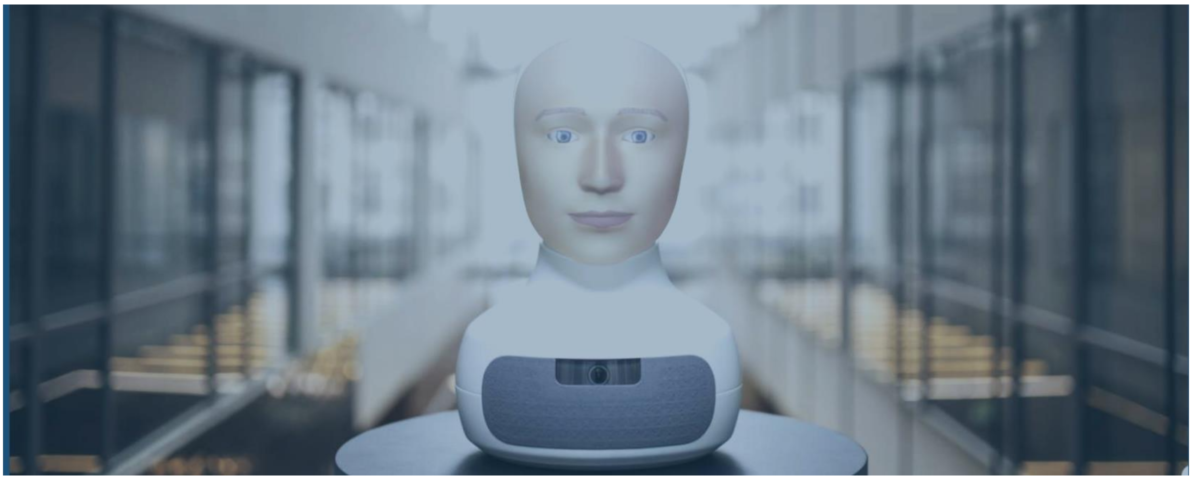
UNIVERSITY OF GOTHENBURG

DEPARTMENT OF APPLIED IT

IN THE PRESENCE OF HUMAN-LIKENESS

An Exploratory Study on How Humans Experience Artificial Agents



Anna Johansson

Elin Örnberg

Thesis:	15 hp
Program:	BSc Cognitive Science
Level:	First Cycle
Year:	2021
Supervisor:	Robert Lowe
Examiner:	Alberto Montebelli
Report nr:	2021:066

Abstract

Interactions with artificial agents are increasing in everyday situations, and social agents might become the next generally adopted interface. However, whether these agents should appear human-like or not is debated as it is still uncertain what effects human-like agents have on humans in an interactive setting. There are several aspects that can affect human responses to artificial agents, such as physical presence and expectations. The uncanny valley hypothesis has been influential in explaining human reactions to human-like agents, but the validity of this hypothesis is still questioned. Earlier studies have mainly been conducted using quantitative measures, which do not seem to capture the whole human experience, hence a wider perspective is needed. Virtual and physical agents might become available in different situations due to financial or technical reasons. Because they differ in physical presence they may be perceived and experienced differently. With a qualitative approach this study explored two scenarios, a physical and a virtual one, using the social robot Furhat. The participants met one version of Furhat in an interactive session. Data was collected from a pre-interaction survey, facial emotion recognition and interviews that was thematically analysed. Due to the study's exploratory approach no generalizations of the results could be made, instead topics for future research were identified. Suggestions from this study is to further investigate feelings of uncanniness in human-robot-interaction, habituation processes, relationships between physical and social presence, and when it is suitable to use human-like design.

Keywords

Human-likeness, Human-robot-interaction, Uncanny valley, Physical presence, Expectations, Emotional reactions, Furhat

Titel

I närvaron av något människolikt: En utforskande studie om hur människor upplever artificiella agenter

Sammanfattning

Interaktioner med artificiella agenter ökar i vardagliga situationer, och social agenter kan bli det nästa allmänt använda gränssnittet. Men om dessa agenter borde vara människolika eller inte är omdebatterat eftersom det fortfarande finns osäkerheter kring hur människolika agenter påverkar människor i en interaktiv situation. Det finns flera aspekter som kan påverka människors reaktioner gentemot artificiella agenter, såsom fysisk närvaro och förväntningar. Hypotesen om kusliga dalen har varit inflytelserik för att förklara människors reaktioner gentemot människolika agenter, men validiteten för denna hypotes är fortfarande ifrågasatt. Tidigare studier har huvudsakligen utförts med kvantitativa metoder, vilket inte verkar fånga hela upplevelsen, och därmed behövs ett bredare perspektiv. Virtuella och fysiska agenter kan komma att bli tillgängliga i olika situationer med anledning av ekonomiska och tekniska orsaker. Eftersom agenterna skiljer sig åt i fysisk närvaro kan de uppfattas och upplevas olika. Med ett kvalitativ angreppssätt ämnar denna studie utforska två scenarion, ett fysiskt och ett virtuellt, av den sociala roboten Furhat. Deltagarna i studien mötte en version av Furhat i en interaktiv session. Data samlades in genom en pre-interaktionsenkät, emotionsigenkänning och intervjuer som var tematiskt analyserade. Med anledning av studiens utforskande angreppssätt kunde inga generaliseringar göras av resultatet, i stället identifierades ämnesområden för framtida forskning. Förslag från denna studie är att vidare utforska upplevda känslor av obehag i människa-robot-interaktioner, invänjningsprocesser, relationen mellan fysisk och social närvaro samt när de är lämpligt att använda människolik design.

Nyckelord

Människoliknande, Människa-robotinteraktion, Kusliga dalen, Fysisk närvaro, Förväntningar, Emotionella reaktioner, Furhat

Foreword

This study was implemented with the support of Volvo Cars whose involvement has been present throughout the process, providing their perspective and some of the material used in the interaction task. Elin Örnberg was responsible for the development of the interaction script, had the main responsibility for the laboratory setup and analysis in FaceReader. Anna Johansson had the main responsibility for the theoretical parts of the work, development of interview questions and the analysis section. The preparatory work, method, data collection, data processing and the remaining sections of the report were a collaborative effort.

Acknowledgement

We want to thank the following persons for their support of this study:

Jenny Wilkie and Jessica Villing, our supervisors at Volvo Cars for their inspiration, ideas, and shared knowledge.

Robert Lowe, our supervisor at University of Gothenburg for the discussions, shared knowledge, technical equipment, and scheduling support.

Bahram Salamat Ravandi for Furhat demonstration and training.

Elin Friberg at Friberg Translation for support with translation of the NARS questionnaire from English to Swedish.

Furhat Robotics for access to their SDK.

Professor Tatsuya Nomura at Ryukoku University, Japan for sharing the article *Measurement of Negative Attitudes Towards Robots*.

Lena Schramm for sharing the article *Warning: This Robot is Not What it Seems! Exploring Expectation Discrepancy Resulting from Robot Design.*

Table of content

Acronyms

API: Application Programming Interface

HHI: Human-Human Interaction

HRI: Human-Robot Interaction

NARS: Negative Attitudes towards Robots Scale

PhyFur: Physical Furhat

SDK: Software Development Kit

TTS: Text-To-Speech

UX: User Experience

VirFur: Virtual Furhat

1. Introduction

1.1 Background

Artificial agents have gone from being an idea in science fiction to becoming common even in regular households through inventions such as robotic vacuum cleaners and voice assistants. Still, the human-robot interaction (HRI) research field continues to raise questions about the human reaction to these agents, especially as social robots with the purpose of being part of our social environment are being developed (Kok & Soh, 2020). Companies wishing to adopt this kind of technology face several challenges as they need to make sure that the technology they use is confidence-inspiring, but also because they need to understand their customers' perception of the technology. The uncanny valley hypothesis (Mori, 1970) is a guiding principle for designing artificial agents (Pollick, 2009), but to this day consecutive support for this hypothesis is missing (see e.g. Wang et al., 2015). Therefore it is still uncertain what effects human-like design has on human perception of artificial agents, and whether the emotional reactions towards them are solely negative.

While the introduction of robots into human life is intriguing for many, virtual agents might be more available due to financial and technical reasons. Both alternatives come with their own set of benefits and limitations, and because of differences in physical presence they may be perceived differently. It is however not clear to what extent, or how they are perceived differently. The social robot Furhat (Furhat Robotics, n.d.a) creates an opportunity to investigate a physical and a virtual version of the same robot, making it possible to explore the two scenarios separately.

Over the past 10 years, the technical advancements have improved the naturalness in different dialogue systems. The landscape has also changed in that consumers are investing in smart speakers for their homes and bringing with them personal voice assistants into their life and into their vehicles (J. Wilkie, personal communication, May 20, 2021). Social robotics may be the future dialogue system, and the potential of this technology needs exploring as these systems could open up for hands- and eye-free interaction, which has attracted interest in the automotive space (J. Wilkie, personal communication, May 20, 2021). By integrating Volvo specific content into the interaction one can create a sense of how this kind of technology can be applied in an information gathering task that could be applied in several employee and customer facing scenarios.

1.2 Purpose and research questions

For this study we are adopting a qualitative approach to examine the feelings triggered in humans when *interacting* with human-like agents. The purpose of the study is to explore the participants' experiences to unfold their emotional reaction, what features in the interaction that might have caused them, and how they perceive the agent they interact with. To do this two interaction scenarios, one using a physical robot and one using a virtual agent of the same robot, is used. By exploring two aspects of the robot, we do not aim to compare, but to explore the scenarios separately. The following research questions will be addressed:

> *Q1. What feelings do participants experience with a human-like agent?*
> *Q2. How are human-like agents perceived?*

As the occurrence of robots and other artificial agents are increasing in everyday situations, the potential for social robotics is highly interesting for companies interested in UX to use this type of technology to craft future user experiences. Social robots are envisioned by some to be the next generally adopted user interface (Wood & Dillenbeck, 2021), but for that to happen it is necessary to understand how humans think about interacting with such agents. In addition to investigating the emotional reaction in subjects that interact with human-like agents this study is also concerned with what the participants think about retrieving information from human-like, social agents, leading to the third research question below:

> *Q3. How do participants experience information retrieval through interacting with a social, human-like agent?*

These questions, although general in scope, will be addressed within a limited setting as described in the theoretical and methodological chapter (section 2 and 3) using the Furhat robot.

2. Theory

2.1 Literature review

2.1.1 The uncanny valley

One theory on human reaction to human-like agents is the uncanny valley hypothesis. This hypothesis was introduced in an essay by the robotics professor Masahiro Mori in 1970 and argues that feelings of discomfort and uncanniness will be predominant when humans are exposed to highly realistic, human-like entities such as prosthetic limbs and robots. Mori (1970) defined the uncanny valley as the relationship between human-likeness and affinity, or more specifically how these constructs create a non-linear relationship as the level of human-likeness of an entity and the affinity for that entity changes. Basically, as we create more human-like agents our affinity increases towards them, but only to a certain point (when one arrives at the valley, see Figure 1). If an artificial agent were to reach a high degree of human-likeness while still being distinguishable from an actual human, Mori (1970) predicted any initial affinity to be replaced with a deeply negative sense of uncanniness, as one would tumble down the uncanny valley. To avoid this valley, and to develop positive feelings towards such an agent it needs to be fully indistinguishable from a healthy human being (Paetzel-Prüsmann, 2020). Hence, highly realistic, human-like design in characters and objects might have a counterproductive effect (Cheetham, 2017), meaning that affinity and acceptance towards robots and computer-generated characters cannot be assumed (Pollick, 2009) as they might fall into the valley.

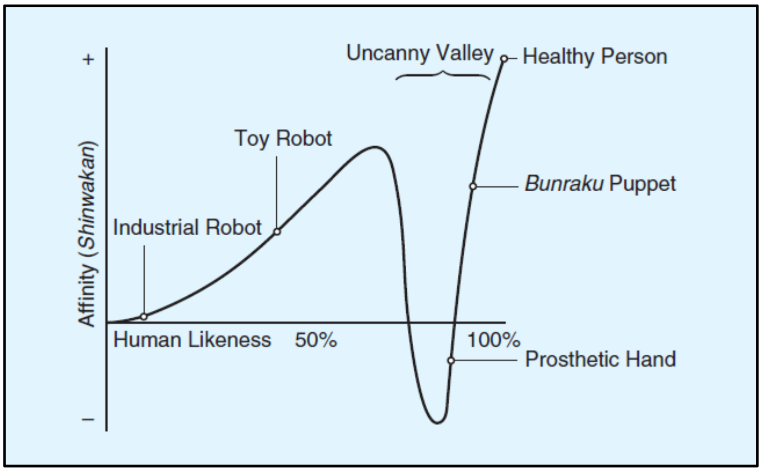


Figure 1: The uncanny valley (Mori, 1970).

