The Persistence of First Impressions: The Effect of Repeated Interactions on the Perception of a Social Robot

Maike Paetzel
maike.paetzel@it.uu.se
Social Robotics Lab
Uppsala University, Sweden

Giulia Perugia
giulia.perugia@it.uu.se
Social Robotics Lab
Uppsala University, Sweden

Ginevra Castellano
ginevra.castellano@it.uu.se
Social Robotics Lab
Uppsala University, Sweden

ABSTRACT

Numerous studies in social psychology have shown that familiarization across repeated interactions improves people's perception of the other. If and how these findings relate to human-robot interaction (HRI) is not well understood, even though such knowledge is crucial when pursuing long-term interactions. In our work, we investigate the persistence of first impressions by asking 49 participants to play a geography game with a robot. We measure how their perception of the robot changes over three sessions with three to ten days of zero exposure in between. Our results show that different perceptual dimensions stabilize within different time frames, with the robot's competence being the fastest to stabilize and perceived threat the most fluctuating over time. We also found evidence that perceptual differences between robots with varying levels of humanlikeness persist across repeated interactions. This study has important implications for HRI design as it sheds new light on the influence of robots' embodiment and interaction abilities. Moreover, it also impacts HRI theory as it presents novel findings contributing to research on the uncanny valley and robot perception in general.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *Natural language interfaces*; • **Computer systems organization** → **Robotics**; • **Computing methodologies** → *Intelligent agents*.

KEYWORDS

Robot Perception, First Impressions, Embodiment, Uncanny Valley

ACM Reference Format:
Maike Paetzel, Giulia Perugia, and Ginevra Castellano. 2020. The Persistence of First Impressions: The Effect of Repeated Interactions on the Perception of a Social Robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3319502.3374786

1 INTRODUCTION

When we meet someone for the first time, a few milliseconds are sufficient to decide whether that person constitutes a threat, is to be trusted or even a potential mate [4][6][57], and a few minutes of conversation are enough to set the ground for the future relationship [27][56]. The mere repeated exposure to a person increases our positive perception [59] and, unless we develop interpersonal conflicts, become bored or disgusted, the more we become familiar with someone, the more our relationship strengthens [21].

When we meet a robot for the first time, we are able to develop a coherent mental model of it within two minutes [47]. The robot's appearance, but also previous exposure to robots in movies, shape our expectations towards its interactive capabilities [15][17][24][25]. In previous work, we highlighted the value of allowing people to interact with a robot before measuring their perception of it [44]. We found that even an interaction of a few minutes improved people's perception of a robot on dimensions like anthropomorphism or likability. Other researchers investigated the effect of interaction time on the perception of a robot and found the perceived warmth of a robot to decrease after giving people the possibility to interact with it for several minutes [10], but also that positively toned interactions could improve a robot's likability over time [60].

While some long-term studies have addressed changes in people's perception of a robot over time, they were mostly conducted in the wild without any control over the number and type of interactions participants had with the robot [16][17][18]. In contrast to these studies, *our work investigates repeated interactions under very controlled conditions, thus allowing to answer the question of whether first impressions could predict long-term perceptions of robots.* This knowledge is very relevant to the Human-Robot Interaction (HRI) community, since most researchers envision robots to be eventually

Figure 1: A participant playing a geographic literacy game with the blended embodiment Furhat.

HRI '20, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-6746-2/20/03…$15.00
https://doi.org/10.1145/3319502.3374786

present and repeatedly used in our everyday life. If findings from social psychology extend to social robots, familiarization through repeated interactions alone should be sufficient to improve people's perception of a robot and strengthen their relationship with it. This means that conclusions we draw from research on first impressions of robots may not be transferable to long-term interactions.

In this paper, we present a study aimed at investigating the effect of repeated interactions with three to ten days of zero exposure in between on the perception of a robot. We designed a collaborative human-robot game scenario focused on promoting geographic literacy (cf. Fig. 1). Participants were asked to play the game with the blended embodiment Furhat [3] in three different sessions. As a robot's anthropomorphism is a major influencing factor for the formation of a mental model, we varied the humanlikeness of Furhat between subjects by creating three different facial textures representing different levels of humanlikeness. Measuring participants' perceived *anthropomorphism*, *competence*, *likability*, *warmth*, *threat* and *discomfort* with the robot at different times throughout the three interaction sessions allowed us to understand the effect of (i) game interaction and (ii) repeated interactions on people's perception of a robot, and the effect of (iii) zero exposure periods on participants' ability to recall such perception over time.

