there was no improvement participants were provided with subtitles that were shown on the robot's tablet. As expected, I observed that participants focused less on the robot's behaviour and more on the subtitles that were displayed on the tablet, as they would provide an answer to a question when they completed reading the question on the tablet, even if the robot had not completed the question yet. Participants that reported not being able to hear the robot without subtitles all indicated it was likely to be their own fault, either claiming they needed a hearing aid if they did not have one or claiming the hearing aid was preventing

them from being able to hear the robot.

Furthermore, their responses seemed less elaborate, as the amount of comments per interaction was lower for participants from Village B that were provided subtitles on the tablet compared to participants that were not provided subtitles. However, looking at the average amount of words used per session, participants that were not provided subtitles used less words on average per session than participants that did receive subtitles. It is possible that the number of comments was higher for participants that did not receive subtitles as they had to ask the robot to repeat itself. On the other hand, participants that received subtitles had a reminder of what question the robot asked which may have allowed them to give more elaborate responses.

5.3.10 Discussion

5.3.10.1 Emotional Deception

To determine the possible occurrence of emotional deception during the interactions, levels of anthropomorphism and to what extent the robot was seen as a social entity was gathered. Furthermore, participants were asked whether they believed the robot was capable of communicating emotions of happiness and sadness. Looking at these measures, it appears that emotional deception may occur, but only at a low level. No evidence was found indicating that participants anthropomorphise the robot differently depending on the behaviour it was displaying. These findings contradict earlier work, where anthropomorphism was higher for a robot that non-verbally displayed emotions compared to a robot that did not (Eyssel et al., 2010). However, Figure 5.8 indicates that anthropomorphism was slightly higher for participants that interacted with the emotive robot. No conclusions can be drawn as this difference was not found significant. However, due to the low number of participants in this study it may be worth investigating further.

Participants did not appear to change their opinion whether Pepper was capable of communicating an emotions of happiness or sadness depending on the behaviour it was displaying. It is possible they did not change their opinion, but this may also occur due to the fact that they were focusing on the subtitles on the tablet and therefore did not notice any differences. However, the scoring for the robot being able to communicate an emotion of happiness or sadness was low to medium (M = 2.73 and M = 2.51 on a

5-point scale respectively). As this approaches the 'medium' level it is possible that some level of deception occurred. Also, participants that interacted with a robot that displayed both emotive and non-emotive behaviour perceived the robot more as a social entity than participants that interacted with a robot that displayed non-emotive behaviour only. This perception of social presence stayed the same over time, where earlier research had found that perceived social presence decreased over time (Leite et al., 2009). However, they argued that the addition of robot 'memory' to the interaction may prevent this decrease. Therefore, memory was included in this study, as the robot would recall what topics were discussed during the previous session. No decrease in social presence over time was found, supporting their argument. However, it should be noted that even though the experiment time-frame was similar, their participants were children and they used a zoomorphic robot, which may both be reasons why the results differ.

It can be concluded that emotional deception may have occurred at a low level. No negative results indicating that participants' mood decreased or they became distressed were found, indicating that the impact of this potential deception appeared minimal. However, this is not unexpected, as AEE was designed to be low, and the chosen interaction topic was neutral and not discussing any personal experiences. This was decided to the impact of 'baseline' AEE, as this impact already may be grave when considering vulnerable populations.

5.3.10.2 Emotional Attachment

The level of attachment stayed the same for the duration of the experiment for all participants. This means however that participants who became attached to the robot early stayed attached to the robot, therefore potentially being at greater risk of experiencing negative consequences. Four participants reported they would miss the robot and nine participants indicated they would be willing to use the robot. According to the results of the attachment questionnaire, most of these participants showed only low levels of attachment to the robot. This raises questions on how to measure emotional attachment. The question whether participants would use the robot if they could in the future was added as frequency of use is mentioned as one way to measure attachment (Li, Browne, and Chau, 2006). However, most of these participants indicated in the attachment questionnaire that they were not very attached to the robot. It should be investigated in the future why this occurs. Perhaps

participants answered positively to the questions in the interview because they felt more inclined to provide a socially acceptable answer as they were answering to me personally instead of writing it down on paper. It is also possible that the indirect measurements of missing the robot and wanting to use it in the future are not suitable as measurements for attachment. However, as the open questions were based on existing literature as well it appears that emotional attachment to the robot is possible, and even though the number of participants was too low to draw any strong conclusions it was mostly participants that interacted with a robot that displayed both emotive and non-emotive behaviour who answered positively to the open questions. Therefore, it should be taken into account that artificial expression of emotion by a robot may increase one's level of attachment towards that robot. Attachment style did not indicate participants' level of attachment to the robot. As mentioned in Section 3.4, it was expected that participants with a fearful attachment style were more likely to become attached to the robot, but this was not found. However, the number of participants was low for this experiment, and only four participants had a fearful attachment style. Therefore, it may still be worth investigating this feature more in future work. No physiological responses or negative changes in participants' mood were found based on participants' level of attachment to the robot. However, as mentioned before arousal was designed to be low and the topic of conversation was chosen to be neutral. This indicates that low levels of AEE during didactic interactions may only have limited consequences, if any. However, this should be investigated further for higher arousal levels and different, more personal interaction topics.

5.3.10.3 Emotional Deception and Emotional Attachment

There were strong correlations between participants' level of attachment and both anthropomorphism and social presence, which indicates there may be a relation between emotional deception and emotional attachment. This finding is supported by the significant impact that participants' level of attachment had on their perception of the robot as a social entity, which was significantly higher for participants that were highly attached to the robot. Therefore, it appears that the construct 'social presence' proves the most useful to measure whether potential ethical risks may occur.

Female participants reported they felt the robot was more capable of communicating emotions than male participants. Also, 75% of the participants that indicated they would

miss the robot during the final interview were female, which was also reflected in the results of the pilot study described in Section 5.2. This implies that female participants may have been more sensitive to artificial expression of emotion by the robot. However, it is also possible that this finding occurred due to the characteristics of the robot (round features, big eyes), the can evoke feelings of maternal care (Vicedo, 2009).

5.3.10.4 HRI Measurements

This research focuses on older adults and although all participants were self-reported healthy, there may be older adults with cognitive or physical impairments in future studies for whom answering questionnaires can be challenging. Furthermore, questionnaires may not provide sufficient insights on the user experience (De Graaf, 2016). Therefore, different measures were added to this experiment to determine whether they can be used for longitudinal field studies, as questionnaires are useful but also have disadvantages. The measures added gathered physiological data, speech prosody data and behaviour data.

Physiological data may provide useful insights as it gathers objective data. However, it depends on the sensors that are used to gather the data whether the findings are reliable. Some sensors require expert knowledge and are difficult to use for experimenters with an HRI background. Other sensors may be more invasive for participants as they may have an allergic reaction. Also in HRI experiments it may be difficult to determine what causes the physiological reaction, as arousal may rise following both a positive and negative experience. Although this can be improved through sensors that may recognise emotions, for example through EEG signals, adding this measurement makes for a more invasive experience for the participant. The overall conclusion is that this type of data may provide useful insights in addition to other types of measurements but should not be used as a main focus in HRI research. However, it should be taken into account that non-invasive sensors that gather physiological data can be unreliable as they are very dependent on movement. Therefore, it depends on the task that participants have to perform whether this can be a suitable addition to their experiment.

Speech prosody could be a useful addition to longitudinal field studies studies to gather richer insights on participants' responses and experiences, without being inconvenient for either the participant or the experimenter. Focusing on the minimum, maximum, mean and

standard deviation of participants' pitch and intensity when interacting with the emotive or non-emotive robot, only few differences were found. Therefore, not all data was analysed. However, the fact that the low emotion manipulation of the robot's behaviour used in this research already resulted in significant differences, indicates that speech prosody data can be a useful metric to gather richer data during field studies. The microphone used to gather speech prosody data was the microphone from the video-camera. This may have impacted the results, as using a better microphone may retrieve better data. Also, as the video-camera was located close to the robot (as can be seen in Figures 5.6 and 5.7), there was a lot of noise from the robot's fans and motors that had to be reduced. The fact that there are differences for participants with limited data available and the use of a camera that was located close to the robot's motors to record the sessions is very promising for the input this type of data can provide when practised accordingly. It is easy to apply for non-expert experimenters and does not require any additional input from participants. The only thing that needs to be taken into account for this data, and for all additional measurements tested in this research, is that the participants' data is protected as biological data is now gathered which makes it easier to identify participants. However, with proper anonymisation methods and data storage this should be an acceptable impediment to overcome for the benefits it can provide.

Behaviour analysis appeared least successful during this field study. It takes a lot of time to code the behaviours, and more than one coder is required to account for potential subjectivity of the observer. These observers need experience in recognising certain behaviours, and the inter-rater reliability of the observers needs to be above a certain threshold to prevent coding bias. As AI advances, it may be possible to use an AI algorithm to code the behaviours. However, even with the use of such an algorithm the findings are difficult to interpret. As found while coding video-recordings, participants sometimes laughed when they felt uncomfortable and not due to the robot's displayed behaviour. This indicates that a lot of knowledge on context is required to be able to properly analyse participants' behaviour. Therefore, this was found to be the least accessible measurement for longitudinal field studies.

Finally, this research explored different metrics that could be useful for HRI research with vulnerable populations in addition to questionnaires. However, looking at all the measures used during this research, it should be noted that the IPANAT (implicit mood

questionnaire) did not provide many useful insights. This questionnaire was included as an implicit measurement of participants' mood, as the PANAS can be susceptible to responses that are deemed socially desirable. However, IPANAT did not provide useful insights in any of the studies conducted in this research. Furthermore, participants of the field study had issues understanding how to complete the questionnaire, as it was very different from the other questionnaires. Therefore, it can be concluded that IPANAT is not necessarily a useful addition to HRI research with vulnerable populations.

5.3.10.5 Insights from Running a Longitudinal Field Study

One strong aspect of this study was that it was conducted at the participants' homes instead of a laboratory, providing the opportunity to gather their responses in a setting where they are more comfortable compared to an artificial laboratory setting, which provides more realistic data. Also, conducting a field study allowed me to recruit participants with limited mobility who would not have participated if they study had been conducted in a laboratory setting. This resulted in a more exemplar participant group of older adults. Furthermore, this study provided many insights on running HRI field studies.

One thing to anticipate when running field studies is that it will not be possible to get a perfect setup for the experiment, as you have to work with what is available. First of all, it is more difficult to create an optimal environment required for the hardware that is used, like a high-speed internet connection that may not be available. Also, using different locations will result in different set-ups of the experiment room. For this study, one village provided me with a small bedroom where the robot and participants would be located between two single beds, and the other village provided me a whole apartment to work with. This results in different environments where distractions may be present, but can also result in people understanding the robot better or not due to acoustics that may differ depending on the size of the room. There are more external factors that may influence the experimental procedure. For example, as people were in their home-environment and not familiar with research settings and the significance of similar conditions for each participant, it happened three times that people would enter the room while an experiment was ongoing. Fortunately, this only occurred when participants were filling in questionnaires and thus limited the impact this had on their perception of the robot, but it should be taken into account that disturbances are more likely when running field studies.

Participants from lab studies were from different age groups that the participants from the field study, so observations made may not be solely due to experiment location. However, it was found that participants from lab studies generally arrived on time, maybe asked some questions at the end, but mostly focused on the experiment and then left. When conducting the field study, participants would arrive early (e.g. 'I was in the neighbourhood' or 'I just came back from shopping'), and would stay longer afterwards, either asking questions that were related to social robotics but not the experiment specifically as they understood I could not inform them on that before the debrief session, or just to chat in general. This means that more time is required to run an experiment in a field study than required when running a lab study.

Overall, these observations make me believe that participants feel more comfortable and free to behave as they usually would when conducting a field study compared to a lab study. They did not mind entering the room when an experiment was ongoing, they would arrive at times that suited them and approximated the agreed time, and it appeared they also felt more at liberty to express their opinion during the experiment. During this research and also my previous education, I have conducted multiple human-robot interactions studies, but this field study was the first time where participants did not mind confronting me personally. As mentioned before participants asked more questions, and even though I tried to limit interactions outside of the experiments it was unethical to not answer questions they had and give them the feeling they were not welcome except for the experiment. This resulted in conversations that deviated from the experiment. This was not unexpected, as participants indicated they looked forward to the sessions and aimed to prolong their experience. However, sometimes participants' comments became personal attacks and violated social norms. Even though this was an uncomfortable experience for me, it is positive that participants felt the liberty to state their opinion. This indicates they felt comfortable during the interactions they had with the robot and their responses were more natural than when the study would have been conducted in a laboratory setting. However, when preparing a field study, a risk assessment should be included on how a researcher should respond in case a participant becomes agitated or behaves inappropriately towards them, and perhaps extra precautions need to be taken depending on the population. This is essential to protect the participant, but also to protect the researcher and ensure they cannot be blamed for consequences of the participant becoming agitated.

5.4 Summary

This chapter aimed to investigate the impact of AEE on older adults, and how this impact could best be measured. First, a pilot study was conducted to guide the design of the longitudinal field study with older adults. It was found that the chosen topic of Wonders of the World appeared suitable for the field study. This study consisted of participants that were on the baseline of being categorised as vulnerable. They were cognitively healthy, but not able to live independently. Hence, they lived in semi-independent retirement villages where support could be provided if needed. The following research questions were investigated through this field study:

1. 'Does AEE impact older adults' human-robot interaction experience?'
2. 'Does the impact of AEE on older adults change over time?'
3. 'Can physiological data, speech prosody data and behavioural data provide valuable insights when used in longitudinal field studies in addition to questionnaires?'

Question 1 was addressed by investigating whether emotional deception and emotional attachment occurred during the interactions. Emotional deception was measured through participants' anthropomorphism, perceived social presence, and the robot's perceived ability to communicate emotions of happiness and sadness. Emotional attachment was measured through the attachment questionnaire, participants' intention to use the robot and their response to the question whether they would miss the robot.

As mentioned in Section 5.3.10, the results indicated that it is possible that older adults were, to a small extent, emotionally deceived by the robot. Emotional attachment did also occur for some participants. Therefore, it can be concluded that AEE can impact the experience for older adults, although this does not always have to occur. No apparent negative consequences were observed through the results of this study. Participants' mood was not negatively impacted by the robot, nor were there any physiological responses found. The differences found in speech prosody data were not consistent enough to determine whether participants were negatively impacted by AEE. Participants were given the opportunity to reach out and ask for another meeting with Pepper if they preferred so. This study has been completed for over a year at the point of writing this, and no

participants have ever reached out to see Pepper again, indicating participants were not attached to the robot to the extent that they became distressed once it was gone.

Question 2 was addressed through gathering data at several points during the experiment. This is presented in Table 5.3 for questionnaires. Other types of data (physiological data, speech prosody data, behavioural data), were gathered for every interaction. No impact of time was apparent from the results, except for the finding that explicit negative mood decreased over time. As negative mood represents feelings such as being nervous or tense, and participants were likely nervous at the start of the experiment, this finding is expected. As this finding was consistent for all participants and did not depend on the condition they were assigned to, it can be concluded that AEE a potential impact of AEE did not change over time.

Question 3 was investigated by gathering different types of data. Besides questionnaires, physiological data, speech prosody data and behavioural data and physiological data were collected. As discussed in Section 5.3.10, speech prosody data appeared the most promising and consistent as an additional measurement for longitudinal field studies that involve vulnerable populations. It depends on the nature of the interaction whether physiological data can be useful, as the sensors required are either sensitive to movement or are more invasive for participants and require a certain level of expertise. Behavioural data may provide rich insights, but is also very time consuming and requires a certain level of expertise when coding the behaviours. Furthermore, it may be difficult to determine whether the behaviour is a consequence of a specific action, or whether the participant reacted to something else, especially when vulnerable people are involved. Therefore, this type of data was found to be the least useful for longitudinal field studies with vulnerable populations.

The overall aim of this research was to develop a framework that can help developers and producers of future social robots to design social robot behaviours that have minimal negative consequences. The findings from all studies described so far, together with general observations, resulted in that framework This will be presented next in Chapter 6.

6 A Framework for Ethical Artificial Expression of Emotion

Chapters 4 and 5 described the user studies that were conducted to investigate participants' responses to AEE. The findings of these studies contributed to the development of the framework on AEE, which will be presented in this chapter. An introduction to existing frameworks is provided in Section 6.1. From here, it is evident that considerations regarding ethical risks that may arise from artificial expression of emotion by robots are limited. The development and presentation of the framework based on the findings of the user studies conducted in this research are discussed in Section 6.2. Even though these user studies lay a foundation for the framework that is being introduced in this chapter, there are some limitations inherent in these studies (e.g. low sample size) that question whether the findings can be generalised. A final online survey was conducted to investigate how the findings from the user studies were reflected in the opinion of the general public. Hence, the research question addressed in this chapter is: *'How do the results found in this research compare to the opinion of the general public?'*. The development and results of this survey, presented in Section 6.3, were used to update the initial framework. The updated framework is presented in Section 6.4. Finally, the framework and conclusions from the survey are summarised in Section 6.5, which also aims to address the research question *'Can the findings of this research be used for the development of a framework on ethical AEE?'*.

6.1 Existing Frameworks

Several frameworks have been developed in the field of autonomous and intelligent systems generally and, more specifically, in the field of social robots. First, some of the more generally known ethical frameworks on autonomous and intelligent systems will be introduced, followed by a presentation of existing frameworks for social robots. It will also be highlighted how the framework provided in this chapter will address gaps in these

existing frameworks.

6.1.1 Frameworks on Autonomous and Intelligent Systems

In the last few years, a large number of ethical documents have been produced that provide principles and guidelines on ethical administration of autonomous and intelligent systems. Examples are Ethically Aligned Design (EAD, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2017), the Guide to the ethical design and application of robots and robotic systems (BS 8611:2016, 2016), and the ethics guidelines for building trustworthy AI that was developed by the High-Level Expert Group on Artificial Intelligence, which was set up by the European Commission in 2018 (AI HLEG, 2019). These documents cover topics ranging from AI to autonomous systems to robotic devices, of which social robots are a subset. The documents cover many ethical concerns (e.g. privacy, transparency, responsibility) that arise with the development and use of autonomous and intelligent systems at a general level, also taking into account the societal impact that these systems can have on humanity. For example, one guideline from the AI HLEG on societal and environmental well-being states that effects of AI systems should be carefully considered and monitored. The core message of this document, that these systems can cause mental harm, is reflected in BS8611 and EAD as well. BS8611 states that robots should never be designed to be deceptive, defining deception as 'the illusion of emotions and intent, which should not be used to exploit vulnerable users' (Boden et al., 2017). EAD highlights potential concerns of artificial expression of emotion by stating that artifacts that facilitate or participate in human society should not cause harm either by damping or amplifying a person's emotional experience.

These guidelines and statements indicate that researchers and other stakeholders of AI and autonomous systems are aware of the fact that autonomous and intelligent systems can have a psychological impact on humanity. However, some additional concerns that arise when deploying social robots have not been considered (Van Maris et al., 2019). For example, assessment provided by AI HLEG includes societal and environmental well-being, among others assessing whether it is clear to the user that the interaction with the system is simulated, and it has no understanding of emotions and feelings. This guideline can be extended to cover deception through AEE. However, in the way that this assessment is phrased, there is no distinction made between intentional and unintentional deception,

which may lead to oversights on the impact of unintentional deception when the assessment is used. Furthermore, in their guideline on technical robustness and safety, it is stated that systems should behave as intended and minimise *unintended and unexpected* harm. Actively searching for ethical issues while the systems are still being developed, will help determine what unintended and unexpected consequences may be. This is a recurring issue in the other documents, where unintended harm may be mentioned but there is no further clarification of how it can be identified in the first place.

Perhaps it is understandable that these general frameworks aim to cover important ethical concerns but do not provide details on how to address them, as this would impact the possibility to generalise the framework. However, a clear strategy is needed to analyse unintended concerns of social robots and AEE before they are fully deployed. This is among others addressed by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. This initiative resulted in several IEEE Standards working groups (P70xx) that focus on drafting new standards on the human during human-technology interaction (Winfield, 2019). For example, P7014 is a working group that specifically investigates potential consequences of emulated empathy in autonomous and intelligent systems [1]. However, in their project description, this group states that systems that can monitor and interact with people at a personal level can potentially be misused. This can result in both predictable and unexpected harm to the user, and standards on empathetic technology can help prevent the exploitation of people. Their topic of investigation is similar to the topic of this research, as they focus on both predictable but also unexpected harm to the user. However, in their interpretation, harm comes from expected behaviour in the form of misuse or exploitation of the technology, where in this research harm is regarded as an unexpected consequence of benevolent behaviour.

Working group P7010 is researching a well-being metric for autonomous and intelligent systems (Musikanski, Havens, and Gunsch, 2018). One of their aims is to establish well-being metrics that relate to human factors which are directly impacted by intelligent and autonomous systems, which is also related to the goal of this research. However, P7010 appears to focus solely on direct impact of these systems, and not in the unexpected and unintended psychological consequences that may occur.

[1] https://standards.ieee.org/project/7014.html

The arguments above indicate that the goal to actively search for ethical concerns in autonomous and intelligent systems is a work in progress. Focusing on AEE, which is in an early stage for social robots, it is essential that the potential concerns (both intentional and unintentional) are determined before these social robots are fully deployed, for the safety of the future users. Therefore, an ethical framework was developed, based on the findings from the user studies that have been discussed in the previous chapters. This framework may help future developers to identify where (un-)intended concerns are more likely to arise, when developing systems capable of AEE.

6.1.2 Frameworks on Robot Applications

Even though there currently appear to be no frameworks on designing specific social robot behaviours, similar works have been presented for the design of social robots and human-robot interactions. For example, the critical 'social' aspects of human-robot interaction design have been presented as robot attributes, user attributes and task structure (Mutlu et al., 2006). A different taxonomy that takes into account factors that may impact the effectiveness of human-robot interactions (Fong, Nourbakhsh, and Dautenhahn, 2003). In this taxonomy, artificial emotions displayed by a robot are presented as having an impact on successful HRI. Even though the survey does not provide underlying concerns for the factors they present, they do question whether we need to consider ethical concerns of socially interactive robots. This indicates that researchers were already aware of potential negative ethical consequences of social robots and their displayed behaviours back in 2003. This taxonomy was extended to include, among others, user populations, sophistication of interaction and the role of the robot (Feil-Seifer and Mataric, 2005), and expanded further to among other things include the ability to assume different roles depending on the user the robot is interacting with (Bertel and Rasmussen, 2013). However, their extensions did not provide more insights on AEE.

The literature even provides a taxonomy on robot deception and potential benefits of this deception on humans (Shim and Arkin, 2013). However, this taxonomy solely focuses on intentional deception.

Another framework focuses more on the well-being of the user than the robot's design (Sorell and Draper, 2014). However, this framework focuses on the impact that a social robot may have on the user's everyday life (e.g. level of autonomy, level of independence,

privacy), and not on the psychological impact that certain behaviours of the robot itself may have on the user.

These frameworks indicate that, within social robotics research, potential concerns of deception and displaying emotions have already been recognised. However, knowledge regarding the impact of these concerns is limited, especially when considering the unintentional psychological impact that AEE by social robots may have on user experience.

6.2 Developing a Framework

Before being able to develop the framework, the important findings from the user studies conducted in this research had to be determined. The main findings from the user studies are the following:

- The oldest participant age group reported the robot appearing less happy or sad than participants of other age groups.
- Depending on a person's personal conditions (e.g. specific needs due to disabilities), they may be more or less inclined to use the robot. This can either be due to the robot not providing the functions required to assist the person, or due to the robot providing too many options that the person does not need yet.
- Participants of the longitudinal field study had issues understanding what the robot was saying when no subtitles were provided.
- It is possible that some level of deception occurred, as some participants reported the robot was moderately capable of communicating happiness or sadness.
- It was not apparent from the qualitative and quantitative measures used in this research that AEE negatively impacted user experience.
- The feeling of attachment was low for most participants, and overall it appeared that participants' level of attachment did not change over time. However, this indicates that the few participants that were highly attached to the robot remained highly attached to the robot for the duration of the experiment.

- The strong positive correlation between participants' level of attachment and the extent to which they anthropomorphise the robot, together with their perception of the robot having a social presence, indicates there may be a relationship between emotional deception and emotional attachment, where deception is more likely when attachment is higher and vice versa.

6.2.1 The SOCRATES Framework

As can be concluded from the literature, there are no set guidelines on what an ethical framework should look like or how it should be developed. Frameworks have been presented in the form of flowcharts, tabular frameworks (e.g. BS8611), lists (e.g. AI HLEG) or as a phased framework (De Graaf, Ben Allouch, and Dijk, 2018). Therefore, it first had to be determined what type of framework would best suit the findings of the user studies conducted in this research. A tabular framework provides a clear overview of the components of the framework with its descriptions. A disadvantage of this type of framework is that these components and their descriptions are often generalised to cover as many applications as possible, which may result in unintended or unexpected interpretations of the components. A flowchart is binary and easy to interpret. Therefore, it was decided to start and develop the framework in the form of a flowchart. The draft of this flowchart can be found in Appendix I. However, it soon became clear that the results from the user studies were not suitable for a binary approach. For example, one factor to be included in the framework is the user's cognitive ability. This is not a binary value where the user is cognitively healthy or not, and different levels of deteriorated health require different approaches. Instead of becoming the framework for this research, the flowchart provided insights on the risk-level for users to experience negative consequences of AEE.

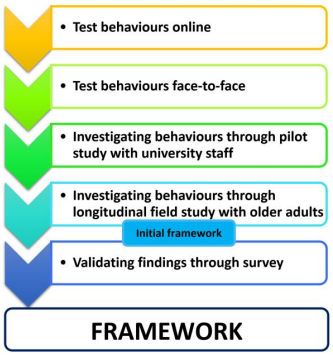


FIGURE 6.1: Current stage of framework development.

The next approach was to develop a tabular framework that presents components important for considering negative consequences of AEE. This approach was inspired by existing ethical toolkits such as the Ethical Toolkit for Engineering/Design Practice (Vallor, 2018). These toolkits include reflective questions that provide richer insights regarding

the technology being developed, as well as suggestions and roadmaps on how potential concerns can be mitigated. As ethical concerns following from AEE can be unintended, it is essential that developers and other stakeholders critically reflect on potential negative consequences of AEE. This can be addressed through contemplative questions at the start of and during development. Therefore, such questions have been included in the framework.

6.2.1.1 Population

The first component of the framework entails the intended target population for the robot that displays AEE. The WHO introduced two main factors that may indicate whether there is a risk of negative ethical consequences through the international classification of functioning, health and disabilities framework (Stucki, 2005). They define disabilities and health as outcomes of an interaction between health conditions (e.g. injuries) and contextual factors, where contextual factors can either be environmental factors such as the physical habitat or personal factors such as demographics. Taking this into account, together with the findings and observations from the user studies conducted where participants responded differently to the robot depending on their age or their personal needs, the first component of the framework can only entail the *population* for which the robot is intended. It is essential to consider the target population before and during development of a social robot, as not the robot's behaviour but the user's response to this behaviour may lead to a negative consequences. The robot's behaviour will likely be predictable. Even if an unexpected event occurs, it is likely that a reason for this unexpected behaviour can be found when transparency guidelines were followed while the robot was being developed (e.g. BS 8611:2016, 2016). However, human behaviour is not always predictable, where people can respond differently in two identical situations depending on their internal factors (e.g. mood). Therefore, the first recommendation of this framework to people involved with the development of a social robot is to *determine the target audience*. This component aims to identify factors that could categorise a person as vulnerable. Many frameworks include vulnerable people and the additional risks for them in their framework, without specifying when a person is vulnerable (Collins, 2017). Furthermore, frameworks often specify additional risks for vulnerable populations, without considering these risks may occur for people without health impairments as well. For example, there are principles

that indicate that deception in robots should not be used to exploit vulnerable users (Boden et al., 2017). However, it is not difficult to imagine that a healthy person that has no experience with social robots may be deceived by its behaviour as well, especially if this deception is an unintended consequence of AEE. Therefore, even though it is true that deception should not be used to exploit vulnerable users, I argue that this guideline applies to any user and not solely vulnerable users. The 'population' component of the framework aims to address these issues by forcing users to specify their target audience.