The uncanny valley is a prominent theoretical assumption within HRI research (Rosenthal-von der Pütten & Weiss, 2015), as well as a respected design principle both in robotics and computer graphics (Pollick, 2009). While it is regarded as an intuitively attractive idea (Wang et al., 2015), the current research paradigm has not been able to confirm its validity (Bartneck et al., 2009; Cheetham, 2017; Kätsyri et al., 2015). Firstly, as Mori (1970) did not provide any empirical evidence for the uncanny valley, researchers have not had a common ground from where they can make their interpretations (Wang et al., 2015). Secondly, because the variables affinity and human-likeness were not clearly defined, different studies have used different operationalizations, making it unclear whether these studies have addressed the same concepts and hypotheses (Kätsyri et al., 2015). This problem has been amplified by the difficulties of translating the original terms *shinwakan* and *bukimi* (that are used to express the affinity-uncanniness dimension)[1], leading to different estimates of their meaning and underlying properties (Bartneck et al., 2007). Thirdly, to the best of our knowledge, the support for the uncanny valley has been inconclusive (Cheetham, 2017; Kätsyri et al., 2015; Wang et al., 2015). Whether this is because of uncertainties about the original hypothesis and how it should be tested, that the phenomenon might be compatible with more than one explanatory model (Kätsyri et al., 2015), or because the original hypothesis is too simplistic (Ho et al., 2008) it has resulted in several hypotheses that aim to explain the nature of the uncanny valley.

Summaries of the current hypotheses on the uncanny valley can be found in e.g. Wang et al. (2015) and Kätsyri et al. (2015). Using Wang and colleagues (2015) distinction the hypotheses can be broadly divided into two categories. The first category views the phenomenon as a result of stimulus-driven perceptual processes, and includes theories regarding pathogen avoidance, mortality salience, and evolutionary aesthetics. The second category connects the uncanny valley phenomenon to cognitive processes. This category contains hypotheses about violation of expectations, categorical uncertainty, and mind perception. Regarding category one, common to these theories is the assumption that human-like agents are viewed by humans as real people, which has not been adequately tested in empirical research. Until this assumption has been tested it is unclear how well these hypotheses account for the uncanny valley phenomenon. Between the hypotheses in category two, good support has been found for perceptually mismatched stimuli as a cause for the uncanny valley (Kärtsyri et al., 2015). There is however an overlap between the cognitive theories on the uncanny valley, indicating that they might not fully capture the uncanny valley phenomenon either (Wang et al., 2015).

[1] *Shinwakan* was first translated to familiarity, but the word affinity has been deemed more suitable and is the word used in the officially approved translation (Paetzel-Prüsmann, 2020). *Bukimi* has been translated to eeriness, creepiness and strangeness (Bartneck et al., 2007; Kätsyri et al., 2015). Eeriness is considered the more accurate translation, but uncanniness is a generally accepted translation as it is close in meaning (Wang et al., 2015).

2.1.2 Interaction and physical presence

An aspect that might affect human reaction to, and interpretation of, human-like agents is whether we are interacting with them, and if they are physically present. With an interactive approach several factors that can affect the psychological and emotional responses to artificial agents are introduced. For example, it has been shown that people interacting with virtual agents demonstrate behaviours that are used in human-human interaction (HHI) (de Borst & de Gelder, 2015), and that the task structure and context can affect a person's interpretation of an agent (Rosenthal-von der Pütten & Weiss, 2015). In a study that used a live interaction paradigm it was found that the interaction in itself had an effect on the participants' emotional response towards the robot (Złotowski et al., 2015). For example, it was observed that a robot's attitude had an effect on the perceived likeability of the robot, and that feelings of uncanniness decreased with an increasing number of interactions (Złotowski et al., 2015).

Physical presence is another factor that has been brought forward as important for understanding human responses towards artificial agents. A meta-review showed that co-present robots (i.e. robots that are physically present in the human's space) were favoured over virtual agents (digital agents made using computer graphics). The review also found that physical presence had a positive effect on the participants' behaviour (e.g. in terms of performance) and attitude (as measured in e.g. enjoyment, trust, utility, and social presence) (Li, 2015). Powers et al. (2007) also found the participants to be more engaged by a collocated robot, and that the robot was more liked than other agents used in the experiment. It was however noticed that the participants tended to remember less information from the interaction with the robot compared to the virtual agent (which could be a result of the participants being distracted by the physical robot), and that they revealed less information about themselves to the robot (Powers et al. 2007). Similar results were found in Kiesler et al. (2008), who reported larger inhibition in the participants interacting with a robot (Kiesler et al., 2008). These examples show that physically present robots can elicit positive feelings in the people interacting with them, but that they also might inhibit the human interactor. Physically present robots have also been shown to cause negative affect. Paetzel-Prüsmann (2020) tested if a collocated robot was perceived as uncannier than a virtual model of the same agent. The results showed that physical presence had a marginal effect on the collocated robot's perceived uncanniness, and that the virtual version was perceived as more uncanny. Both versions were however considered uncannier than video recordings of the robot and of a human (Paetzel-Prüsmann, 2020), indicating that physical presence changes something in the way observers perceive uncanniness (e.g. physical presence might increase the agent's social presence or the arousal levels in the interactor, making the robot seem uncannier).

2.1.3 Expectation

Expectations are another aspect that factors into reactions to, and interpretation of, artificial agents. As most people do not have extensive experience of interacting with such agents it has been suggested that humans rely on social cues and knowledge of HHI to approach interactive situations with artificial agents, applying human-to-human scripting to make sense of the situation (Craig & Edwards, 2021). Similarly, Bartneck et al. (2020) found that humans tend to have expectations based on experiences from HHI when interacting with human-like robots, and Schramm et al. (2020) has observed that a human-like robot generates mental models that include expectations of human abilities. Mental models including human capability can cause an expectation discrepancy when interacting with robots in case there are limitations to the robot's abilities. This discrepancy may cause misplaced trust, disappointment, or issues with accepting the robot (Schramm et al., 2020).

The actual use of the robot, how it is introduced and how it presents itself can affect a person’s expectations (Schramm et al., 2020). If a robot has high-performance skills humans tend to expect the robot to also have more basic skills as well, as it is difficult to imagine that a living being would have high-performing abilities in isolation (Malle et al., 2020). Because even face gestures can send signals that the robot may have e.g. emotional systems or complex interaction abilities, it is hard to create social robot interactions without creating high expectations in the observer (which might get violated) (Schramm et al., 2020). Expectations affect important aspects such as trust in the agent. A robot’s design and how it is portrayed has been shown to greatly impact how trustworthy it appears, but trustworthiness can also be lost if a user experiences too big of a gap between the expectations and the actual interactive experience (Kok & Soh, 2020).

2.1.4 Feelings towards artificial agents

To understand our reactions towards artificial agents, and how they affect how we perceive them we need to understand not only the cognitive processes involved, but also the emotional responses. While cognitive processes are needed to create an understanding for a situation, emotional processes are needed to give it value and to inform us about the situation's conditions (Norman, 2013). Emotions are important for preparing humans for situations that concern our wellbeing (Ekman, 2003), and they are also an integral part in the way we communicate with each other (Craig & Edwards, 2021). These aspects might all factor into our interactions with artificial agents as our conception of them might be influenced by our social-emotional capabilities (Craig & Edwards, 2021).

What emotions are is still a contested question (Ho et al., 2008; Russell, 2003). To distinguish and contrast emotions from each other dimensional models have been developed. One of these conceptual frameworks is the core affect model by Russell (2003). This model distributes emotions in a two-dimensional, circular space using arousal (activation-deactivation) and valence (pleasure-displeasure) as its dimensions (see Figure 2). The outer ring represents feelings/arousal at their extremes, while the centre represents a neutral point (or adaptation level). Rather than portraying humans' emotional life as separate episodes of emotions and non-emotions, this model captures fluctuations in core affects (what we understand as feelings), creating a more general approach to emotion (Russell, 2003).

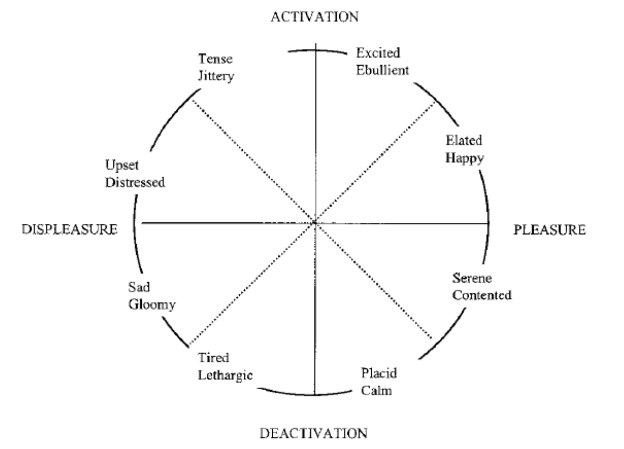


Figure 2: The core affect model (Russell, 2003)

Although it has not been found that human-likeness manipulations automatically lead to feelings of uncanniness (Kätsyri et al., 2015), the uncanny valley has been pivotal to explain emotional reactions towards human-like agents. The emotional components to the uncanny valley are nevertheless quite unknown (Wang et al., 2015). Mori's (1970) definition of feelings of uncanniness are the absence of affinity in a reaction towards uncanny things. This definition might be a bit too simplistic however, as feelings of uncanniness are of a peculiar nature that is both strange and difficult to describe (Mangan, 2015). Looking into what emotion terms are predictive of attributions of uncanny and creepy, Ho et al. (2008) found that fear, but also disgust, shock and nervousness were related to these concepts. They also found that these visceral reactions captured the sense of uncanny better than feelings of strangeness or confusion, and that these attributions were emotional rather than cognitive (Ho et al., 2008). Still, feelings of uncanniness are hard to grasp as uncanniness is often related to a sense of something sinister or strange, but without a clear reason to what is causing it. Mangan (2015) refers to this sensation as a fringe experience, i.e. an emotional experience to which we have limited introspective access and sensory content. Positioned at the fringe of consciousness this feeling combines aspects of familiarity, wrongness, and threat, creating an experience of ill-defined discomfort (Mangan, 2015).

2.2 Empirical approaches

Human reactions and interpretations of artificial agents, especially from the perspective of the uncanny valley, have been studied primarily using quantitative measures. Qualitative studies on the subject are rare (Paetzel-Prüsmann, 2020), meaning that there has been a missed opportunity to create an understanding of why people might find artificial agents to be uncanny (Złotowski et al., 2015). To the best of our knowledge one of the few qualitative, user-centric studies has been done by Lee et al. (2016). They found that the subjects, when asked to describe the robots used in the study, did not focus on how human-like the robots were as a whole. Rather the robots were described using specific human-like features. It was also noted that the robots were interpreted based on contextual and social factors (such as social roles, race, gender, and the context in which the robot could be used), and not based on the affinity-uncanniness dimension suggested by the uncanny valley logic (Lee et al., 2016). This study indicates that other aspects beyond feelings of uncanniness factor into the human experience of artificial agents.

Paetzel-Prüsmann (2020) found that most studies examining emotional responses to human-like agents have relied on judgements of pictures or videos rather than interactions with such agents. These studies rarely exposed their participants to the agents for longer than a few seconds, which might have limited their ability to assess the agents and to create a proper representation of their abilities and demeanour. So far it has not been established whether the conclusions made in these studies can be generalized to interactive scenarios (Paetzel-Prüsmann, 2020). There is also uncertainty about what part the uncanny valley plays in interactive situations with human-like agents (Złotowski et al., 2015). Hence only limited conclusions can be drawn on the impact of interaction in HRI with human-like agents.

It has also been noted that a more holistic approach is needed to study the human experience with human-like agents, that includes aspects such as attitudes towards artificial agents, task context, personality traits, and the technological background of the human interactors (Rosenthal-von der Pütten & Weiss, 2015). Based on these observations and owing to the uncanny valley being context-sensitive, a qualitative approach seems reasonable to explore the factors that might affect feelings and attitudes towards artificial agents in an interactive setting (as they might change over time). Since there are comparatively few interactive studies conducted that examines the impact of the uncanny valley this approach might identify subjects for future studies. Although the results cannot be generalised this study may improve our understanding on how we are affected by human-like agents in an interactive, task-specific context.