2 RELATED WORK

First Impressions and Their Persistence

In *zero acquaintance contexts*, situations in which we are confronted with previously unknown persons, we make judgements within a few milliseconds. We need just about 39 ms of exposure to decide if another person is a threat [6] and 100 ms to judge someone's attractiveness, likability, trustworthiness, competence, and aggressiveness [57]. Even personality traits can be inferred from appearance [40]. These rapid judgements are based on "shared stereotypes about particular physical appearance characteristics" [4]. However, someone's personality can also be judged from listening to short voice samples [36].

According to Wood, *first impressions* do not only refer to millisecond-long encounters, but to any impression "that is made quickly, usually within 5 minutes from meeting someone for the first time" [58]. Another term used for such multimodal encounters is *thin slice impressions* [5]. Ambady et al. showed how accurate people can be in judging someone's internal state, personality, interaction motives and social relations from first impressions [5]. In general, traits that can be inferred from behavior, like extroversion, are assessed more accurately than less observable traits like openness [58]. The accuracy of judgements generally increases with repeated exposures. However, positive perceptions usually require longer exposure times, while negative ones can be accurately judged after 5 s of exposure [12].

The (Dis)Advantage of Familiarization

Overall, first impressions have shown to correctly predict the development of interpersonal relationships [27][56]. However, the mental models derived from first impressions can change over time if novel or unexpected information is received [5]. In social psychology, the question of whether familiarizing with someone through repeated interactions will lead to negative or positive perceptions is still unanswered. In 1968, Zajonc introduced the *mere exposure effect*, which suggests that repeated exposures to a stimulus will enhance people's attitude towards it [59]. Twenty years later, a meta-analysis by Bornstein found overwhelming evidence for a positive exposure-affect relationship [11]. Empirical evidence suggests that affect is usually at its highest after 10 to 20 exposures, and that a delay between exposures can result in a stronger raise of affective rating [11]. One explanation of this effect is reduced uncertainty which in turn leads to positive affect [30].

While the evidence regarding the exposure effect is rarely debated, skepticism is high when it comes to the influence of repeated interactions. Multiple empirical and theoretical evidence point towards an increase in affect when becoming familiar with someone. Since most social encounters are positive, an increased number of social encounters should lead to higher emotional reward (conditioning) and more possibilities to gain favorable impressions [50]. In addition, repeated interactions reduce uncertainty and increase perceptual fluidity [50], which is generally linked with an increase in affect [49]. Finkel et al. suggests, however, that familiarity undermines attraction if the target person becomes unappealing, the context more competitive, if boredom increases, or interpersonal conflict arises [21]. Familiarization does not only reduce uncertainty, but also leads to *habituation*. Dijksterhuis and Smith found that affective reactions towards extreme stimuli decreased with repeated encounters regardless of the positive or negative nature of the stimulus [19]. In the field of Human-Computer Interaction, this habituation is known as the *novelty effect*, a decrease in the engagement with a stimulus after its initial novelty has worn off.

Liking and Engaging with Robots

Human-Robot relationships are complex and research on the likability and engagement with robots is not conclusive yet. Appearance is one factor that has been identified as crucial to promote likability and engagement of a robot and to predict people's approach to talking to and working with it [24][34]. In 1970, Masahiro Mori proposed the theory that people feel more familiar with humanlike robots. However, once a robot reaches a critical point of extreme but not complete humanlikeness, people develop uncanny feelings [38]. In their review of the so-called *uncanny valley effect*, Kätsyri et al. found most evidence for the *perceptual mismatch theory* [28]. According to this theory, negative perceptions increase if inconsistencies in the robot's level of anthropomorphism exist. Bartneck et al., however, pointed out that the perception of the uncanny valley might depend on more dimensions than humanlikeness alone [8].