Health Impairments One factor that can possibly result in a negative outcome is if the user has health impairments, as such impairments may potentially disrupt intended outcomes of robot functions. Impairments can be either physical, cognitive or both.

Cognitive Impairments Once the target audience has been determined, one should consider *whether this audience includes people with cognitive impairments*. Consider a person that has dementia. Due to their impairment, they may be less able to realise that the robot is a machine that artificially expresses emotion. Therefore, they be more easily deceived by AEE. This can potentially lead to false expectations and the person over-trusting the robot, which may result in a negative outcome. It is found that people with dementia may misread the relationship with their caregivers resulting in higher levels of dependency (Edberg and Edfors, 2008). It is not difficult to imagine something similar happening when people with dementia interact with a social robot. Even more so if this robot is capable of AEE, which may give the appearance of empathy and a consciousness (James, Watson, and MacDonald, 2018). Due to the gradient nature of cognitive impairments it is difficult to assign a level of risk of negative outcomes when people with dementia are involved. However, it is assumed that the level of risk increases with the severity of the impairment. Therefore, it is essential that *cognitive abilities are assessed* and *a domain expert is consulted.* If the level of risk is high, it should be *considered whether the benefits of using the robot outweigh the disadvantages*, which can be measured through *quality of life* measurements (e.g. World Health Organization, 2013) and *additional measures* such as the ones used in this research. As cognitive impairments may deteriorate over time, these *assessments should be performed periodically*.

Physical Impairments Besides cognitive impairments, it should be considered *whether the target audience includes people with physical impairments.* If a person has a physical impairment, it is possible they cannot benefit from the support the robot can provide. They may feel the robot is not capable of providing the physical assistance required. This may lead to feelings of hurt and negligence, as they can use assistance but are disadvantaged over people with no impairments. The person should be approached regarding their needs and expectations, to *determine whether the robot is the best fit for them.* Compared to people with cognitive impairments, people with physical impairments may be more capable of determining the robot's true abilities and realise themselves that it may not provide the support they need. Therefore, the risk level of negative outcomes is lower for this group than it is for people with cognitive impairments. However, as mentioned there may be a negative psychological impact of the realisation that the robot cannot meet their needs. Therefore, the risk level of negative outcomes is medium for this user group. Some types of physical impairment may deteriorate over time. Therefore, it needs to be *periodically assessed whether the robot can still meet the person's needs*.

Cognitive and Physical Impairments People that have both cognitive and physical impairments are most at risk of experiencing negative outcomes of AEE. Where people with physical impairments may realise that the robot may not be able to meet their needs, this may be harder for participants that also have cognitive impairments. People of this user group should be *monitored closely* when interacting with the robot, to ensure negative outcomes are prevented.

Contextual Factors As mentioned before, not only health impairments but also contextual factors can impact a person's health. These contextual factors can be defined as personal factors or environmental factors. No indication can be given regarding the risk of negative consequences for contextual factors, as there can be large interpersonal differences for these factors.

Personal Factors Examples of personal factors are demographic characteristics and personality. For example, as the user studies conducted in this study have indicated, user's age can influence their response to and perception of the robot's behaviour. Furthermore, personal factors can be an indicator of potential health impairments. For example, age

may be a predictor of the occurrence of cognitive decline. However, even without any health conditions, personal factors may indicate the likeliness of negative consequences of AEE. For example, in the longitudinal field study, female participants reported more often they felt the robot was capable of communicating an emotion of happiness and sadness. In the pilot study that was conducted to develop the longitudinal field study it appeared that female participants became more attached to the robot than male participants (Section 5.2). As mentioned before, there may be a relationship between emotional attachment and emotional deception. Therefore, the findings indicate that female participants may have been more easily deceived by AEE than male participants. However, these factors should only be treated as indicators and *their impact should be assessed*.

It should be ensured that these factors do not bias expectations of developers and stakeholders on what users are potentially at higher risk of negative consequences. Building assumptions based on personal factors can lead to biased conclusions and is ethically incorrect to do. For example, people over the age of 65 years old are often categorised as 'older adults', and older adults are often categorised as 'vulnerable'. However, people of this age may be in good health and not have any health impairments. Therefore, it would be ethically incorrect to treat a person as vulnerable based on the fact that they are 65 years of age or older. Hence, *conclusions on a person's health should never be drawn solely on personal factors.*

Environmental Factors Finally, external factors that may influence a person's health should be considered. For example, a social robot can help people live independently for longer, specially if these people have no family to support them. However, receiving no support besides from the robot may result in users becoming overly dependent on the robot. They may ask the robot for support in every task, where family members may encourage them to do some tasks themselves. Furthermore, some domestic environments may not be suitable for robots, resulting in the user not being able to benefit from the support the robot can provide. Therefore, *the living situation of the user should be determined*, to ensure providing them with a robot is in their best interest.

In summary, it is essential to first determine the target audience when developing a social robot. Users are the least predictable factor of human-robot interactions, and depending on their health situation they may have very different needs and expectations.

Therefore, their expectations should be guided to be as true to the robot's abilities as possible, to prevent negative outcomes.

6.2.1.2 Artificial Expression of Emotion

This research focuses on artificial expression of emotion by a robot through the way it behaves, and the effects that this behaviour can have on the user. Therefore, it would have been reasonable to introduce AEE as the first component of the framework. However, first considering the target population and potential vulnerabilities of this audience facilitates the consideration whether AEE could raise concerns. Hence, AEE is introduced as the second component of this framework.

AEE through Verbal Interactions One of the reasons why the term 'Artificial Expression of Emotion' was used in this research is that it is possible to artificially express emotions through different means. One of these means is behaviour, as investigated in this research. However, artificial emotions can be expressed through verbal interaction as well. For example, imagine someone meeting a robot for the first time and the robot says: 'Hello, my name is X, I am pleased to meet you'. This is not difficult to imagine, and has likely been used in research already. After all, this is a universal introductory sentence when meeting someone for the first time. However, it should be considered that, even when human-like interactions with a robot are preferred, it *is still a machine and not a person*. Therefore, we are not meeting *someone* new and it is questionable whether the robot should behave as such. This behaviour may lead to expectations of the robot having some form of consciousness, which can result in emotional deception. The findings of the longitudinal field study with older adults indicated that a low level of deception may have occurred. This experiment was based on AEE through behaviour and the robot did not provide any personal opinions or preferences. Therefore, deception can only have occurred through the robot's displayed behaviour. If this is combined with a misrepresentation of the robot's abilities through verbal interactions, the risk of deception and ensuing negative ethical consequences increases. Taking this into account, and as expressed at the International Conference on Robot Ethics and Standards in 2019 (Van Maris et al., 2019), one suggestion is to avoid the use of 'I' combined with either an emotion or opinion where possible (e,g. 'I am happy', 'I think'). However, informative statements regarding the robot's actions

should not provide issues (e.g. 'I am going to my charging station').

Unintended AEE Accordingly, once the target audience has been established, one should ask themselves: *'Does AEE occur?'*. As was the case in this research, it may be the intention to make the robot artificially express emotion. However, it may also be possible that AEE occurs (mostly through verbal interactions) without it being intentional. Therefore, for each interaction it should be considered whether AEE occurs, and whether this is intentional. If it is not intentional, one should consider whether the desired outcome can be reached without the use of AEE. This can be addressed by considering whether it is likely that AEE will lead to false expectations, for example when the robot states an opinion, which is a misrepresentation of its internal state. If AEE is required, potential negative consequences may be prevented by providing additional information on the robot's abilities that can guide user expectations. This can be provided before an interaction, but it is possible that this information is repressed once the interaction starts. Therefore, it may also be an option to include an additional statement of the robot that fits within the interaction but provides more information on its internal state.

To summarise, one should consider for each interaction whether AEE occurs, as this may also occur unintentionally. If it occurs, the reason why should be established, and whether the desired outcome can be reached without AEE. If it is essential for the interaction, the target audience has already been determined, so steps to address potential negative outcomes can be taken accordingly.

6.2.1.3 Accessibility

One of the most important findings of the longitudinal field study with older adults was the incidental finding that participants had difficulties understanding what the robot was saying. This was not found before in experiments conducted at the lab, but does raise concerns about the robot's accessibility and users' accessibility needs (Van Maris et al., 2020a). This issue impacted participants' ability to interact with the robot. Subtitles were provided, as otherwise participants would not have been able to interact with the robot. Therefore, accessibility is an essential component of the framework. It should be noted that accessibility is not only a concern when considering AEE by social robots, but for assistive robots in general. One might argue that 'accessibility' should be a

sub-category of the component 'population' that was discussed before, as it revolves around the target population and possible impairments that the target population may have. However, as mentioned before there are different severity levels of both physical and cognitive impairments that require different approaches. Finally, the 'population' component addresses issues that may arise from the target population, where accessibility issues come from an issue in the robot that needs to be addressed.

Modalities Used for AEE That participants had issues interacting with the robot without subtitles, highlights the importance of addressing accessibility issues for successful human-robot interactions. Therefore, it is treated as a separate component of this framework. To establish potential accessibility issues, one should ask themselves *'What modalities are used to display AEE?'*. It is important that the robot is capable of adjusting to issues by using different and/or multiple modalities. This will allow the robot to be flexible in addressing and minimising accessibility issues. However, for all modalities separately, but also their combined use, it should be determined whether there are potential concerns. Similar to addressing health impairments, it is essential to consider that accessibility issues may occur at a later stage. *Recurring tests* should indicate whether a robot's flexibility and ability to address accessibility needs is sufficient over time, taking into account that for example a patient's level of dementia may progress to a more severe stage, resulting in an inability to use certain modalities.

6.2.1.4 The Initial SOCRATES Framework

The components of this framework are based on the findings from the longitudinal field study with older adults as described in Chapter 5. This initial framework is presented in Table 6.1. However, one impediment of conducting human-robot interaction studies, especially with a specific target group like older adults, is recruiting enough participants to ensure the conclusions that are drawn from the data are reliable. Furthermore, several observations were made that were worth investigating further and possibly include in the framework. Therefore, an online survey was conducted to investigate whether these findings were reflected and the observations supported in the opinion of the general public, which will be discussed next. Following from this, the framework could be updated and its validity increased.

Component	General recommendations	Reflective questions	Specific recommendations	Level of risk
Population	Determine the target audience.	Can the target audience include people with cognitive impairments?	Periodically assess level of impairment, consult domain expert.	Gradient, from medium to high
	Ensure that the benefits provided through AEE outweight negative outcomes.	Can the target audience include people with physical impairments?	Periodically assess whether the robot provides appropriate suppoert.	Gradient, from low to medium
		Can the target audience include people with both cognitive and physical impairments?	Continuous monitoring of mental health.	High
		Are there personal factors that may indicate vulnerability of AEE?	Assess factors that may indicate vulnerability of AEE	Dependent on personal factors
		Are there environmental factors that may indicate vulnerability of AEE?	Assess whether the user's environment is appropriate for social robot support.	Dependent on user environment
		Will the benefits of AEE outweigh potential negative outcomes?	Assess quality of life.	Risk increases with the number of potential negative outcomes
Artificial Expression of Emotion	Consider why AEE is used.	Does unintended AEE occur (either through behaviour or speech)?	Avoid statements where the robot expresses to experience an emotion or have an opinion.	Dependent on context
		Is the use of AEE needed?	Ensure benefits outweigh disadvantages.	ND
Accessibility	Consider what modalities are used for AEE.	Can modalities used for AEE lead to accessibility issues?	Periodically assess whether the robot's flexibility and ability to address accessibility issues is appropriate.	Dependent on population

TABLE 6.1: The Initial SOCRATES Framework

6.3 Evaluation of the Framework through the Opinion of the General Public

An online survey was conducted to allow for a large and diverse target audience. First, the survey was developed based on the findings and observations from the longitudinal field study with older adults. Next, it was distributed to a diverse audience to gather the opinion of the general public on these matters. Finally, the framework was updated to include the findings from this survey.

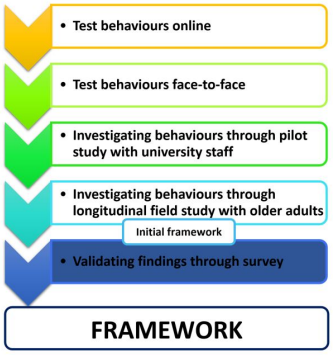


FIGURE 6.2: Current stage of framework development.

6.3.1 Development of the Survey

In order to build the survey, multiple statements were generated. Following from this, items were piloted with both lay people and experts in HRI. Based on this feedback, the survey questions were refined (e.g. 'It is acceptable for users to perceive a social robot as a social entity' became 'It is ok for people to think of a social robot as a living being') to ensure they were easy to understand for participants with no prior knowledge of social robots.

When the statements were finalised, it had to be determined what scale would be best for participants to indicate their level of agreement with the statements. The first option to consider was whether the scale should have a midpoint or not (e.g. 4-point scale versus 5-point scale). The expectation with an odd number of answers is that the average will always be adjacent to the central point of the scale (e.g. a mean of 3 on a 5-point scale with 1 = strongly disagree, 2 = disagree, 3 = neither disagree nor agree, 4 = agree, 5 = strongly agree), as this 'middle' answer provides participants with the option to pick the neutral option and not think about the statements. This issue can be accounted for when using a scale with an even number of answers, as the 'neutral' answer is left out in this case (e.g. a 4-point scale: 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). This forces participants to think more thoroughly and 'choose a side'. However, as this survey investigates a topic that participants may not have considered before, it is possible that participants are truly neutral about the topic and do not agree nor disagree with the

statement. Besides, excluding a midpoint forces participants to choose a direction (Tsang, 2012). Therefore, it was decided to use a 5-point scale.

A slider was used to try and minimise the impact that the aesthetics of the scale may have on participants' responses. For example, participants may agree relatively strongly with a sentence, but feel insecure about choosing the 'strongest' option and would therefore choose number '4' instead of '5' on a five-point scale. To account for this, the slider gave the impression the scale was continuous, even though the responses were recorded as numbers on a five-point scale. The starting position of the slider needed to be established next, as any starting position could potentially influence participants' responses. To try and prime them as little as possible through the location of the slider, it was originated at the centre of the scale. To account for this 'neutral' bias, participants only saw the labels 'strongly disagree' and 'strongly agree' on the ends of the scale, and no label above the starting position of the slider, which would be 'neither agree nor disagree'. See Figure 6.3 for an example.

FIGURE 6.3: Example survey question with a 'continuous' scale from strongly disagree to strongly agree, with the origin of the slider at the centre of the scale.

6.3.2 Participants

Initially, the survey was distributed using an online recruitment platform. However, it was found that the participants gathered through this platform were mostly young adults (age $M = 26$, $SD = 9$). As the earlier user studies had shown that participants of different ages perceived the robot differently and the aim was to get the opinion of the general public, it was important to ensure there was an even spread of age between the participants. Therefore, the survey was also distributed via email to university staff and friends and family, asking them to complete the survey and distribute it further while also ensuring them that participation was voluntary. In the end, 243 participants completed the survey. Four of these participants were excluded from analysis as they gave the same answer to

two or more questions that were both positively and negatively phrased. Therefore, data of 239 participants was used for data analysis. Participants' age ranged from 18 to 77 years old ($M = 41$, $SD = 18$). The distribution of participants per age group is shown in Figure 6.4. In total 137 participants reported being male, 99 female and 3 who either preferred not to respond or did not identify as male or female.

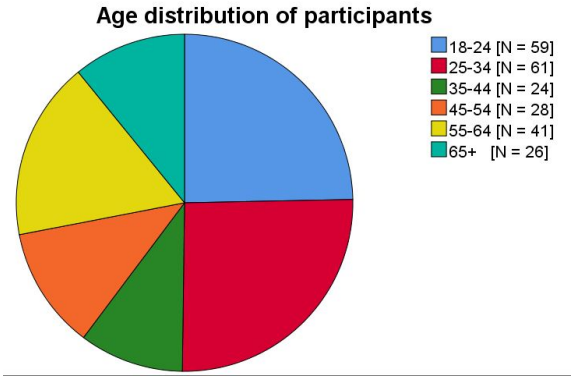


FIGURE 6.4: Distribution of participants per age group.

Participants were not very familiar with social robots ($M = 1.90$ on a 5-point scale, $SD = 0.90$), where 77% of participants reported being not or slightly familiar with social robots. 16% of participants reported being moderately familiar with social robots and 7% reported being very or extremely familiar with social robots. Although familiarity with robotic technologies was a bit higher, this was still relatively low ($M = 2.31$, $SD = 1.01$), with 62% of participants reporting being not or slightly familiar with robotic technologies. Furthermore, 26% reported being moderately familiar with robotic technologies and 12% reported they were very or extremely familiar with robotic technologies. Participants' familiarity with social robots and robotic technologies can be found in Figure 6.5. It should be noted that participants recruited through Prolific received a small amount of money for their participation, where participants recruited through email distribution did not. This may have resulted in different motivations to participate, which may have impacted the findings.

6.3.3 Materials

Qualtrics was used for the initial survey where the data was tested and also the final survey to investigate whether the findings from the user studies were reflected in the opinion of the general public. however, for the final survey participants were gathered through the

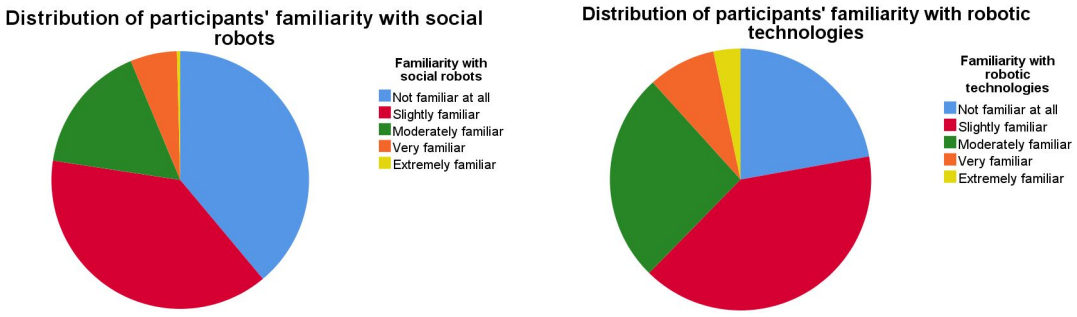


FIGURE 6.5: Distribution of participants' level of familiarity with social robots (left) and robotic technologies (right).

online platform Prolific.

6.3.4 Measures

First, the general demographics questions that were used for each study (age, gender, familiarity with social robots and robotics technologies) were presented to participants, with additional questions asking for participants' highest completed level of education, their current job, their nationality and their activity on social media. These were added to gather understanding of the participants and whether this reflected the general public.

As mentioned before, the statements were divided into five categories. The order of these categories did not change and was displayed in the same order as below. The statements in each category were randomly ordered.

The first category, 'Social robots displaying emotions', investigated whether participants found it acceptable for a robot to express artificial emotions, without taking into account specific considerations yet, such as specific target audiences. The category distinguishes between human-like (e.g. Pepper) and animal-like (e.g. AIBO, PARO) robots as there are advantages and disadvantages for both types of robot and both types of robot have shown to be useful (e.g. Kang et al., 2019; Allegra et al., 2018). No images of robots were provided so participants could rate the statements based on their mental model of social robots without being primed to one specific embodiment. Additional statements for animal-like robots and not providing images allowed generalisation of the findings from the longitudinal field study and ensuring the framework would not solely apply to the robot and didactic interaction style that was used during the user studies in this research. Furthermore, this category investigated what potential robot roles are accepted (companion,

friend, pet) and whether participants think that people will respond in a certain way to robots that express artificial emotions. These statements investigated if a certain level of emotional deception may have occurred during the longitudinal field study with older adults. The introduction and the final statements of this category that were provided in the survey are as follows:

- **Social robots displaying emotions:** Social robots can show emotions when interacting with people. For example, they can appear happy or sad. Showing emotions may result in a more pleasant experiences for the people interacting with the robot; however, they can also lead to people misunderstanding the robot's ability. Please rate to what extent you agree with the following statements:
  - Social robots can show emotions.
  - Social robots should show emotions.
  - People will trust a social robot that shows emotions.
  - People will overestimate a social robot that shows emotions.
  - It is acceptable for people overestimate a social robot's abilities.
  - It is acceptable for people to think of a social robot as their friend.
  - It is acceptable for people to think of a social robot as a companion.
  - It is acceptable for a social robot to look like a human.
  - It is acceptable for a social robot to behave like a human.
  - It is acceptable for people to treat a social robot as a human.
  - It is acceptable for a social robot to look like a pet.
  - It is not acceptable for people to think of a social robot as their friend.
  - It is acceptable for a social robot to behave like a pet.
  - It is acceptable for people to treat a social robot as a pet.
  - It is okay for people to think of a social robot that shows emotions as a living being.
  - It is okay for people to think of a social robot as a living being.

The following category was based on the finding that some participants of the longitudinal field study did become attached to the robot. This category asked participants about people becoming attached to a robot and whether they believe the level of attachment to a robot changes if it expresses artificial emotions.

- **Attachment to social robots:** If people interact with a social robot a lot, they may become attached to it; where they develop an affection for the robot. Please rate to what extent you agree with the following statements:
  - It is okay for people to become attached to a social robot.
  - It is okay for people to become attached to a social robot that shows emotions.
  - People become more easily attached to a social robot when they trust it.
  - It is acceptable for a person to become attached to a social robot as long as this person benefits from the robot.
  - It is not acceptable for a person to become attached to a social robot.
  - It is acceptable for a person to become dependent on a social robot when they are highly attached to it.

After taking into account people's level of attachment to a social robot, participants were asked to consider the effect that artificial expression of emotion may have on potentially vulnerable people. An example is people whose cognitive function is not optimal, as suggested by some observations I made as discussed in Section 5.3.9.

- **Vulnerable users using social robots:** Popular target groups for social robot development include potentially vulnerable people such as those with autism or older adults with possible dementia. The tasks that a social robot could perform would include for example: helping children with autism to develop their social interaction skills, and remind older adults with memory loss to drink enough and take their medicine. Please rate to what extent you agree with the following statements:
  - These people will overestimate a social robot's abilities if it shows emotions.
  - It is acceptable that these people overestimate a social robot's abilities.
  - It is acceptable for a social robot to show emotions when it is interacting with these people.

  - It is acceptable that older adults become dependent on a social robot if that means they can live independently for longer.
  - It is acceptable for these people to become attached to a social robot.
  - It is not acceptable for these users to overestimate a social robot's abilities.
  - It is acceptable that these people who are highly attached to a social robot become dependent on it.
  - It is acceptable that these people who are highly attached to a social robot treat it as a human.
  - It is acceptable that these people who are highly attached to a social robot treat it as a pet.
  - It is acceptable that these people perceive a social robot that shows emotions as a living being.
  - It is acceptable that these people perceive a social robot as a living being.

When asking participants of the longitudinal field study what kind of role they would like for a social robot during the final interview, some of them responded they would like for the robot to monitor their health and share it with their GP so they would not need to go visit them in person. This raised some questions regarding data collection and what people would like to be shared with others and what not. Therefore, the following statements regarding data collection were added to the survey:

- **Data collection by social robots:** Potential tasks for a social robot can be to remind people drink enough, take their medicine and monitor progress of rehabilitation. The robot can alert other people, such as family or the doctor, if needed. To do this, the robot will require access to personal data. Please rate to what extent you agree with the following statements:
  - It is acceptable for a social robot to store data about a person if this benefits the person over the long term.
  - It is acceptable for the anonymised personal data that the social robot uses to be stored and used by the robot developer for future design.
  - It is acceptable for a social robot to be used to monitor the progress and health of a person.

  - A social robot can be used to encourage older people to interact with others.
  - It is not okay for a social robot to be used to monitor the progress and health of a person.

Finally, the potential role of social robots, especially in a close proximity to vulnerable users, should consider how a robot communicates with the user. This whole research focuses on unintentional deception through artificial expression of emotion by showing certain behaviours. However, there is also *the context of what the robot says* that may be deceptive, some would classify this as lying, and should be investigated in more detail in the future. Nonetheless, the participants were shown an event where a deception through speech occurred, to determine the general public's opinion on a robot telling so called 'white lies'.

- **Lying social robots:** Currently social robots are being developed to assist caregivers with their jobs. There are situations where caregivers may lie to their patients. For example, imagine an older person who has dementia. This person might ask the caregiver when a loved one (who is deceased) will be visiting them. So as to not upset the patient, the caregiver may say that their loved one will visit later that day. Now imagine that a social robot is keeping the patient company while the caregiver is out of the room. The patient now asks the robot when their loved one will visit and the robot says that their loved one will visit later that day. Please rate to what extent you agree with the following statements:
  - There are situations where it is acceptable for a social robot to lie to a person.
  - There are situations where it is acceptable for a social robot to lie to a vulnerable person.
  - A social robot should never lie to the person they are interacting with, even if it appears to benefit that person.
  - A social robot should lie to the person it is interacting with f it appears to benefit that person.
  - A social robot should never lie to a vulnerable person, even if it appears to benefit that person.

After rating the statements from the five categories, participants were asked to answer the following open questions:

- Have you ever interacted with a social robot?
- If you would not use a social robot, what would be the primary reason and/or concern?
- If you would use a social robot, what would be the primary reason to do so?

And finally, participants were asked to indicate whether they would use a social robot (yes/no) and provide an explanation why they chose that option for the following questions:

- 'Would you use a social robot if....'
  - You had dementia?
  - You had a physical disability?
  - You had complex medical needs requiring medication?
  - You felt lonely?
  - Your family/partner recommended it?
  - Your family/partner did not like the robot?

The complete survey can be found in Appendix H.

6.3.5 Experimental Procedure

Participants were first provided with the information sheet and asked for their consent. They were then provided with the statements, which were divided into five categories that each represented findings and/or observations from the longitudinal field study with older adults:

- **Social robots displaying emotions:** As Artificial Expression of Emotion is the main topic investigated in this research, it was essential to start the survey with statements on this topic, also providing statements on potential deception.
- **Attachment to social robots:** As results indicated that some participants became highly attached to the robot, participants of the survey were asked whether they believed this to be a problem.

- **Vulnerable users using social robots:** The target group of this research was older adults, who may be more vulnerable to negative consequences of certain robot behaviours when there is a cognitive decline. Therefore, participants were asked to provide their opinion whether they believed potentially vulnerable users would respond differently to a robot and whether this would present issues.

- **Data collection by social robots:** The final interview of the longitudinal field study with older adults indicated that participants would not mind the robot gathering personal data from them, if that resulted in positive outcomes such as not needing to physically meet the doctor when health measurements could be forwarded by the robot. Even though data and privacy is not a main focus of this research, it was deemed important to investigate further whether the general public was of the same opinion.

- **Lying social robots:** Lying is a form of deception. Emotional deception in this research is a form of unintentional deception, where lying is a form of intentional deception. However, there may be situations where lies and therefore intentional deception are beneficial, an example will be provided later. As population is one of the components of the framework and lying may potentially be beneficial when a robot interacts with vulnerable people, this was included in the survey to investigate in more detail.

The order in which these categories were presented in the survey was always the same. However, the order of the statements in each category was randomised. For three statements, a reverse worded version was added to the survey in order to identify participants who were potentially not reading every question. Whilst there are potential issues with reverse wording, as this survey was administered online, it was important to include these items. It was decided to add three statements as it would be possible participants would misread or misunderstand a statement while still paying attention.

The survey ended with some open questions where participants were asked whether they would use a social robot in different scenarios. After the survey was completed, participants were asked once again whether they consented to their data being used for analysis. If they selected ‘no’, their data was deleted.

6.3.6 Results

First, internal consistency analyses were conducted to determine the reliability of the survey. Results showed high reliability for the categories 'social robots displaying emotions', 'attachment to social robots', 'vulnerable users using social robots' and 'lying social robots' and acceptable reliability for the category 'data collection by social robots'. The Cronbach α for these categories can be found in Table 6.2.