3. Method

3.1 Participants

For this study eight participants were recruited (four women and four men, age span 22-36). Six participants were students at University of Gothenburg from three different faculties. The two remaining participants were non-academic professionals. All participants spoke Swedish fluently and were highly proficient in English and therefore able to conduct conversations in both languages. The participants' involvement in this study was fully based on their willingness to partake, and they were not given any compensation for their attendance.

Seven out of eight participants reported having some experience of interacting with artificial agents. These seven had interaction experiences with voice assistants, and one of the participants also had interaction experience with household robots. In the pre-interaction survey (section 3.2.4) six participants described their experience with voice assistants as limited; two described themselves as more frequent users. None of the participants had interactive experience with industrial, social, or other physically present robots, but three participants reported that they had theoretical knowledge of such agents. Only one participant had no prior interaction experience with artificial agents. Since most of the participants had limited or no experience with HRI the sample was deemed suitable as the study aims at investigating the respondents' unmediated reactions to a robot/virtual version of a robot.

3.2 Material and instruments

3.2.1 Furhat

For this study a Furhat Robot (Furhat Robotics, 2018) was used. Furhat is a social robot head with a back-projected face on a plastic facially formed mask that enables both phonemical lip synchronization and temporally accurate voice. The back-projection creates several possibilities to change the facial appearance and to create detailed expressions without any facial, sound generating movement machinery (Paetzel et al., 2016) while also giving it eye-movements that are more naturalistic avoiding the Mona Lisa effect, which is the feeling of being stalked, usually experienced in 2D pictures/screens since it lacks eye gaze direction (Moubayed et al., 2012). Using a projected face also enables movements without complex robotic actuators that might be slow, or easily broken. Furhat's neck has three degrees of freedom enabling human-like movements like nodding or changing gaze with head position (Furhat Robotics, n.d.a). The Furhat robot was chosen for this study due to its highly human-like face without any of the drawbacks that usually follows with highly realistic robot faces (such as sound generating machinery, static movements, and limited facial expression). Because the study wants to explore two aspects of the same agent (Furhat), we use a physical Furhat (PhyFur) and a virtual version of Furhat (VirFur), which is an animated version within the Furhat software development kit (SDK) (Furhat Robotics, 2021) (Figure 3).

Exploring two aspects of Furhat

The skill script was the same for both PhyFur and VirFur, meaning that the whole interaction specification including the appearance, voice, content, gestures, and dialog flow was equal for both versions. For VirFur a larger computer screen was used together with additional speakers to generate a more similar head size, volume and sound compared to PhyFur. Even if the appearances were supposed to be equal due to the same settings it was noted that the back-projected animation on PhyFur was slightly different compared to the graphic animation on the computer screen. Another difference was found in Furhat's gaze and head position. The physical robot's gaze and head movement is connected to a face recognition system, in which it can find faces and change head position to look directly onto the person it is talking to. This automatic face recognition feature was not included in the virtual version, but the robot's gaze and head position could be adjusted manually. To make sure that the VirFur was placed correctly and looking at the participant it was adjusted prior to every interaction step.

The chosen appearance for Furhat was a human-like male face with a texture called *Marty*. The available female faces usually included attributed makeup and did not have as naturalistic appearance as the male faces. Given that Furhat is bald, which usually is a more common haircut among men than women within the Scandinavian culture, it was also considered to be more suitable to go with a male version since we could not add a hat or hair in the virtual version, and we wanted the VirFur and PhyFur to be as similar as possible. The chosen male voice (called *Matthew*) is a text-to-speech (TTS) voice with an American English accent, that was used due to its realistic sound. This is one of the voices that comes with the Furhat software (Furhat Robotics, 2021) and allows for fine tuning adjustments e.g. pitch, volume, and speed regulations of the speech.

Figure 3: VirFur within the SDK (left) and PhyFur (right).

3.2.2 Interaction

Before interacting with the robot, the participants were given a small, non-interactive presentation of Furhat to make sure all were given the same introduction. To guide the interaction a goal driven interaction assignment was used, as a measure to control the interaction structure without limiting the participants too much. For the assignment the participants were handed flash cards (Appendix 3) to be used to ask Furhat questions. The flash cards contained short statements from which the participants were asked to formulate a question. The participants were handed these cards before the interaction started so that they could become familiar with the assignment. If the participant found it difficult to formulate a question, the backside of the flash card contained an example question. The participants could ask their questions in any order, since the dialogue script was built as an open question session. This assignment was used as it was easy to learn and understand, and as it simulated a potential situation where Furhat could be used for information gathering. The decision to limit the participants' assignment to reading flash cards was made to ensure that the participants focused on the robot rather than the task.

The whole interaction was in English and consisted of three phases (see Figure 4). The first phase was a general knowledge question session, in which the participants learned to interact with Furhat, to become familiar with the situation, and to establish a first impression. The second phase started after a short interview (section 3.2.5). It contained a specific dialog where the participants had to navigate a conversation with Furhat while answering Furhat's questions. The second phase was directly connected to the third phase, in which Furhat introduced a content specific question session. The interaction was done in English because of some differences within the coding of different languages, and due to the application programming interface (API) calling an external knowledge base that required English key words.

1. Furhat Presentation
2. Interaction Phase 1: General question session
3. *First impression interview*
4. Interaction Phase 2: Content specific dialogue
5. Interaction Phase 3: Content specific question session
6. *Final interview*

Figure 4: Progression diagram.

Furhat Presentation

The presentation script was provided by Furhat Robotics (n.d.b) and included information about Furhat as a social robot, and the abilities to change appearance, voice and showing expressions.

General knowledge question session

The first question session script was built with an introduction where Furhat first asked for the participant's name, and then opened the conversation for the question session (e.g. "I have gathered a lot of information lately, do you have any questions?"). When the participant asked a question the code broke down the input to key words and called a knowledge base (Wolfram Alpha LLC, 2009) via an API. The framework for the API-calls to Wolfram Alpha was built by Furhat Robotics (n.d.b). The dialog around the responses from the database was rebuilt and customised for this study to include some additional responses, variation in speech, pauses and randomisation of the responses in between utterances to generate a more natural conversation with Furhat.

Content specific dialog and question session

The second and third phase included a dialogue session and a question session, respectively, that contained content that specifically focused on topics related to Volvo Cars. The choice of Volvo material was to simulate a potential area for application where HRI is used to retrieve information. The dialog session started out with Furhat asking the participants if they have a driver's license. The following dialog was based on Furhat asking participants to choose between two main topics, "safety" or "electrification". Furhat then shared information about the topic and asked for input from the participants where they were requested to choose between some subtopics. The content specific question session started directly after the dialogue session where Furhat introduced it by saying "Do you have some questions about Volvo features?". Key words from the participants' questions had predefined answers within the script, so if the participants formulated a question containing the keywords "Pilot assist" (e.g. "Can you tell me what Pilot assist is?") Furhat would share information about the Pilot assist function within a Volvo car. This interaction also included variation, pauses and randomisation of the responses. Both the content specific dialog and question session were built and designed for this study, and Volvo content was collected from Volvo Car's webpage (https://www.volvocars.com/se) and car manuals provided by Volvo Cars (J. Wilkie, personal communication, April 19, 2021).

3.2.3 FaceReader

To analyse the participants' emotional expressions FaceReader (Noldus, 2020) was used. FaceReader is a software program that can analyse a total of seven facial expressions, from a combination of facial features based on Ekman's (2003) six emotional states which includes surprise, sadness, anger, happiness, scared, disgust, a neutral state (a state where the participant shows no significant emotion) and, optionally, contempt. FaceReader can also measure continuous psychological affect states (valence and arousal) (Noldus, n.d.). For the analysis recorded material was used from the interaction phases and analysed with mean valence. The measure of emotional valence is displayed between -1 to +1, representing a scale between a negative and a positive emotional valence with a neutral valence around 0. The material prior to, and after, the interaction was not analysed because other stimuli could have caused the participants' emotional expressions.

A grey screen was placed behind the participants to create a similar background for all participants and extra lighting was included in the setup to highlight the participants' faces (Figure 5 and Figure 6). Analysing facial expressions were used as an objective measure to complement the subjective data from the interviews, and as a measure to discover potential deviations or trends in the data.

Figure 5: The setup environment VirFur (left) and PhyFur (right).

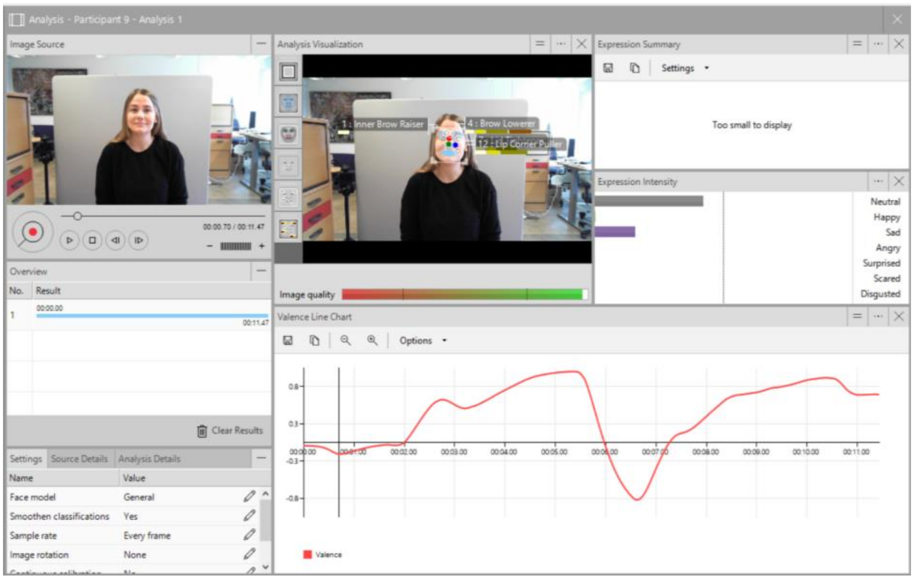


Figure 6: Example of FaceReader analysis (one of the researchers as model).

3.2.4 Survey

For this study the participants were asked to fill in a survey prior to the interaction session. The pre-interaction survey consisted of three parts. In the first part the participants were asked to give their consent to the study. Here they were informed about their right to withdraw from the study and their right to confidentiality, together with a statement that clarified that the data collected was only to be used for the purpose of this specific study.

In the second part of the survey the participants were asked to fill in some general information (age and gender) and a brief description of their earlier experiences with artificial agents. This was done to get an impression of their technical background and if they were accustomed to that type of technology. The third part consisted of the Negative Attitudes towards Robots Scale (NARS) (Nomura et al., 2006). NARS is a questionnaire that consists of three subscales and a total of fourteen questionnaire items (Appendix 2). Subscale 1 consists of six items regarding negative attitudes towards interactive situations with robots; subscale 2 contains five items regarding negative attitudes towards robots' social influence; and subscale 3 has three items that concerns negative attitudes towards emotions in interactions with robots. In the questionnaire the participants judge their level of agreement on various claims using a Likert scale of 1-5 (strongly disagree - strongly agree). The answers in subscales 1 and 2 are summarized according to the Likert scale (1=1 point and 5=5 points), but subscale 3 are reversed items, and therefore summarized in opposite to the scale (1=5 points and 5=1 point). The higher the summarized points in a subscale, the more negative is the attitude towards the robot in that area (Nomura et al., 2006). As NARS does not have an official Swedish translation the questionnaire was translated from English by the researchers with the help of a translator (see acknowledgement). In the survey the questions were displayed in both Swedish and English to make sure nothing got lost in translation. As feelings towards robots can be influenced by one's attitudes towards robots (Paetzel-Prüsmann, 2020), this measure was included to create an understanding about what the participants think about robots prior to meeting one.