While creating an initial engagement with robots is often easy due to the novelty effect [54], keeping engagement over a longer period of time has shown to be difficult, mostly because users become bored by the robot's narrow or repetitive interaction capabilities [33]. Even if embodiment cues are consistent, an anthropomorphic appearance may lead to deception or even rejection of the robot, as it sets high expectations on the robot's capabilities, which might not be met in the actual interaction [15][17]. Initial expectations towards robots are often "science-fiction-movie-character-like, and do not resemble those developed in laboratories" [25]. Horstmann and Krämer thus highlight the importance of contact with robots to reduce the fear elicited by "bad" fictional representations [26]. When it comes to designing such contacts, it has been shown that the number of modalities of a virtual agent can positively impact its perceived believability, warmth and competence [41]. Mattar and Wachsmuth suggested that the inclusion of personal topics in human-agent conversations might have a positive impact on the construction of interpersonal relationships [35], although self-disclosure is not sufficient to increase a robot's likability [20].

Familiarization with Robots

In comparison to the amount of research dedicated to first impressions of robots, long-term investigations are still comparably sparse. If existent, research has mostly focused on relationship-maintaining behaviors and adaptation when it comes to task performance and engagement in long-term scenarios (see [33] for a survey). For example, emotional [39] and empathetic behavior [32], as well as personalization [31] and adaptation [1] have been shown to have a positive influence on people's relationship with a robot. Lack of enjoyment and utility or concerns regarding the robot's intelligence were found to be important factors behind people's refusal to use a robot after a short period of time [18], whereas perceived adaptability and sociability were crucial to determine continuous use and adoption [18]. Kertesz and Turunen reported lack of interactive capabilities and autonomy to be important motives to discontinue using a self-purchased AIBO robot [29].

Research studying the influence of repeated interactions on human-robot relationship comes to conflicting conclusions. While some found an indication for decreased engagement due to familiarization [52], others reported an increase in social engagement [2] and time spent with the robot [22]. De Graaf et al. found indications for what they reported to be a mere exposure effect with increased rating of users interacting with personal robots for several months [16][17]. While the robot and its behavior remained constant, due to the nature of the study, judging the impact of the number and type of interaction with the robot on people's attitude is difficult.

To our knowledge, only a few empirical studies have specifically investigated whether the first impression of a robot persist after longer interactions. Bergmann et al. found that a robot's warmth decreased between a first impression (15 s) and a longer multimodal interaction with it [10]. The mental model of the robot's competence, however, was persistent. Stubbs et al. reported employees' perception of the anthropomorphism of a robot installed in their museum to increase over several months of exposure, while the perceived competence fluctuated over time [55]. Złotowski et al. did not find evidence for a mere exposure effect in HRI, but suggested that "a positively toned interaction" is required to increase likability. They also reported that uncanny robots could not recover from initial uncanny feelings even after a positively toned interaction [60]. Haring et al. [23] found people's perception of a robot to change from their first impression to after their first interaction, but to be persistent afterwards. In a previous study, we investigated whether first impressions persisted when people were gradually exposed to a robot's multimodal interaction capabilities and found both an increase in the positive and a decrease in the negative perception of the robot [44]. With an exception of Stubbs et al., all previously discussed studies investigated repeated exposure or interaction within the time frame of a day. This paper, on the contrary, is focused on repeated interactions over a longer period of time with multiple days of zero exposure in between.

3 RESEARCH QUESTIONS

Following up on our previous research, the study presented in this paper is aimed at better understanding how a collaborative game interaction could influence initial perceptions of a robot.

(RQ1) How does a collaborative game interaction influence people's first impressions of a social robot?

To answer this question, we measured participants' perceptions after a two-minute social chat and after their first game interaction.

In addition, we aimed at advancing the related work by studying the effect of repeated interactions on the perception of a robot.

(RQ2) How do repeated interactions influence the perception of a social robot?

With respect to previous work, in this work we used true repeated interactions with multiple days of zero exposure in between, instead of multiple measurements within one interaction session [10][23][44][60]. We also kept the type, structure and length of the interaction constant across three sessions to reduce the confounding factors observed in other long-term investigations.