	Cronbach α after data collection	Cronbach α after item removal
Social robots displaying emotions	0.89	N/A
Attachment to social robots	0.79	0.81
Vulnerable users using social robots	0.83	0.86
Data collection by social robots	0.67	0.70
Lying social robots	0.85	N/A

TABLE 6.2: Reliability (Cronbach α) of each survey category before and after item removal. N/A indicates no items were removed for that category.

These analyses indicated that for three categories it would be worth considering to remove a question, as the correlation of this question with other questions was low and removal would increase overall reliability for these categories. The items that were removed from analysis were the following:

- **Attachment to social robots:** 'People become more easily attached to a social robot when they trust it.'
- **Vulnerable users using social robots:** 'These people will overestimate a social robot's abilities if it shows emotions.'
- **Data collection by social robots:** 'It is not okay for a social robot to be used to monitor the progress and health of a person.'

Even though reliability was already high for the categories 'Attachment to social robots' and 'Vulnerable users using social robots' (Cronbach $\alpha > 0.70$ is usually perceived as high reliability), one statement was excluded for both categories as this further increased reliability and also reduced the length of the survey, making it more efficient. For the categories 'Social robots displaying emotions' and 'Lying social robots' no items were removed as removal would either decrease reliability or increase it such that Cronbach α

$>$ 0.90, which may indicate redundancy and therefore would result in the survey being less efficient. No total score of all statements was calculated as the categories entail different topics and it would not make sense to sum all statements together. Therefore, reliability of all statements combined was not analysed.

6.3.6.1 Descriptive Results for Each Category

The averages for most categories of the survey are close to the centre of the scale (3; neither agree nor disagree), except for the category 'Data collection by social robots' that leans more towards agreement with the statements that were provided in this category. These findings are visualised in Figure 6.6.

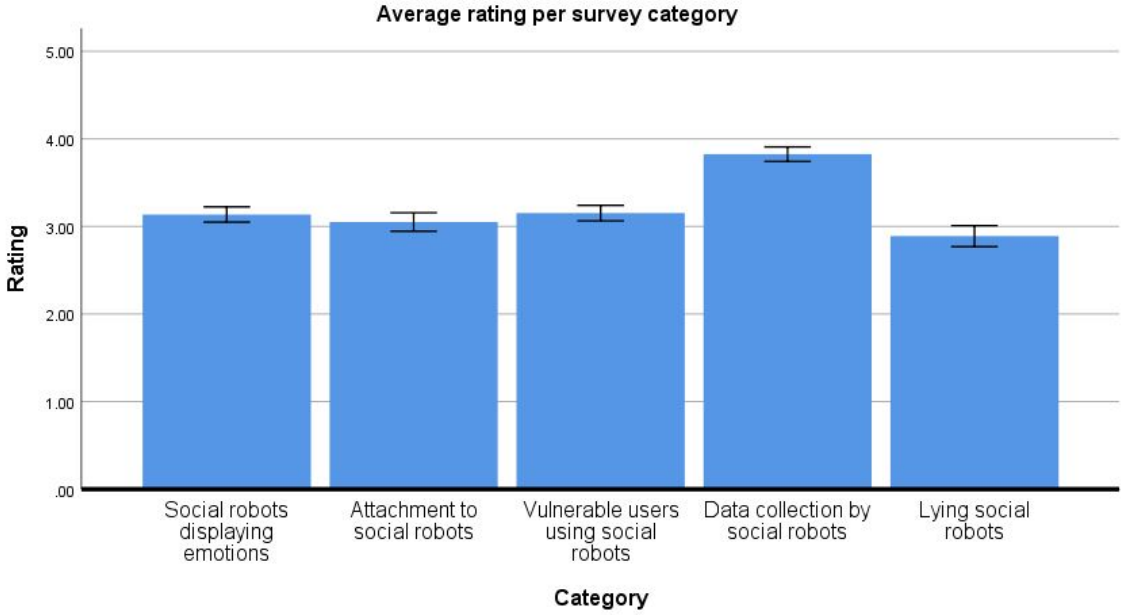


FIGURE 6.6: Mean and standard deviation for each survey category. A low rating means a low level of agreement (i.e. 1 = strongly disagree, 5 = strongly agree).

Pearson product-moment correlations were conducted to investigate the relationships between the survey categories. Significant positive correlations were found between all categories except between 'Social robots displaying emotions' and 'Lying social robots'. The results can be found in Table 6.3. Specific *p*-values for each correlation can be found in Appendix H.

Pearson product-moment correlations and multiple linear regression analyses were conducted in order to investigate individual differences. Age, gender and familiarity with social robots and robotic technologies were analysed. Whilst the intention had been to

	1	2	3	4	5
1: Social robots displaying emotions	1	0.72**	0.61**	0.20**	0.10
2: Attachment to social robots		1	0.63**	0.16*	0.14*
3: Vulnerable users using social robots			1	0.39**	0.32**
4: Data collection by social robots				1	0.15*
5: Lying social robots					1

*p $<$ 0.05, **p $<$ 0.01, N = 239

TABLE 6.3: Pearson product-moment correlations between survey categories.

explore ethnicity and level of education, the sample was not diverse enough to analyse this. Of all participants, 59% came either from the United Kingdom of The Netherlands. The other 41% was divided into many different countries such as Australia, Spain, Italy, Brazil, USA, Canada, Poland and more. As this resulted in a highly unequal distribution, I decided to not analyse the impact of cultural background on participants' responses.

Participants were asked to provide their highest completed degree as an open question. Taking into account that participants from different countries with different education systems took part in this experiment, it was not possible to clearly categorise participants with the knowledge I had on the education systems in these countries. However, taking into account that participants recruited through Prolific are often well-educated and the survey was distributed further through university emailing lists, it can be assumed that the education level of participants overall was relatively high.

6.3.6.2 Age

In order to investigate the relationship between participants' age and their responses to the survey categories, Pearson product-moment correlations were conducted. A strong negative correlation between the category 'Social robots displaying emotions' and participants' age was found ($r(239)$ = -0.21, $p < 0.01$), indicating that level of agreement with the statements in the category 'Social robots displaying emotions' decreased as age increased. A one-way ANOVA supported this finding as by a weak influence of age on participants' level of agreement with the statements in this category was found ($F(5, 233) = 2.67$, $p = 0.02$, $\eta_p^2 = 0.05$).

Post-hoc pairwise comparisons indicated that participants aged 65+ agreed significantly

	Correlation		**One-way ANOVA**		
	r(239)	*p*	*F*(5, 233)	*p*	η_p^2
Social robots displaying emotions	-0.21	**0.01****	2.67	**0.02***	0.05
Attachment to social robots	-0.06	0.33	1.33	0.25	0.03
Vulnerable users using social robots	-0.10	0.11	1.89	0.10	0.04
Data collection by social robots	0.01	0.96	0.29	0.92	0.01
Lying social robots	0.10	0.12	1.18	0.32	0.03

*p $<$ 0.05, **p $<$ 0.01

TABLE 6.4: Relationship between age and the survey categories (correlation) and age differences by survey category.

less with the statements of the category 'Social robots displaying emotions' than participants from all other age groups, except for the participant group aged 55-64. Furthermore, participants aged between 55 and 64 agreed significantly less with the statements from this category than participants aged 18-24. No other significant differences were found. Average ratings per age group for each category can be found in Figure 6.7.

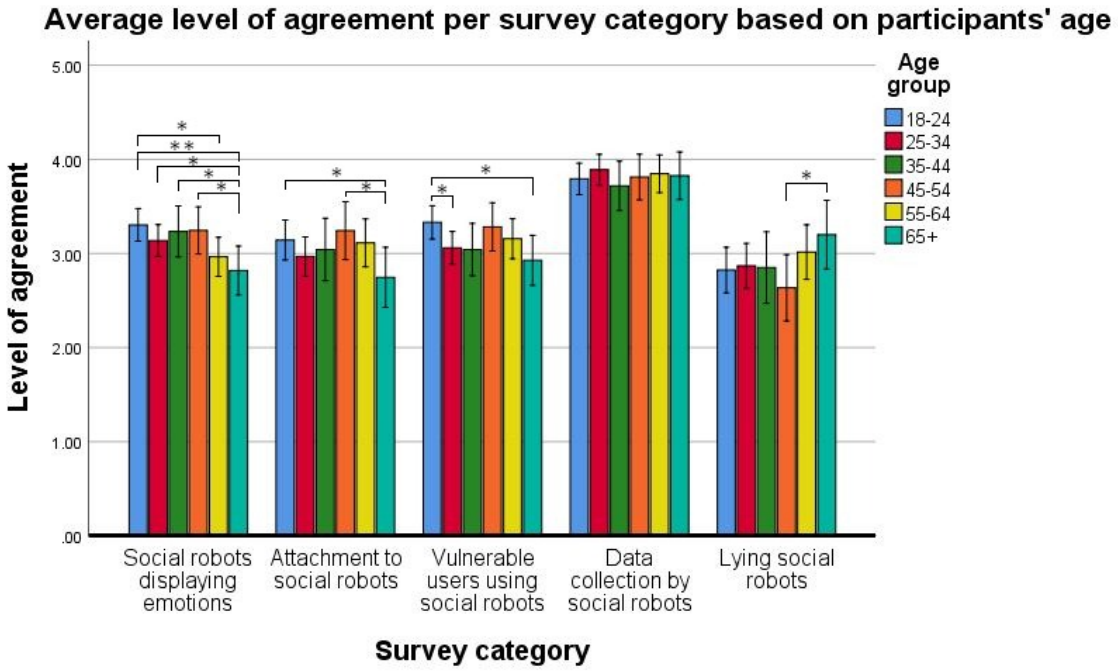


FIGURE 6.7: **$p < 0.01$, *$p < 0.05$; Average ratings for each survey category divided by age range on a 5-point scale from strongly disagree (1) to strongly agree (5).

No other significant relationship between age and any of the other categories was found,

nor was there an impact of age on any of the other categories from the survey (see Table 6.4). These results, as well as all other pairwise comparisons involving age, can be found in Appendix H.

6.3.6.3 Gender

In total, 3 participants identified as 'other' or preferred not to say. As this number was too low to draw conclusions compared to the number of participants that identified as male and female (137 and 99 respectively), these participants were excluded from analysis for gender comparisons only. Independent samples t-tests were conducted to identify whether participants' gender impacted their responses to the survey categories. It was found that responses differed based on participants' gender for the category 'Lying social robots' ($t(234) = -2.42$, $p = 0.02$), where male participants agreed significantly less ($p = 0.02$; $M = 2.77$, $SD = 0.99$) with the statements in this category than female participants ($M = 3.07$, $SD = 0.84$). None of the other categories were rated differently depending on participants' gender. These results are provided in Appendix H.

6.3.6.4 Familiarity with Social Robots and Robotic Technologies

One-way MANOVA was performed to investigate whether participants' familiarity with social robots or their familiarity with robotic technologies impacted their responses to the survey categories. It was found that familiarity with social robots had a weak effect on several of the survey categories, and familiarity with robotic technologies did not significantly impact participants' responses, as presented in Table 6.5. Post-hoc pairwise comparisons for these categories indicate that participants that reported being 'moderately familiar' with social robots scored higher for these categories than participants that reported being less or more familiar with social robots, as can be found in Figure 6.8. Details of the pairwise comparisons are provided in Appendix H.

6.3.6.5 Individual Differences

Pearson correlations were conducted to investigate the relationship between the survey categories and participants' age, gender, familiarity with social robots and familiarity with robotic technologies. The results are provided in Table 6.6.

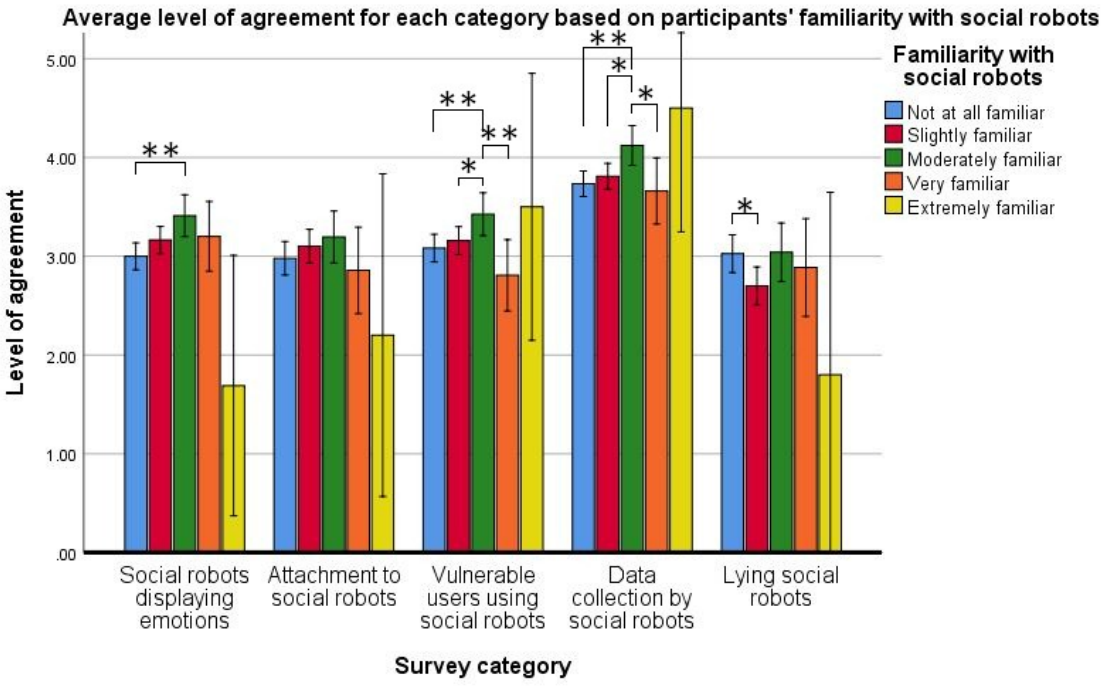


FIGURE 6.8: $^{**}p < 0.01$, $^{*}p < 0.05$; Average ratings for each survey category divided by participants' familiarity with social robots on a 5-point scale from strongly disagree (1) to strongly agree (5).

	Familiarity with social robots			**Familiarity with robotic technologies**		
	$F(4, 234)$	p	η_p^2	$F(4, 234)$	p	η_p^2
Social robots displaying emotions	3.85	**0.01****	0.06	1.36	0.25	0.02
Attachment to social robots	1.01	0.40	0.02	0.36	0.84	0.01
Vulnerable users using social robots	2.75	**0.03***	0.05	0.84	0.50	0.01
Data collection by social robots	3.13	**0.02***	0.05	1.01	0.40	0.02
Lying social robots	2.02	0.09	0.03	0.47	0.76	0.01

*p < 0.05, **p < 0.01

TABLE 6.5: Impact of familiarity with social robots and robotic technologies on the survey categories.

As these findings indicated a relationship between the survey categories and participant characteristics, follow-up multiple linear regression analyses were conducted to investigate whether or not these individual differences predicted the scores of the survey categories. The predictor variables entered are age and gender. Furthermore, dummy variables were used to code familiarity of social robots and robotic technologies (4 dummy variables). The individual differences were all entered into the model. Collectively, the variables significantly explained 10.7% of the of the variability in the category 'Social robots displaying emotions' ($F(10,235) = 2.69$, $p < 0.01$, $R^2 = 0.107$). The analysis showed that age was a significant predictor of participants' level of agreement with the statements of this

	Social robots displaying emotions		**Attachment to social robots**		**Vulnerable users using social robots**		**Data collection by social robots**		**Lying social robots**	
	r(239)	*p*	*r(239)*	*p*	*r(239)*	*p*	*r(239)*	*p*	*r(239)*	*p*
Age	-0.21	0.01**	-0.06	0.33	-0.10	0.11	0.01	0.96	0.10	0.12
Gender	-0.05	0.43	-0.07	0.31	-0.05	0.42	-0.10	0.13	0.13	0.04*
Familiarity with social robots	0.14	0.03*	0.02	0.71	0.06	0.38	0.13	0.05	-0.05	0.44
Familiarity with robotic technologies	0.11	0.08	0.06	0.39	0.07	0.29	0.09	0.19	0.01	0.94

*$p < 0.05$, **$p < 0.01$

TABLE 6.6: Relationship between user characteristics and the survey categories. These relationships indicate there may be some individual differences that predict the scores of the survey categories.

category (β = -0.01, p = 0.01). Being moderately familiar with social robots significantly predicted participants' responses to the statements of this category as well (β = 0.35, p = 0.02). No other variables significantly predicted participants' level of agreement with the category 'Social robots and emotions'. The other survey categories were not significantly predicted by any of the independent variables. These statistics, together with the β- and p-values for all independent variables for all survey categories can be found in Appendix H.

6.3.6.6 Comparison of Specific Statements of the Survey

As mentioned before, statements of the survey were phrased such that it would be possible to investigate the opinion of the general public more broadly than solely for the conditions used in this research. For example, there were different statements regarding the embodiment or the role of the robot. These statements were compared to investigate the opinion of the general public on these robot characteristics. Paired samples t-tests were performed for specific statements from the category 'Social robots displaying emotions'. This category contains statements regarding whether a robot should look/behave/be treated as a human or an animal, which helps me determine the impact of the aesthetics of the robot. The statements 'It is okay for people to think of a social robot that shows emotions as a living being' and 'It is okay for people to think of a social robot as a living being', and the statements 'Social robots can show emotions' and 'Social robots should show emotions' were compared as well, to investigate participants' opinion regarding artificial expression of emotion by robots in more detail.

For the category 'Attachment to social robots', the statements 'It is okay for people to become attached to a social robot' and 'It is okay for people to become attached to a

social robot that shows emotions' were compared, which could provide insights on whether participants thought people would become more or less attached to a robot that displays emotions compared to a robot that does not.

No statements were compared for the categories 'Vulnerable users using social robots' and 'Data collection by social robots' (see Table 6.7), as these categories investigates participants' general opinion on these matters and there were no specific statements that could be compared against one another. Finally, for the category 'Lying social robots', the statements 'There are situations where it is acceptable for a social robot to lie to a person' and 'There are situations where it is acceptable for a social robot to lie to a vulnerable person' were compared through paired samples t-tests, as well as the statements 'A social robot should never lie to the person they are interacting with, even if it appears to benefit that person' and 'A social robot should never lie to a vulnerable person, even if it appears to benefit that person'. All compared statements can be found in Table 6.7. Significant differences are indicated in this table, the exact significant values can be found in Appendix H.

	r(239)	***t(239)***
It is acceptable for a social robot to look like a *(human/animal)*.	0.31**	-3.72**
It is acceptable for a social robot to behave like a *(human/animal)*.	0.51**	-3.12**
It is acceptable for people to treat a social robot as a *(human/animal)*.	0.42**	-6.41**
It is okay for people to think of a *(social robot/social robot that shows emotions)* as a living being.	0.73**	3.08**
It is acceptable for people to think of a social robot as their *(friend/companion)*.	0.65**	-5.56**
Social robots *(can/should)* show emotions.	0.63**	0.07
It is okay for people to become attached to a *(social robot/social robot that shows emotions)*.	0.75**	0.00
There are situations where it is acceptable for a social robot to lie to a *(person/vulnerable person)*.	0.78**	-2.34*
A social robot should never lie to *(the person they are interacting with/a vulnerable person)*, even if it appears to benefit that person.	0.35**	-0.66

*p $<$ 0.05, **p $<$ 0.01

TABLE 6.7: Correlations and differences comparing specific statements of the survey. The italic part between brackets indicate the difference between two statements.

As Table 6.7 shows, there were strong positive correlations for each compared pair of statements, indicating that if agreement for one statement increased, it increased for the statement it was compared to as well. Paired samples t-tests indicated a significant difference for many compared statements, (also shown in Table 6.7). For example, for the first three statements in this table participants agreed significantly more with statements

	M	***SD***	***M***	***SD***
		human		*animal*
It is acceptable for a social robot to look like a *(human/animal)*.	3.30	1.03	3.59	1.02
It is acceptable for a social robot to behave like a *(human/animal)*.	3.18	1.05	3.40	1.08
It is acceptable for people to treat a social robot as a *(human/animal)*.	2.69	1.20	3.22	1.18
		social robot		*social robot that shows emotions*
It is okay for people to think of a *(social robot/social robot that shows emotions)* as a living being.	2.58	1.16	2.41	1.16
It is okay for people to become attached to a *(social robot/ social robot that shows emotions).*	3.09	1.12	3.09	1.10
		friend		*companion*
It is acceptable for people to think of a social robot as their *(friend/companion)*.	3.00	1.20	3.35	1.08
		can		*should*
Social robots *(can/should)* show emotions.	3.26	1.05	3.26	1.05
		person		*vulnerable person*
There are situations where it is acceptable for a social robot to lie to a *(person/vulnerable person)*.	2.93	1.20	3.05	1.22
		the person they are interacting with		*a vulnerable person*
A social robot should never lie to *(the person they are interacting with/a vulnerable person)*, even if it appears to benefit that person.	2.78	1.17	2.84	1.21

TABLE 6.8: Mean and standard deviation for each pair of compared statements. Means are calculated from a 5-point scale. Bold statements indicate that level of agreement differed significantly for the compared statements.

if they involved 'animal' compared to 'human'. All means and standard deviations for compared statements can be found in Table 6.8.

6.3.6.7 Final Questions

Based on observations from the user studies and in order to gain further information on the opinion of the general public regarding the use of social robots, several yes/no questions (provided in Table 6.9) were presented to participants at the end of the survey.

Would you use a social robot if...:	**% yes**	**% no**
You had dementia?	71.9	28.1
You had a physical disability?	81.4	18.6
You had complex medical needs requiring medication?	72.9	27.1
You felt lonely?	44.3	55.7
Your partner/family recommended it?	54.8	45.2
Your partner/family did not like the robot?	33.0	67.0

TABLE 6.9: Percentage of participants responding 'yes' or 'no' to whether they would use a social robot in specific circumstances.

When asked for clarification when answering 'yes' or 'no', example responses to use the robot were 'It can be helpful in daily activities' and 'It could really help', where 'I would not trust it' and 'A robot cannot substitute a human' were example responses to not use the robot. Finally, participants were asked for the primary reason why they would or would not use a social robot. Reasons that were mentioned to not use the robot were limited abilities of the robot ('They will not be able to match what humans can do') and privacy concerns ('They may be controlled by someone from the outside'). Example reasons why people would use the robot are for physical support ('To make life easier and assist me with day-to-day tasks') and lighten care tasks of others ('To make the burden for my family lighter'). Some participants provided reasons to use a robot, but specified the tasks they would use the robot for, like 'Something assistive, not replacing a human' and 'As an additional support in a system of care could be useful. Otherwise I would aim at not using'.

6.3.7 Discussion

This online survey was designed and conducted to determine whether the findings and observations from the user studies that were conducted in this research are reflected in the opinion of the general public. This was deemed essential, as it provides support whether the findings can be generalised such that they can be used to develop a framework. User studies always have limitations, such as a small number of participants, or restricted interactions with limited freedom for participants to explore the robot's abilities. Therefore, when findings from such studies are used to develop a framework that aims to advice future development, it needs to be justified whether these are representative of the general public, and not just the participants that took part in the experiment. First, it can be concluded that the results from the survey were safe to analyse as the reliability was high, but not so high that it could indicate redundancy or duplication, which may occur when Cronbach $\alpha > 0.90$. In total three statements were excluded from analysis, one statement for the categories 'Attachment to social robots', 'Vulnerable users using social robots' and 'Data collection by social robots' each. In hindsight it makes sense that the statements for the first two categories did not fit with the remaining statements, as these two statements were the only two statements that did not state whether something was acceptable, but whether something could occur or not (e.g. 'People become more easily attached to a

social robot when they trust it' where all other sentences started with 'It is okay/acceptable for people to...'). The statement for the category 'Data collection by social robots' was the only negative statement in this category. The category has the statements 'It is not okay for a social robot to be used to monitor the progress and health of a person' and 'It is acceptable for a social robot to be used to monitor the progress and health of a person', and the first one was excluded from analysis in the end. This statement was added to the survey as an attention check - if participants answered positive or negative for both statements they were excluded from analysis as it would indicate they had not been paying attention. As the sentences were not exactly the opposite of one another ('It is not okay' versus 'It is acceptable') I decided to reverse code the negative statement and test for reliability. If the statement were not reverse coded, it would make sense it would not match with the other statements. Why it stood out now is not clear to me. However, more obvious attention checks such as 'Please select strongly disagree' instead of adding reverse-worded statements will be used in the future.

6.3.7.1 Results of Each Category

Looking at Figure 6.6, it can be concluded that participants did not strongly agree or disagree with most of the categories. Taking into account the relatively high number of participants that were not at all or slightly familiar with social robots (77%) and robotic technologies (64%), it is possible that participants often selected 'neither agree nor disagree' as the survey presented situations they had not encountered before and therefore had not formed an opinion on yet. The positive conclusion from this finding is that the participants did not strongly disagree with many of the statements, which indicates that the findings from the user studies provide valuable insights, even though the number of participants was low.

It is interesting that overall participants agreed more with the statements in the category 'Data collection by social robots' compared to the other categories. It was expected that participants would disagree with these statements as they entail the use and storage of their personal data, potentially by third parties. It is possible that participants were not very concerned about the use of their data, are unaware how their data may be collected, or they may not have recognised the potential concerns. It is also possible that the introduction to the section on data collection biased participants, as it only discussed positive outcomes of

a robot using and/or storing their personal data. However, this result indicates the need for future research on people's awareness of data collection and data storage, and what is deemed acceptable and what not. Also, further research on whether people would think differently of a social robot collecting and storing their data compared to other technologies such as computers and smart phones could provide useful insights on how social robots should be used.

6.3.7.2 Age

Looking in more detail at the four categories that averaged all around the centre of the scale, Figure 6.7 reveals that participants from the older age group (65+) usually agreed less with statements than participants from other age groups, except for the category 'Lying social robots'. It is not clear why this is the case, but perhaps these participants were more able to identify with the introduction to this specific category as it entailed an older person who has dementia, where overall they may feel less familiar with social robots and new technologies and therefore may be less agreeable with the use of them than participants from other age groups.

6.3.7.3 Gender

The results showed that male participants agreed significantly less with statements from the category 'Lying social robots' than female participants. However, the effect size of this difference was very small, so no strong conclusions can be drawn from this result.

6.3.7.4 Cultural Background and Level of Education

As mentioned before I decided to not investigate whether participants' cultural background and education had an impact on their responses. Even though I aimed to gather the opinion of the 'general public', I only partially succeeded in this. The number of participants was decent and there was a decent distribution of participants from different age groups. However, looking at cultural background and also education there was not a great distribution as most participants were European with a higher education. As this research mainly focuses on the use of social robots for older adults and therefore concentrates on different approaches depending on age, participant recruitment was deemed successful for this survey. However, future work should investigate if and how cultural background and

also different education levels may influence people's opinion of artificial expression of emotion by social robots. This is another reason why the component 'Population' of the framework is extremely important and should be considered before any other components.

6.3.7.5 Comparison of Specific Statements

Comparing specific statements regarding the robot looking/behaving/being treated as a human or an animal showed participants agreed slightly more with these statements when they considered an animal with respect to a human. I decided to not provide a description nor a picture of a social robot in the survey so participants would not be influenced by the appearance of the robot that was provided in the survey but had to rely on their own mental model of what a social robot should look like. Therefore, the difference between the animal or human component indicates that the appearance of the robot plays a role in what is deemed acceptable behaviour and what is not. As a result of this, robot characteristics need to be added as a component to the framework. This applies to the role of the robot as well, as participants agreed more with people thinking of the robot as a companion than as a friend. This again indicates that the role of the robot is essential and therefore will be added as a component to the framework as well. This will be discussed in more detail in the next section.

6.3.7.6 Insights of Final Survey Questions

The final questions of the survey indicated that participants would use a social robot if it were in an assistive context, as over 70% of participants reported they would use a social robot if they had dementia, a physical disability or complex medical needs. However, more than half of the participants indicated they would not use the robot if they felt lonely, which indicates they would prefer to not use a robot for companionship. Finally, it appears that participants found the opinion of people close to them important as more than half reported they would use the robot if their partner or family recommended it and two-thirds of the participants indicated they would not use a social robot if their partner or family did not like the robot. This finding coincides with earlier work that I conducted with a colleague, where we found that participants found the opinion of others important when deciding to use a social robot (Bishop et al., 2019). As we concluded then, inclusion of the people close to the intended users is important for acceptance of the robot and consequentially

successful human-robot interactions.