3.2.5 Interviews

To collect data about the participants' subjective experience with Furhat semi-structured interviews were used. Two interviews were conducted. The first interview was performed after the general knowledge question session (section 3.4) to capture the respondents' initial reflections about the agent. The participants were asked about how their first experience with Furhat felt, how they would describe the agent, and what abilities they would ascribe to it. Finally, they were asked more generally what they thought of Furhat, and what feelings the agent had elicited in them. The second interview was performed after the second and third interaction task. These interviews were more in depth as the participants were given more questions and more space to elaborate. A set of questions had been prepared as a base for the interview (Appendix 4). Depending on the respondent's answers other questions were added (e.g. follow-up questions, questions to clarify the meaning of an answer, or request aimed at addressing contradictions), or the participant was asked to go into more detail. Some questions focused on the elicited feelings in the participant when they interacted with Furhat. As earlier research has expressed a need for clarifications about which aspects of the robot (e.g. appearance, movement, interaction) contribute to a person's feelings (Ho et al., 2008) these questions were followed up for more specific answers. Because of the role expectations play in human interactions we also addressed the participants' expectations of the agent and the situation, and how they related the interaction with Furhat to other experiences they had with artificial agents (depending on what they had written in the questionnaire). A separate type of question dealt with the actual assignment, and what the participants thought about using Furhat.

During the interviews the participants were shown clips from their video recording and asked to elaborate on those clips. The clips chosen were ones where the participant had an explicit reaction to something in the interaction. Before concluding the interview all participants were asked if they had other comments they wanted to share with the interviewer. Both interviews were conducted in Swedish to make it easier for participants to express themselves freely.

All interviews were transcribed manually word for word. Pauses, stutters, filler sounds and phrases were left out as this study did not focus on what behaviour the researchers observed or interpreted in the participants, but on the aspects that the participants shared verbally after the interaction. The transcriptions were then categorised into themes by sorting citations from the interviews. Here categorization refers to the deconstruction and rearrangement of data into themes on which the analysis is based (Maxwell, 1996). Three general themes were defined based on our research questions: *self-reported feelings (Q1), interaction experience (Q2)* and *dialogue (Q3)*. These themes were broken down into subcategories when commonalities had been identified (Appendix 5).

3.4 Procedure

The participants were recruited by asking around and through the social media website Facebook. A notice of interest was published in a Facebook-group (Kognitionsvetenskap 2020-2023, whose members consists of first year bachelor students in cognitive science at University of Gothenburg) and on the Facebook-page of one of the researchers. In the notice of interest there was information about the language requirements (that the participants had to be fluent in both Swedish and English), but that no other prior knowledge was needed to participate. The participants were also asked about their current occupation, level of education, and if they have interaction experience with artificial agents (Appendix 1). All participants were asked to fill out the *pre-interaction survey* two days before attending the interactive session.

Prior to the interactive session all participants were informed about the task and reminded about their right to confidentiality, anonymity, and right to withdraw from the study if they wished to do so. The participants were shown to a chair, which was placed in front of either VirFur or PhyFur. Half of the participants interacted with VirFur, while the other half was presented to PhyFur. Allocating half of the participants to each version of Furhat, rather than letting them interact with both, was done to prevent novelty and recency effects from affecting their interpretation of the agent. Because of prior findings on how differences in physical and virtual presence affect uncanniness perception (see section 2.1.2), and because it is problematic to compare two agents that differs in type of presence (collocated vs. virtual) (Li, 2015) the division was done to make the scenarios clearly distinctive from one another.

The participants were first introduced to the session by the non-interactive *Furhat presentation*, which was approximately 1 minute long. After the presentation the participants were informed about the first task and handed 10 flashcards. They were given time to become familiar with the questions before starting the interaction. The first interaction part started with the *general knowledge question session*. The duration of the first interaction was intended to be around 5 minutes to give the participants enough time to create a first impression (Paetzel-Prüsmann, 2020), but could vary +-2 minutes due to individual differences (such as time to understand the task, adaptability to conversation, or attempts to explore Furhat's functions). Before continuing with the second and third interaction phase there was a *first impression interview* which took between 4-8 minutes to carry out. Then the participants were informed about the upcoming session and handed 10 new, content specific, flashcards to read through. The second interaction contained a *content specific dialogue* and was followed by the *content specific question session.* The duration for this interaction session was estimated to last for 7 minutes, but similarly to the first interaction session this could vary +-2 minutes due to individual differences. The interaction was recorded using a camera that was placed in front of the participants (see Figure 5 for camera placement). This recording was used as input in FaceReader and for transcription of the first impression interview. The interaction session was concluded with a semi-structured interview that was conducted in the same room. The interviews were recorded using an iPhone 7 and an application called Diktafon (Dorada App Software, 2020). The iPhone was placed out of sight of the interviewer and the participant so as not to disturb the conversation, but the participants were informed about its presence. The interviews lasted between 15-32 minutes, and the total duration of the study was approximately 45-70 minutes, depending on the length of the last interview.

The methods for analysing the data were mean valence output from FaceReader, the NARS results, and the thematic analysis of the transcribed interviews. The summarized subscale scores in NARS were categorized into three levels: low, medium, and high (Figure 7). The transcriptions were deconstructed and categorized into three major themes, and further into subcategories. The content in each subcategory were then summarized, and further deconstructed and analysed to find patterns in the overall data. The feelings identified in the transcripts were categorized using Russell's (2003) core affect model, and divided into four categories: high-arousal, positive-valence; low-arousal, positive-valence; high-arousal, negative-valence; and low-arousal, negative-valence.

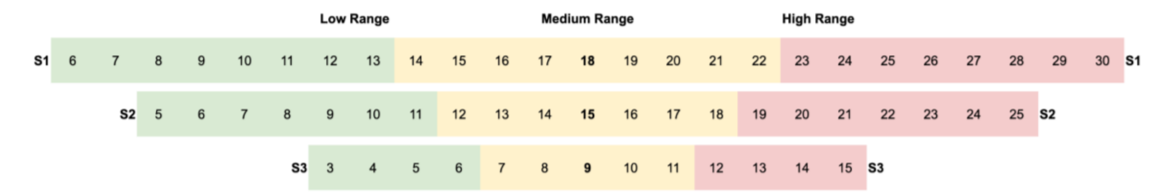


Figure 7: Ranging low/medium/high score according to minimum and maximum scores S1: 6-30; S2: 5-25; S3: 3-15. A high score in NARS indicates a negative attitude in relation to the subscale's topic (Nomura et al., 2006).

With this methodological setup the study intended to capture the emotional reactions in naive users of social, human-like robots. To address our research questions the thematic analysis of the interviews was combined with the results from NARS and FaceReader with the purpose of finding potential patterns within each scenario, from which topics for future studies could be identified.

4. Result

4.1 Scenario 1: PhyFur

4.1.1 NARS and FaceReader output

The NARS questionnaire showed that most of the participants scored either low or medium in the different subscales. One participant (P02) had a low score throughout all subscales, while two participants (P04, P06) had a low score in subscale 1 and medium score in subscale 2 and 3. One participant (P08) scored medium in subscale 1 and a high score in subscale 2 and 3.

Throughout the session the participants meeting PhyFur had a positive total mean valence (referring to the mean value for the whole interaction). The participants started out on a higher mean valence in the presentation, with a decreasing trend throughout the interaction phase 1-3. Participant P06 was excluded from the data because FaceReader could not pick up the participant's expressions due to facial hair.

Scenario 1:	**Pre-interaction survey NARS** *Negative Attitudes towards Robots Scale*			**Interaction FaceReader** Mean Valence (min -1 , max +1)			
PhyFur Participant	**Subscale 1** *Situations of Interaction with Robots*	**Subscale 2** *Social Influence of Robots*	**Subscale 3** *Emotions in Interaction with Robots*	**Presentation**	**Phase 1**	**Phase 2 + 3**	**Total**
P02	6	8	5	0.703	0.328	0.167	**0.185**
P04	7	13	7	0.585	0.384	0.127	**0.247**
P06	11	18	7	n/a	n/a	n/a	**n/a**
P08	16	19	13	0.333	0.151	0.061	**0.109**

Table 1: Result from NARS and FaceReader for each participant in scenario 1.

4.1.2 Interview theme 1: Self-reported feelings

Pleasant feelings

All four participants (P02, P04, P06, P08) expressed high-arousal, positive-valence feelings in terms of having fun, being excited, happy and that they found it interesting to interact with PhyFur. One participant (P06) expressed feeling safe (low-arousal, positive-valence) while interacting with the robot, and another one (P08) explicitly mentioned not being deterred from interacting with robots again. These pleasant feelings were described by the participants as being due to the experience of talking (P02, P06), connecting (P04, P06), and cooperating (P04) with someone. The interaction being a new kind of experience (P02, P08) and PhyFur's ability to move its head (P08), joke, and smile (P04) also contributed to the pleasant feelings.

Unpleasant feelings
Three participants (P02, P04, P08) expressed that they felt high-arousal, negative-valence feelings including discomfort, eeriness, shock, frustration, and annoyance. Feelings of discomfort, eeriness and shock were referred to PhyFur's human-like appearance (P04, P08), while annoyance and frustration emerged when the interaction encountered problems (P02, P04). One participant (P08) expressed that the discomfort was not related to the interaction, but to PhyFur looking like a doll that blinks and imitates facial expressions. Participant P04 commented on the feeling of eeriness being connected to the robot changing appearance in the Furhat presentation, but also to how quickly s/he accepted the changes and that a "new agent" was present.

Other
Feeling bewildered (high-arousal, neutral-valence) was mentioned by one of the participants (P04) as a response to seeing something so human-like that clearly is not human. One participant (P06) reported having no strong feelings and feeling neutral towards PhyFur.

4.1.3 Interview theme 2: Interactive experience

Appearance and traits
All four participants' (P02, P04, P06, P08) descriptions of PhyFur related to the face. The other components of the robot were rarely mentioned. When describing PhyFur the participants often focused on specific features, such as the eyes and the eye movements, the facial expressions, the movements in the skin, and the contours in the face. Two participants (P04, P08) described the appearance of the robot as human-like; two participants pointed out that the eye movements and its ability to move its head made it look naturalistic, genuine (P06) and more real (P08). PhyFur was also described as kind, helpful, safe (P02) and inviting (P08). The robot's emotional expressions were however described by two participants (P06, P08) as exaggerated and static (more like an impression of how a human facial expression looks).

Social presence
PhyFur was described by three participants (P02, P04, P06) as socially present. Two of them (P04, P06) said that the realistic face made the robot feel like a living being. Participant P02 commented on the facial expressions, and that they created a sense of talking, if not *with*, at least *to* someone that embodies the voice. The facial expressions and the head movements were described by one participant (P04) as a bridging factor, making it feel like being two agents in a conversation rather than talking to a tool. Two participants (P06, P08) felt the need to say thank you to PhyFur, to be polite in case PhyFur remembered what was said (P06) and to make the conversation run smoothly (P08).

Expectation on ability
The human-like features were expected to be reflected in some of the robot's abilities by the participants. PhyFur's conversational skills were mentioned by three participants (P02, P04, P08) as less developed than expected, leading them to adjust their way of communicating. Two participants (P02, P08) talked about the robot not being able to understand what they thought were relatively easy questions, and one of them (P08) pointed out that Furhat was breaking conversational rules (such as talking when not expected and interrupting). Participant P06 thought PhyFur would talk more casually due to its human-like face. It was also pointed out by three participants (P02, P06, P08) that the robot's gaze deviated from human gaze. That PhyFur continuously looked at participant P06 was described as annoying, and participant P02 commented that the experience was not the same as interacting with a human as the robot did not move its gaze to e.g. indicate that it was thinking.

Participants P02 and P04 stated that they experienced a clash between the expectation and the actual experience of PhyFur. Participant P04 said that the clash might be due to the rational thoughts of seeing a plastic thing with a projector, meeting the emotional side that sees something human, and the jumping back and forth between these impressions due to minor deviations in the appearance and interaction. The same participant also expressed that the combination of a human-like appearance and robot-like abilities felt bad, as it takes you out of the illusion of interacting with someone.

4.1.4 Interview theme 3: Dialogue

Positive feedback on the interaction task and user-interface
With respect to the specific interaction task and its content it was mentioned by one participant (P02) that the information was short, concise, and interesting, but that the specific Volvo content came off as a bit odd at first, but then became a good example of how this technology can be used. Participant P06 liked the first interaction phase but felt like the robot was trying to sell stuff in the remaining session. When PhyFur included humour in the dialogue (the phrase "Nice, huh?" with a wink) it was experienced as genuinely funny compared to when it made jokes (P04), and it was recommended that more stuff like that is used in future dialogue (P06). One participant (P06) said that the robot, as a user-interface, was neat.