Our experimental design also allows to investigate how the perception of a social robot changes over periods of zero exposure.

(RQ3) How accurate are people in recalling their perception of a social robot over multiple days of zero exposure?

We investigated this question by asking people to recall their perception of the robot after multiple days of zero exposure before showing them the robot again.

As discussed in Section 2, there are many factors influencing people's perception of robots. As anthropomorphism is one of these, we manipulated it to understand if it influenced people's perceptions.

(RQ4) Does the robot's level of anthropomorphism influence people's perception of it in repeated interactions?

When it comes to anthropomorphism, interesting types of robots are those that carry mismatching cues in their humanlikeness since these robots usually elicit uncanny feelings. We designed one of the facial textures of Furhat to be a morph between a humanlike and a mechanical robot, thus embodying these mismatching cues and likely eliciting initial negative feelings.

(RQ5) Can a robot with mismatching cues recover from initial negative impressions through repeated interactions?

Following findings from social psychology, we expected repeated interactions to increase the perceptual fluidity of the morph robot and thus positively influence its perception over time.

4 METHODOLOGY

The study followed a 3x3 mixed experimental design with *embodiment* (humanlike, mechanical, and morph) as between-subject and *repeated interactions* (session S1, S2 and S3) as within-subject factor. Between each session, participants had three to ten days of zero exposure (S1-S2: $M = 6.63$, $SD = 1.52$; S2-S3: $M = 6.55$, $SD = 1.11$).

4.1 Participants

Sixty students were recruited from an international Master's course at Uppsala University, Sweden, to participate in the experiment. Two participants were excluded because the system had technical failures, four as they had interacted with the robot before, two since they suspected the agent to be remote-controlled, other two as they interacted with the robot only once, and one because the break between S1 and S2 was longer than 10 days. The results reported in this paper were obtained from the remaining 49 participants. All 49 participants (M= 36; F= 13) interacted with the robot at least twice (Human: $N = 16$, Mechanical: $N = 17$, Morph: $N = 16$) and 40 of them completed all three interaction sessions (Human: $N = 14$, Mechanical: $N = 13$, Morph: $N = 13$). Participants were aged between 19 and 50 years ($M = 24:78$, $SD = 4:70$) and all were or had been enrolled in a Computer Science or related program. The study was approved by the regional ethics board, and participants were compensated with credits for their time.

4.2 Scenario

The collaborative game scenario was designed to (a) enable a *fun and engaging experience* with a social robot, (b) create an *in-group identity* between the human and the robot, and (c) elicit mutual *disclosure of personal information.* In each session, the majority of time was devoted to an interaction involving a cooperative dialogue-based game designed to improve geographic literacy (see [45] for a detailed description). Within the game, the participant took the role of the *tutor* and was tasked with helping the robot to correctly locate *target countries* on a world map by verbally describing them. The robotic agent acted as a *learner* with extremely limited geographic knowledge. The goal of the human-robot team was to score as many points as possible in the given time of ten minutes.

Before and after the game, each participant had a two-minute social chat with the robot. The social chat was identical for all participants, and the order of topics was predefined. The chat could revolve around the team's performance at the game, but the robot used it also to ask participants questions, for instance, which countries they had visited in the past, or where they came from. While the game interaction was the same across sessions (apart from different target countries), the topic of the pre- and post-game interaction changed. The robot remembered facts from previous sessions, and asked follow-up questions.

4.3 Embodiment

In this work, we used the robotic head Furhat [3]. Furhat is equipped with a rigid mask of a male face on which a facial texture is projected from within. Unlike traditional robots, this platform allows the facial texture to be changed while keeping the overall embodiment fixed. It still differs from virtual agents due to its physical torso and the two degrees of freedom that allow it to orientate its head in the three-dimensional space.

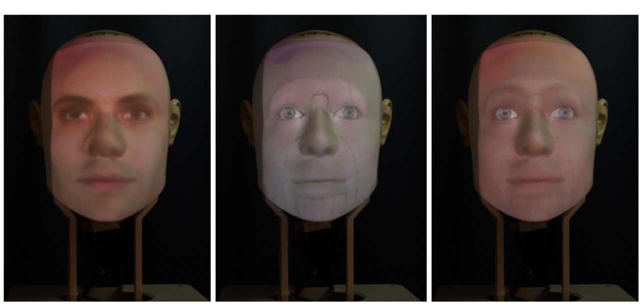

Figure 2: The Furhat robot with the humanlike (left), mechanical (center) and morph (right) facial texture applied.