Participants reported privacy as a potential reason to not use social robots. This is interesting, as agreement to the statements in the category 'Data collection by social robots' was relatively high, indicating participants preferred the benefits the robot could provide over privacy concerns. However, it is possible they were influenced by the phrasing of the statements in this category, or perhaps they do not relate to data collection when considering privacy but the use of cameras and being physically observed instead. Nevertheless, it is imperative to investigate people's opinion towards data collection by a social robot in more detail.

Finally, it appeared that participants preferred a physically assistive robot over a companion. This concurs with the responses from participants of the longitudinal field study with older adults, where one-third of the participants reported they would like a robot as a companion and two-third of the participants thought it would be useful as a helper.

Overall it can be concluded that this survey can be used to gather opinions on robots that are capable of artificial expression of emotion. It depends on the demographics of participants to what extent the findings can be generalised to fit the 'general public'. However, when using this questionnaire it should be taken into account that many people have not considered artificial expression of emotion by robots and what impact it may have, therefore it is possible that results will be similar to results from this survey even though participants may actually feel differently about the topic when considering it in more detail.

6.4 The Framework Revised

The initial framework consisted of three components: target population, occurrence of AEE, and accessibility issues. This framework will be extended to include the findings of the final survey. This survey investigated the opinion of the general public, to determine whether the findings from the user studies conducted in this research can be generalised to be converted into components for a framework.

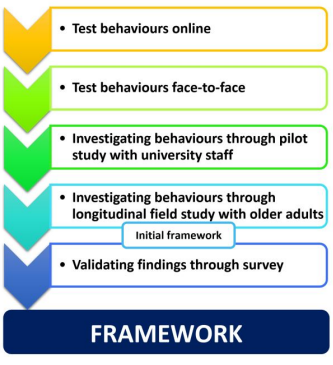


FIGURE 6.9: Final stage of framework development.

Following from the findings of the online survey, there

were several findings that stood out, resulting in an extension of the framework. The most important findings of the survey are the following:

- Participants did not have strong opinions of the survey categories. It is possible that the survey presented a topic they had not considered much as they were relatively unfamiliar with it.
- Participants expressed it is acceptable for a robot to collect, store and use data if this is in the benefit of the user.
- Participants expressed it is more acceptable for a robot to look, behave and be treated like an animal compared to a human.
- Participants expressed it is more acceptable to think of a social robot as a companion compared to a friend.
- Participants expressed it is more acceptable for a social robot to lie to a vulnerable person compared to a person without health impairments.
- Participants expressed they would use a social robot in an assistive context, but less often as a companion.

Several of these findings resulted in components being added to the framework, to ensure it raises awareness of concerns of AEE. However, not all results were included in the updated framework, as not all of them applied to AEE specifically (e.g. results on data collection). The added components, together with the final completed framework are discussed in the following sections.

6.4.1 Robot Role

The results of the survey indicated that different robot roles may need different levels of awareness to prevent negative outcomes. Participants preferred for the robot to be perceived as a companion compared to a friend. This difference was also supported by the results of the field study. Some participants indicated they would prefer the robot as a companion, but most reported they saw the robot as a tool to provide support. The impact

of the robot role on human-robot interactions has also been addressed in the literature (Dautenhahn et al., 2005). It has also been presented as a factor impacting the effectiveness of these interactions in a framework for social robots (Feil-Seifer and Mataric, 2005). However, this framework aimed to guide social robot design, and not on the psychological impact that the role of the robot may have on the people using it, as investigated in this research.

6.4.1.1 AEE as a Requirement for the Robot Role

The risk of negative consequences can be higher when AEE is required for successful completion of the robot's task. Imagine a robot that is supporting somebody go through physical rehabilitation. The robot can be expertise focused during the interactions and explain why certain exercises will help rehabilitation. However, it was found that people were persuaded to perform a physical exercise for longer when the robot showed a form of empathy (Winkle et al., 2019), which can be expressed using AEE (Tapus and Mataric, 2007; James, Watson, and MacDonald, 2018). Furthermore, they found that persuasion was higher when the robot indicated similar preferences to the participants' preferences. However, this indicates the robot requires to provide a 'personal' opinion, which was recommended against earlier in this framework, in the component 'Artificial Expression of Emotion'. When aiming to persuade somebody, the use of AEE can have negative consequences as it may lead to emotional blackmail. It may also lead to physical harm as users may perform more repetitions than they are capable of at that point to please the robot.

An example role where AEE is less essential, is providing reminders to hydrate and take medication. In cases of early stage dementia people may forget they need to take their medication, but are willing to do so when they are reminded. In this case only the reminder is required and no persuasion is needed. Therefore, the need of AEE is not essential for successful completion of the task and the risk of negative outcomes is low.

Due to the diversity of assistive robot roles it is not possible to provide a general risk-level of negative outcomes. However, it appears more likely that negative outcomes may happen when AEE is essential for the task. Therefore, one should ask themselves whether *AEE is required for successful completion of the robot's task.* This can be addressed by

assessing how AEE impacted user experience for each task.

6.4.2 Level of Exposure

Negative outcomes of AEE may be more likely to happen when the level of exposure to AEE is high. High exposure may more easily lead to emotional attachment (Ilicic and Webster, 2011)and potential negative effects like overtrust and high levels of dependence. The level of exposure can depend on the robot role, where exposure to a companion robot can be higher than exposure to a robot that provides support during rehabilitation. It is likely that social robots will be able to provide multiple tasks (e.g. provide companionship but also provide physical assistance). Therefore, it may be possible that people ask the robot to something they could have done themselves, increasing their exposure to AEE and also their level of dependence on the robot. Therefore, *the user's abilities should be known*, so it can be subtly suggested the robot's support is not needed for that specific task. Furthermore, it should be considered that *the benefit of AEE may change* when exposure is high and behaviours become more predictable. To understand the consequences of exposure, it should be *assessed to what extent people become dependent on the robot, or become attached to it*.

6.4.3 Ownership

The level of exposure is likely to be higher when the robot has a single owner compared to when several users share the robot. Therefore, it should be *considered whether the robot can be shared by multiple owners.* It should be considered *how the response to AEE may differ based on ownership*. If a robot appears capable of experiencing emotions and conveying empathy, users may be more likely to entrust secret to the robot, especially if they belong to a vulnerable population. However, if users have to share the robot, they may entrust less to the robot in fear of it revealing their secrets. In that case, limited AEE may be preferable. In the case of multiple ownership, it may also be possible that users become jealous of other users, especially when they are vulnerable and may misunderstand the relationship they have with the robot (Edberg and Edfors, 2008). Therefore, the *aspects of the 'population' component of the framework should be considered* for all users.

6.4.4 Robot Appearance

As only the robot Pepper was used during the user studies conducted in this research, not much can be said about the impact of the robot's appearance. However, the literature indicates that it could influence the effectiveness of human-robot interactions (Fong, Nourbakhsh, and Dautenhahn, 2003). Therefore, it was included in the survey that was conducted to investigate the opinion of the general public. As this survey did not specify any robot characteristics when discussing a social robot, participants had to use their own mental model of a robot during the survey, without being primed by descriptions or images of a social robot's appearance. The survey category 'Social robots displaying emotions' contained statements that distinguished between human- and animal-like robots. Even though these statements were positively correlated. This indicates that if participants agreed highly with a statement involving an animal-like robot, they also agreed highly with that same statement involving a human-like robot. However, pairwise comparisons indicated that participants agreed more with statements concerning animal-like robots, compared to statements concerning human-like robots (see Table 6.8). This result indicates that the robot's characteristics impact participants' expectations. Therefore, the robot's characteristics will also likely impact expectations regarding AEE. Therefore, it should first be *determined what characteristics are required for the robot's intended functionality*, followed by considering the question *'Does the robot's appearance represent its abilities?'*. This may help ensure the appearance is developed such that expectations are guided to correctly represent the robot's abilities. More specifically, it should be considered whether the robot's appearance will increase the perception of AEE, and *assessed whether false expectations are raised due to the robot's appearance*. For example, if AEE through speech is essential for the robot role, the robot should not have an animal-like appearance, as it is not expected that animals will talk.

6.4.5 Stakeholders

Finally, stakeholders are an important factor to consider for the use of AEE. It is possible that other people besides the target population will interact with the robot. Therefore, it should be considered whether AEE may impact these stakeholders as well. Back in 2006 it was already addressed that interactions with a robot may also lead to interactions with

other people (Kidd, Taggart, and Turkle, 2006). However, the impact of AEE on these other people was not discussed, although it has been highlighted that not only impact on care receivers but also caregivers needs to be considered (Vallor, 2011). The answer to the question *'Are stakeholders involved in successful completion of the robot's tasks?'* may help determine users outside of the target population, and their risk of negative consequences when exposed to AEE. One might argue it is not essential to determine the impact on stakeholders. However, the scenario provided in the survey, where a social robot is used to ensure a patient with dementia does not become stressed while the caregiver is getting their medication, can prove otherwise. If a family member patient is present while the robot lies to the patient to keep them calm, this may psychologically impact this family member and lead to distrust. This may result in them advising the patient against using the robot, which is a negative outcome for the patient as they can no longer benefit from the service provided by the robot. Therefore, it is essential to *consider the impact of AEE on other stakeholders* when determining the robot role.

6.4.6 Naming the SOCRATES Framework

Now that the framework has been finalised, it needs a name. As a reference to the project that I have been allowed to be a part of over the last three years, I decided to go for the acronym SOCRATES, which stands for **SOC**ial **R**obots and **A**r**T**ificial **E**motion**S**. This name feels appropriate as it is a nod towards the research project called SOCRATES (SOcial Cognitive Robot Agents in The European Society) which has given me the opportunity to experience this journey. Also, it refers to the Greek philosopher Socrates who was known for his influence on moral virtues and ethics and as a forerunner of reflective thinking. This is very fitting, as the goal of this research is the prompt reflective thinking and make people realise they should pause and consider the effect that their potential design of artificial expression of emotion by a robot can have, as seemingly innocent interactions between robots and their users can have harmful and/or negative consequences that were not foreseen during development. The SOCRATES framework with supporting guidelines is provided in Table 6.10.

Component	General recommendations	Reflective questions	Specific recommendations	Level of risk
Population	Determine the target audience.	Can the target audience include people with cognitive impairments?	Periodically assess level of impairment, consult domain expert.	Gradient, from medium to high, increases with severity of impairment
	Ensure that the benefits provided through AEE outweigh negative outcomes.	Can the target audience include people with physical impairments?	Periodically assess whether the robot provides appropriate support.	Gradient, from low to medium, increases with severity of impairment
		Can the target audience include people with both cognitive and physical impairments?	Continuous monitoring of mental health.	High
		Are there personal factors that may indicate vulnerability of AEE?	Assess factors that may indicate vulnerability of AEE	Dependent on personal factors
		Are there environmental factors that may indicate vulnerability of AEE?	Assess whether the user's environment is appropriate for social robot support.	Dependent on user environment
		Will the benefits of AEE outweigh potential negative outcomes?	Assess quality of life.	Risk increases with number of potential negative outcomes
Artificial Expression of Emotion	Consider why AEE is used.	Does unintended AEE occur (either through behaviour or speech)?	Avoid statements where the robot expresses to experience an emotion or have an opinion.	Dependent on context
		Is the use of AEE needed?	Assess whether benefits outweigh disadvantages .	Dependent on other framework components
Accessibility	Consider what modalities are used for AEE.	Can modalities used for AEE lead to accessibility issues?	Periodically assess whether the robot's flexibility and ability to address accessibility issues is appropriate.	Dependent on population
Robot Role	Consider the need of AEE for the robot's tasks	Is AEE required for successful completion of the robot's tasks?	Assess response to AEE for each task	Dependent on robot task
Level of Exposure	Consider the user's abilities.	Does AEE remain relevant when the level of exposure increases?	Assess impact of exposure on level of dependency on and attachment to the robot.	Gradient increase with the increase of exposure
Robot Ownership	Determine single or multiple ownership	How is the response to AEE impacted by ownership?	Consider 'population' component for each owner.	Higher for single ownership
Robot Appearance	Determine appearance requirements for robot functionality	Does the robot's appearance represent its abilities?	Evaluate the robot's appearance does not raise false expectations regarding AEE	Risk increases when appearance deviates from abilities
Stakeholders	Determine whether stakeholders may interact with or observe the robot.	Are stakeholders involved in successful completion of the robot's tasks?	Periodically assess impact of AEE on stakeholders.	Dependent on stakeholder's health situation.

TABLE 6.10: The SOCRATES Framework

6.4.7 Limitations of the SOCRATES Framework

No framework is perfect from the start, as it needs to be tested and validated before its true value can be determined and potential adaptations can be made. In this case, the framework is based on findings from human-robot interaction studies that were conducted during this research, and as with each human-robot interaction experiment, there were some limitations. To begin with, the number of people that participated was limited for some studies, and the findings may only apply to the specific research setting that were used for each experiment. For example, the content of the interactions was didactic, where it is expected that participants may share personal experiences with social robots when this robot would provide some form of companionship in addition to providing information. This may lead to different expectations of the robot's behaviour and should be investigated further. This was partially addressed by evaluating the findings of the user studies through the online survey, that investigated the opinion of the general public on these findings. However, it is possible that changes in the experiments conducted in this research may lead to different results, which again may impact the SOCRATES framework.

6.5 Summary

First, this chapter presented existing ethical frameworks for autonomous systems and social robotics, and introduced how the SOCRATES framework may bridge certain gaps. This framework was initially based on the findings from the user studies described in Chapters 4 and 5. It was designed to provide guidance on how to develop artificial expression of emotion by robots that has limited negative ethical consequences. To strengthen the foundation of this framework and not solely base it on user studies that may yield different results when introducing small changes, the opinion of the general public was gathered in the form of an online survey. The research questions investigated in this chapter are: *'How do the results found in this research compare to the opinion of the general public?'* and *'Can the findings of this research be used for the development of a framework on ethical AEE?'*. The results from the survey indicated that it is not clear whether the findings of the studies are represented in the opinion of the general public, as most results indicated that participants neither agreed or disagreed with the findings presented in the survey. This may be due to the fact that many participants were unfamiliar with social robots and

robotic technologies, which meant they had not considered the situations presented to them in the survey before and therefore did not have a strong opinion. It does however answer the question whether the findings can be used to develop the framework, as findings indicated participants overall did not disagree with the findings which may have indicated the findings were too specific for the experimental procedure used to be used for the framework. Finally, the framework was updated to include findings of the survey that were appropriate to strengthen the framework. Both the SOCRATES framework and the survey are being transformed into journal paper submissions at the time of writing.

7 Discussion

In this work, I highlighted that there is little knowledge of potential negative outcomes regarding the use of social robots. Researchers have shown that such robots can have positive impact on people's lives (e.g. Hutson et al., 2011; Hung et al., 2019), and others have raised potential concerns of the use of such robots (e.g. Sharkey and Sharkey, 2012; Sullins, 2012; Sparrow and Sparrow, 2006; Turkle, 2006; Coeckelbergh, 2010). However, even though these concerns have been discussed in the literature, they are often not analysed further (Vandemeulebroucke, Casterlé, and Gastmans, 2018). This analysis is important for two reasons. First of all, it allows us to better understand potential negative consequences of these concerns and for them to be addressed. Also, even though concerns identified in the literature may discourage poor practice, not identifying whether these concerns are justified can avoidably constrain research and development of new technologies.

The concerns that were highlighted and investigated in this work are emotional deception and emotional attachment. The reason to investigate these concerns, is that little is known regarding the potential negative outcomes of the use of social robots in assistive contexts. Two main research questions were investigated in this work:

RQa: *'Can artificial expression of emotion by a social robot lead to emotional deception and emotional attachment?'*

RQb: *'Can artificial expression of emotion by a social robot result in negative consequences?'*

Several user studies were conducted in this research to address these questions, which have been discussed in Chapters 4, 5 and 6. Based on the findings from these studies, the answer to **RQa** is that artificial expression of emotion can potentially lead to emotional deception and emotional attachment, though it may not always occur. As mentioned in Chapter 2, this was the expected outcome, as it was addressed often in the literature (e.g. Sharkey and Sharkey, 2011; Borenstein and Arkin, 2019).

No negative consequences of AEE were apparent from the qualitative and quantitative measures that were used in this research, as questioned in **RQb**. However, as I encountered several obstacles and limitations while conducting the user experiments, it cannot be concluded that AEE will never result in negative consequences, and thus needs to be investigated further. The literature often discusses impact on populations vulnerable as a result of cognitive impairments, indicating that negative consequences are more likely for these populations (e.g. Sharkey, 2014). However, even though participants in this research had no cognitive impairments, they were categorised as vulnerable, as they were not able to live independently. They lived in semi-independent retirement villages, so support was close when needed. Therefore, it was expected that no negative consequences would be found in this research.

The findings from the research performed to answer the main research questions of this work were used to develop an ethical framework on AEE, namely the SOCRATES framework. This framework may help developers and other stakeholders to consider emotional deception and emotional attachment as concerns of AEE by social robots, and support these people in limiting negative consequences before social robots are widely deployed into the world.

However, even though I was able to come up with answers for the research questions of this research, it is still difficult to draw clear conclusions, as the answers are vague and raise more questions due to the obstacles and limitations I encountered. This will be discussed in the remainder of this chapter, before concluding with the contributions of this research and directions for future work. These insights are being transformed into a journal paper submission at the time of writing.

7.1 Insights from Research Findings

Even though I was limited in the conclusions I could draw from this work due to obstacles I encountered, I still made several interesting observations that provide new insights or raise new questions. For example, even though all studies indicated that the implemented behaviours were perceived as designed, it was found that age impacted participants' perception of these emotions, with a negative correlation for happy and neutral behaviour and a positive correlation for sad behaviour. This indicates that as participants' age

increased, their perception of the behaviour became more 'neutral'. This observation raises some interesting questions. For example, do these findings indicate that older adults are less likely to be impacted by AEE, if solely focusing on age and not considering other potential factors such as cognitive decline? Is the impact of AEE influenced by the severity with which an emotion is perceived, or is it a binary phenomenon where consequences either occur or not? Addressing these questions will allow us to better understand the impact of AEE by social robots.

7.1.1 Insights from the Longitudinal Field Study with Older Adults

Both positive and negative emotions were displayed for the emotive condition of the longitudinal field study. However, these emotions were investigated separately during the study with psychology students that was conducted, to determine whether the implemented behaviours were perceived as intended. Even though the different emotions did not significantly impact emotional deception (measured as anthropomorphism and social presence), a trend was observed that both factors were scored higher when the robot displayed sad behaviour compared to happy behaviour. This suggests a possibility that the sad behaviour in this research led to higher levels of deception than the happy behaviour, and it should be investigated further whether negative emotions are more likely to result in deception than positive emotions. This is a possibility, as a display of negative emotions may suggest an understanding of context by the robot that is not necessarily required for it to be able to display happy emotions.

The occurrence of deception was low to medium for the longitudinal field study. This may have been due to the participants not being deceived by AEE. However, this finding can also have been impacted by the use of the tablet to provide subtitles. As observed in the pre-study with university staff, participants were focusing on the tablet and not paying attention to the behaviour that the robot was displaying. This was observed again during the field study with older adults, where they would sometimes interrupt the robot to answer a question presented on the tablet while the robot had not completed the question yet. Consequently, the impact of the robot's behaviour (emotive / non-emotive) may have been less clear than when participants had not been distracted. More importantly, this insight raises questions regarding robot accessibility to different populations, which I aimed to address in the SOCRATES framework by suggesting periodical assessment of

the successful use of modalities. The impact of this tablet is strengthened by the result that participants of the test group, and thus interacted with the emotive robot, perceived the robot more as a social entity than participants of the control group, that only saw the non-emotive robot behaviour.

Besides the impact that the use of the tablet may have had on the results, the need for this use raises some concerns. Participants in earlier studies had not had any issues understanding the robot, indicating a reduced accessibility of the robot for older adults with ageing-related issues. Also, they may be less able to benefit from social robots as a consequence of these accessibility issues. These concerns have been published in more detail in other work (Van Maris et al., 2020a).

As discussed in Chapter 2, there were several reasons to use a questionnaire on object attachment to determine participants' level of attachment to the robot. The results of this questionnaire indicated that attachment was low for most participants. However, additional questions to measure attachment, such as participants' willingness to use the robot in the future and whether they would miss it, provided different results and indicated attachment was higher. This mismatch in results indicates that human-robot attachment is not fully understood yet and existing questionnaires may not be sufficient. This indication is strengthened by the argument that social robots can be identified as a new ontological class (De Graaf, 2016; Kahn Jr and Shen, 2017).

The results from the field study suggested that there may be a relationship between emotional deception and emotional attachment, as participants' level of attachment positively correlated with both the extent to which they anthropomorphised the robot, as well as to what extent they perceived the robot as a social entity. Future work can increase our understanding whether this relationship is unidirectional as suggested by some (Sharkey and Sharkey, 2020) or bidirectional, which will allow us to then more sufficiently address the potential negative consequences of deception and attachment.

Furthermore, the results (or lack thereof) from the technologies used in this research have indicated that these technologies are not advanced enough yet. As I was unable to draw strong conclusions from the speech prosody data, physiological data (HRV and EDA) and behavioural data, the devices used did not provide additional understanding of user experiences during human-robot interactions. This raises questions regarding the research

that is currently being performed in HRI. Are we asking the right research questions? Should we even continue research in this area if we cannot sufficiently measure potential consequences of current developments? Or should we perhaps shift, if not at least expand, the focus from developing social robots to include research in the best approaches to measure the impact of human-robot interactions on individuals and society?

7.1.2 Insights from the Survey that was Presented to the General Public

The final survey that was conducted to determine the extent to which the findings of the other studies could be generalised, resulted in a new approach to social robot surveys. Participants with contextual scenarios to allow them to better understand what they are presented with is a novel approach in HRI research. Results of this survey indicate that observations from the other studies are generalisable, as participants were more agreeable to use the robot in an assistive function. This was also fed back by participants of the longitudinal field study. These results suggest that the requirement to have an assistive function applies to the general public and not solely older adults. Furthermore, participants indicated they were more likely to use the robot if family members recommended it, and less likely to use it if family members were not happy with it. This was also found in other research that I have been involved in that has not been discussed in detail in this work (Bishop et al., 2019), and again appears generalisable and not only applies to older adults.

Finally, it was expected that answers in this survey would often be neutral. As discussed in Chapter 6 the option to provide a 'neutral' answer was deliberately provided to prevent unintended bias, which led to the possibility of participants choosing this option. However, it is also a possibility that participants responded neutral to most items, as they were presented with scenarios they had not considered, and therefore had not formed an opinion on yet. The only items where participants leaned more towards being agreeable than being neutral, were statements on data collection by social robots. This raises a plethora of questions regarding people's perception of data protection and privacy that provides enough input for a whole new research project and will not be discussed further here.

7.2 Ambiguities in Research Results

In this work, I identified that emotional deception and emotional attachment had been considered in the literature as potential concerns that may lead to negative consequences when artificial emotions are expressed by a robot, and that these concerns had never been investigated. It is essential to do so, as it will increase understanding of such concerns and their potential negative consequences, and also establishes whether the concerns are justified and reflected in practice.

As mentioned in Chapter 2, there exists no specific measurement for people's attachment to social robots. After considering human attachment, pet attachment and object attachment I established that a measurement for object attachment would be most suitable for this research. However, as mentioned in the discussion of Chapter 5, the findings for these measurements did not support each other. The object attachment questionnaire used in this work indicated that attachment to the robot was low for most participants, where people that scored low on attachment did admit that they would miss the robot. This indicates that perhaps measurements on object attachment are not sufficient to measure attachment to robots and it requires a completely new measurement. This could be possible as it has been argued before that social robots are a new technological genre (De Graaf, 2016) that requires new approaches as well.

This applies to the measurement of emotional deception as well. There is no universally agreed measurement for deception, which is understandable as there are gradations of deception and distinctions such as intentional and unintentional deception. In this work, anthropomorphism and social presence were used to measure the potential occurrence of emotional deception, as high levels of anthropomorphism and social presence during AEE by a robot may indicate that participants perceive the robot as a social entity with human-like features. This means they may be deceived by the robot's displayed behaviour. I decided to use questionnaires often used in HRI research to measure these aspects. However, it is possible, and perhaps even likely, that the use of different measurements for anthropomorphism and social presence will result in different outcomes.

Looking at the final online survey, that was conducted to gather the opinion of the general public, there are some ambiguous statements that may have resulted in participants

perceiving a statement differently than intended. For example, there are several statements that discuss a robot 'showing' emotions. The statements in this survey were updated several times, and written such that they would be easy to understand for people with no knowledge of social robots. However, in hindsight the word 'showing' perhaps should have been replaced with 'displaying' or 'communicating', as 'showing' may be perceived as the robot having some intrinsic motivation, whilst this is less likely for the other options.

One conclusion that can be drawn from these limitations, is that we currently do not have sufficient measurements to address these ethical concerns. This indicates that perhaps the research focus in social robotics should include, if not shift, to how we can successfully measure these concerns before social robots are truly deployed in the world.

7.3 Limitations and Obstacles

First of all, it should be noted that AEE was designed to be low in this research, as the robot was not displaying high levels of emotional manipulation. This approach was taken to investigate whether low manipulation could already lead to emotion detection and identification, and to determine whether low levels of emotion manipulation could already influence human-robot interactions. The findings indicate that emotion detection and identification occurred for the low level of emotion manipulation used in this research. This manipulation did not seem to greatly impact participants' user experience, except for the fact that the robot was perceived more as a social entity by the participants that saw emotive robot behaviour, compared to the participants that saw non-emotive robot behaviour only. However, as human-robot interactions are likely to involve higher levels of emotion manipulation than expressed in this research, and robots may interact with users that experience a higher level of vulnerability than the participants of this research, it can be concluded that more research is needed to determine whether AEE can lead to negative consequences.

7.3.1 Limitations of this Research

Design considerations had to be made based on access to the target population, which resulted in decisions that may not be optimal in terms of validity and reproducibility. A mixed between-within subject approach was used for the longitudinal field study with

older adults. However, as the number of participants was limited for this study, no strong conclusions could be drawn from the results. Nonetheless, these findings are valuable to guide future work. Furthermore, not all personal factors were taken into consideration for this study, such as participants' socio-economic status. The research also focused on one-on-one interactions between the participants and the robot. As mentioned in the SOCRATES framework, other stakeholders may not only impact participants' responses but they may be affected by AEE themselves as well. This was not considered in this research and provides an interesting basis for future work.

Another design consideration involved the type of interaction. It was decided to use didactic interactions for this research, as the aim was to gather a baseline understanding of the impact AEE could have. Encouraging more personal interactions would have led to increased interpersonal differences and weaker conclusions in this research. However, as social robots will likely interact with people on a personal level, for example when providing companionship, this study does not provide any insights on people's responses to more personal interactions. This should be investigated further in future work.

The research aimed to identify useful measures for longitudinal field studies with vulnerable populations. However, it is likely that the findings were impacted by a lack of experience in using the sensors and analysing the data. Nonetheless, this did lead to insights regarding the experience required to use these types of measurements for HRI experiments.

Furthermore, even though a lot of effort went into recruiting a diverse participant group for the final survey that established the opinion of the general public on AEE, the population was still not as diverse as initially intended. Furthermore, the decision was made to not provide any visuals or descriptions of social robots so as to not prime participants in their responses. However, as overall familiarity with social robots was relatively low, it is likely that these participants' mental model of a social robot was based on reports from popular media. This may have lead to different responses than in cases where visuals or descriptions were shown that would provide an accurate overview of a social robot's abilities.

Also, the attention check used in the final survey provided ambiguous results, as discussed in Chapter 6. In the future, instead of adding statements that are the opposite

of other statements to test whether participants will give the same answer once results or these statements are reverse-coded, specific attention-checks like 'Please select *strongly disagree*' will be used to determine whether participants paid attention during the survey.