Negative feedback on the interaction task and user-interface
PhyFur was perceived to have somewhat limited capacity, both in terms of its ability to understand and answer the participants, and its knowledge. The speech was described as commando-based, forced, and unnatural (P06), pre-programmed (P08) and a bit mechanical (P02). Having to repeat phrases (P08), not responding as a human would do (P02, P08), being quite specific with the choice of words and having to speak relatively clearly (P04) was considered problematic. Participant P06 did not like how the information was given, and participant P08 got tired of the robot when it did not understand what was said and decided to

try out its head movements instead. Two participants (P02, P04) described uncertainty about PhyFur's knowledge. One suggested that it might have limited access to data (P02) and the other said that it was hard to judge whether its knowledge was limited to the Volvo topics (P04). Two participants (P02, P08) expressed that interacting with an actual human is preferable in some cases. One of them (P08) found the purpose of the robot unclear.

Two participants (P04, P06) talked about PhyFur having a robot-like setup and sounding monotone at the same time as looking human-like. This was perceived to affect how the information was received. Participant P06 stated that it was hard to understand and memorize the information when the robot talked about technical information in a robotic way, especially as s/he did not have any previous knowledge about the topic. Participant P04 commented on the differences between information that comes from a digital source and a human source. Depending on the source s/he talked about having different approaches to how one interprets the information. PhyFur having contradictory qualities (being robot-like while looking human) made the participant almost stop listening as it became too much to take in at the same time. Both participants (P04, P06) emphasized that the presentation needed something more to become interesting (e.g. more natural speech or other stimuli). Participant P04 also expressed the importance of having a real conversation rather than just asking the robot questions.

For an overall summary of the interview results for scenario 1 see Table 3.

4.2 Scenario 2: VirFur

4.2.1 NARS and FaceReader output

The NARS questionnaire showed that most of the participants scored either low or medium in the different subscales. Participant P01 had a low score in both subscale 1 and 3, and a medium score in subscale 2. Participant P03 had a low score in subscale 1 and scored medium on both subscale 2 and 3. One participant (P05) scored low throughout the subscales and the last participant (P07) scored high in subscale 3 and medium in both subscale 1 and 2.

Participants meeting VirFur showed a decreasing trend in mean valence from the presentation and throughout phase 1-3. While three participants (P01, P03, P07) had a negative total mean valence close to 0, participant P05 deviated by having a positive mean valence throughout the whole session (referring to the mean value for the whole interaction).

Scenario 2:	**Pre-interaction survey NARS** *Negative Attitude towards Robots Scale*			**Interaction FaceReader** Mean Valence (min -1, max +1)			
VirFur Participant	**Subscale 1** *Situations of Interaction with Robots*	**Subscale 2** *Social Influence of Robots*	**Subscale 3** *Emotions in Interaction with Robots*	**Presentation**	**Phase 1**	**Phase 2 + 3**	**Total**
P01	12	15	6	0.345	0.010	-0.082	**-0.009**
P03	11	12	10	0.052	0.031	-0.040	**-0.001**
P05	8	9	4	0.694	0.444	0.136	**0.294**
P07	15	18	14	0.284	-0.190	-0.194	**-0.150**

Table 2: Result from NARS and FaceReader for each participant in scenario 2.

4.2.2 Interview theme 1: Self-reported feelings

Pleasant feelings

All four participants (P01, P03, P05, P07) communicated high-arousal, positive-valence feelings in terms of having fun, being excited and interested. These feelings were expressed together with descriptions of the interaction (P01, P05, P07), that it felt like having a conversation with someone (P03), and that the situation felt like practice for future interactions with similar agents (P01). Reported low-arousal, positive-valence feelings were safe (P01), comfortable (P03, P05, P07) and calm (P03, P05). One participant (P03) pointed out that the interaction with VirFur got more comfortable over time. Two participants (P03, P05) said that they appreciated meeting a virtual agent instead of a physical agent, and P05 also explicitly said that VirFur was not scary. Participant P07 found the agent to be not as scary as expected.

Unpleasant feelings

Three participants (P01, P03, P07) described high-arousal, negative-valence feelings such as nervousness (P01, P03), frustration (P01), tension (P01, P03), annoyance (P03), creepiness (P03, P07) and discomfort (P03, P07). Nervousness, frustration, tension, and annoyance were connected to problems related to the interaction (P01, P03), that the situation was new (P01, P03), and that the participants felt insecure about the situation and about how they should behave towards VirFur (P01, P03, P07). Two participants (P03, P07) expressed uncertainty about VirFur's emotional capacity, which made them unsure about how they should behave. This uncertainty was described as creepy by participant P03. Participant P07 commented on feeling a bit exposed by VirFur due to its intense gaze, and therefore perceived it as creepy. One participant (P07) described feeling ashamed (low-arousal, negative-valence) for not knowing when to respond to VirFur. Participant P05 did not express any negatively valanced feelings.

4.2.3 Interview theme 2: Interactive experience

Appearance and traits

VirFur was described as smart, analysing, observant, and cold (P07); kind (P01, P07), polite (P03) and helpful (P01); nice, inviting, friendly (P01); and social (P05). All four participants (P01, P03, P05, P07) described the agent as human-like due to its expressive face with its distinct eyebrow- and eye movements. Not having eyelashes, naturalistic skin colour (P07), and not being attached to a human body (P01, P05) made the agent appear non-human. VirFur being shown on a screen also made the experience feeling less real (P05). Two participants (P05, P07) described VirFur as being robot-like and unnatural due to jerky facial movements. One of them (P05) stated that although the movements were not "completely off", that more and smoother features could be added.

Two of the participants (P03, P07) commented on the mouth as a distinct feature that appeared mechanical and non-human-like. One of them (P07) said that the agent did not look happy even though it smiled, and that the smile was stiff and forced. At the same time participant P07 expressed that the mouth was more comfortable to look at compared to the eyes. The eye movements were expressed by three participants (P01, P03, P07) as being uncomfortable to look at and weird looking, but P03 also described the eyes as happy and warm. Participant P05 found the appearance to be very neutral. P01 commented on feeling distracted by the face.

Social presence

Three participants (P01, P03, P05) said that it felt like they were talking to someone or something when interacting with VirFur. Its social presence was based on the appearance, the expressions, and the way of talking VirFur was equipped with. The knowledge that VirFur was not a living being made the interaction uncomfortable for participant P03. Participant P01 assumed VirFur would understand her casual way of talking just because of VirFur's human-like appearance, and when those expectations did not match VirFur's actual abilities it changed her mind about its perceived human-likeness. VirFur's trustworthiness was commented on by three participants (P03, P05, P07). One of them (P07) felt that the agent's eyes made it feel like s/he was observed, and that the gaze was intrusive. This participant was also concerned that the agent might record the session, and that it would not forget the interaction. This led to an initial impression of having to be alert all the time when interacting with it, but this impression was described by P07 as fading over time as s/he got more comfortable. Participant P03 expressed an initial concern that VirFur would behave surprising and unreliable, but overtime as s/he got more comfortable with the interaction this concern was reduced. Two participants (P03, P05) said that they had preferred to have access to the correct answers so that they could have compared VirFur's answers.

Adjusted behaviour
Two participants (P03, P07) expressed being insecure about how they should behave towards VirFur. Both had expected the interaction to be harder than it was, either due to lack of experience (P03) or due to feeling cautious towards AI and social robots (P07). Participant P03 talked about wanting feedback from VirFur (in terms of showing understanding and affirming behaviour) to get a better understanding on how to adjust to the interaction. Three participants (P03, P05, P07) said that they had to adapt their speech to make VirFur understand what they wanted. P03 and P05 rephrased their questions and P07 said s/he started to talk louder and more harshly. Three participants (P03, P05, P07) wanted to talk to VirFur like they would talk to a human. One participant (P03) wanted to act social because it felt natural but said that VirFur was just listening and not really interacting. Another participant (P07) wanted to reply to VirFur for a more human-like interaction but decided to inhibit that instinct. Participant P05 said that VirFur needed some fine tuning because it was hard to know when an utterance started/stopped, but s/he was overall impressed by VirFur’s comprehension abilities.

4.2.4 Interview theme 3: Dialogue

Positive feedback on the interaction task and user-interface
All four participants (P01, P03, P05, P07) expressed positive opinions on receiving information from VirFur. More specifically the interaction was described as fun (P03), nice (P05), and that it contained good content (P01, P05, P07). One participant (P07) was convinced that VirFur could be used as an interface to retrieve information, but it was unclear in what contexts. Two participants (P03, P07) explicitly said that the Volvo content was entertaining and educational, while another (P01) expressed that s/he would be more receptive to the Volvo content if s/he had been interested in buying a car at the time.

Negative feedback on the interaction task and user-interface
Two participants (P01, P07) mentioned that they wanted to ask VirFur follow-up questions, but neither of them tried to do so during the interaction as they were not sure if they were allowed to do that. Problems with retrieving information concerned some of VirFur's way of phrasing answers (P01, P05), talking too fast (P07) and distraction by the facial movements (P01).

For an overall summary of the interview results for scenario 2 see Table 4.

	Interview Theme: Feelings			Interview Theme: Interaction			Interview Theme: Dialogue	
	Pleasant	**Unpleasant**	**Other**	**Appearance and traits**	**Social Presence**	**Expectations on ability**	**Positive feedback**	**Negative feedback**
Physical Furhat	• Fun • Excited • Happy • Interesting • Safe • Connecting • Cooperating • New experience • Head moving • Smiling and joking	• Discomfort • Eeriness • Shock • Human-like appearance • Frustration • Annoyance • Interaction problems • Doll-like appearance • Changing appearance	• Neutral • Bewildered	• Face • Eyes and eye-movements • Human-like • Naturalistic • Genuine • Kind • Helpful • Safe • Inviting • Exaggerated expressions	• Socially present • Realistic face • Living being • Talking to someone • Embodied voice • Bridging factor • Politeness	• Expectations of human-like features reflected in abilities • Less developed conversational skills • Adjusted behaviour • Breaking rules of conversation • Casual talk • Deviated gaze behaviour • Clash between expectations and abilities	• Concise and interesting • Good example of application • Appreciated content • Selling content • Humour • Neat interface	• Limited understanding and knowledge • Commando-based • Mechanical • Conversation issues • Unclear purpose • Robot-like setup with human-like appearance • Monotone voice • Information interpretation

Table 3: Keyword summary of interview result for scenario 1 (PhyFur)

	Interview Theme: Feelings		Interview Theme: Interaction			Interview Theme: Dialogue	
	Pleasant	**Unpleasant**	**Appearance and traits**	**Social Presence**	**Adjusted behaviour**	**Positive feedback**	**Negative feedback**
Virtual Furhat	• Fun • Excited • Interested • Conversation with someone • Practice for the future • Safe • Comfortable • Calm • Changes over time • Virtual agent preferred • Not scary	• Nervous • Frustration • Tention • Annoyance • Creepiness • Discomfort • New situation • Insecure about situation and how to behave • Unsure of emotional capacity • Intrusive gaze • Ashamed	• Smart • Analyzing • Observant • Cold • Kind • Polite • Helpful • Nice • Inviting • Friendly • Social • Human-like appearance • Expressful face • Eyes and eye-movements • No body • Inhuman • Screen presence • Robot-like expressions • Stiff mouth • Uncomfortable eyes • Neutral appearance • Distracted by face	• Talking to someone or something • Socially present • Expectation on casual speech • Violation of expectations • Questioned trustworthiness and reliability • Intrusive gaze • Comfortable over time	• Insecure of how to behave • Lacking feedback • Adapted speech • Casual talk • Not really interacting • Inhibited replies • Unclear start/stop	• Fun • Nice interaction • Good content • Entertaining • Educational	• No follow up questions • Information interpretation • Phrasing answers • Talking to fast • Distracted by face

Table 4: Keyword summary of interview result for scenario 2 (VirFur)

5. Discussion

5.1 Scenario 1: PhyFur

5.1.1 Complex feelings

While the uncanny valley phenomenon predicts feelings of uncanniness to dominate the emotional reaction towards human-like entities (Mori, 1970), our interviews show something else. Not only did the participants express feelings that are located on both sides of Russell's (2003) core valence states, but these feelings were also related to other aspects than just the agent's appearance. One the one hand, that the participants expressed both pleasant and unpleasant feelings might be because the robot was not human-like enough to bring the participants to the uncanny valley (which would be in accordance with the classic hypothesis). On the other hand, while components of affinity (the sense of interacting with a social entity) and uncanniness (reactions to human-like appearance) are present, these were not the only ones. This could be an indication that the participants interpreted their experience partly outside of the affinity-uncanniness dimension, and what we get is a more complex rendering of the emotional experience that is not captured by the uncanny valley hypothesis. The latter interpretation goes in line with an understanding of emotions as fluctuating rather than static (Russell, 2003) and has similarities to Lee et al.'s (2016) findings that human evaluations in relation to robots can be done outside of the uncanny valley logic.