Following the approach by McDonnell et al. [37], we created three different facial textures with varying degree of humanlikeness using the FaceGen modeller and the Paint.NET digital photo editing package. The texture of the *humanlike* face was created from a photograph of a real human (cf. Fig. 2 left), while the *mechanical* face was based on the picture of a robot (cf. Fig. 2 center). The third facial texture was created as a *morph* between the humanlike and the mechanical face (cf. Fig. 2 right) in which both humanlike and mechanical features were still visible, but more subtle [44].

During the experiment, the robot was remotely controlled by an operator [51] on an interface (cf. Fig. 3 left-top) that consisted of a world map from which to select the robot's guess, and a number of buttons to control the robot's speech, gaze (towards the user or specific regions of the map) and facial expressions (smiling and frowning). The operator was tasked with replacing the Natural Language Understanding unit of the system. Hence, s/he merely categorized the responses of the participants. In the game, the operator had specific rules regarding the agent's previous knowledge, which information about a country was sufficient to make a guess, and how to react to scoring or not scoring points, among others. The initial set of responses for the interface was gained from the analysis of human game partners playing the game, and then expanded after analyzing 50 online human-operator game interactions (see [45] for details). The online study also served to train the operator responsible for all interaction sessions. During the game interaction, participants were led to believe that the robot was autonomous. The debriefing on the remote-controlled nature of the interaction was performed as soon as the entire experiment was concluded.

4.4 Measures

Before their first encounter with the robot, participants filled out questionnaire (Q1) covering their demographics, personality traits [48], and negative attitude towards robots (NARS) [42].

The second questionnaire (Q2) was designed to measure participants' perception of the robot and included the sub-scale *anthropomorphism* from the Godspeed questionnaire (5 items on a 5 point Likert scale, $\alpha = :91$) [9], two 5-item sub-scales to measure *likability* and *perceived threat* (5 point Likert scale, likability: $\alpha = :83$, perceived threat $\alpha = :89$) [53], and the Robotic Social Attributes Scale covering *warmth, competence* and *discomfort* (RoSoAS; 18 items on a 7 point Likert scale; warmth: $\alpha = :92$; competence: $\alpha = :95$; discomfort: $\alpha = :90$) [13].

The questionnaire (Q3) was presented to participants at the end of each session. It included the same sub-scales of Q2, questions to evaluate the *satisfaction* with the game from the author's previous work (e.g., "In the end I felt satisfied with our score" and "I talked to the robot the way I normally talk to another person"; 5 point Likert scale) [46], and the sub-scale *involvement* of the User Engagement Questionnaire (UES; 3 items on a 5 point-Likert scale, $\alpha = :71$) [43], which was completed for the robot and the game.

As the study was part of a larger project, we made further recordings of participants with two RGB webcams, a Kinect, a RealSense