7.3.2 Obstacles Encountered in this Research

Especially interesting for researchers in the area of deception, attachment and AEE are the obstacles I encountered regarding the means to measure these phenomena. As stated in Chapter 2, there are no established questionnaires for neither deception nor attachment, nor are there any other types of measurement often used for them. This is not unexpected as I highlighted that these issues are often addressed in the literature but there is little empirical research on them. However, this did make it difficult to determine whether the measurements used were suitable. For example, as also mentioned in Chapter 2, anthropomorphism and social presence can both be indicators that deception occurred. However, even though questionnaires often used in HRI research were used to measure these characteristics, it may be that there are other more extensive questionnaires on these topics that would have provided richer insights than the ones used in this work. Furthermore, it is possible that the questionnaire is suitable but there is just no impact, which can be difficult to conclude if uncertain whether the questionnaire was suitable.

Besides the use of questionnaires, I also found that the devices I used to gather physiological data were not sufficient to draw strong conclusions. As the sensors I used have successfully been used in other research it can be concluded that they are not advanced enough to support HRI research. This raises the need for more advanced technologies, and also raises the question whether development of social robots should continue while we do not have sufficient means to investigate potential negative consequences of these technologies.

Furthermore, the use of new technologies can lead to accessibility issues. This is more likely for vulnerable populations, and in situations where sensitive topics are being researched that may have a psychological impact. Such issues may not have been experienced before and require adaptations that may impact the research.

Finally, as discussed in Chapter 5, I experienced an unpleasant encounter where a participant addressed me personally, violating social norms. This is something that should

be considered when running experiments on sensitive topics such as AEE, that can have a psychological impact on participants. This is even more important when the participant group involves vulnerable people. For such studies, a risk assessment should be performed. Not only to ensure that participants are protected from potential harm, but also to ensure the researcher is protected and cannot be held responsible in case a participant would become agitated.

7.4 Gaps Discovered in this Research

Several gaps have been highlighted in this work, and addressing these can provide guidance on developing better tools, materials and frameworks required for empirical investigation of ethical concerns in HRI. The first gap highlighted in this work is that there are ethical concerns regarding the use of social robots, namely emotional deception and emotional attachment. These concerns have been raised in the literature, but there is very little empirical evidence that supports them, and helps us understand their potential negative consequences.

Furthermore, it was found that there are no established methods, either in the form of questionnaires or through the use of sensors, to investigate emotional deception and emotional attachment, or any indirect and/or unintended consequences of the use of social robots in general. This raises the question whether the development of social robots should continue while we do not have the means to investigate potential negative consequences of their use. Perhaps a shift in research focus to include the development of metrics on unintended and/or indirect consequences of social robots can address this question.

Finally, as also presented in other work (Van Maris et al., 2019), existing frameworks do not sufficiently cover social robotics research, as the psychological impact of social robots is far greater than the issues addressed in existing frameworks due to their multi-modal interaction abilities. More research regarding these abilities and their (psychological) impact on different populations is required for us to be able to understand the potential concerns and address them accordingly. I want to highlight the use of the term *different populations*, as a final gap addressed in this work entails the fact that frameworks often address ethical concerns for vulnerable populations only, where I argue that many psy-

chological consequences of social robots can apply to anyone and not solely vulnerable people.

These gaps highlight that the tools, measurements available are not sufficient to investigate and determine potential negative consequences of AEE during human-robot interactions. Furthermore, the frameworks currently available are not designed specifically for social robots, resulting in gaps that can lead to negative outcomes if not addressed. Addressing these gaps will lead to the development of better materials, tools and frameworks that are more fit for empirical investigation of ethical concerns in HRI.

7.5 Contributions to Ethical HRI

As mentioned in the Introduction, this work contributed to the field of HRI in several ways. These contributions are summarised in more detail below:

Ethical HRI Literature - The importance of research regarding the impact of ethical AEE was highlighted, which can support successful human-robot interactions. Psychological and physiological effects of AEE were considered, as well as potential measures to investigate these factors. The following conclusions were drawn:

Physiology: Physiological data can be useful to gather insights on whether participants are aroused or not. It depends on the task of the experiment whether non-invasive equipment can be used for data gathering, as non-invasive sensors are sensitive to movement and require stationary tasks. Also, elevated arousal does not indicate whether the physical response is positive or negative so additional measurements besides physiological devices are needed for useful insights.

Speech: Speech prosody data can also provide insights on whether participants are aroused or not. As only a microphone is required as additional measurement that does not depend on the nature of the interaction this is the most promising feature that can be used to gather understanding of people's responses to robots besides questionnaires.

Behaviour: Behaviour analysis can provide useful observations on people's responses to social robots. However, expert knowledge is required to anal-

yse people's behaviour, and it will not be possible to determine whether the behaviour occurs due to the robot's behaviour or other external factors.

Ethical HRI Framework - The SOCRATES Framework was developed, that can support developers and producers of future social robots with designing social robot behaviours, to ensure there are minimal negative consequences for vulnerable users. This framework aimed to address the fact that often no specification is provided for when a person is categorised as vulnerable by forcing users to determine their target audience. Furthermore, reflective questions and recommendations were presented that evoke critical thinking and can assist in the identification of ethical concerns of social robot behaviours.

Ethical HRI Assessment - A survey was developed to assess the opinion of the general public on the findings from this research. Due to its high reliability, the survey is suitable for future use. The novel approach that was taken in the survey, where participants were provided with real contextual scenarios, allowed for participants to make more informed decisions.

Ethical HRI Evaluations - Two online surveys and three formal studies were conducted in this research. These experiments reported on participants' experiences, which led to a richer understanding of the impact that AEE can have on user experience during human-robot interactions, and the effect of longitudinal field-studies on older adults over time. This longitudinal field study resulted in several methodology recommendations on running a longitudinal field study:

1. Use non-invasive sensors for gathering data where possible, as to not distract the participant from the interaction task.
2. Include qualitative data in the analysis. The number of participants is likely to be low, but is likely to better represent the target audience than when participants have to visit a laboratory. Therefore, including qualitative data may provide richer insights and better understanding of the user experience.
3. As there are currently no established means to measure emotional deception and emotional attachment, a large arrangement of measurements should be used to acquire an understanding of these phenomena.

7.6 Future Work

This work provides many directions for future work. As presented in Section 7.4, several gaps have been discovered that should be addressed to ensure successful development of ethical social robots. Furthermore, there are opportunities to improve the SOCRATES framework and expand the research presented in this work, that will be discussed next.

7.6.1 Improving the SOCRATES Framework

First of all, the framework should be tested. This may lead to the need to update recommendations or reflective questions, or components may need to be added. Furthermore, it should be considered whether descriptions provided are clear enough for the stakeholders involved in the development and use of social robots and AEE.

The final survey was developed to gather the opinion of the general public, which was then integrated into the framework. This can be extended by sending the survey to experts and include their opinions as well. The framework can then be updated accordingly.

Furthermore, it can be explored how existing frameworks can be used to provide more detailed guidelines on their implementation. Frameworks provide general guidance, but more detailed information may be required for developers to consider potential negative consequences. This can be done in the form of a rule-based knowledge system. The frameworks can be used as inputs, and findings from HRI- and HCI-studies can be used to build a rule-based knowledge system that can provide *clear and concise* guidance on the development of ethical social robots. For example, emotion can be artificially expressed by social robots as investigated in this research, but also non-social robots may be capable of doing this, and even chat-bots can express emotions through text. In this case, the SOCRATES framework could apply to AEE by a social robot, where BS8611 (BS 8611:2016, 2016) is more useful for other robotic technologies and AIHLEG may be the best option for chat-bots as they can be perceived as artificial intelligence systems (AI HLEG, 2019). Therefore, depending on the technology that is being developed, combining all frameworks and outcomes in one rule-based knowledge system can provide researchers with clear guidance that can be applied to different but related technologies. Inspiration for the development of this rule-based knowledge system can be drawn from the 'safe HRI'

research field, where the COVR toolkit for collaborative robotics was introduced.This toolkit provides insights on what may lead to harmful situations in collaborative HRI (Bessler et al., 2018). One advantage of such a system is that results from existing work can be included to gain a more complete understanding of people's responses to new technologies, as well as incidental findings to try and limit unintentional harm. This should help guide a risk assessment when designing ethically safe technologies.

7.6.2 Extension of this Research

There are several possibilities to address the other limitations mentioned for this work. This includes more research regarding the use of physiological sensors, speech prosody and behaviour analysis for longitudinal field studies with vulnerable populations. Furthermore, the intensity of AEE and the type of interactions can be updated to be more representative of everyday interactions.

The context of the experiments in this study was restricted to a didactic setting and low emotion manipulation of the robot's behaviour, to gain a baseline understanding of whether participants could be emotionally deceived by a robot, and become attached to it if it did not interact with them at a personal level. However, it is worth investigating whether and how people's responses differ when the robot is capable of interacting with them at a personal level, and there is a higher level of emotion manipulation. It should then be considered how this would influence the SOCRATES framework.

Also, this research distinguished between positive and negative emotions that were defined as happy and sad respectively. Identifying more emotions and investigating people's responses could provide more insights regarding the ethical concerns of AEE. Perhaps only negative emotions lead to emotional deception as it may evoke empathetic responses from participants. This makes for interesting and essential future research.

Besides the type of emotion, the way this emotion is displayed could also influence people's responses to them. It is possible that there are specific modalities that are more likely to result in deception or attachment. More research on the impact of modalities used can provide insights on how to implement AEE with minimal ethical risks to ensure users only experience the benefits of AEE.

7.7 Concluding Remarks

In this work I aimed to bridge ethical perspectives and empirical methodologies, to gain a rich understanding of the psychological impact that artificial expression of emotion by social robots may have on older adults. Through conducting several user studies, including a longitudinal field study that resulted in a different and perhaps more realistic participant group than a lab-based study, I gained insights on user experiences when a social robot expresses artificial emotions. But far more importantly, I identified several gaps that indicate it is nearly impossible to successfully determine potential negative consequences of the use of social robots. This raises the question whether development of social robots should continue, if we do not have the means to understand their impact yet. Furthermore, I argued that psychological risks of social robots and their expression of artificial emotion are likely unintended and possibly also indirect, making it difficult to understand their impact and requiring more research to address this. I developed the SOCRATES framework to help developers and other stakeholders understand potential concerns of their technological advances, and to prompt them to critically reflect on what potential (unintended) negative consequences of their technological advances can be. This research and future work that will follow from it can help shape both policy and practice of human-robot interactions, to support a successful transition from 'the era of social robotics research' to 'the era of social robotics applications'.

References

Adar, E., Tan, D. S., and Teevan, J. (2013). Benevolent deception in human computer interaction. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, pp. 1863–1872.

AI HLEG (2019). *Ethics guidelines for trustworthy AI*. High-Level Expert Group on Artificial Intelligence, the European Commission.

Allegra, D., Alessandro, F., Santoro, C., and Stanco, F. (2018). Experiences in Using the Pepper Robotic Platform for Museum Assistance Applications. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 1033–1037.

Anderson, S. L. and Anderson, M. (2007). The consequences for human beings of creating ethical robots. In: *Human implications of human-robot interaction: Papers from the 2007 AAAI workshop*, pp. 1–4.

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., and Stolcke, A. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In: *Seventh International Conference on Spoken Language Processing*.

Aristoteles (1960). *The Rhetoric of Aristotle: an expanded translation with supplementary examples for students of composition and public speaking*. Appleton-Century-Crofts.

Arkin, R. C., Ulam, P., and Wagner, A. R. (2012). Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception. *Proceedings of the IEEE*. 100 (3), pp. 571–589.

Asaro, P. M. (2006). What should we want from a robot ethic? *The International Review of Information Ethics*. 6, pp. 9–16.

Asimov, I. (1941). Three laws of robotics. *Asimov, I. Runaround*.

Baek, H. J., Cho, C.-H., Cho, J., and Woo, J.-M. (2015). Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. *Telemedicine and e-Health*. 21 (5), pp. 404–414.

Baevsky, R. and Berseneva, A. (2008). Methodical recommendations use Kardivar system for determination of the stress level and estimation of the body adaptability - Standards of measurements and physiological interpretation. (no place), pp. 1–42.

Baisch, S., Kolling, T., and Knopf, M. (2017). Factors impacting on older and younger peoples' perceptions of elderly robot users. *Innovation in Aging*. 1 (1), p. 1190.

Banks, M. R., Willoughby, L. M., and Banks, W. A. (2008). Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. *Journal of the American Medical Directors Association*. 9 (3), pp. 173–177.

Barrett, E., Burke, M., Whelan, S., Santorelli, A., Oliveira, B. L., Cavallo, F., Dröes, R. M., Hopper, L., Fawcett-Henesy, A., Meiland, F. J., et al. (2019). Evaluation of a companion robot for individuals with dementia: quantitative findings of the MARIO project in an Irish residential care setting. *Journal of Gerontological Nursing*. 45 (7), pp. 36–45.

Bartholomew, K. (1990). Avoidance of intimacy: An attachment perspective. *Journal of Social and Personal relationships*. 7 (2), pp. 147–178.

Bartholomew, K. and Horowitz, L. M. (1991). Attachment styles among young adults: a test of a four-category model. *Journal of personality and social psychology*. 61 (2), pp. 226–244.

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*. 1 (1), pp. 71–81.

Baumann, M. F., Brändle, C., Coenen, C., and Zimmer-Merkle, S. (2019). Taking responsibility: A responsible research and innovation (RRI) perspective on insurance issues of semi-autonomous driving. *Transportation research part A: policy and practice*. 124, pp. 557–572.

Beauchamp, T. L. and Childress, J. F. (1979). Principles of biomedical ethics. *Ethics*. 4.

Bechade, L., Dubuisson-Duplessis, G., Pittaro, G., Garcia, M., and Devillers, L. (2019). Towards metrics of evaluation of pepper robot as a social companion for the elderly. In: *Advanced Social Interaction with Agents*. Springer, pp. 89–101.

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., and Sommavilla, G. (2013). Interpretation of emotional body language displayed by a humanoid robot: A case study with children. *International Journal of Social Robotics*. 5 (3), pp. 325–334.

Beer, J. M., Prakash, A., Mitzner, T. L., and Rogers, W. A. (2011). *Understanding robot acceptance*. Tech. rep. Georgia Institute of Technology, pp. 1–45.

Bente, G., Senokozlieva, M., Pennig, S., Al-Issa, A., and Fischer, O. (2008). Deciphering the secret code: A new methodology for the cross-cultural analysis of nonverbal behavior. *Behavior Research Methods*. 40 (1), pp. 269–277.

Bertel, L. B. (2011). *Peers: persuasive educational and entertainment robotics*. PhD thesis. Aalborg University.

Bertel, L. B. and Rasmussen, D. M. (2013). On being a peer: What persuasive technology for teaching can gain from social robotics in education. *International Journal of Conceptual Structures and Smart Applications (IJCSSA)*. 1 (2), pp. 58–68.

Bessler, J., Schaake, L., Bidard, C., Buurke, J. H., Lassen, A. E., Nielsen, K., Saenz, J., and Vicentini, F. (2018). Covr–towards simplified evaluation and validation of collaborative robotics applications across a wide range of domains based on robot safety skills. In: *International Symposium on Wearable Robotics*. Springer, pp. 123–126.

Betella, A. and Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one*. 11 (2), pp. 1–11.

Bishop, L., Maris, A. van, Dogramadzi, S., and Zook, N. (2019). Social robots: The influence of human and robot characteristics on acceptance. *Paladyn, Journal of Behavioral Robotics*. 10 (1), pp. 346–358.

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., et al. (2017). Principles of robotics: regulating robots in the real world. *Connection Science*. 29 (2), pp. 124–129.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.* 5 (9), pp. 341–345.

Borenstein, J. and Arkin, R. (2019). Robots, ethics, and intimacy: The need for scientific research. In: *On the cognitive, ethical, and scientific dimensions of artificial intelligence*. Springer, pp. 299–309.

Bowlby, J. (1958). The nature of the child's tie to his mother. *International journal of psycho-analysis*. 39, pp. 350–373.

Bowlby, J. (1969). *Attachment and Loss: Sadness and Depression.-NY: Basic Books, 1980.-Xv, 472 S (DLBÅ)*. Hogarth Press.

Breazeal, C. (2011). Social robots for health applications. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, pp. 5368–5371.

Breazeal, C. and Scassellati, B. (1999). How to build robots that make friends and influence people. In: *Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on*. Vol. 2. IEEE, pp. 858–863.

Brennan, K. A., Clark, C. L., and Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. *Attachment theory and close relationships*, pp. 46–76.

Bryson, J. J. (2018). Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology*. 20 (1), pp. 15–26.

BS 8611:2016 (2016). Robots and Robotic Devices: Guide to the Ethical Design and Application of Robots and Robotic Systems.

Cacioppo, J. T. and Patrick, W. (2008). *Loneliness: Human nature and the need for social connection*. New York: WW Norton & Company.

Callén, B., Domènech, M., López, D., and Tirado, F. (2009). Telecare research:(Cosmo) politicizing methodology. *ALTER-European Journal of Disability Research/Revue Européenne de Recherche sur le Handicap*. 3 (2), pp. 110–122.

Carlberg, C. (2017). *Predictive Analytics: Microsoft® Excel 2016*. Que Publishing.

Carros, F., Meurer, J., Löffler, D., Unbehaun, D., Matthies, S., Koch, I., Wieching, R., Randall, D., Hassenzahl, M., and Wulf, V. (2020). Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12.

Chen, H., Park, H. W., and Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education*. 150, pp. 1–19.

Clabaugh, C. and Matarić, M. (2018). Robots for the people, by the people: Personalizing human-machine interaction. *Science Robotics*. 3 (21), pp. 1–3.

Coeckelbergh, M. (2010). Health care, capabilities, and AI assistive technologies. *Ethical theory and moral practice*. 13 (2), pp. 181–190.

Coeckelbergh, M. (2011). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics*. 3 (2), pp. 197–204.

Coeckelbergh, M. (2012). Can we trust robots? *Ethics and information technology*. 14 (1), pp. 53–60.

Coeckelbergh, M. (2020). How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance. *International Journal of Social Robotics*, pp. 1–10.

Cohen, A. S., Minor, K. S., Najolia, G. M., and Hong, S. L. (2009). A laboratory-based procedure for measuring emotional expression from natural speech. *Behavior research methods*. 41 (1), pp. 204–212.

Collins, E. C. (2017). Vulnerable users: deceptive robotics. *Connection Science*. 29 (3), pp. 223–229.

Collins, N. L. and Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of personality and social psychology*. 58 (4), p. 644.

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of nonverbal behavior*. 28 (2), pp. 117–139.

Crumpton, J. and Bethel, C. L. (2016). A survey of using vocal prosody to convey emotion in robot speech. *International Journal of Social Robotics*. 8 (2), pp. 271–285.

D'mello, S. and Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*. 2 (4), pp. 1–39.

Danaher, J. (2020). Robot Betrayal: a guide to the ethics of robotic deception. *Ethics and Information Technology*, pp. 1–12.

Dautenhahn, K. (2007a). Methodology & themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*. 4 (1), pp. 103–108.

Dautenhahn, K. (2007b). Socially intelligent robots: dimensions of human–robot interaction. *Philosophical transactions of the royal society B: Biological sciences*. 362 (1480), pp. 679–704.

Dautenhahn, K. and Billard, A. (1999). Bringing up robots or—the psychology of socially intelligent robots: From theory to implementation, pp. 366–367.

Dautenhahn, K., Woods, S., Kaouri, C., Walters, M. L., Koay, K. L., and Werry, I. (2005). What is a robot companion-friend, assistant or butler? In: *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, pp. 1192–1197.

De Graaf, M. M. (2016). An ethical evaluation of human–robot relationships. *International journal of social robotics*. 8 (4), pp. 589–598.

De Graaf, M. M., Allouch, S. B., and Van Dijk, J. (2015). What makes robots social?: A user's perspective on characteristics for social human-robot interaction. In: *International Conference on Social Robotics*. Springer, pp. 184–193.

De Graaf, M. M., Allouch, S. B., and Dijk, J. A. van (2016). Long-term evaluation of a social robot in real homes. *Interaction studies*. 17 (3), pp. 462–491.

De Graaf, M. M., Ben Allouch, S., and Dijk, J. A. van (2018). A phased framework for long-term user acceptance of interactive technology in domestic environments. *new media & society*. 20 (7), pp. 2582–2603.

De Graaf, M. M. A. (2015). *Living with robots: investigating the user acceptance of social robots in domestic environments*. PhD thesis. University of Twente.

Dehais, F., Sisbot, E. A., Alami, R., and Causse, M. (2011). Physiological and subjective evaluation of a human–robot object hand-over task. *Applied ergonomics*. 42 (6), pp. 785–791.

Deng, E., Mutlu, B., and Mataric, M. (2019). Embodiment in socially interactive robots. *Foundations and Trends in Robotics*. 7 (4), pp. 1–85.

Derrick, B. and White, P. (2017). Comparing two samples from an individual Likert question. *International Journal of Mathematics and Statistics*. 18 (3).

Dragan, A., Holladay, R., and Srinivasa, S. (2015). Deceptive robot motion: synthesis, analysis and experiments. *Autonomous Robots*. 39 (3), pp. 331–345.

Edberg, A.-K. and Edfors, E. (2008). Nursing care for people with frontal-lobe dementia-difficulties and possibilities. *International psychogeriatrics*. 20 (2), p. 361.

Erden, M. S. (2013). Emotional postures for the humanoid-robot nao. *International Journal of Social Robotics*. 5 (4), pp. 441–456.

Esco, M. R. and Flatt, A. A. (2014). Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. *Journal of sports science & medicine*. 13 (3), pp. 535–541.

Eyssel, F., Hegel, F., Horstmann, G., and Wagner, C. (2010). Anthropomorphic inferences from emotional nonverbal cues: A case study. In: *19th international symposium in robot and human interactive communication*. IEEE, pp. 646–651.

Feil-Seifer, D. and Mataric, M. J. (2005). Defining socially assistive robotics. In: *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.* IEEE, pp. 465–468.

Feil-Seifer, D. and Mataric, M. J. (2011). Socially Assistive Robotics-Ethical Issues Related to Technology. *IEEE Robotics and Automation Magazine*. 18 (1), pp. 24–31.

Fernaeus, Y., Håkansson, M., Jacobsson, M., and Ljungblad, S. (2010). How do you play with a robotic toy animal? A long-term study of Pleo. In: *Proceedings of the 9th international Conference on interaction Design and Children*, pp. 39–48.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*. 42 (3-4), pp. 143–166.

Frennert, S. and Östlund, B. (2014). Seven matters of concern of social robots and older people. *International Journal of Social Robotics*. 6 (2), pp. 299–310.

Frey, S. and Pool, J. (1976). *A new approach to the analysis of visible behavior*. Departement of Psychology, University of Bern.

Fulmer, I. S., Barry, B., and Long, D. A. (2009). Lying and smiling: Informational and emotional deception in negotiation. *Journal of Business Ethics*. 88 (4), pp. 691–709.

Galaz, Z., Mekyska, J., Mzourek, Z., Smekal, Z., Rektorova, I., Eliasova, I., Kostalova, M., Mrackova, M., and Berankova, D. (2016). Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease. *Computer methods and programs in biomedicine*. 127, pp. 301–317.

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., Das, M., and Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30. 1.

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannan, P., Diebold Jr, A. R., Durbin, M., Edmonson, M. S., Fischer, J., Hymes, D., Kimball, S. T., et al. (1968). Proxemics [and comments and replies]. *Current anthropology*. 9 (2/3), pp. 83–108.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*. 53 (5), pp. 517–527.

Häring, M., Bee, N., and André, E. (2011). Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In: *2011 RO-MAN*. IEEE, pp. 204–209.

Heerink, M., Krose, B., Evers, V., and Wielinga, B. (2008). The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users. In: *RO-*

MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp. 695–700.

Heerink, M., Kröse, B., Evers, V., and Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: the almere model. *International journal of social robotics*. 2 (4), pp. 361–375.

Huang, L., Varnado, T., and Gillan, D. (2013). An exploration of robot builders' attachment to their LEGO robots. In: *Proceedings of the Human Factors and Ergonomics Society annual meeting*. Vol. 57. 1. SAGE Publications Sage CA: Los Angeles, CA, pp. 1825–1829.

Huber, A., Lammer, L., and Vincze, M. (2014). Do socially assistive robots compromise our moral autonomy? In: *Proc. of the Int. Conf. Going Beyond the Laboratory-Ethical and Societal Challenges for Robotics*.

Hung, L., Liu, C., Woldum, E., Au-Yeung, A., Berndt, A., Wallsworth, C., Horne, N., Gregorio, M., Mann, J., and Chaudhury, H. (2019). The benefits of and barriers to using a social robot PARO in care settings: a scoping review. *BMC geriatrics*. 19 (1), p. 232.

Hursthouse, R. and Pettigrove, G. (2020). Virtue ethics in stanford encyclopedia of philosophy. *Retrieved January*. 17, p. 2021.

Hutson, S., Lim, S. L., Bentley, P. J., Bianchi-Berthouze, N., and Bowling, A. (2011). Investigating the suitability of social robots for the wellbeing of the elderly. In: *International Conference on Affective Computing and Intelligent Interaction*. Springer, pp. 578–587.

Hyman, R. (1989). The psychology of deception. *Annual Review of Psychology*. 40 (1), pp. 133–154.

Ilicic, J. and Webster, C. M. (2011). Effects of multiple endorsements and consumer–celebrity attachment on attitude and purchase intention. *Australasian Marketing Journal (AMJ)*. 19 (4), pp. 230–237.

Jackson, J. (1991). Telling the truth. *Journal of medical ethics*. 17 (1), pp. 5–9.

James, J., Watson, C. I., and MacDonald, B. (2018). Artificial empathy in social robots: An analysis of emotions in speech. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 632–637.

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. 1 (9), pp. 389–399.

Johnson, D. O. and Cuijpers, R. H. (2019). Investigating the effect of a humanoid robot's head position on imitating human emotions. *International Journal of Social Robotics*. 11 (1), pp. 65–74.

Jung, M. and Hinds, P. (2018). Robots in the wild: A time for more robust theories of human-robot interaction. *ACM Transactions on Human-Robot Interaction*. 7 (1), pp. 1–5.

Kahn Jr, P. H. and Shen, S. (2017). NOC NOC, Who's There? A New Ontological Category (NOC) for Social Robots. *New Perspectives on Human Development*, pp. 13–142.

Kaminski, M. E., Rueben, M., Smart, W. D., and Grimm, C. M. (2016). Averting robot eyes. *Md. L. Rev.* 76, p. 983.

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*. 19 (1-2), pp. 61–84.

Kang, H. S., Makimoto, K., Konno, R., and Koh, I. S. (2019). Review of outcome measures in PARO robot intervention studies for dementia care. *Geriatric Nursing*. 41 (3), pp. 207–214.

Keefer, L. A., Landau, M. J., Rothschild, Z. K., and Sullivan, D. (2012). Attachment to objects as compensation for close others' perceived unreliability. *Journal of Experimental Social Psychology*. 48 (4), pp. 912–917.

Keltner, D. and Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*. 13 (5), pp. 505–521.

Khaksar, S. M. S., Khosla, R., Chu, M. T., and Shahmehr, F. S. (2016). Service innovation using social robot to reduce social vulnerability among older people in residential care facilities. *Technological Forecasting and Social Change*. 113, pp. 438–453.

Kidd, C. D., Taggart, W., and Turkle, S. (2006). A sociable robot to encourage social interaction among the elderly. In: *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, pp. 3972–3976.

Kirby, R., Forlizzi, J., and Simmons, R. (2010). Affective social robots. *Robotics and Autonomous Systems*. 58 (3), pp. 322–332.

Kory-Westlund, J. M. and Breazeal, C. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Frontiers in Robotics and AI*. 6 (81), pp. 1–17.

Kühnlenz, B., Erhart, M., Kainert, M., Wang, Z.-Q., Wilm, J., and Kühnlenz, K. (2018). Impact of trajectory profiles on user stress in close human-robot interaction. *at-Automatisierungstechnik*. 66 (6), pp. 483–491.

Laborde, S., Mosley, E., and Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting. *Frontiers in psychology*. 8, p. 213.