When describing the reasons for their emotional reactions the participants referred to different parts of the interaction, both related to the robot and the situation itself. Appearance, as often used as an explanation to human emotional reaction towards robots (Paetzel-Prüsmann, 2020), appears in this case to be one factor, but not the only one, that plays a part in the participant's emotional experience. As Złotowski et al. (2015) have observed, the interaction also introduces another dimension that shapes the emotional reaction. In this case features of the interaction contribute to both positive and negative feelings according to the participants themselves. As a positive mean valence is visible in the FaceReader output for this scenario, it appears as the reaction towards this robot is not necessarily negative. It could also be the case that uncanniness is not necessarily perceived as negative and limiting, e.g. that a robot could be perceived as somewhat uncanny at the same time as the person has a positive interactive experience. If so, uncanniness might not be such a limiting factor as the uncanny valley hypothesis makes it. Hence, to get a fuller understanding on how interactions with human-like agents affects humans the positive experiences need to be captured together with the unpleasant ones and examined together.

5.1.2 Human-likeness creates expectations

PhyFur's degree of human-likeness was a general pattern in the participants' description of the robot. Like the findings in Lee et al. (2016) the participants in this case highlighted specific features in the appearance that were human-like rather than the robot as whole, but they also ascribed it personal qualities and social presence.

The impression of PhyFur as a human-like, social entity seemingly seeped into the participants' expectations of what the robot should be able to do, and what it should know. As none of the participants had prior experience with social robots this could be a result of them referencing experiences from interactions with humans (Bartneck et al., 2020; Craig & Edwards, 2021; Schramm et al., 2020), but as none of them made any explicit comments about making such references this interpretation remains tentative. Another interpretation of the situation is that the robot's appearance and presence indicated high-performance skills (Malle et al., 2020) that resembles those of a human. Being met with a behaviour that did not match their impression of the robot seems to have caused the participants frustration and created a clash between expectation and experience, indicating that something in the participants' approach were challenged. Looking at the mean valence output (section 4.1.1), this might be an explanation for the decreasing values, but explicit testing is needed to confirm this observation.

While human-likeness was central in their descriptions, the robot was not perceived as human in its entirety. Two participants explicitly talked about being torn between two impressions (something that is not human, and something that appears human), and how this led to a decreased capability to interpret the information given from the robot. This begs the question to what extent human-like design serves its purpose as it might lead to assumptions of the robot's capability, but also confusion in the human interactor on how to approach this type of technology.

5.1.3 General findings

Although the participants were met with some issues during the interaction, all of them still found the experience to be an overall positive and interesting one, as interpreted from the interview responses and the FaceReader output. Looking at subscale 1 in the NARS (section 4.1.1), this sample consisted of persons that had less negative attitude towards interactions with robots, which might have affected how the experience is perceived. The robot being physically present could also influence the participants' attitude (Li, 2015; Powers et al., 2007), and so could its ability to react to the participants movements. However, as this is not controlled for in this study we cannot make any conclusions about the impact of these aspects.

Based on the transcripts from this scenario PhyFur as a user-interface has potential, but to be engaging it needs to be perceived as capable. For example, it is not enough for the robot to have good things to say, the content needs to be presented in a way that aligns with the visual impression of the agent (in this case Furhat). This indicates that how the information is conveyed (e.g. that it is presented in a socially fluid way) might be more important than the actual content.

5.2 Scenario 2: VirFur

5.2.1 Contradicting experiences of the interaction

The participants in this scenario experienced both pleasant and unpleasant feelings during the session, but seemingly more because of the interaction than the agent based on the participants' transcripts. It is not clear what this pattern implies, but it could be that something in the interaction was more prominent in the participants' recall compared to aspects related to the agent. One guess would be prior exposure to human-looking animations, but as this was not addressed in the study it remains speculative.

That both pleasant and unpleasant feelings were reported by the participants as related to the interaction comes off as contradictory. It could be that some aspects of the interaction were considered pleasant while others were not, or that the overall experience was perceived as pleasant while specific parts of the session were unpleasant. Either way, this indicates a more complex pattern that might be partly explained by the participants' attitude towards interacting with the agent (subscale 1 in NARS indicate a less negative attitude, see section 4.2.1), and partly by other factors such as the task structure or context (Rosenthal-von der Pütten & Weiss, 2015).

Although uncanniness was not used by any of the participants to describe their experience, other emotion terms that have been attributed to this feeling (Ho et al., 2008) were reported. This could indicate that this agent elicited feelings of uncanniness too (as observed by Paetzel-Prüsmann, 2020), but it is difficult to state for certain that these emotion terms were coupled with the specific ill-defined discomfort that comes along with the feeling of uncanniness (Mangan, 2015). Contrary to the uncanny valley hypothesis (Mori, 1970) these feelings were not accounted for by VirFur's appearance. Rather, the discomfort of interacting with VirFur emerged from the novelty of the situation and a limited understanding of the agent's emotional capability and memory capacity, leading to a worry of how reliable the agent is.

5.2.2 Uncertain expectations and adjusted behaviour

The expectations on VirFur were primarily shaped by the agent's unknown nature and its communicative abilities. The uncertainty affected both how the interaction felt (see above), and the behaviour of the participants that experienced it. The behavioural outcome was adjusted and inhibited speech, even though participants wanted to talk more casually and act like they would do in an HHI. Inhibition has been seen in interactions with robots (Powers et al., 2007; Kiesler et al., 2008), but this scenario could mark similar effects in interactions with virtual agents as well. While the uncertainty faded over time, the adjusted behaviour remained throughout the interaction as the participants found a strategy that worked for them. Finding what they thought was the agent's limits might have made the interaction less exciting, and it could have decreased their affinity towards the agent. This could be why we see a lowered emotional valence response towards the agent (see FaceReader output section 4.2.1).

The participants also expressed uncertainty about VirFur's trustworthiness. This was an unexpected finding that can be interpreted in several ways. Firstly, this could be due to the participants scoring in the medium range in NARS's second subscale (section 4.2.1, participant P05 excluded), meaning that they might carry some worry about the agent's social influence. Secondly, it could be connected to feelings of uncanniness caused by an expectation gap (Kok & Soh, 2020) between what the participant anticipated and what they did not know about the agent's behaviour. The conclusion here is that pieces of information are missing as the observation is open for more than one interpretation, meaning that this study did not capture how the participants perceived trust.

5.2.3 General findings

Like the observations of Lee et al. (2016) VirFur's appearance was described using specific features. While some features were seen as human-like, most comments were on the features that did not measure up to its human counterpart. VirFur's appearance has a simpler design than other types of animations (e.g. some game- and movie graphics have reached a high level of realism), so it could be that the participants have been exposed to more realistic animations prior and therefore expected something more. This might be why the mean valence was comparatively low (except for one participant, see section 4.2.1) – the VirFur stimulus might not have been exciting enough. Despite this, based on the participant's accounts it appears as though the agent was still giving an impression of having some social presence, although that impression could vary.

The participants in this scenario expressed more interest for the interaction rather than the agent, and had positive feedback on VirFur as a user-interface for retrieving information. Apart from one participant that got distracted by the robot's features, the others liked and commented on the content in the task. As no recall test was performed it is unclear if the participants remember more from the interaction with the virtual agent (such as in Powers et al., 2007), but they found themselves being entertained by the task and thought it was a good way of receiving information, although the phrasing needed some fine tuning. The focus on the interaction and information given could perhaps be related to the today common practice of receiving information from screens (e.g. through television, gaming, and smartphones). The stimulus not being a completely novel one might have led to the interaction and content becoming a central aspect to concentrate on, but as neither the notice of interest nor the survey captured such experiences this interpretation cannot be confirmed.

5.3 Suggestions for future studies

This study suggests that interactions with human-like agents can lead to both pleasant and unpleasant feelings. To our knowledge, positive affect in relation to human-like artificial agents is less researched than negative affect. Similarly to how Ho et al., (2008) investigated attributions of uncanniness, one could start with looking at what kind of emotion terms are expressed in relation to affinity. This would not only help understanding the human experience, but it could also help with the operationalisation of the concept. Since prior studies seem to have focused on uncanniness based on appearances it could also be worthwhile to examine whether this uncanniness extends into the interaction, and what is causing uncanny interactions with artificial agents. Another approach would be to see whether uncanniness is necessarily limiting for the interaction, or if there are aspects that can overshadow the sense of uncanniness. As Mangan (2015) points out, uncanniness is a peculiar feeling that goes beyond what might be measurable through introspection. Therefore other methods should be used to complement self-report measures to capture the essence of the experience, and what is causing it. Finding objective measures to investigate this feeling might be a way forward.

The FaceReader mean valence for all participants in this study decreased between each interaction phase. This observation might be of interest for further examination as something seems to change over time within this interaction (similar to observations made by Złotowski et al., 2015). It could be the result of a habituation process as novelty effects wear off, but if so it is unclear what factors influenced that process; whether this adjustment is permanent or if it changes between interactions; or if it is related to the specific agent or to the interactor's overall approach towards artificial agents. Either way, further examination could be fruitful as this might say something about how human-like agents can be deployed in human spaces, and what is needed to learn humans to cooperate with such agents.

Both interactions with the VirFur and PhyFur seemingly created a sense of social presence for the participants, but it remains unsure if this is the same kind of social presence, what factors contributed to that presence, and how it might be related to feelings of uncanniness and affinity. As physical presence factors into behaviour and attitudes towards artificial agents (Li, 2015) there is a need to examine how physical presence affects social presence in human-like agents, and how this influences the relationship between artificial agents and humans. Furthermore, it is also of interest to examine whether humans' attitude towards robots' social influence also factor into this relationship.

Ecological validity can be enhanced by investigating potential application scopes for HRI in environments that match the interaction topic. Putting the participant in a more real-life scenario could influence how the agent and the content is perceived, and it opens up possibilities to try how different interactions work in different settings. What people might expect when interacting with a human-like agent raises the question in what scenarios it might be helpful to have a human-like design in the interaction, and when it may be necessary to have another appearance that better represents the agent's actual abilities and purposes.

5.4 Trustworthiness and ethical considerations

To ensure the study’s trustworthiness different types of data were used. Following the data collection some observations were made. Firstly, the FaceReader output showed consistently higher valence for all participants in the first part of the session (the presentation). This could be the result of the participants only listening at this stage, making it easier for FaceReader to identify their expressions. It could also be because they did not encounter any issues at this stage, leading to a higher score. As it cannot be deduced which of these explanations are correct, it is unclear if the mean valence decrease actually represents an emotional change in the participants. The problems that the participants encountered in the interaction were mainly due to the coding script, meaning that future studies using other code might not get the same result. A Wizard of Oz-technique could be a solution to this problem, but this technique comes with its own issues.

In this study we specifically asked for the participants' emotional experiences. Since interviews are self-reports and therefore subjective, a problem can be that the participants might exclude parts of information they do not want to share or believe to be irrelevant. The participants might also be unable to describe what they felt in a way that explains the underlying causes to that reaction, or they might find talking about feelings uncomfortable. Taking this into account we let people speak freely from open, unloaded questions, and unless a participant directly expressed that they had experienced a certain emotion they were not pressed for answers. In the case that a participant’s answer was ambiguous they were given the opportunity to describe and clarify, and because the interactions with Furhat were recorded it was possible to replay sections of the interview to the respondent, making it possible to address specific episodes in the interaction.

For this study a simple interaction between a human and a human-like agent was simulated in a laboratory-like environment, with the researchers present in the room. To ensure that all participants experienced the same kind of interaction for a similar duration of time, the interaction task was given a structure with specific interaction patterns. This was also done to make it easier for the participants as this type of interaction was a first-time experience. To avoid demand characteristics the participants were not observed directly by the researchers, the interview questions were piloted, and both researchers were present at all stages of the data collection. We acknowledge that the artificial setup might have limited the participants’ experience, and that this reduced the ecological validity of this study. It is also recognized that the choice of method limits the possibility to generalise the results as the sample size is small and because causal relations cannot be identified. This is however compensated by gaining richer and more diversified answers that gives insight to the complex phenomenon that the uncanny valley is.