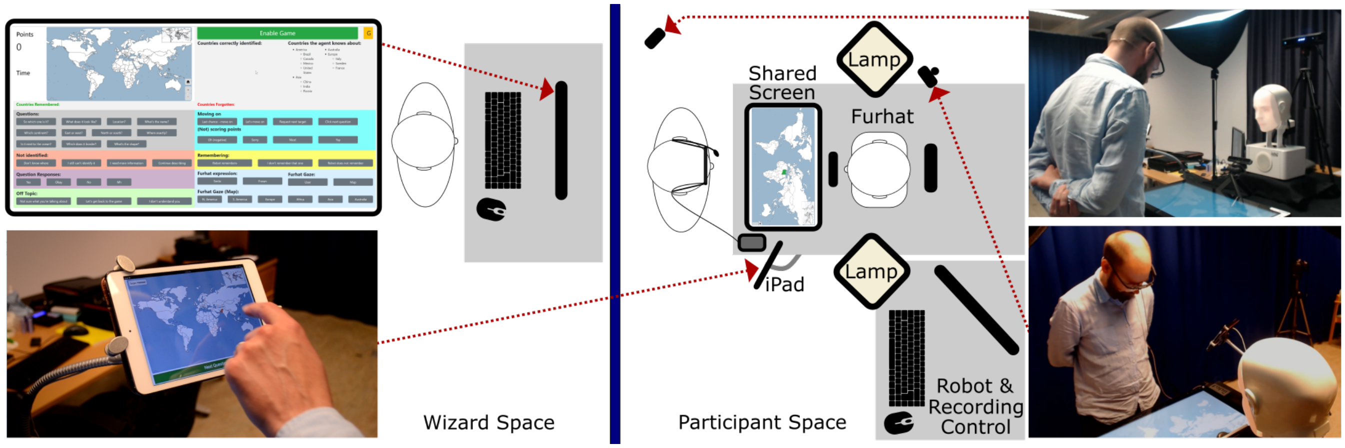

Figure 3: Schematics of the experimental setup during the geographic game, including the operator interface (top left), the Director screen on the iPad (bottom left) and the two RGB cameras' angles of view (right).

camera, Tobii eyetracking glasses, and a close-range Sennheiser microphone. As these recordings were used to answer other research questions, their analysis is not presented in this paper.

4.5 Experimental Setup and Procedure

In the beginning of the first session, participants were asked to give their informed written consent for participation and read the game rules. They were then brought to the game table and asked to fill out Q1 while the robot was still covered by a blanket. As depicted in Fig. 3, Furhat was placed on a table facing the participant who was standing on the other side of a shared touchscreen. The iPad on which the questionnaires were filled out and the game information was displayed was placed to the participant's right. Once Q1 was completed, the experimenter started the recordings and uncovered the robot. At this point, participants had a two-minutes social chat with the robot, after which they filled out Q2. The robot was coupled with the questionnaire system, so that it would autonomously enable the game on completion of Q2. After the ten-minute game, the robot started a two-minute post-game social chat with the participant before asking them to fill out Q3.

In S2 and S3, participants were directed to the game table right away. Before the robot was uncovered, they filled out Q2 based on what they could recall from their previous interaction. After uncovering the robot, participants had their pre-game interaction (the social chat) followed by the game interaction and the post-game chat with the robot without a break. Again, Q3 was filled out once the entire interaction session had finished.

5 RESULTS

To understand whether the game interaction changed participants' first impression of the robot (RQ1), we ran a factorial ANOVA (2x3) with *game interaction* as within-subject factor (Q2 and Q3 at S1) and level of humanlikeness as between-subject factor (human, morph, and mechanical robot). Furthermore, to disclose changes in the perception of the robot across *repeated interactions* (RQ2), especially with reference to embodiment (RQ4/5), we performed a factorial ANOVA (3x3) with repeated interactions as within-subject factor (Q3 at S1, S2, S3) and level of humanlikeness as between-subject factor. To comprehend whether participants correctly recalled their previous perception of the robot at the beginning of S2 and S3 (RQ3), we carried out a factorial ANOVA (2x3) with *memory* as within-subject factor and level of humanlikeness as between-subject factor. For this analysis, we compared Q3 at S1 with Q2 at S2, and Q3 at S2 with Q2 at S3. To ensure that the number of days of zero exposure did not confound our results, we ran a Spearman's rank correlation between the days of zero exposure and the dependent variables collected at Q2 in the subsequent sessions. Moreover, we performed the same ANOVAs described above as ANCOVAs, using the days of zero exposure as covariate. Both tests revealed that the days of zero exposure did not have a significant influence on our results. A post-hoc G*power analysis showed that the power of all statistically significant results in our study was > 0.98. According to Cohen [14], the effect sizes of our results ranged from medium (f= 0.36) to large (f= 1.11) with an average of f= 0.57 (large).