Leite, I., Martinho, C., and Paiva, A. (2013). Social robots for long-term interaction: a survey. *International Journal of Social Robotics*. 5 (2), pp. 291–308.

Leite, I., Martinho, C., Pereira, A., and Paiva, A. (2009). As time goes by: Long-term evaluation of social presence in robotic companions. In: *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, pp. 669–674.

Leong, B. and Selinger, E. (2019). Robot eyes wide shut: Understanding dishonest anthropomorphism. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 299–308.

Li, D., Browne, G. J., and Chau, P. Y. (2006). An empirical investigation of web site use using a commitment-based model. *Decision Sciences*. 37 (3), pp. 427–444.

Lindblom, J. and Andreasson, R. (2016). Current challenges for UX evaluation of human-robot interaction. In: *Advances in ergonomics of manufacturing: Managing the enterprise of the future*. Springer, pp. 267–277.

Mandler, G. (1980). The generation of emotion: A psychological theory. In: *Theories of emotion*. Elsevier, pp. 219–243.

Marmpena, M., Lim, A., and Dahl, T. S. (2018). How does the robot feel? perception of valence and arousal in emotional body language. *Paladyn, Journal of Behavioral Robotics*. 9 (1), pp. 168–182.

Matthias, A. (2015). Robot lies in health care: When is deception morally permissible? *Kennedy Institute of Ethics Journal*. 25 (2), pp. 169–162.

McColl, D. and Nejat, G. (2014). Recognizing emotional body language displayed by a human-like social robot. *International Journal of Social Robotics*. 6 (2), pp. 261–280.

Michel, A. and Carpenter, J. (Oct. 2013). *The professor of robot love*. Available from: `https://dronecenter.bard.edu/interview-professor-robot-love/`.

Mircioiu, C. and Atkinson, J. (2017). A comparison of parametric and non-parametric methods applied to a Likert scale. *Pharmacy*. 5 (2), p. 26.

Misselhorn, C., Pompe, U., and Stapleton, M. (2013). Ethical considerations regarding the use of social robots in the fourth age. *GeroPsych*. 26, pp. 121–133.

Moshkina, L., Park, S., Arkin, R. C., Lee, J. K., and Jung, H. (2011). TAME: Time-varying affective response for humanoid robots. *International Journal of Social Robotics*. 3 (3), pp. 207–221.

Mugge, R. (2004). Personalizing product appearance: the effect on product attachment. In: *Proceedings of 2004 International Conference on Design and Emotion. Ankara, Turkey*.

Murphy, J., Gretzel, U., and Pesonen, J. (2019). Marketing robot services in hospitality and tourism: the role of anthropomorphism. *Journal of Travel & Tourism Marketing*. 36 (7), pp. 784–795.

Murphy, R. and Woods, D. D. (2009). Beyond Asimov: the three laws of responsible robotics. *IEEE intelligent systems*. 24 (4), pp. 14–20.

Murray, K. (2011). *Microsoft Office 365: Connect and collaborate virtually anywhere, anytime*. Microsoft Press.

Musikanski, L., Havens, J., and Gunsch, G. (2018). IEEE P7010 Well-being Metrics Standard for Autonomous and Intelligent Systems. *IEEE, New York, NY, Tech. Rep*.

Mutlu, B., Osman, S., Forlizzi, J., Hodgins, J., and Kiesler, S. (2006). Task structure and user attributes as elements of human-robot interaction design. In: *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, pp. 74–79.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., and Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*. 53 (4), pp. 695–699.

Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., and Takayama, L. (2005). Improving automotive safety by pairing driver emotion and car voice emotion. In: *CHI'05 extended abstracts on Human factors in computing systems*, pp. 1973–1976.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, C. (1995). Can computer personalities be human personalities? In: *Conference companion on Human factors in computing systems*. ACM, pp. 228–229.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 72–78.

Neave, N., Tyson, H., McInnes, L., and Hamilton, C. (2016). The role of attachment style and anthropomorphism in predicting hoarding behaviours in a non-clinical sample. *Personality and Individual Differences*. 99, pp. 33–37.

Niculescu, A., Dijk, B. van, Nijholt, A., Li, H., and See, S. L. (2013). Making social robots more attractive: the effects of voice pitch, humor and empathy. *International journal of social robotics*. 5 (2), pp. 171–191.

Niemelä, M., Heikkilä, P., Lammi, H., and Oksman, V. (2019). A social robot in a shopping mall: studies on acceptance and stakeholder expectations. In: *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*. Springer, pp. 119–144.

Nikolić, M., Vente, W. de, Colonnesi, C., and Bögels, S. M. (2016). Autonomic arousal in children of parents with and without social anxiety disorder: A high-risk study. *Journal of Child Psychology and Psychiatry*. 57 (9), pp. 1047–1055.

Ostrowski, A. K., DiPaola, D., Partridge, E., Park, H. W., and Breazeal, C. (2019). Older adults living with social robots: promoting social connectedness in long-term communities. *IEEE Robotics & Automation Magazine*. 26 (2), pp. 59–70.

Paiva, A., Leite, I., and Ribeiro, T. (2014). Emotion modeling for social robots. *The Oxford handbook of affective computing*, pp. 296–308.

Palan, S. and Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*. 17, pp. 22–27.

Pandey, A. K. and Gelin, R. (2018). A mass-produced sociable humanoid robot: pepper: the first machine of its kind. *IEEE Robotics & Automation Magazine*. 25 (3), pp. 40–48.

Picard, R. W. (2000). *Affective computing*. MIT press.

Pot, E., Monceaux, J., Gelin, R., and Maisonnier, B. (2009). Choregraphe: a graphical tool for humanoid robot programming. In: *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, pp. 46–51.

Prescott, T. J. (2017). Robots are not just tools. *Connection Science*. 29 (2), pp. 142–149.

Prescott, T. J. and Robillard, J. M. (2020). Are Friends Electric? The Benefits and Risks of Human-Robot Relationships. *iScience*, p. 101993.

Price, W. N. and Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature medicine*. 25 (1), pp. 37–43.

Pripfl, J., Körtner, T., Batko-Klein, D., Hebesberger, D., Weninger, M., Gisinger, C., Frennert, S., Eftring, H., Antona, M., Adami, I., et al. (2016). Results of a real world trial with a mobile social service robot for older adults. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 497–498.

Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, S. for, Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., and Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*. 49 (8), pp. 1017–1034.

Pu, L., Moyle, W., Jones, C., and Todorovic, M. (2019). The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. *The Gerontologist*. 59 (1), e37–e51.

Qualtrics (2013). Qualtrics. *Provo, UT, USA*.

Quirin, M., Kazén, M., and Kuhl, J. (2009). When nonsense sounds happy or helpless: the implicit positive and negative affect test (IPANAT). *Journal of personality and social psychology*. 97 (3), pp. 500–516.

Read, R. G. and Belpaeme, T. (2010). Interpreting non-linguistic utterances by robots: studying the influence of physical appearance. In: *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, pp. 65–70.

Reeves, B. and Nass, C. (1996). Media equation theory. *Retrieved March*. 5, p. 2009.

Reis, H. T., Collins, W. A., and Berscheid, E. (2000). The relationship context of human behavior and development. *Psychological bulletin*. 126 (6), p. 844.

Rosenthal-von der Pütten, A. M., Krämer, N. C., and Herrmann, J. (2018). The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics*. 10 (5), pp. 569–582.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., and Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*. 5 (1), pp. 17–34.

Russell, D. C. (2013). *The Cambridge companion to virtue ethics*. Cambridge University Press.

Saerbeck, M. and Bartneck, C. (2010). Perception of affect elicited by robot motion. In: *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 53–60.

Scassellati, B., Boccanfuso, L., Huang, C.-M., Mademtzi, M., Qin, M., Salomons, N., Ventola, P., and Shic, F. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*. 3 (21), pp. 1–9.

Scheutz, M. (2011). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. *Robot ethics: The ethical and social implications of robotics*, pp. 205–222.

Schifferstein, H. N. and Zwartkruis-Pelgrim, E. P. (2008). Consumer-product attachment: Measurement and design implications. *International journal of design*. 2 (3), pp. 1–13.

Schrum, M. L., Johnson, M., Ghuy, M., and Gombolay, M. C. (2020). Four Years in Review: Statistical Practices of Likert Scales in Human-Robot Interaction Studies. In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 43–52.

Schuller, B., Rigoll, G., and Lang, M. (2003). Hidden Markov model-based speech emotion recognition. In: *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).* Vol. 2. IEEE, pp. 1–4.

Sharkey, A. (2014). Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. *Ethics and Information Technology*. 16 (1), pp. 63–75.

Sharkey, A. and Sharkey, N. (2011). Children, the elderly, and interactive robots. *IEEE Robotics & Automation Magazine*. 18 (1), pp. 32–38.

Sharkey, A. and Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics and information technology*. 14 (1), pp. 27–40.

Sharkey, A. and Sharkey, N. (2020). We need to talk about deception in social robotics! *Ethics and Information Technology*, pp. 1–8.

Sharkey, N. (2017). Why robots should not be delegated with the decision to kill. *Connection Science*. 29 (2), pp. 177–186.

Shim, J. and Arkin, R. C. (2013). A Taxonomy of Robot Deception and its Benefits in HRI. In: *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. IEEE, pp. 2328–2335.

Sorell, T. and Draper, H. (2014). Robot carers, ethics, and older people. *Ethics and Information Technology*. 16 (3), pp. 183–195.

Sparrow, R. (2002). The march of the robot dogs. *Ethics and information Technology*. 4 (4), pp. 305–318.

Sparrow, R. (2016). Kicking a robot dog. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 229–229.

Sparrow, R. (2020). Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *International Journal of Social Robotics*, pp. 1–7.

Sparrow, R. and Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*. 16 (2), pp. 141–161.

SPSS, IBM (2017). IBM SPSS Statistics for Windows, version 25. *Armonk, NY: IBM SPSS Corp.[Google Scholar]*.

Steinfeld, A., Jenkins, O. C., and Scassellati, B. (2009). The oz of wizard: simulating the human for interaction research. In: *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 101–108.

Stucki, G. (2005). International Classification of Functioning, Disability, and Health (ICF): a promising framework and classification for rehabilitation medicine. *American journal of physical medicine & rehabilitation*. 84 (10), pp. 733–740.

Suh, K.-S., Kim, H., and Suh, E. K. (2011). What if your avatar looks like you? Dual-congruity perspectives for avatar use. *MIS Quarterly*. 35 (3), pp. 711–729.

Sullins, J. P. (2012). Robots, love, and sex: the ethics of building a love machine. *IEEE transactions on affective computing*. 3 (4), pp. 398–409.

Tao, J. and Tan, T. (2005). Affective computing: A review. In: *International Conference on Affective computing and intelligent interaction*. Springer, pp. 981–995.

Tapus, A. and Mataric, M. J. (2007). Emulating Empathy in Socially Assistive Robotics. In: *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, pp. 93–96.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., and Karjalainen, P. A. (2014). Kubios HRV–heart rate variability analysis software. *Computer methods and programs in biomedicine*. 113 (1), pp. 210–220.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2017). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, Overview: Version 2*.

Thimmesch-Gill, Z., Harder, K. A., and Koutstaal, W. (2017). Perceiving emotions in robot body language: Acute stress heightens sensitivity to negativity while attenuating sensitivity to arousal. *Computers in Human Behavior*. 76, pp. 59–67.

Trivers, R. (2011). *The folly of fools: The logic of deceit and self-deception in human life*. Basic Books (AZ).

Tsang, K. K. (2012). The use of midpoint on Likert Scale: The implications for educational research. *Hong Kong Teachers' Centre Journal*. 11 (1), pp. 121–130.

Turkle, S. (2006). A nascent robotics culture: New complicities for companionship. *American Association for Artificial Intelligence Technical Report Series AAAI*.

Turkle, S. (2007). Authenticity in the age of digital companions. *Interaction studies*. 8 (3), pp. 501–517.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

Unbehaun, D., Aal, K., Carros, F., Wieching, R., and Wulf, V. (2019). Creative and Cognitive Activities in Social Assistive Robots and Older Adults: Results from an Exploratory Field Study with Pepper. In: *Proceedings of the 17th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing and the Design of Cooperation Technologies - Demos and Posters, Reports of the European Society for Socially Embedded Technologies*. European Society for Socially Embedded Technologies (EUSSET), pp. 1–4.

Vallor, S. (2018). An Ethical Toolkit for Engineering/Design Practice. *Markkula Center for Applied Ethics, Santa Clara University, https://www. scu. edu/ethics-in-technology-practice/ethical-toolkit*.

Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. *Philosophy & Technology*. 24 (3), pp. 251–268.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Vallor, S. and Bekey, G. A. (2017). Artificial intelligence and the ethics of self-learning robots.

Van Maris, A., Zook, N., Studley, M., and Dogramadzi, S. (2019). The need for ethical principles and guidelines in social robots. *Artificial intelligence, robots and ethics: Proceedings of the 4th international conference on Robot Ethics and Standards (ICRES 2019)*, pp. 19–24.

Van Maris, A. (2018). The Effect of Affective Robot Behaviour on the Level of Attachment After One Interaction. In: *International PhD Conference on Safe and Social Robotics (SSR-2018)*, pp. 1–4.

Van Maris, A., Dogramadzi, S., Zook, N., Studley, M., Winfield, A., and Caleb-Solly, P. (2020a). Speech Related Accessibility Issues in Social Robots. In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 505–507.

Van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., and Dogramadzi, S. (2018). Ethical considerations of (contextually) affective robot behaviour. In: *Hybrid Worlds: Societal and Ethical Challenges - Proceedings of the International Conference on Robot Ethics and Standards (ICRES 2018)*. CLAWAR Association Ltd, pp. 13–19.

Van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., and Dogramadzi, S. (2020b). Designing ethical social robots—a longitudinal field study with older adults. *Frontiers in Robotics and AI*. 7.

Van Wynsberghe, A. (2012). *Designing Robots with Care*. PhD thesis. University of Twente.

Vandemeulebroucke, T., Casterlé, B. D. de, and Gastmans, C. (2018). The use of care robots in aged care: A systematic review of argument-based ethics literature. *Archives of gerontology and geriatrics*. 74, pp. 15–25.

Vasco, V., Willemse, C., Chevalier, P., De Tommaso, D., Gower, V., Gramatica, F., Tikhanoff, V., Pattacini, U., Metta, G., and Wykowska, A. (2019). Train with Me: A Study Comparing a Socially Assistive Robot and a Virtual Agent for a Rehabilitation Task. In: *International Conference on Social Robotics*. Springer, pp. 453–463.

Vescio, B., Salsone, M., Gambardella, A., and Quattrone, A. (2018). Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating

heart rate variability in healthy subjects in short-and long-term recordings. *Sensors*. 18 (844), pp. 1–14.

Vicedo, M. (2009). The father of ethology and the foster mother of ducks: Konrad Lorenz as expert on motherhood. *Isis*. 100 (2), pp. 263–291.

Vinciarelli, A., Pantic, M., Bourlard, H., and Pentland, A. (2008). Social signal processing: state-of-the-art and future perspectives of an emerging domain. In: *Proceedings of the 16th ACM international conference on Multimedia*, pp. 1061–1070.

Voiculescu, A. (2017). Reflections on the EPSRC Principles of Robotics from the new far-side of the law. *Connection Science*. 29 (2), pp. 160–169.

Wada, K. and Shibata, T. (2007). Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. *IEEE transactions on robotics*. 23 (5), pp. 972–980.

Wagner, A. R. and Arkin, R. C. (2011). Acting deceptively: Providing robots with the capacity for deception. *International Journal of Social Robotics*. 3 (1), pp. 5–26.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*. 54 (6), pp. 1063–1070.

Weijers, S. (2013). *Exploring Human-Robot Social Relations*. MA thesis. University of Twente.

Wilson, C. (2017). Is it love or loneliness? Exploring the impact of everyday digital technology use on the wellbeing of older adults. *Ageing & Society*, pp. 1–25.

Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*. 2 (2), pp. 46–48.

Winkle, K., Lemaignan, S., Caleb-Solly, P., Leonards, U., Turton, A., and Bremner, P. (2019). Effective persuasion strategies for socially assistive robots. In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 277–285.

Wood, L. J., Zaraki, A., Robins, B., and Dautenhahn, K. (2019). Developing kaspar: a humanoid robot for children with autism. *International Journal of Social Robotics*, pp. 1–18.

World Health Organization (2013). *The World Health Organization Quality of Life (WHO-QOL)*.

World Health Organization (2015). *World report on ageing and health*. World Health Organization.

Wu, Y.-h., Wrobel, J., Cornuet, M., Kerhervé, H., Damnée, S., and Rigaud, A.-S. (2014). Acceptance of an assistive robot in older adults: a mixed-method study of human–robot interaction over a 1-month period in the Living Lab setting. *Clinical interventions in aging*. 9, pp. 801–811.

A Questionnaires

A1 Adapted Godspeed questionnaire

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly
Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive
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
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
Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Surprised	1	2	3	4	5	Quiescent

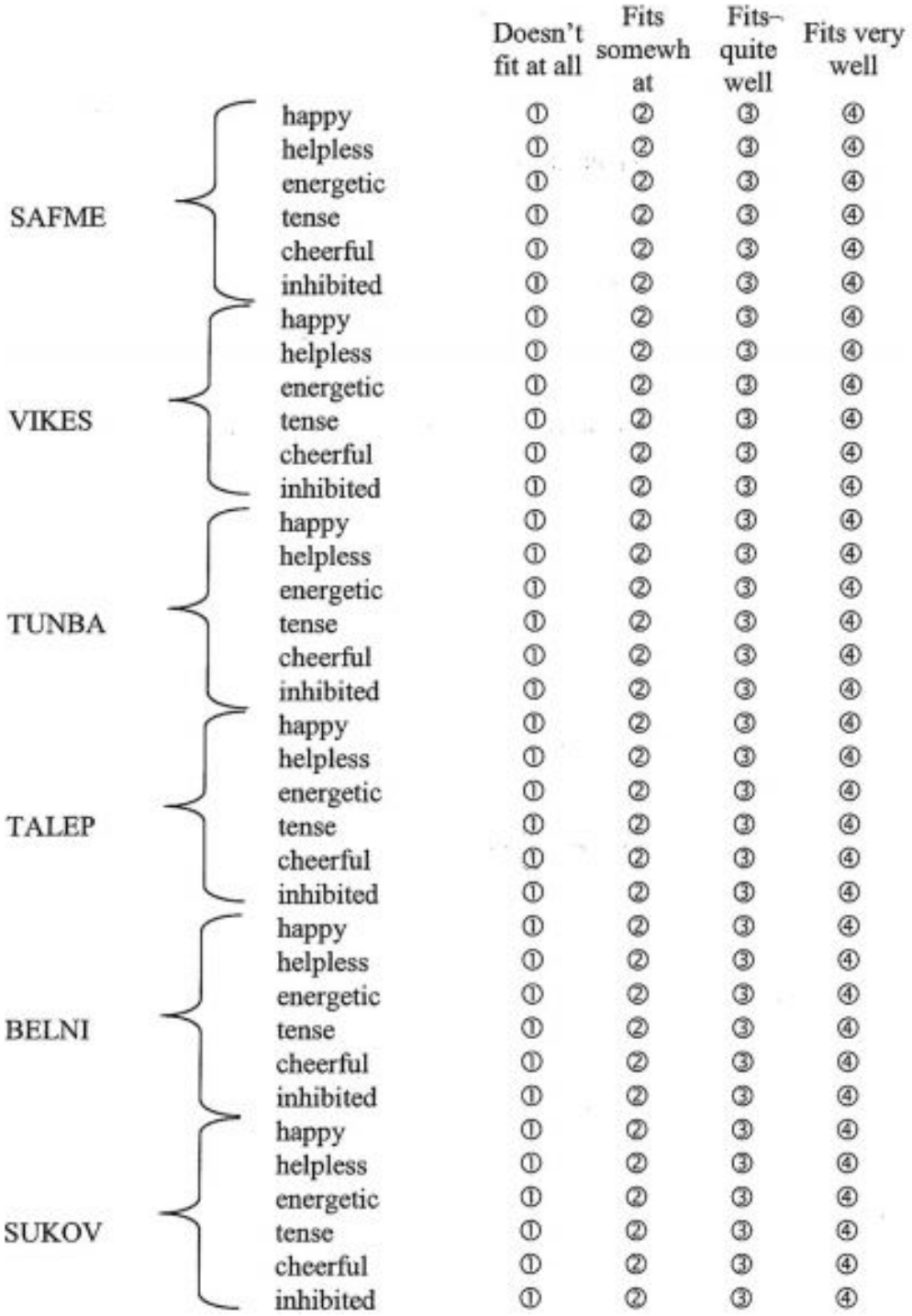
A2 Adapted Almere model of trust questionnaire

If I should use the robot, I would be afraid to make mistakes with it	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
If I should use the robot, I would be afraid to break something	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I find the robot scary	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I find the robot intimidating	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think it's a good idea to use the robot	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
The robot would make life more interesting	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
It's good to make use of the robot	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I enjoy the robot talking to me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I enjoy doing things with the robot	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I find the robot enjoyable	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I find the robot fascinating	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I find the robot boring	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think I will know quickly how to use the robot	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I find the robot easy to use	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think I can use the robot without any help	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think I can use the robot when there is someone around to help me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think I can use the robot when I have a good manual	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I consider the robot a pleasant conversational partner	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I find the robot pleasant to interact with	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I feel the robot understands me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think the robot is nice	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think the robot is useful to me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
It would be convenient for me to have the robot	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think the robot can help me with many things	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think the staff would like me using the robot	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I think it would give a good impression if I should use the robot	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
When interacting with the robot I felt like I'm talking to a real person	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
It sometimes felt as if the robot was really looking at me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I can imagine the robot to be a living creature	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I often think the robot is not a real person	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Sometimes the robot seems to have real feelings	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I would trust the robot if it gave me advice	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I would follow the advice the robot gives me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree

A3 PANAS - explicit mood

Interested	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Distressed	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Excited	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Upset	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Strong	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Guilty	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Scared	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Hostile	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Enthusiastic	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Proud	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Irritable	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Alert	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Ashamed	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Inspired	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Nervous	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Determined	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Attentive	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Jittery	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Active	Very slightly or not at all	A little	Moderately	Quite a bit	A lot
Afraid	Very slightly or not at all	A little	Moderately	Quite a bit	A lot

A4 IPANAT - implicit mood

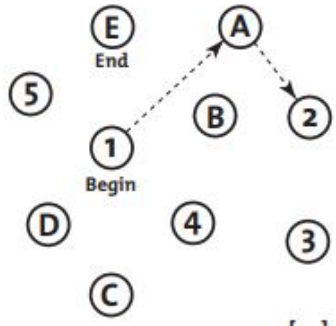


A5 Montreal Cognitive Assessment - MOCA

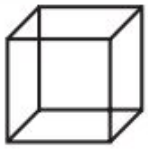
MONTREAL COGNITIVE ASSESSMENT (MOCA)

NAME :
Education : Date of birth :
Sex : DATE :

VISUOSPATIAL / EXECUTIVE — POINTS



Copy cube



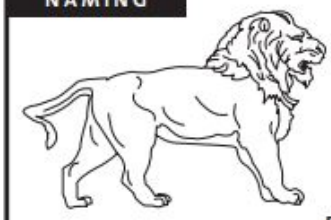
[ ] [ ]

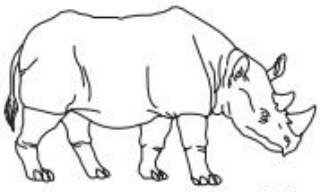
Draw CLOCK (Ten past eleven) (3 points)

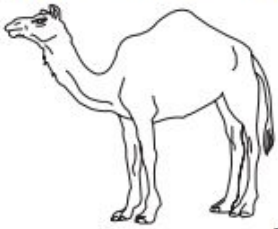
[ ] Contour [ ] Numbers [ ] Hands

__/5

NAMING







[ ] [ ] [ ]

__/3

MEMORY — Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.

	FACE	VELVET	CHURCH	DAISY	RED
1st trial					
2nd trial					

No points

ATTENTION — Read list of digits (1 digit/ sec.).

Subject has to repeat them in the forward order [ ] 2 1 8 5 4

Subject has to repeat them in the backward order [ ] 7 4 2

__/2

Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors

[ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B

__/1

Serial 7 subtraction starting at 100 [ ] 93 [ ] 86 [ ] 79 [ ] 72 [ ] 65

4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt**

__/3

LANGUAGE — Repeat : I only know that John is the one to help today. [ ]

The cat always hid under the couch when dogs were in the room. [ ]

__/2

Fluency / Name maximum number of words in one minute that begin with the letter F [ ] ____ (N ≥ 11 words)

__/1

ABSTRACTION — Similarity between e.g. banana - orange = fruit [ ] train – bicycle [ ] watch - ruler

__/2

DELAYED RECALL

	Has to recall words WITH NO CUE	FACE	VELVET	CHURCH	DAISY	RED	Points for UNCUED recall only
		[ ]	[ ]	[ ]	[ ]	[ ]	
Optional	Category cue						
	Multiple choice cue						

__/5

ORIENTATION [ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] City

__/6

© Z.Nasreddine MD Version November 7, 2004

www.mocatest.org

Normal ≥ 26 / 30

TOTAL __/30

Add 1 point if ≤ 12 yr edu

A6 Attachment style questionnaire

I prefer not to show people that are dear to me how I feel deep down	SD	D	Neither A nor D	A	SA
I worry about being abandoned	SD	D	Neither A nor D	A	SA
I am very comfortable being close with people that are dear to me	SD	D	Neither A nor D	A	SA
I worry a lot about my relationships	SD	D	Neither A nor D	A	SA
Just when somebody who is dear to me starts to get close, I find myself pulling away	SD	D	Neither A nor D	A	SA
I worry that people that are dear to me won't care about me as much as I care about them	SD	D	Neither A nor D	A	SA
I get uncomfortable when someone that is dear to me wants to be very close	SD	D	Neither A nor D	A	SA
I worry a fair amount about losing people that are dear to me	SD	D	Neither A nor D	A	SA
I often wish that the feelings of someone dear to me were as strong as my feelings for him/her	SD	D	Neither A nor D	A	SA
I want to get close to people that are dear to me, but I keep pulling back	SD	D	Neither A nor D	A	SA
I often want to merge completely with people that are dear to me, and this sometimes scares them away	SD	D	Neither A nor D	A	SA
I am nervous when someone dear to me gets too close with me	SD	D	Neither A nor D	A	SA
I worry about being alone	SD	D	Neither A nor D	A	SA
I feel comfortable sharing my private thoughts and feelings with people that are dear to me	SD	D	Neither A nor D	A	SA
My desire to be very close sometimes scares people away	SD	D	Neither A nor D	A	SA
I try to avoid getting too close with people that are dear to me	SD	D	Neither A nor D	A	SA
I need a lot of reassurance that I am loved by people that are dear to me	SD	D	Neither A nor D	A	SA
I find it relatively easy to get close to people that are dear to me	SD	D	Neither A nor D	A	SA
Sometimes I feel that I force people that are dear to me to show more feeling, more commitment	SD	D	Neither A nor D	A	SA
I find it difficult to allow myself to depend on people that are dear to me	SD	D	Neither A nor D	A	SA
I do not often worry about being abandoned	SD	D	Neither A nor D	A	SA
I prefer not to be too close to people that are dear to me	SD	D	Neither A nor D	A	SA
If I can't get people that are dear to me to show interest in me, I get upset or angry	SD	D	Neither A nor D	A	SA
I tell people that are dear to me just about everything	SD	D	Neither A nor D	A	SA
I find that people that are dear to me don't want to get as close as I would like	SD	D	Neither A nor D	A	SA
I usually discuss my problems and concerns with people that are dear to me	SD	D	Neither A nor D	A	SA
When I'm not involved in a relationship, I feel somewhat anxious and insecure	SD	D	Neither A nor D	A	SA
I feel comfortable depending on people that are dear to me	SD	D	Neither A nor D	A	SA
I get frustrated when people that are dear to me are not around as much as I would like	SD	D	Neither A nor D	A	SA
I don't mind asking people that are dear to me for comfort, advice or help	SD	D	Neither A nor D	A	SA
I get frustrated if people that are dear to me are not available when I need them	SD	D	Neither A nor D	A	SA
It helps to turn to people that are dear to me in times of need	SD	D	Neither A nor D	A	SA
When people that are dear to me disapprove of me, I feel really bad about myself	SD	D	Neither A nor D	A	SA
I turn to people that are dear to me for many things, including comfort and reassurance	SD	D	Neither A nor D	A	SA
I resent it when people that are dear to me spend time away from me	SD	D	Neither A nor D	A	SA

Abbreviations are used for layout reasons here, full options were provided in the study.