Interacting with a robot can be considered as uncomfortable for some people, therefore the participants were made aware of what kind of study they were signing up for (the notice of interest stated that it was concerning HRI, and the pre-interaction survey also included this information). The interactive situation was also designed to exclude situations that explicitly could trigger negative feelings (e.g. scaring the participant). Recording and analysing the participants' facial expressions and their interview responses can be considered a breach of their integrity. Therefore all participants were asked to consent to the study and the use of the collected data before partaking in the interactive session.

6. Conclusion

This exploratory study shows that findings based on quantitative studies may not be enough to capture the complex feelings and the expectations that follow from actual interaction with artificial agents. As an answer to Q1 (*what feelings do participants experience with a human-like agent?*) the results suggest that humans can have a range of feelings under the same interactive setting, and that these feelings might not be completely isolated from each other (e.g. aspects of uncanniness might not preclude positive interaction experiences). Regarding Q2 (*how are human-like agents perceived?*) the physical agent has different challenges than the virtual agent; although looking and acting similar, they are perceived differently. Both versions induce certain expectations, and if these are not met some negative affect is to be expected. Having one's expectations violated is not the same as entering into the uncanny valley though, it could even be the other way around – identifying an agent's limitations might actually make the interaction more comfortable. For Q3 (*how do participants experience information retrieval through interacting with a social, human-like agent?*) this case indicates that content cannot be separated from how it is presented, and the context it is presented in. It might be that human-like design offer some benefits in HRI, but to reach its full potential the whole spectrum of what it is to have a good interaction (e.g. presence and responsiveness) needs to be captured.

The two main contributions of this study indicates that expectations play a role when interacting with a social robot, and that this forms human behaviour as well as the emotional reaction towards the agent. The results also point to human feelings towards artificial agents being more multifaceted than previous research suggests.

We recognise that this study has some limitations due to its explorative design: the small sample, the general scope of research questions, the use of a specific robot, and the design of the interaction all contribute to a result from which only limited conclusions can be drawn. However, this creates an opportunity to research this subject further. This study has identified several topics for future research, such as the impact of feelings of uncanniness in HRI, habituation processes, the relationship between physical and social presence and when to use human-like design. Hopefully, these insights can contribute to the HRI research field and the understanding of the workings of the uncanny valley in interactive settings.

References

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Sabanovic, S. (2020). *Human-Robot Interaction – An Introduction*. Cambridge: Cambridge University Press.

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). *Is the Uncanny Valley an Uncanny Cliff?* In Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007, Jeju, Korea, pp. 368-373. 10.1109/ROMAN.2007.4415111

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). *My Robotic Doppelgänger - A Critical Look at the Uncanny Valley.* In Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2009, Toyama, Japan, pp. 269-276. 10.1109/ROMAN.2009.5326351

Cheetham, M. (2017). Editorial: The Uncanny Valley Hypothesis and beyond. *Frontiers in Psychol*ogy, 8(1738). https://doi.org/10.3389/fpsyg.2017.01738

Craig, M. J. A., & Edwards, C. (2021): Feeling for Our Robot Overlords: Perceptions of Emotionally Expressive Social Robots in Initial Interactions. *Communication Studies*, 72(2), 251-265. https://doi.org/10.1080/10510974.2021.1880457

de Borst, A. W., & de Gelder, B. (2015). Is it the Real Deal? Perception of Virtual Characters Versus Humans: an Affective Cognitive Neuroscience Perspective. *Frontiers in Psychol*ogy, 6(576). https://doi.org/10.3389/fpsyg.2015.00576

Dorada App Software. (2020). *Diktafon Voice Recorder* (Version 1.16.10) [Mobile Application]. App Store. https://apps.apple.com/us/app/dictaphone-voice-recorder/id834938465

Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life.* Times Books: New York

Furhat Robotics. (2018). *Furhat Robot* (Generation 2) [Physical Robot].

Furhat Robotics. (2021) *Furhat SDK* (Version 1.23.1) [Software Development Kit].

Furhat Robotics. (n.d.a). *Furhat Robot*. Retrieved from https://furhatrobotics.com/furhat-robot/

Furhat Robotics. (n.d.b). *Furhat Robotics/example-skills*. GitHub. Retrieved from https://github.com/FurhatRobotics/example-skills

Ho, C.-C., MacDorman, K. F., & Pramono, Z. A. (2008). *Human Emotion and the Uncanny Valley: A GLM, MDS, and Isomap Analysis of Robot Video Ratings*. In Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, HRI '08, New York, USA,. 169-176. http://dx.doi.org/10.1145/1349822.1349845

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic Interactions With a Robot and Robot-Like Agent. *Social Cognition*, 26(2), 169-181. https://doi.org/10.1521/soco.2008.26.2.169

Kok, B.C., & Soh, H. (2020). Trust in Robots: Challenges and Opportunities. *Current Robotics Reports*, 1**,** 297–309. https://doi.org/10.1007/s43154-020-00029-y

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A Review of Empirical Evidence on Different Uncanny Valley Hypotheses: Support for Perceptual Mismatch as one Road to the Valley of Eeriness. *Frontiers in Psychology* 6:390. https://doi.org/10.3389/fpsyg.2015.00390

Lee, H. R., Šabanović, S., & Stolterman, E. (2016). How Humanlike Should a Social Robot Be: A User-Centered Exploration. *AAAI Spring Symposia Series*, North America. https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12751

Li, J. (2015). The Benefit of Being Physically Present: A Survey of Experimental Works Comparing Copresent Robots, Telepresent Robots and Virtual Agents. *International Journal of Human-Computer Studies*, 77, 23-37. http://dx.doi.org/10.1016/j.ijhcs.2015.01.001

Malle, B.F., Fischer, K., Young, J.E., Moon, A.J., & Colling, E. (2020). Trust and the Discrepancy Between Expectations and Actual Capabilities of Social Robots. In D. Zhang and B. Wei (Eds.), *Human-robot interaction: Control, analysis, and design*. Cambridge Scholars Publishing: New York

Mangan, B. (2015). The Uncanny Valley as Fringe Experience. *Interactions Studies* 16(2), 193-199. 10.1075/is.16.2.05man

Maxwell, J. A. (1996). *Qualitative Research Design: An Interactive Approach* (Applied Social Research Methods Series, 42). Sage Publications: Thousand Oaks, California

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7, 33–35. Transl. K. F. MacDorman and N. Kageki (2012), *IEEE Robotics & Automatization Magazine* 19, 98–100. 10.1109/MRA.2012.2192811

Moubayed, S.A., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A Back-projected Human-like Robot Head for Multiparty Human-Machine Interaction. *Cognitive Behavioural Systems*, 7403, 114-130. 10.1007/978-3-642-34584-5_9

Noldus. (2020). *FaceReader* (Version 8.1) [Computer software].

Noldus. (n.d.). *FaceReader Resources.* Retrieved from https://www.noldus.com/facereader/resources

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of Negative Attitudes Toward Robots. *Interaction Studies*, 7(3), 437-454. 10.1075/is.7.3.14nom

Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition.* Basic Books: New York

Paetzel, M., Peters, C., Nyström, I., & Castellano, G. (2016). *Preliminary Results from Using a Back-Projected Robot Head in Uncanny Valley Research*. In Proceedings of the 25th International Symposium on Robot and Human Interactive Communication, RO-MAN 2016, Piscataway, NJ: IEEE, pp. 944-945. http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-308428

Paetzel-Prüsmann, M. (2020). *The Novelty in the Uncanny: Designing Interactions to Change First Impressions* (PhD dissertation, Uppsala University, Uppsala, Sweden). http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-418921

Pollick, F. E. (2009). In Search of the Uncanny Valley. In: P. Daras, O.M. Ibarra (Eds.), *User Centric Media, UCMEDIA 2009*. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 40. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12630-7_8

Rosenthal-von der Pütten, A., & Weiss, A. (2015). The Uncanny Valley Phenomenon - Does it Affect All of Us? *Interaction Studies*, 16(2), 206-214. 10.1075/is.16.2.07ros

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, 110(1), 145-172. 10.1037/0033-295x.110.1.145

Schramm L. T., Dufault D., & Young J. E. (2020). *Warning: This Robot is Not What it Seems! Exploring Expectation Discrepancy Resulting from Robot Design*. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, HRI '20, Cambridge, UK, pp. 439–441. https://doi.org/10.1145/3371382.3378280

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The Uncanny Valley: Existence and Explanations. *Review of General Psychology,* 19(4), 393-407. https://doi.org/10.1037/gpr0000056

Wolfram Alpha LLC. (2009). *Wolfram|Alpha* (n/a) [Knowledgebase]. http://www.wolframalpha.com/input/?i=2%2B2

Wood, C., & Dillenbeck, S. (2021). *What are Social Robots? An Introduction to the Furhat Robot.* Furhat Robotics. Retrieved from https://furhatrobotics.com/blog/what-are-social-robots/

Złotowski, J.A., Sumioka, H., Nishio, S., Glas, D.F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the Uncanny Valley: the Influence of Repeated Interactions and a Robot's Attitude on its Perception. *Frontiers in Psychology*, 6(883). https://doi.org/10.3389/fpsyg.2015.00883

Appendix 1: Notice of interest

Intresseanmälan

Vi söker deltagare till en robotinteraktionsstudie som utgör en del av vårt examensarbete på Kandidatprogrammet i kognitionsvetenskap vid Göteborgs Universitet.

Studien kommer att genomföras i slutet av april/början av maj, och innebär fysiskt deltagande i Patricia-huset på Campus Lindholmen. Deltagandet är individuellt, och vi kommer att vidta särskilda åtgärder för att säkerställa att miljön är säker för alla deltagare (bland annat genom rengöring av alla kontaktytor mellan varje deltagare, samt användning av munskydd).

Om du är intresserad av att delta är du välkommen att fylla i formuläret nedan, så återkommer vi med mer information om aktuella datum. Observera att deltagandet i denna studie förutsätter att du bor i Göteborg, och att du kan tala både svenska och engelska flytande. Inga andra förkunskaper krävs.

Vänligen,
Anna Johansson och Elin Örnberg

*Required

För- och efternamn *

Your answer

E-mail *

Your answer

Telefonnummer

Your answer

Utbildning - högst avslutade studienivå *

Your answer

Nuvarande sysselsättning *

Your answer

Har du har någon erfarenhet av interaktion med artificiella agenter? Kryssa i nedan (flerval)

- ☐ Nej, jag har ingen erfarenhet av att interagera med artificiella agenter
- ☐ Ja, jag har erfarenhet av röstassistenter (ex. Siri, Alexa, Google Home)
- ☐ Ja, jag har erfarenhet av hushållsrobotar (ex. dammsugare, gräsklippare)
- ☐ Ja, jag har erfarenhet av industrirobotar (ex. tillverkningsrobotar)
- ☐ Ja, jag har erfarenhet av sociala robotar (ex. Pepper, Nao, Furhat)
- ☐ Other:

Kommentarer

Your answer

Submit

Appendix 2: Survey

Part 1 of the survey: consent

Interaktionsstudie

Denna enkät utgör den första delen av vår robotsinteraktionsstudie som utförs som en del av vårt examensarbete.

För att delta i studien behöver du samtycka till ditt medverkande och användningen av din data för denna studie. Som deltagare har du själv rätt att bestämma över din medverkan, och du kan om du så önskar välja att avbryta ditt deltagande under studien. Alla uppgifter som du lämnar kommer att hanteras med konfidentialitet, och den information som sedan används i examensarbetet kommer att anonymiseras. Data som samlas in under denna studie kommer enbart att användas i detta examensarbete. Denna studie sker i samarbete med Volvo Cars, vilket innebär att studiens resultat kommer att förmedlas till dem.