5.1 Embodiment

5.1.1 Manipulation check. We found a significant main effect of condition on the robot's perceived anthropomorphism ($F(2, 37) = 4.734, p = .015$), warmth ($F(2, 37) = 6.117, p = .005$), competence ($F(2, 37) = 5.102, p = .011$), and the way participants talked to the robot ($F(2, 37) = 4.323, p = .021$). In general, the humanlike robot was rated more favorably than the morph. It was perceived as more anthropomorphic (humanlike: $M = 3.65$, $SD = 0.67$; morph: $M = 2.97$, $SD = .69$, $p = .022$), warmer (humanlike: $M = 4.84$, $SD = 1.01$; morph: $M = 3.44$, $SD = 1.08$, $p = .005$), more competent (humanlike: $M = 5.45$, $SD = 0.79$; morph: $M = 4.37$, $SD = 1.09$, $p = .012$), and, albeit just being a trend, more likable (humanlike: $M = 3.50$, $SD = .70$; morph: $M = 2.87$, $SD = .65$, $p = .058$). Participants also reported they talked to the robot more like with another person in the humanlike condition (humanlike: $M = 4.21$, $SD = .75$; morph robot: $M = 3.31$, $SD = .95$, $p = .031$). These differences were not observed between the morph and the mechanical robot, and were only a trend when comparing the humanlike and the mechanical robot on anthropomorphism (humanlike: $M = 3.65$,

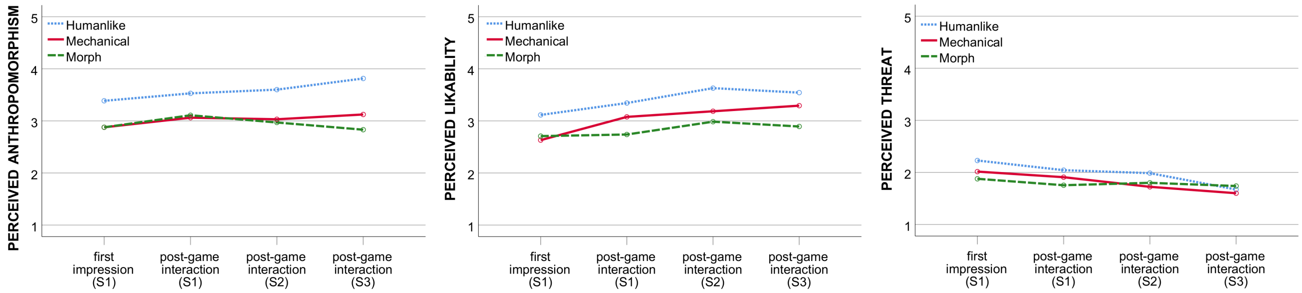

Figure 4: The perceived anthropomorphism (left), likability (center) and threat (right) of the robot from its first impression to after the last interaction session with it.

SD = 0:67; mechanical: *M* = 3:07, *SD* = :77, $p = {:}063$). This means that the manipulation of the robot's anthropomorphism worked particularly well for the humanlike and the morph robot, and only partially for the humanlike and the mechanical robot. Contrary to our design goal, the perception of the morph and the mechanical texture partially overlapped, suggesting that the mechanical features were too dominant in the morphed face.

5.1.2 Correlations. Across sessions, anthropomorphism was significantly positively correlated with likability (S1: $r^1 47^o = {:}747$, $p < {:}001$; S2: $r^1 47^o = {:}856$, $p < {:}001$; S3: $r^1 38^o = {:}759$, $p < {:}001$), warmth (S1: $r^1 47^o = {:}730$, $p < {:}001$; S2: $r^1 47^o = {:}832$, $p < {:}001$; S3: $r^1 38^o = {:}851$, $p < {:}001$), competence (S1: $r^1 47^o = {:}649$, $p < {:}001$; S2: $r^1 47^o = {:}805$, $p < {:}001$; S3: $r^1 38^o = {:}805$, $p < {:}001$), involvement with the robot (S1: $r^1 47^o = {:}549$, $p < {:}001$; S2: $r^1 47^o = {:}506$, $p < {:}001$; S3: $r^1 38^o = {:}400$, $p = {:}010$), and involvement with the game (S1: $r^1 47^o = {:}549$, $p < {:}001$; S2: $r^1 47^o = {:}417$, $p = {:}003$; S3: $r^1 38^o = {:}400$, $p = {:}010$). This shows that the more anthropomorphic participants found the robot, the more they liked it and engaged with it.