A7 Attachment questionnaire

I have a personal bond with Pepper	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Pepper is very dear to me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Peper has a special role in my life	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Pepper does not emotionally affect me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I have no feelings for Pepper	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I feel emotionally connected to Pepper	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Pepper has no special meaning to me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Pepper means a lot to me	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I feel emotionally attached to Pepper	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
If Pepper could be with you as an informative companion, how often would you use it?	Daily	Weekly	Monthly	Not at all	

A8 Final Interview

The following questions were asked in a random order during the final interview:

- What role would you like Pepper to play in your life? Please elaborate.
- Was there anything you did not like about Pepper or its behaviour? Please elaborate.
- Do you think you would get bored of Pepper, if you could use it whenever you want? Please elaborate.
- Will you miss Pepper? Please elaborate.
- Do you find the interactions with Pepper had an influence on your mood? Please elaborate.
- There is a growing concern amongst some roboticists that social robot behaviours are deceptive, as robots do not and cannot feel emotions, nor do they have any real interest in the person they are interacting with. Do you think Pepper was deceptive? If so, do you feel that deception was acceptable? Please elaborate.
- Is there anything else you would like to add?

B Implementation of AEE

Pre-determined movements used to convey neutral behaviour:

- ShowNearDistanceLeft_01
- BothArmsBumpInFront_01
- ShowNearDistanceRight_01
- GoToStance_Space&Time_LeanRight
- GoToStance_Space&Time_Center
- GoToStance_Space&Time_LeanLeft
- GoToPosture(Stand)

Pre-determined movements used to convey happy behaviour:

- GoToStance_Exclamation_LeanRight
- GoToStance_Exclamation_LeanBack
- GoToStance_Exclamation_LeanLeft
- StrongNodGatherArmsInFront_01
- Excited
- FastPointAtUserLeftArm_01
- CircleBothArmsLeaningFront_01
- StrongArmsAndLegsBump_LeanRight_01
- OfferBothHands_HeadNod_LeanLeft_01
- StrongArmsOpen_FlexEnd_01

Pre-determined movements used to convey sad behaviour:

- GoToStance_Exclamation_LeanRight
- GoToStance_Exclamation_LeanBack
- GoToStance_Exclamation_LeanLeft
- BothArmsUpAndDown_HeadShake_01
- SlowlyOfferBothHands_01

- FancyRightArmCircle_LeanRight_01
- RightArmUpAndDownWithBump_HeadShake_01

Predetermined animated-say-boxes used:

- Stand/Gestures/Salute_1
- Stand/Gestures/Hey_2
- Stand/Gestures/Explain_1
- Stand/Gestures/Explain_11
- Stand/Gestures/Explain_6
- Stand/Gestures/Explain_2
- Stand/Gestures/Explain_10

An example sequence for a neutral interactions is shown below. If it was aimed to convey happiness or sadness, the item to which the emotion applied would be exchanged for one of the pre-determined movements for happiness or sadness.

ShowNearDistanceLeft_01
GoToPosture(Stand)
Wait(0.5)
GoToStance_Space&Time_LeanRight
BothArmsBumpInFront_01
GoToPosture(Stand)
Wait(0.5)
Stand/Gestures/Explain_11
GoToStance_Space&Time_Center
GoToPosture(Stand)

C Online evaluation of implemented behaviours - additional results

Results of non-parametric tests used

A Friedman test was conducted to compare whether the three implemented robot behaviours were rated significantly differently from each other. $\chi^2(2) = 218.0$, $p < 0.001$. Post-hoc analysis using Wilcoxon tests with Bonferroni correction applied (new $p = 0.017$) indicated that happy behaviour was rated significantly higher than neutral behaviour (Z = -7.21, $p < 0.001$) and sad behaviour (Z = -10.60, $p < 0.001$), and neutral behaviour was rated significantly higher than sad behaviour (Z = -10.38, $p < 0.001$).

Age

A Spearman's rank-order correlation was conducted to determine the relationship between participants' age and their ratings of the different robot behaviours. Strong, negative correlations were found between participants' age and their rating of happy behaviour ($r_s(159)$ = -0.41, $p < 0.001$) and neutral behaviour ($r_s(159)$ = -0.33, $p < 0.001$). A positive correlation was found between participants' age and their rating of happy behaviour ($r_s(159) = 0.17$, $p = 0.03$).

A Kruskal-Wallis H test was conducted to investigate whether participants' age had an impact on their rating of the implemented behaviours. It was found that age influenced rating of happy ($H = 30.25$, $p < 0.001$) and neutral ($H = 19.19$, $p = 0.002$) behaviour, but not of sad behaviour ($H = 5.24$, $p = 0.39$). Details for differences between age groups following Mann Whitney U tests can be found in Table C.1. The ratings per age group are visualised in Figure 4.7.

Gender

Mann Whitney U test indicated that there was no influence of participants' gender on their rating of the emotive robot behaviours, as can be found in Table C.2.

	Happy		Neutral		Sad	
	Z	*p*	*Z*	*p*	*Z*	*p*
18-24 / 25-34	-0.03	0.97	-0.23	0.82	-0.09	0.93
18-24 / 35-44	0.01	**0.011**	0.08	0.09	1.00	1.00
18-24 / 45-54	-2.99	**0.003**	-1.95	0.051	-0.85	0.40
18-24 / 55-64	-3.90	**0.000**	-2.74	**0.006**	-1.40	0.16
18-24 / 65-74	0.07	0.09	0.04	**0.04**	0.18	0.20
25-34 / 35-44	-2.57	**0.01**	-2.09	**0.04**	-0.10	0.92
25-34 / 45-54	-3.21	**0.001**	-2.42	**0.02**	-1.03	0.30
25-34 / 55-64	-4.44	**0.000**	-3.58	**0.000**	-1.69	0.09
25-34 / 65-74	0.09	0.10	0.02	**0.02**	0.19	0.20
35-44 / 45-54	-0.16	0.88	-0.03	0.98	-0.87	0.39
35-44 / 55-64	-1.02	0.31	-0.67	0.50	-1.31	0.19
35-44 / 65-74	0.59	0.60	0.39	0.41	0.23	0.25
45-54 / 55-64	-1.14	0.25	-0.73	0.47	-0.29	0.77
45-54 / 65-74	0.67	0.70	0.34	0.36	0.52	0.56
55-64 / 65-74	0.89	0.90	0.52	0.55	0.57	0.58

TABLE C.1: Comparison of ratings of emotive robot behaviours (happy, neutral, sad) per participants' age range.

	Happy	Neutral	Sad
Z	-0.81	-0.59	-0.58
p	0.42	0.56	0.57

TABLE C.2: Comparison of ratings of emotive robot behaviours (happy, neutral, sad) per participants' gender.

Familiarity with social robots

Kruskal-Wallis H test was conducted to investigate the impact of participants' familiarity with social robots on their ratings. A significant impact of familiarity was found for ratings of happy behaviour ($H = 16.04$, $p = 0.003$) and sad behaviour ($H = 10.42$, $p = 0.03$). The impact on rating of neutral behaviour was not significant ($H = 6.11$, $p = 0.19$). Post hoc Mann Whitney U tests indicated exact differences, which can be found in Table C.3.

	Happy		Neutral		Sad	
	Z	*p*	*Z*	*p*	*Z*	*p*
not familiar at all / slightly familiar	-0.92	0.36	-0.02	0.98	-1.46	0.15
not familiar at all / moderately familiar	-3.13	**0.002**	-2.05	**0.04**	-2.81	**0.005**
not familiar at all / very familiar	-2.86	**0.004**	-0.83	0.41	-1.73	0.08
not familiar at all / extremely familiar	0.21	0.23	0.52	0.54	0.42	0.44
slightly familiar / moderately familiar	-2.07	**0.04**	-1.85	0.07	-1.71	0.09
slightly familiar / very familiar	-0.08	**0.04**	-0.88	0.38	-1.14	0.26
slightly familiar/ extremely familiar	0.31	0.33	0.60	0.62	0.64	0.68
moderately familiar / very familiar	0.76	0.77	0.04	**0.04**	0.84	0.85
moderately familiar / extremely familiar	1.00	1.00	0.81	0.82	0.68	0.72
very familiar/ extremely familiar	0.76	0.79	0.33	0.35	0.90	0.96

TABLE C.3: Comparison of ratings of emotive robot behaviours (happy, neutral, sad) per participants' familiarity with social robots.

Additional results of parametric analysis

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.74	0.02*	0.01**	0.01**	0.06
25-34		-	0.02*	0.01**	0.01**	0.06
35-44			-	0.58	0.24	0.80
45-54				-	0.53	0.91
55-64					-	0.62

*$p < 0.05$, **$p < 0.01$

TABLE C.4: Significant differences for rating happy behaviour based on participants' age range.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.74	0.03*	0.02*	0.01**	0.02*
25-34		-	0.02*	0.01*	0.01**	0.02*
35-44			-	0.96	0.61	0.39
45-54				-	0.62	0.40
55-64					-	0.56

*$p < 0.05$, **p

TABLE C.5: Significant differences for rating neutral behaviour based on participants' age range.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.94	0.95	0.45	0.20	0.22
25-34		-	0.99	0.31	0.08	0.16
35-44			-	0.39	0.16	0.19
45-54				-	0.59	0.44
55-64					-	0.63

TABLE C.6: Significant differences for rating sad behaviour based on participants' age range.

	not at all familiar	**slightly familiar**	**moderately familiar**	**very familiar**	**extremely familiar**
not at all familiar	-	0.25	0.01**	0.01**	0.14
slightly familiar		-	0.05	0.05	0.31
moderately familiar			-	0.76	0.98
very familiar				-	0.87

$^{**}p < 0.01$

TABLE C.7: Significant differences for rating happy behaviour based on participants' familiarity with social robots.

	not at all familiar	**slightly familiar**	**moderately familiar**	**very familiar**	**extremely familiar**
not at all familiar	-	0.48	0.01*	0.53	0.34
slightly familiar		-	0.07	0.31	0.50
moderately familiar			-	0.02*	0.82
very familiar				-	0.24

$^{*}p < 0.05$

TABLE C.8: Significant differences for rating neutral behaviour based on participants' familiarity with social robots.

	not at all familiar	**slightly familiar**	**moderately familiar**	**very familiar**	**extremely familiar**
not at all familiar	-	0.28	0.01**	0.05	0.33
slightly familiar		-	0.08	0.25	0.58
moderately familiar			-	0.78	0.75
very familiar				-	0.89

$^{**}p < 0.01$

TABLE C.9: Significant differences for rating sad behaviour based on participants' familiarity with social robots.

D Long-term field study with older adults - additional results

Speech prosody pitch analysis

	min				max				mean				sd			
	t	*p*	*NE*	*E*	*t*	*p*	*NE*	*E*	*t*	*p*	*NE*	*E*	*t*	*p*	*NE*	*E*
1	-0.51	0.62	146.64	139.16	-0.01	0.99	433.47	433.11	-0.39	0.70	257.67	254.26	1.00	0.33	68.65	75.78
2	-0.85	0.41	132.34	143.17	-0.16	0.87	450.64	456.76	1.09	0.29	248.06	269.31	-0.69	0.50	77.98	88.19
3	0.70	0.49	166.82	157.03	1.37	0.19	450.33	392.09	0.81	0.43	268.68	254.23	1.05	0.31	77.44	64.23
4	-1.47	0.17	84.83	89.71	-0.78	0.45	517.46	546.46	-0.25	0.80	231.41	241.39	-0.12	0.91	144.14	146.94
5	-0.71	0.48	124.49	130.57	-0.31	0.76	343.66	360.91	-0.96	0.35	193.55	213.10	-0.02	0.99	66.17	66.53
6	0.38	0.71	105.26	102.42	1.99	0.07	407.80	297.23	1.92	0.08	181.57	154.10	2.37	**0.04**	85.01	44.08
7	0.60	0.56	107.60	101.66	-0.23	0.82	436.50	447.79	-6.75	**0.00**	211.62	447.79	0.39	0.70	89.85	83.33
8	-1.41	0.17	120.75	135.58	0.54	0.60	420.88	394.65	-0.97	0.35	204.72	221.11	-0.01	0.99	75.55	75.74

TABLE D.1: Values for the minimum, maximum, mean and standard deviation of participants' pitch during interactions with the robot.

Speech prosody intensity analysis

	min				max				mean				sd			
	t	*p*	*NE*	*E*	*t*	*p*	*NE*	*E*	*t*	*p*	*NE*	*E*	*t*	*p*	*NE*	*E*
1	-1.75	0.1	15.36	21.52	-2.62	0.02	72.34	75.21	-1.37	0.19	46.29	50.14	1.29	0.22	16.35	14.63
2	-2.37	**0.03**	11.46	19.81	-0.39	0.70	65.34	65.94	-1.62	0.12	39.71	43.41	1.23	0.23	14.94	13.35
3	0.34	0.74	19.80	18.47	0.97	0.35	72.08	70.66	1.70	0.11	49.72	46.30	-0.75	0.47	15.05	16.27
4	-2.93	**0.01**	8.27	21.00	0.34	0.74	62.31	61.45	-1.81	0.10	36.30	41.02	2.04	0.06	15.08	11.60
5	0.12	0.91	17.26	16.79	3.41	**0.00**	65.95	59.82	1.23	0.23	42.05	38.59	1.60	0.12	14.17	12.17
6	1.16	0.27	17.13	11.68	0.31	0.76	67.22	66.73	0.89	0.39	43.77	40.25	-1.20	0.25	14.61	16.17
7	0.22	0.83	13.06	12.17	-0.38	0.71	71.30	71.91	-0.34	0.74	43.37	44.62	-1.61	0.12	15.99	18.23
8	-0.39	0.70	12.14	13.18	-9.14	**0.00**	69.55	395.65	2.15	**0.04**	43.09	38.70	1.37	0.19	16.99	15.18

TABLE D.2: Values for the minimum, maximum, mean and standard deviation of participants' speech intensity during interactions with the robot.

E Online evaluation of framework - additional results

Correlations between categories

Specific p-values for correlations between survey categories.

		1	2	3	4	5
1: Social robots displaying emotions	*r*	1	0.72	0.61	0.20	0.10
	p		0.00**	0.00**	0.00**	0.12
2: Attachment to social robots	*r*		1	0.63	0.16	0.14
	p			0.00**	0.02*	0.03*
3: Vulnerable users using social robots	*r*			1	0.39	0.32
	p				0.00**	0.00**
4: Data collection by social robots	*r*				1	0.15
	p					0.02*
5: Lying social robots	*r*					1
	p					

*p $<$ 0.05, **p $<$ 0.01, N = 239.

TABLE E.1: Pearson product-moment correlations between categories, including *p*-values.

Age

Pearson correlation was conducted to investigate the relationship between age and the survey categories. One-way MANOVA was conducted to examine the impact of age on participants' agreement with the survey categories. Results for both analyses can be found in Table E.2.

	Relationship with age		Impact of age		
	r(243)	*p*	*F(5, 233)*	*p*	*eta*
Social robots displaying emotions	-0.23	0.00**	2.67	0.02*	0.05
Attachment to social robots	-0.11	0.10	1.33	0.25	0.03
Vulnerable users using social robots	-0.12	0.06	1.89	0.10	0.04
Data collection by social robots	-0.09	0.17	0.29	0.92	0.01
Lying social robots	0.09	0.18	1.18	0.32	0.03

**$p < 0.01$, *$p < 0.05$

TABLE E.2: Relationship between survey categories and age, and impact of age on survey categories.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.18	0.67	0.71	0.01*	0.01**
25-34		-	0.55	0.48	0.21	0.05*
35-44			-	0.95	0.12	0.03*
45-54				-	0.09	0.02*
55-64					-	0.39

*p < 0.05, **p < 0.01

TABLE E.3: Significant differences in participants' responses to the category 'Social robots displaying emotions' based on their age.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.25	0.62	0.60	0.86	0.04*
25-34		-	0.71	0.15	0.39	0.26
35-44			-	0.38	0.74	0.21
45-54				-	0.52	0.03*
55-64					-	0.08

*p < 0.05

TABLE E.4: Significant differences in participants' responses to the category 'attachment to social robots' based on their age.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.03*	0.92	0.16	0.49	0.01*
25-34		-	0.92	0.16	0.49	0.41
35-44			-	0.21	0.52	0.56
45-54				-	0.46	0.06
55-64					-	0.19

*p $<$ 0.05

TABLE E.5: Significant differences in participants' responses to the category 'Vulnerable users using social robots' based on their age.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.42	0.64	0.89	0.68	0.82
25-34		-	0.28	0.61	0.75	0.68
35-44			-	0.61	0.44	0.56
45-54				-	0.83	0.94
55-64					-	0.90

TABLE E.6: Significant differences in participants' responses to the category 'Data collection by social robots' based on their age.

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**
18-24	-	0.74	0.91	0.39	0.32	0.09
25-34		-	0.93	0.28	0.45	0.14
35-44			-	0.42	0.50	0.19
45-54				-	0.10	0.03*
55-64					-	0.43

*p $<$ 0.05

TABLE E.7: Significant differences in participants' responses to the category 'Lying social robots' based on their age.

Gender

Independent samples t-tests were conducted to investigate whether gender impacted participants' agreement with the survey categories. The results are provided in Table E.8.

	Impact of gender		Male		Female	
	t(234)	*p*	*M*	*SD*	*M*	*SD*
Social robots displaying emotions	1.61	0.11	3.19	0.69	3.04	0.67
Attachment to social robots	1.65	0.10	3.12	0.83	2.94	0.82
Vulnerable users using social robots	1.51	0.13	3.20	0.70	3.06	0.68
Data collection by social robots	1.48	0.25	3.87	0.66	3.78	0.61
Lying social robots	-2.42	0.02*	2.77	0.99	3.07	0.84

*$p < 0.05$

TABLE E.8: Impact of gender on participants' level of agreement with survey categories, as well as descriptive results for these categories based on gender.

Familiarity

	not at all familiar	slightly familiar	moderately familiar	very familiar	extremely familiar
not at all familiar	-	0.10	0.01**	0.29	0.05
slightly familiar		-	0.06	0.85	0.03*
moderately familiar			-	0.32	0.01*
very familiar				-	0.03*

TABLE E.9: Significant differences in participants' responses to the category 'Social robots displaying emotions' based on their familiarity with social robots.

	not at all familiar	slightly familiar	moderately familiar	very familiar	extremely familiar
not at all familiar	-	0.45	0.01**	0.16	0.55
slightly familiar		-	0.04*	0.08	0.62
moderately familiar			-	0.01**	0.92
very familiar				-	0.33

TABLE E.10: Significant differences in participants' responses to the category 'Vulnerable users using social robots' based on their familiarity with social robots.s

	not at all familiar	slightly familiar	moderately familiar	very familiar	extremely familiar
not at all familiar	-	0.42	0.01**	0.69	0.23
slightly familiar		-	0.01*	0.42	0.28
moderately familiar			-	0.02*	0.56
very familiar				-	0.20

TABLE E.11: Significant differences in participants' responses to the category 'Data collection by social robots' based on their familiarity with social robots.

Compared statements

	Correlation		Paired samples t-test	
	r	*p*	*t(238)*	*p*
It is acceptable for a social robot to look like a *(human/animal)*.	0.31**	0.00	-3.72	0.00
It is acceptable for a social robot to behave like a *(human/animal)*.	0.51**	0.00	-3.12	0.00
It is acceptable for people to treat a social robot as a *(human/animal)*.	0.42**	0.00	-6.41	0.00
It is okay for people to think of a *(social robot/social robot that shows emotions)* as a living being.	0.73**	0.00	3.08	0.00
It is acceptable for people to think of a social robot as their *(friend/companion)*.	0.65**	0.00	-5.56	0.00
Social robots *(can/should)* show emotions.	0.63**	0.00	0.07	0.94
It is okay for people to become attached to a *(social robot/social robot that shows emotions)*.	0.75**	0.00	0.00	1.00
There are situations where it is acceptable for a social robot to lie to a *(person/vulnerable person)*.	0.78**	0.00	-2.34	0.02
A social robot should never lie to *(the person they are interacting with/a vulnerable person)*, even if it appears to benefit that person.	0.35**	0.00	-0.66	0.51

**p < 0.01

TABLE E.12: Correlations and differences comparing specific statements of the survey. The italic parts indicate how two statements differed.

Regression

Multiple regression analyses were conducted to investigate how age, gender, familiarity with social robots and familiarity with robotic technologies impacted the survey categories. Table E.13 shows the impact of the variables on these categories, and Table E.14 shows to what extent each variable separately impacted the level of agreement with the survey categories.

	F(10, 235)	*p*	*R^2*
Social robots displaying emotions	2.69	0.01**	0.11
Attachment to social robots	0.98	0.46	0.04
Vulnerable users using social robots	1.55	0.12	0.07
Data collection by social robots	1.57	0.12	0.07
Lying social robots	1.84	0.06	0.08

$^{**}p < 0.01$

TABLE E.13: Impact of the variables age, gender, familiarity with social robots and familiarity with robotic technologies on participants' level of agreement with the survey categories.

	Social robots displaying emotions		Attachment to social robots		Vulnerable users using social robots		Data collection by social robots		Lying social robots	
	β	p	β	p	β	p	β	p	β	p
Age	-0.007	0.007**	-0.002	0.563	-0.003	0.318	0.001	0.635	0.006	0.082
Gender	-0.142	0.117	-0.172	0.129	-0.137	0.144	-0.071	0.410	0.290	0.022*
Slightly familiar with social robots	0.128	0.241	0.068	0.621	-0.011	0.923	0.060	0.568	-0.353	0.021*
Moderately familiar with social robots	0.351	0.020*	0.143	0.448	0.266	0.089	0.397	0.006**	-0.081	0.698
Very familiar with social robots	-0.008	0.974	-0.440	0.165	-0.412	0.118	-0.006	0.981	-0.179	0.613
Extremely familiar with social robots	-1.577	0.031*	-1.224	0.180	0.297	0.694	0.542	0.437	-1.175	0.249
Slightly familiar with robotic technologies	-0.131	0.291	-0.078	0.618	0.073	0.572	0.066	0.579	0.149	0.393
Moderately familiar with robotic technologies	-0.100	0.502	-0.088	0.639	0.072	0.642	0.069	0.631	0.246	0.239
Very familiar with robotic technologies	-0.038	0.863	0.256	0.355	0.108	0.638	-0.128	0.548	0.513	0.098
Extremely familiar with robotic technologies	0.002	0.995	0.282	0.479	0.041	0.900	0.244	0.422	0.303	0.495

**$p < 0.01$, *$p < 0.05$

TABLE E.14: Impact of each variable on participants' level of agreement with the survey categories.

F Example table to code behaviours

Example coding table to analyse participants' behaviour during the interactions.

PARTICIPANT ID:	XXXXX	**INTERACTION #:**	X				
TIMESLOT (SEC)	Hands (+duration)	Mouth (+duration)	Face (+duration)	Nod (+duration)	Shake (+duration)	Smile (+duration)	Reposition (+duration)
0 - 60			I (1), I (2)	I (1)			
61 - 120		I (7)	I (5), I (1), I (1)	I (1)	I (1)	I (8)	
121 - 180		I (10), I (3), I (7)	I (1), I (1)	I (3)	I (1), I (6), I (1)	I (5)	
181 - 240		I (5)	I (17), I (8)			I (3)	
241 - 300		I (2), I (2)	I (8), I (5), I (7)	I (1), I (1)	I (3), I (2)	I (1)	
301 - 360			I (1)		I (1)	I (1)	
361 - 420							
421 - 480							

Notes from coder: Hearing difficulties affected communication between the participant and robot.

Explanation of behaviours:

- **Hands:** Participant moving hands (touching one hand unintentionally with other hand, play with jewellery or hair, play with fingers)
- **Mouth:** Cover mouth by placing finger(s) or hand over lips
- **Face:** Touch face any other way as described by 'mouth' (e.g. rub chin, scratch head, wipe eyebrow)
- **Nod:** Move head up and down (or vice versa) at least once
- **Shake:** Move head left to right (or vice versa) at least once
- **Smile:** One or both corners of mouth go up
- **Reposition:** Participant repositions themselves in the chair (cross/uncross legs, move weight, move arms to a different position, lean forward, lean backwards)

G HRV Report

Introduction

Research in social robots and human-robot interaction and how they can be employed in the real world is expanding. An example target group that can benefit from social robots is older adults. Social robots can potentially help them live independently for longer. It may also provide opportunities to address loneliness as it can be a conversation starter between people and also interact with people itself.

However, before these robots can successfully integrate into society we first have to make sure we fully understand what people expect of these robots and how they respond to them, as robots that do not meet expectations or do not behave as preferred may be misused or not be used at all (Hancock et al., 2011). This would be unfortunate, as not only time, money and effort would be lost on the development robots that are not being used, but also the intended users are missing out on the benefits that the robot could provide.

To ensure successful integration into society it is essential that we as researchers fully understand what intended users expect from the robot and how they respond to them. Therefore, gathering rich data that will provide useful and rich insights during human-robot interaction (HRI) studies is imperative.

Measurements to gather this data consist of questionnaires, recordings (both video and audio) and physiological data, all with their own advantages and disadvantages. A combination of all may likely produce the richest data to analyse HRI studies. This report will discuss the use of heart rate variability (HRV) as a potential metric for HRI.

In the context of my thesis, I am investigating whether participants respond differently to a social robot that displays emotive behaviour compared to a robot that displays neutral behaviour. What is defined as emotive and neutral behaviour is not essential for this report and is described in earlier work (Van Maris et al., 2018). I will investigate participants' responses through a combination of questionnaires, recordings and physiological data,

including HRV. However, before I can use HRV to analyse participants' responses, I first will need to identify whether HRV is a relevant and useful metric for HRI experiments. This is what this report will discuss. Usually, HRV is measured in a clinical setting, for example for research on heart disease (Vescio et al., 2018). However, my goal of using HRV in HRI experiments is to investigate whether there is a physiological response to the interaction with the robot, more specifically whether there is a emotional reaction.

The research question that will be researched in this report is: *'Can HRV be used reliably to better understand people's responses during human-robot interactions?'*

The remainder of this report consists of an overview of HRV, what it represents and how it can be measured, followed by an initial analysis of HRV data that I gathered during a human-robot interaction experiment and a discussion whether HRV is a reliable metric for HRI.

Heart Rate Variability

One way to investigate people's experiences during human-robot interactions is through measuring their level of emotional arousal and/or stress. Perceptions of threat results in higher stress levels (Thayer et al., 2012). Stress can be measured by looking at activity of the Autonomic Nervous System (ANS) as changes in ANS activity can indicate changes in stress levels (Dandu et al., no date, Jobbágy et al., 2017). The ANS has a sympathetic (SNS) and parasympathetic (PNS) branch. The sympathetic branch is responsible for activation, for example a fight-or-flight response, and the parasympathetic branch is responsible for relaxation and a state of rest and recovery. When sympathetic activity increases, stress increases as well; whereas stress decreases when parasympathetic activity increases (Jobbágy et al., 2017).

The ANS influences heart rate and heart rate variability (Thayer et al., 2012). HRV is the variation of time between heart beats, often expressed in milliseconds. Parasympathetic effects are fast (expressed in milliseconds), while sympathetic effects are slower (expressed in seconds). Therefore, only parasympathetic effects can impact HRV (Thayer et al., 2012).