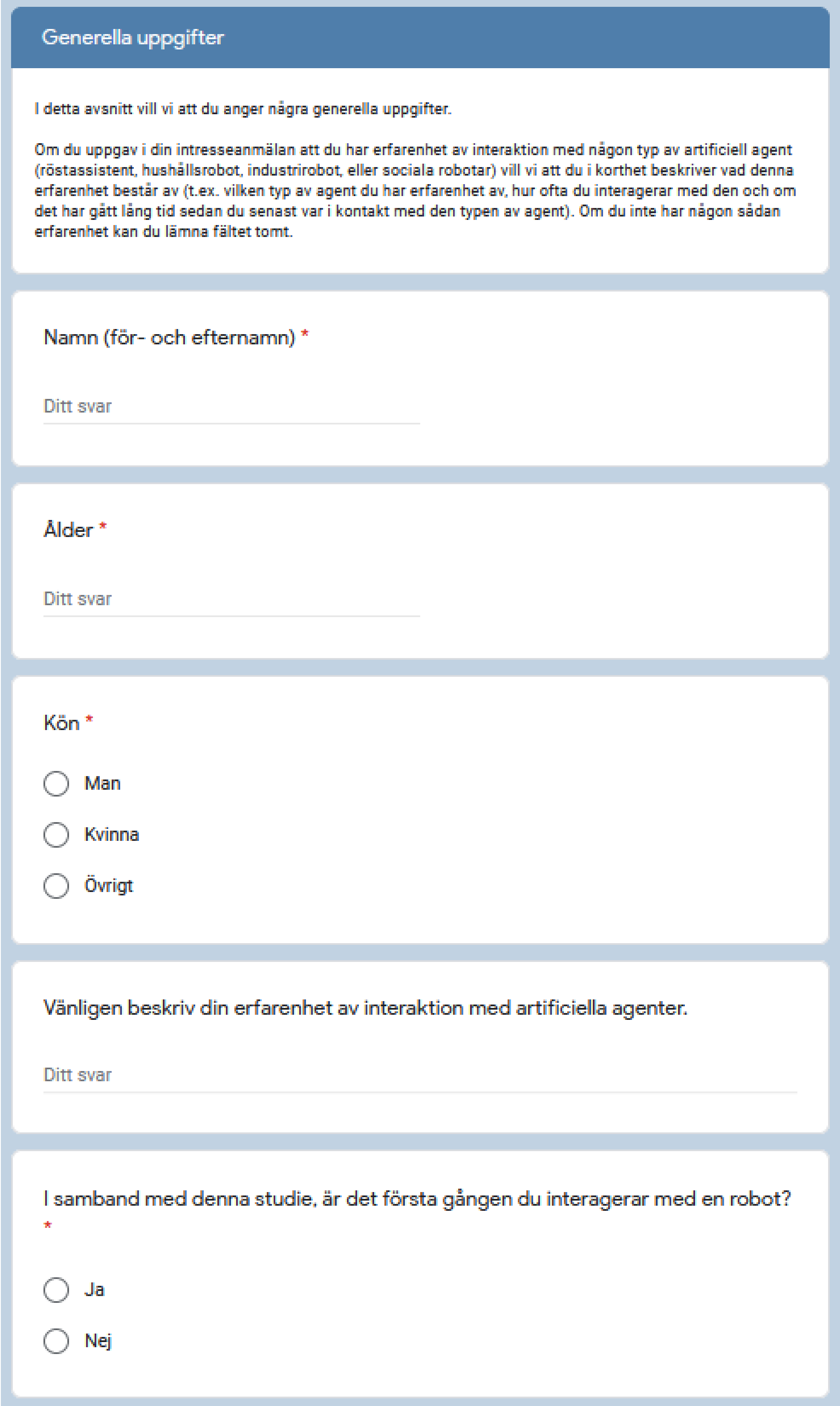
Observera att ditt deltagande i interaktionen kommer att filmas, och vi kommer att göra en ljudupptagning i samband med en avslutande intervju.

*Obligatorisk

Samtycker du till ditt medverkande och användningen av din data för denna studie? *

☐ Ja

Part 2 of the survey: general information



Generella uppgifter

I detta avsnitt vill vi att du anger några generella uppgifter.

Om du uppgav i din intresseanmälan att du har erfarenhet av interaktion med någon typ av artificiell agent (röstassistent, hushållsrobot, industrirobot, eller sociala robotar) vill vi att du i korthet beskriver vad denna erfarenhet består av (t.ex. vilken typ av agent du har erfarenhet av, hur ofta du interagerar med den och om det har gått lång tid sedan du senast var i kontakt med den typen av agent). Om du inte har någon sådan erfarenhet kan du lämna fältet tomt.

Namn (för- och efternamn) *

Ditt svar

Ålder *

Ditt svar

Kön *

- Man
- Kvinna
- Övrigt

Vänligen beskriv din erfarenhet av interaktion med artificiella agenter.

Ditt svar

I samband med denna studie, är det första gången du interagerar med en robot? *

- Ja
- Nej

Part 3 of the survey: NARS

1. **S2** Jag skulle känna obehag om robotar verkligen hade känslor / I would feel uneasy if robots really had emotions
2. **S2** Någonting dåligt skulle kunna hända om robotar utvecklades till levande varelser / Something bad might happen if robots developed into living beings
3. **S3** Jag skulle känna mig avslappnad i en konversation med robotar / I would feel relaxed talking with robots
4. **S1** Jag skulle känna obehag om jag gavs ett jobb där jag behövde använda robotar / I would feel uneasy if I was given a job where I had to use robots
5. **S3** Om robotar hade känslor skulle jag kunna bli vän med dem / If robots had emotions, I would be able to make friends with them
6. **S3** Jag känner mig bekväm tillsammans med robotar som har känslor / I feel comforted being with robots that have emotions
7. **S1** Ordet "robot" betyder ingenting för mig / The word ‘‘robot’’ means nothing to me
8. **S1** Jag skulle känna mig nervös om jag manövrerade en robot framför andra människor / I would feel nervous operating a robot in front of other people
9. **S1** Jag skulle avsky tanken på att robotar eller artificiella intelligenser gjorde bedömningar om saker / I would hate the idea that robots or artificial intelligences were making judgments about things
10. **S1** Jag skulle känna mig väldigt nervös bara av att stå framför en robot / I would feel very nervous just standing in front of a robot
11. **S2** Jag känner att om jag är för beroende av en robot kan någonting dåligt hända / I feel that if I depend on robots too much, something bad might happen
12. **S1** Jag skulle känna mig paranoid av att tala med en robot / I would feel paranoid talking with a robot
13. **S2** Jag är orolig för att robotar skulle ha ett dåligt inflytande på barn / I am concerned that robots would be a bad influence on children
14. **S2** Jag känner att samhället i framtiden kommer att domineras av robotar / I feel that in the future society will be dominated by robots

Figure 8: The NARS question items (Nomura et al., 2006) used in the pre-interaction survey, displayed both in Swedish and English.
S1 = subscale 1; S2 = subscale 2; S3 = subscale 3

Appendix 3: Interaction flashcards

96 $ in swedish krona	Example question: How much is 96 dollars in swedish krona?
Brother to Angelina Jolie	Example question: Who is Angelina Jolie's brother? Who is the brother of Angelina Jolie?
Size of New York Chinatown	Example question: What is the size of New York's Chinatown? How large is Chinatown in New York?
275 / 32 =	Example question: What is 275 divided by 32? How much is 275 divided by 32?
Ask Furhat for a joke	Example question: Can you tell me a joke? Do you have a joke for me?

Age of Steven Spielberg	Example question: How old is Steven Spielberg
Sister to Albert Einstein	Example question: Who is Albert Einstein's sister? Who is the sister of Albert Einstein?
65 x 52 =	Example question: What is 65 times 52? How much is 65 times 52?
Capital of Zimbabwe	Example question: What is the capital of Zimbabwe?
16th President of United States	Example question: Who was the sixteenth president of the United States?

VOLVO ELECTRIFICATION Charging	*Example question:* How do I charge my car? Can you tell me about charging?
VOLVO ELECTRIFICATION Battery life span	*Example question:* What is the estimated battery life span? How long can a battery last?
VOLVO ELECTRIFICATION Drive distance with 50 % battery	*Example question:* How far can a drive with 50% battery? What is the driving distance with a half charged battery?
VOLVO ELECTRIFICATION Preconditioning the car	*Example question:* What is preconditioning? Can you tell me about preconditioning?
VOLVO ELECTRIFICATION Long term storing	*Example question:* How can I store my car long term? What about long term storing?

VOLVO SAFETY Pilot Assist function	*Example question:* What is Pilot Assist? What does Pilot Assist?
VOLVO SAFETY Electric cars safety	*Example question:* What about safety within electric cars? Are electric cars safe?
VOLVO SAFETY Lane keeping aid	*Example question:* What is lane keeping aid? Tell me about Lane keeping aid?
VOLVO SAFETY Safety for children	*Example question:* Can you tell me about safety for children? What about child safety?
VOLVO SAFETY Connected safety	*Example question:* What is connected safety? How does connected safety work?

Appendix 4: Interview questions

First impression interview

- Hur känns det?
- Vad är ditt första intryck av Furhat?
    - Hur skulle du beskriva Furhats utseende?
    - Baserat på vad du ser, vad skulle du säga att Furhat kan göra?
    - Finns det några förmågor eller egenskaper du skulle säga att roboten har?
- Vad tycker du om Furhat?
- Vad får Furhat dig att känna?

Post-interaction interview

- Du angav i enkäten att du har tidigare erfarenhet av interaktion med [det som angetts i enkäten]. Hur skulle du jämföra den erfarenheten från denna interaktion?
- Hur kändes det att träffa Furhat?
- Vad hade du för förväntningar när du kom hit?
    - Vad fick du för förväntningar när du väl såg Furhat?
    - Var det något i situationen som gick emot eller överensstämde med dina förväntningar? Vad?
- Vad fick du för upplevelse när du interagerade med Furhat?
    - Vad var det som fick dig att uppleva det?
- Kan du beskriva vad som hände här [visar klipp från videoinspelning]?
- Vad fick du för tankar medan du interagerade med Furhat?
- Hur skulle du beskriva Furhat efter att ha interagerat med den lite längre?
    - Hur förändrades ditt intryck av Furhat?
- Vad tyckte du om att interagera med Furhat?
    - Är detta ett gränssnitt du kan se dig själv använda i framtiden?
        - I vilket situation kan du se dig använda detta gränssnitt?
- Hur kändes det att sitta och prata med Furhat medan vi var i rummet?
- Vad tyckte du om uppgiften du utförde?
    - Vad tycker du om idén att använda den här typen av teknik för att få information?
- Har du några andra kommentarer eller synpunkter du vill dela med dig av?

Appendix 5: Themes and subcategories

Feelings
Theme includes participants explicitly expressing feelings categorised using Russel's core affect model (2003) generating four subcategories: high-arousal, positive-valence (*AP*); low-arousal, positive-valence (*DP*); high-arousal, negative-valence (*AU*); and low-arousal, negative-valence (*DU*). Theme content intends to categorize data for Q1 (*what feelings do participants experience with a human-like agent?*)

Interaction Experience
To capture the width of the interaction, citations were collected and divided into the following subcategories: Furhat's *Appearance*, *Abilities*, *Presence*, *Expectations*, *Problems*, and participants' *Behaviour*. The theme's content intends to categorize data for Q2 (*how are human-like agents perceived?*)

Dialogue
This theme summarizes the conversational content where the subcategories included were: *Content*, *Context*, *Conversation* and *Language*. Theme content intends to categorize data for Q3 (*how do participants experience information retrieval through interacting with a social, human-like agent?*)

Appendix 6: Result/participant

	Pre-interaction survey			**FaceReader Interaction** (mean valence)				**Interviews** (mentioned subtopics)		
Participant	**S1**	**S2**	**S3**	**Presen-tation**	**Phase 1**	**Phase 2+3**	**Total**	**Feelings**	**Interaction Experience**	**Dialogue**
P01-VirFur	12	15	6	0.345	0.010	-0.082	-0.009	AP DP AU	Abilities Appearance Presence	Content Context Conversation Language
P02-PhyFur	6	8	5	0.703	0.328	0.167	0.185	AP DP AU	Abilities Appearance Behaviour Expectation Presence Problem	Content Context Conversation
P03-VirFur	11	12	10	0.052	0.031	-0.040	-0.001	AP DP AU	Abilities Appearance Behaviour Expectation Presence Problem	Content Context
P04-PhyFur	7	13	7	0.585	0.384	0.127	0.247	AP AU	Abilities Appearance Behaviour Expectation Presence Problem	Content Context Conversation
P05-VirFur	8	9	4	0.694	0.444	0.136	0.294	AP DP	Abilities Appearance Behaviour Expectation Presence Problem	Content Context Language
P06-PhyFur	11	18	7	n/a	n/a	n/a	n/a	Neutral AP DP	Abilities Appearance Behaviour Expectation Presence	Content Context Conversation
P07-VirFur	15	18	14	0.284	-0.190	-0.194	-0.150	AP DP DU AU	Abilities Appearance Behaviour Expectations Presence	Content Conversation
P08-PhyFur	16	19	13	0.333	0.151	0.061	0.109	AP AU	Abilities Appearance Behaviour Presence Problem	Context Conversation Language

Table 5: Overview participants in NARS score, FaceReader and interview themes with subtopics.