The influence of HRV and the parasympathetic system is bidirectional (Thayer et al., 2012). While parasympathetic effects can impact HRV, HRV has also been found to be a good indicator for changes in the parasympathetic branch, for example during emotional

situations (De Jonckheere et al., 2012). It has been claimed that HRV can be used as a tool to better understand emotion in social processes (Dandu et al., no date). If HRV is low, this means that sympathetic activity is high and stress and arousal are high as well. Conversely, high HRV indicates activity of the parasympathetic system and low levels of stress and arousal.

Gathering HRV Data

The preferred way to measure heart rate variability is by using electrocardiography which produces electrocardiograms (ECG). ECGs are recordings that shows the electrical activity of the heart (Vescio et al., 2018). This is done using electrodes that are placed on the torso. However, the use of ECG provides some issues as these electrodes need to be placed on bare skin and people may be allergic to the conductive gel that is applied to the electrodes (Jong, Horng, et al., 2017, Vescio et al., 2018). Also, ECG devices are difficult to use for a layperson as incorrect placement of an electrode can result in an erroneous recording. Therefore, photoplethysmography (PPG) sensors that can be used in wearable devices have been introduced as alternative for ECG devises and are being used more often to measure heart rate variability (Vescio et al., 2018). Where ECG measures the electrical activity of the heart, PPG uses light sensors to measure the rate of the blood flow. These sensors are often placed either on the fingers, wrist or on the earlobe (Jong, Horng, et al., 2017, Vescio et al., 2018, Morelli et al., 2018). However, it is better to place the sensor on the earlobe, as the sensor is sensitive to movement and most daily activities require the use of the hand which includes fingers and wrist.

Both ECG and PPG measure the time between beats. ECG does this through measuring beat-to-beat intervals (RR intervals), from which normal-to-normal (NN) intervals are extracted by removing artefacts. These NN intervals are then used for HRV analysis. PPG measures the changes in blood volume and the peaks and valleys in these changes, called PP intervals. These PP intervals can be associated with RR intervals from ECG devices, even though there is a brief delay as it takes time for the blood to travel through the vessels. Just as for RR intervals, artefacts needs to be removed from PP intervals before analysis.

Several studies have investigated whether PPG is a reliable alternative for ECG. One study compared PPG data gathered through an earlobe sensor to ECG data. They found that the PPG sensor provided similar results to the ECG sensor (Vescio et al., 2018). They

gathered a relatively short recording of 20 minutes and a long recording of 24 hours. For both recordings data from 10 participants was gathered. A weakness of this paper is that the terms RR intervals and NN intervals are used interchangeably, without explaining the difference (exclusion of artefacts or not). However, it is clarified how artefacts are removed for both the PPG and ECG device, indicating that abnormal beats are removed from the data. The largest errors between the two devices were found in pNN50 (a metric for HRV analysis; this will be discussed in more detail in the next section).

A different study investigated whether pNN50 values were similar for PPG and ECG data that was recorded in parallel (Jobbágy et al., 2017). They found that NN intervals determined through the ECG sensor differed from intervals determined through the PPG sensor. This is similar to the finding of the previously described study where the errors between the PPG and ECG devices were largest for pNN50. However, in this study the ECG sensor was placed at the participants' back, where the PPG sensor was placed at the fingertip. As mentioned before, sensors placed on the hand are very sensitive to motion. This can also have resulted in the differences found between the two sensors.

In summary, research has shown that, PPG devices can be a reasonable substitute for ECG as long as the PPG sensor is places such that it will not be disrupted by motion. An advantage of using PPG devices is that it is non-invasive as there is no electrical interaction with the human body. They also require less maintenance and are cheaper than ECG devices (e.g. Tamura et al., 2014, Bolanos, Nazeran, and Haltiwanger, 2006).

Analysing HRV Data

HRV can be described either using time-domain, frequency domain and non-linear measurements. Time-domain measurements are based on the time between heart beats (inter-beat intervals or IBI). The frequency-domain entails four frequency bands that correlate with certain physiological responses. For example, blood pressure regulation falls under the low-frequency band (Shaffer and Ginsberg, 2017). Non-linear measurements are used to measure the level of predictability of a time series. HRV recordings can have a long-term (24 hours), short-term (5 min) or ultra-short-term (<5 min) duration. However, as longer recording periods are able to better represent the response to different workloads and stimuli, long-term and short-term values can not be used interchangeably. Also, the duration of the HRV recordings influences what metric can be used best, as some metrics require

a certain minimum duration to provide reliable output. With this in mind, time-domain measurements are most suitable for the data I gathered and will be discussed in more detail in this report.

Measurements in the time-domain are all based on normal-to-normal intervals between heart beats. Indices in the time-domain of HRV measure the amount of HRV that is observed. Most often used are the SDNN, RMSSD and pNN50.

SDNN stands for the standard deviation of the inter-beat intervals of *normal* sinus beats and is measured in milliseconds. SDNN values are derived from SDRR values, which stands for the standard deviation of the time between heart beats of *all* sinus beats (also measured in milliseconds). This means that abnormal beats from SDRR (also known as artefacts) have been removed to get SDNN, which is used for HRV analysis. Abnormal beats should be removed as they can reflect noise or cardiac dysfunction. However, the terms RR interval and NN interval are often used interchangeably in the literature. For example, in some papers the terms 'RR' and 'NN' are introduced as the same thing (e.g. Vescio et al., 2018, Morelli et al., 2018). 'NN' is also defined as the difference between the lengths of successive heart beats and not the *normal* successive heart beats (Jobbágy et al., 2017).

Other metrics that take into account the normal-to-normal intervals and provide more information on HRV are SDANN (the standard deviation of the average normal-to-normal intervals for all 5 minute segments in a 24 hour recording) and the SDNN Index, which is the average of all 5 minute segments of the SDANN. These metrics are most reliable for 24-hour recordings as a longer time period provides a better representation of the slower processes of the cardio-vascular system and the response to workload and environmental stimuli (Shaffer and Ginsberg, 2017). Therefore, even though especially SDNN is the most commonly used metric for heart rate variability (Jong, Horng, et al., 2017), they are not the most useful metrics to analyse heart rate variability in my experiment, as interactions and thus the recordings lasted between 5 and 8 minutes (more information on these interactions will be provided in Section 3.).

NN50 is the total number of NN intervals where the difference between two adjacent intervals is larger than 50 milliseconds. pNN50 is the percentage of adjacent intervals where the difference is more than 50 milliseconds. A decrease in pNN50 can indicate an increase in a person's stress level (Jobbágy et al., 2017). At minimum a 2 minute

recording is required for both metrics, making them suitable metrics to analyse HRV for my experiment. However, as discussed before pNN50 is not always a reliable metric when using a PPG device to gather HRV data. Besides that, there are also arguments that other differences than 50 milliseconds (e.g. 0-20 ms and 20-50 ms) can be better measures. More specifically, it was found that a combination of the three differences mentioned above (pNN0-20, pNN20-50 and pNN50, combined also known as pNNtri) provide a better measure for stress than just pNN50, especially in older adults (Jobbágy et al., 2017). Studies showed that the biggest difference between ECG and PPG devices appeared in pNN50 (Jeyhani et al., 2015), which is another reason why I decided to not use it in my HRV analysis.

RMSSD reflects the difference between heartbeats and stands for the root mean square of successive differences, measured between *normal* heartbeats. To determine the RMSSD, first the time difference between each adjacent heartbeat needs to be measured in milliseconds. These differences are squared and the average of these squared differences is calculated. Finally, the square root of the total is calculated. The usual minimum recording for RMSSD is 5 minutes, however, shorter periods of 10 seconds, 30 seconds and 60 seconds have been researched as well (Salahuddin et al., 2007, Baek et al., 2015, Esco and Flatt, 2014), making this a suitable metric for me to analyse HRV. As mentioned before, HRV is influenced by the parasympathetic system. RMSSD reflects influence of the parasympathetic system as well (Thayer et al., 2012), supporting my decision to use RMSSD for analysis.

Other metrics in the time-domain are the difference between minimum and maximum heart rate, the HRV Triangular Index (HTI) and the Triangular Interpolation of the NN Interval Histogram (TINN). However, even though they are described in an overview of HRV metrics (Shaffer and Ginsberg, 2017), they are not mentioned in several other papers investigating HRV (e.g. Morelli et al., 2018, Matic et al., 2012, Vescio et al., 2018) while the metrics discussed above are. As these metrics are less commonly used in HRV analysis and I will not use them for my analysis either, they will not be discussed any further in this report.

Methodology and Initial Results

Experimental Setup

The goal of the experiment that I ran was to investigate whether older adults were influenced by a robot displaying different emotions. Older adults interacted eight times with a social robot, each interaction lasted between 5 and 8 minutes. During these interactions, the robot would inform participants on the Seven Wonders of the World and would do this while either displaying emotive or neutral behaviour. Besides administering questionnaires, video-recordings were taken to analyse participants' behaviour and physiological data was gathered in the form of galvanic skin response and heart rate variability. HRV was not gathered for clinical purposes in this study; it was gathered to investigate whether there was a physiological response to the interaction with the robot. More specifically, it was intended to measure the emotional context during the interaction; whether people were emotionally aroused or stressed by the robot. This is different from HRV studies in clinical settings, where they use stimuli to evoke a physiological response in the participants. Instead, we wanted to see whether there was a physiological response in a 'natural' setting.

Materials

The social robot that was used for this experiment was Pepper, developed by SoftBank Robotics (Pandey and Gelin, 2018). The sensor that was used to gather PPG data is from the company Shimmer Sensing[1]. Sensors from Shimmer Sensing are used for HRV experiments (e.g. Matic et al., 2012). For this experiment, the Shimmer3 GSR+ was used to gather physiological data. This device provides two finger sensors to measure galvanic skin response (which will not be discussed further in this report) and an earlobe sensor to measure HRV.

The raw PPG data that was gathered needed to be pre-processed before it could be analysed. To pre-process and analyse the data, the standard version of the software Kubios was used [2]. This software was used by others as well to analyse PPG data and remove artefacts (Vescio et al., 2018). The software provides artefact correction algorithms. The

[1] http://www.shimmersensing.com/

[2] https://www.kubios.com/

user guide suggests that for each individual it is decided what threshold value for artefact detection fits best, ranging from very low to very strong. For each interaction, it was tested when the software would detect artefacts. However, it would be ensured that the number of artefacts found was not too high, as to not accidentally identify normal-to-normal RR intervals as artefacts. The software provided feedback as to when this would be the case. Following this procedure, the threshold was set to either medium or strong for most participants.

Initial Results

As mentioned before, the most useful HRV metric for me to look at was RMSSD from the time-domain. Table G.1 shows the RMSSD values for each interaction of ten participants and Figure H.1 shows the RMSSD zones that indicate whether the values are low, normal or high. Blank spaces in the table indicate faulty recordings. A low RMSSD value indicates that parasympathetic activity is inhibited and therefore HRV is low, which indicates stress and/or arousal.

	Interaction							
Participant	1	2	3	4	5	6	7	8
1	24	24	26	22	28	31	22	18
2	18	26	22	23	26	23	24	28
3	19	17	18	13	19	15	18	12
4	73	70	54	45	35	53	62	59
5	79	129	132	154	74		118	
6	9	43	20		26	17	25	23
7		16	41	43	34	25	35	17
8		45	47	55		32	92	35
9	21	51	54	55	30	43		
10	30	55	43	100	68	59	81	69

TABLE G.1: RMSSD values for each interaction for all participants.

Besides the time- and frequency-domain metrics, the Kubios software also provides a stress index, which is based on Baevsky's stress index (Baevsky and Berseneva, 2008).

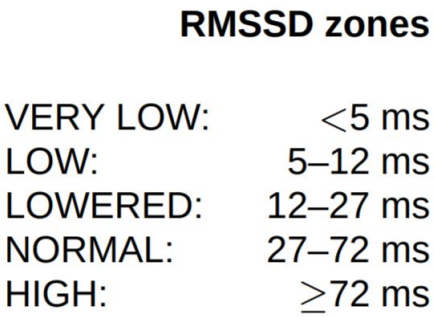


FIGURE G.1: RMSSD zones provided by Kubios User Guide

More specifically, the values from Kubios as the square roots from Baevsky's stress indeces. The square roots are used for normal distribution purposes. The stress indeces can be found in Table G.2, together with an overview of which stress zones indicate low, normal or high stress levels (Figure H.2).

	Interaction							
Participant	1	2	3	4	5	6	7	8
1	13	12	14	15	8	10	16	15
2	14	12	14	14	14	16	15	12
3	14	16	16	15	18	17	17	22
4	8	8	9	10	12	12	9	9
5	8	3	5	5	8		5	
6		21	13	16	14	19	14	20
7		20	12	12	17	24	12	24
8		10	9	10		13	4	15
9	12	9	9	8	11	10		
10	10	7	8	7	7	7	6	6

TABLE G.2: Stress level for each interaction for all participants.

Following the values provided in the Kubios user guide, these tables and figures indicate that several participants have lowered RMSSD and elevated stress levels (e.g. participants 2 and 3). However, When Baevsky introduced the Stress Index on which the Kubios Stress Index is based, they also indicated that the values that indicate elevated stress levels are

		Stress zones
	($\sqrt{SI}$)	(Baevsky's SI)
VERY HIGH:	$\geq$30	($\geq$900)
HIGH:	22.4–30	(500–900)
ELEVATED:	12.2–22.4	(150–500)
NORMAL:	7.1–12.2	(50-150)
LOW:	<7.1	(<50)

FIGURE G.2: Stress levels and stress zones from Baevsky (Baevsky and Berseneva, 2008).

observed by older adults in rest (Baevsky and Berseneva, 2008). Values between 300 and 500 (or between 17.3 and 22.4 when using Kubios) are observed by older adults when they are resting, where these values are found in healthy people during emotional load and physical work. This shift is likely to be true for RMSSD as well, as normal RMSSD values decrease with age (e.g. Antelmi et al., 2004). Therefore, it appears that there was no physiological response or participants to the interactions with the robot, indicating they were not stressed and/or emotionally aroused during the interactions with the robot.

Conclusions

The answer to the question *'Can HRV be used reliably to better understand people's responses during human-robot interactions?'* is: it appears that HRV can be used as a metric for HRI, but more work is needed before this can be stated with confidence. The HRV sensor appears to work, but due to the limited number of participants and the demographics of the participants no conclusions can be drawn. However, the results can give an indication on participants' responses to the robot.

Results show that HRV appeared to be lowered, which indicates that participants may have been stressed or emotionally aroused during the interactions with the robot. However, all participants were of older age. As discussed before, age may require a shift in HRV zones to correctly represent the physiological state of the participants. The use of older adults as participants is important for my thesis. However, it is a disadvantage for this report as it prevents me from being able to fully answer the research question due to the fact that their age may have influenced the results. Also, it is more likely due to their age that participants take medication for heart disease which may have impacted the

results. The HRV values found in my experiment appear fitting for older adults (as they are slightly shifted from 'normal' HRV values which appear realistic for people of older age), especially as it was expected that HRV would be normal as there was no stimulus to actively evoke a response. Therefore, I consider the data reliable enough for me to investigate further for my thesis. However, it should be noted that reliability of HRV data for HRI experiments depends on what sensors will be used, as PPG sensors are sensitive to motion and may not be suitable for certain forms of human-robot interaction. As I used an earlobe sensor and participants were required to sit in a chair and interact with a robot, this was not a problem for my experiment. The fact that HRV gives normal values is promising for the reliability of HRV as a metric in HRI experiments. However, due to the low number and demographics of the participants, together with the fact that there was no specific stimulus to evoke a reaction, this needs to be investigated further.

Observations from the Literature

To better understand HRV and the way it was calculated, I started with reading a technical paper that provided an overview of HRV metrics (Shaffer and Ginsberg, 2017). This provided me with a decent basis for me to be able to understand what authors were talking about when reading papers where HRV was analysed. However, it appears that distinctions that are made in the technical paper are disregarded in practice. One example is the use of RR intervals versus NN intervals, which I already mentioned earlier in this report. The terms are often used interchangeably in papers, or when introducing RR intervals it is actually stated they are known as NN intervals as well (e.g. Vescio et al., 2018, Morelli et al., 2018). Reading these papers more thoroughly it appears that they do remove artefacts but it would be more clear for laypeople if the terminology was consistent. Terminology is not always consistent for frequency-domain analysis either. I did not discuss this type of analysis in detail in this report as it is less relevant for me, but reading about it made me realise there is inconsistency in terminologies used. The frequency-domain is divided into four frequency bands (ultra-low-frequency, very-low-frequency, low-frequency and high-frequency) that each represent different physiological processes like breathing patterns or blood pressure regulation. Therefore, I was very surprised when I read in a paper that the ULF band was introduced as 'the frequencies below LF' (Morelli et al., 2018). All ULF is indeed below LF, but not all frequencies below LF are ULF, as they can also be

VLF. However, as LF and HF are used most frequency for HRV analysis and ULF and VLF not so much this was not a recurrent finding like the one discussed before about the use of RR and NN. Another observation from the literature is that SDNN is used often for short recordings. It is stated that SDNN is more precise when calculated over recordings of 24 hours compared to shorter recordings (Shaffer and Ginsberg, 2017). However, often SDNN is calculated for shorter recordings as well (e.g. Jong, Horng, et al., 2017, Vescio et al., 2018, Matic et al., 2012, Jobbágy et al., 2017, Morelli et al., 2018).

However, as I am not an expert in the field, it may be that I am looking at this from a too technical approach. For example, within HRI research we use many Likert type questionnaires from which we then calculate the mean, where the correct way would be to calculate the median. But as many studies now calculate the mean it appears to have become the norm for analysing questionnaires. Perhaps something similar is the case in HRV research as well, where there is a general understanding that artefacts are removed so the terms RR interval and NN interval can indeed be used interchangeably. In the case of calculating SDNN, it may be decided that SDNN is always measured as this is the most used HRV metric together with RMSSD, or even though SDNN is more accurate for long-term recordings it is still a reliable measurement for shorter recordings.

Future Work

As discussed above there are many opportunities for future work. For my thesis specifically, I intend to look more in depth into these results. For example, why is the HRV of participant 5 so high and their stress level so low? What makes them stand out from the others? I will look into this further by looking at their questionnaire data and video-recordings, which I assume will provide me with interesting insights. Following that analysis, I intend to argue in my thesis that a combination of questionnaires, recordings and physiological data provides the best insights in our understanding of human-robot interactions.

References

Antelmi, I., De Paula, R. S., Shinzato, A. R., Peres, C. A., Mansur, A. J., and Grupi, C. J. (2004). Influence of age, gender, body mass index, and functional capacity on heart

rate variability in a cohort of subjects without heart disease. *The American journal of cardiology*. 93 (3), pp. 381–385.

Baek, H. J., Cho, C.-H., Cho, J., and Woo, J.-M. (2015). Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. *Telemedicine and e-Health*. 21 (5), pp. 404–414.

Baevsky, R. and Berseneva, A. (2008). Methodical recommendations use Kardivar system for determination of the stress level and estimation of the body adaptability - Standards of measurements and physiological interpretation. (no place), pp. 1–42.

Bolanos, M., Nazeran, H., and Haltiwanger, E. (2006). Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals. In: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 4289–4294.

Dandu, S. R., Gill, G., Americas, S., and Siefert, C. (no date). Quantifying Emotional Responses of Viewers based on Physiological Signals. (no place).

De Jonckheere, J., Rommel, D., Nandrino, J., Jeanne, M., and Logier (2012). Heart rate variability analysis as an index of emotion regulation processes: Interest of the Analgesia Nociception Index (ANI). In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 3432–3435.

Esco, M. R. and Flatt, A. A. (2014). Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. *Journal of sports science & medicine*. 13 (3), pp. 535–541.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*. 53 (5), pp. 517–527.

Jeyhani, V., Mahdiani, S., Peltokangas, M., and Vehkaoja, A. (2015). Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 5952–5955.

Jobbágy, Á., Majnár, M., Tóth, L. K., and Nagy, P. (2017). HRV-based stress level assessment using very short recordings. *Periodica Polytechnica Electrical Engineering and Computer Science*. 61 (3), pp. 238–245.

Jong, G.-J., Horng, G.-J., et al. (2017). The PPG Physiological Signal for Heart Rate Variability Analysis. *Wireless Personal Communications*. 97 (4), pp. 5229–5276.

Matic, A., Cipresso, P., Osmani, V., Serino, S., Popleteev, A., Gaggioli, A., Mayora, O., and Riva, G. (2012). Sedentary work style and heart rate variability: A short term analysis. In: *Proceedings of the 2nd International Workshop on Computing Paradigms for Mental Health, MindCare 2012, in Conjunction with BIOSTEC 2012*, pp. 96–101.

Morelli, D., Bartoloni, L., Colombo, M., Plans, D., and Clifton, D. A. (2018). Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. *Healthcare technology letters*. 5 (2), pp. 59–64.

Pandey, A. K. and Gelin, R. (2018). A mass-produced sociable humanoid robot: pepper: the first machine of its kind. *IEEE Robotics & Automation Magazine*. 25 (3), pp. 40–48.

Salahuddin, L., Cho, J., Jeong, M. G., and Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In: *2007 29th annual international conference of the ieee engineering in medicine and biology society*. IEEE, pp. 4656–4659.

Shaffer, F. and Ginsberg, J. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*. 5 (258), pp. 1–13.

Tamura, T., Maeda, Y., Sekine, M., and Yoshida, M. (2014). Wearable photoplethysmographic sensors—past and present. *Electronics*. 3 (2), pp. 282–302.

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers III, J. J., and Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*. 36 (2), pp. 747–756.

Van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., and Dogramadzi, S. (2018). Ethical considerations of (contextually) affective robot behaviour. In: *Hybrid Worlds: Societal and Ethical Challenges - Proceedings of the International Conference on Robot Ethics and Standards (ICRES 2018)*. CLAWAR Association Ltd, pp. 13–19.

Vescio, B., Salsone, M., Gambardella, A., and Quattrone, A. (2018). Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short-and long-term recordings. *Sensors*. 18 (844), pp. 1–14.

H Survey to gather the opinion of the general public

Demographic information:

- Age
- Gender
- Nationality
- Highest completed level of education
- Familiarity with social robots
- Familiarity with robotic technologies

Social robots displaying emotions: Social robots can show emotions when interacting with people. For example, they can appear happy or sad. Showing emotions may result in a more pleasant experiences for the people interacting with the robot; however, they can also lead to people misunderstanding the robot's ability. Please rate to what extent you agree with the following statements:

Social robots can show emotions.	SD	D	Neither A nor D	A	SA
Social robots should show emotions.	SD	D	Neither A nor D	A	SA
People will trust a social robot that shows emotions.	SD	D	Neither A nor D	A	SA
People will overestimate a social robot that shows emotions.	SD	D	Neither A nor D	A	SA
It is acceptable for people overestimate a social robot's abilities.	SD	D	Neither A nor D	A	SA
It is acceptable for people to think of a social robot as their friend.	SD	D	Neither A nor D	A	SA
It is acceptable for people to think of a social robot as a companion.	SD	D	Neither A nor D	A	SA
It is acceptable for a social robot to look like a human.	SD	D	Neither A nor D	A	SA
It is acceptable for a social robot to behave like a human.	SD	D	Neither A nor D	A	SA
It is acceptable for people to treat a social robot as a human.	SD	D	Neither A nor D	A	SA
It is acceptable for a social robot to look like a pet.	SD	D	Neither A nor D	A	SA
It is not acceptable for people to think of a social robot as their friend.	SD	D	Neither A nor D	A	SA
It is acceptable for a social robot to behave like a pet.	SD	D	Neither A nor D	A	SA
It is acceptable for people to treat a social robot as a pet.	SD	D	Neither A nor D	A	SA
It is okay for people to think of a social robot that shows emotions as a living being	SD	D	Neither A nor D	A	SA
It is okay for people to think of a social robot as a living being	SD	D	Neither A nor D	A	SA

Attachment to social robots: If people interact with a social robot a lot, they may become attached to it; where they develop an affection for the robot. Please rate to what extent you agree with the following statements:

It is okay for people to become attached to a social robot.	SD	D	Neither A nor D	A	SA
It is okay for people to become attached to a social robot that shows emotions.	SD	D	Neither A nor D	A	SA
People become more easily attached to a social robot when they trust it.	SD	D	Neither A nor D	A	SA
It is acceptable for a person to become attached to a social robot as long as this person benefits from the robot.	SD	D	Neither A nor D	A	SA
It is not acceptable for a person to become attached to a social robot.	SD	D	Neither A nor D	A	SA
It is acceptable for a person to become dependent on a social robot when they are highly attached to it.	SD	D	Neither A nor D	A	SA

Vulnerable users using social robots: Popular target groups for social robot development include potentially vulnerable people such as those with autism or older adults with possible dementia. The tasks that a social robot could perform would include for example: helping children with autism to develop their social interaction skills, and remind older adults with memory loss to drink enough and take their medicine. Please rate to what extent you agree with the following statements:

These people will overestimate a social robot's abilities if it shows emotions.	SD	D	Neither A nor D	A	SA
It is acceptable that these people overestimate a social robot's abilities.	SD	D	Neither A nor D	A	SA
It is acceptable for a social robot to show emotions when it is interacting with these people.	SD	D	Neither A nor D	A	SA
It is acceptable that older adults become dependent on a social robot if that means they can live independently for longer.	SD	D	Neither A nor D	A	SA
It is acceptable for these people to become attached to a social robot.	SD	D	Neither A nor D	A	SA
It is not acceptable for these users to overestimate a social robot's abilities.	SD	D	Neither A nor D	A	SA
It is acceptable that these people who are highly attached to a social robot become dependent on it.	SD	D	Neither A nor D	A	SA
It is acceptable that these people who are highly attached to a social robot treat it as a human.	SD	D	Neither A nor D	A	SA
It is acceptable that these people who are highly attached to a social robot treat it as a pet.	SD	D	Neither A nor D	A	SA
It is acceptable that these people perceive a social robot that shows emotions as a living being.	SD	D	Neither A nor D	A	SA
It is acceptable that these people perceive a social robot as a living being.	SD	D	Neither A nor D	A	SA

Data collection by social robots: Potential tasks for a social robot can be to remind people drink enough, take their medicine and monitor progress of rehabilitation. The robot can alert other people, such as family or the doctor, if needed. To do this, the robot will require access to personal data. Please rate to what extent you agree with the following statements:

It is acceptable for a social robot to store data about a person if this benefits the person over the long term.	SD	D	Neither A nor D	A	SA
It is acceptable for the anonymised personal data that the social robot uses to be stored and used by the robot developer for future design.	SD	D	Neither A nor D	A	SA
It is acceptable for a social robot to be used to monitor the progress and health of a person.	SD	D	Neither A nor D	A	SA
A social robot can be used to encourage older people to interact with others.	SD	D	Neither A nor D	A	SA
It is not okay for a social robot to be used to monitor the progress and health of a person.	SD	D	Neither A nor D	A	SA

Lying social robots: Currently social robots are being developed to assist caregivers with their jobs. There are situations where caregivers may lie to their patients. For example, imagine an older person who has dementia. This person might ask the caregiver when a loved one (who is deceased) will be visiting them. So as to not upset the patient, the caregiver may say that their loved one will visit later that day. Now imagine that a social robot is keeping the patient company while the caregiver is out of the room. The patient now asks the robot when their loved one will visit and the robot says that their loved one will visit later that day. Please rate to what extent you agree with the following statements:

There are situations where it is acceptable for a social robot to lie to a person.	SD	D	Neither A nor D	A	SA
There are situations where it is acceptable for a social robot to lie to a vulnerable person.	SD	D	Neither A nor D	A	SA
A social robot should never lie to the person they are interacting with, even if it appears to benefit that person.	SD	D	Neither A nor D	A	SA
A social robot should lie to the person it is interacting with f it appears to benefit that person.	SD	D	Neither A nor D	A	SA
A social robot should never lie to a vulnerable person, even if it appears to benefit that person.	SD	D	Neither A nor D	A	SA

Final questions:

- Have you ever interacted with a social robot?
- If you would not use a social robot, what would be the primary reason and/or concern?
- If you would use a social robot, what would be the primary reason to do so?
- Would you use a social robot if...:
  - You had dementia?
  - You had a physical disability?
  - You had complex medical needs requiring medication?
  - You felt lonely?
  - Your family/partner recommended it?
  - Your family/partner did not like the robot?

I Ethics framework flowchart draft