5.2 Engagement

5.2.1 Manipulation check. Participants found the robot and the game highly engaging (involvement with the robot at S1 *M* = 3:99, *SD* = :51; S2 *M* = 3:95, *SD* = :61; S3 *M* = 3:97, *SD* = :57; involvement with the game at S1 *M* = 4:11, *SD* = :59; S2 *M* = 4:07, *SD* = :54; S3 *M* = 4:04, *SD* = :64) and we did not find any significant effect of the repeated interactions on involvement with the robot ($F^1 2; 36^o = {:}163; p = {:}851$) and involvement with the game ($F^1 2; 36^o = {:}506; p = {:}607$). Participants' engagement thus remained stable across sessions and cannot be considered a confounder.

5.2.2 Correlations. Across sessions, the correlation between involvement and perceived threat, and involvement and discomfort changed. Involvement with the robot was significantly negatively correlated with perceived threat in S2 ($r^1 47^o = \ {:}351$, $p = {:}014$) and S3 ($r^1 38^o = \ {:}449$, $p = {:}004$), and with discomfort in S3 ($r^1 38^o = \ {:}444$, $p = {:}004$). Similarly, involvement with the game was significantly negatively correlated with perceived threat in S2 ($r^1 47^o = \ {:}482$, $p < {:}001$) and S3 ($r^1 38^o = \ {:}346$, $p = {:}029$), and with discomfort in S3 ($r^1 38^o = \ {:}320$, $p = {:}044$). Interestingly, involvement with the game was also significantly positively correlated with warmth in S1 ($r^1 38^o = {:}493$, $p < {:}001$), but not in S2 and S3. This shows that when the robot was perceived as less threatening and elicited less discomfort, engagement improved.

5.3 Effect of Repeated Interactions

5.3.1 Score. We found a significant main effect of repeated interactions on the score of participants ($F^1 2; 36^o = 22{:}021; p < {:}001$). Participants' score increased significantly between S1 (*M* = 23:15, *SD* = 9:41) and S2 (*M* = 28:45, *SD* = 10:43, $p < {:}001$), S2 and S3 (*M* = 30:77, *SD* = 10:39, $p = {:}014$), and consequently between S1 and S3 ($p < {:}001$). We found a significant main effect of repeated interactions on participants' satisfaction with the score ($F^1 2; 36^o = 3{:}680; p = {:}035$), with a descrease in satisfaction from S1 (*M* = 3:90, *SD* = 0:90) to S2 (*M* = 3:42, *SD* = 1:08, $p = {:}034$) but not from S2 to S3 ($p = 1{:}000$), and from S1 to S3 ($p = {:}261$).

5.3.2 Anthropomorphism. Playing the game with the robot had a significant effect on the perception of anthropomorphism ($F^1 1; 46^o = 9{:}107; p = {:}004$), which increased from the first encounter with the robot (*M* = 3:11, *SD* = :66) to the first game interaction with it (*M* = 3:33, *SD* = :67; see Fig. 4 left). The perceived anthropomorphism did not change across repeated interactions ($F^1 2; 36^o = {:}166; p = {:}847$), while its correlation with perceived threat did. Perceived anthropomorphism was significantly negatively correlated with perceived threat in S1 ($r^1 47^o = \ {:}283$, $p = {:}049$) and S2 ($r^1 47^o = \ {:}402$, $p = {:}004$), but the correlation disappeared in S3. This result is interesting if we consider that the correlation between anthropomorphism and the score of the participant was only significant in the third session ($r^1 38^o = \ {:}382$, $p = {:}015$). While the perception of the robot's anthropomorphism stabilized after the the first game interaction, the time of zero exposure significantly decreased participants' ratings of its anthropomorphism between S1 and S2 ($F^1 1; 46^o = 20{:}204; p < {:}001$) and between S2 and S3 ($F^1 1; 37^o = 8{:}304; p = {:}007$). Participants recalled the robot as less anthropomorphic at the beginning of S2 (*M* = 3:05, *SD* = :74) than they perceived it to be at the end of S1 (*M* = 3:33, *SD* = :67), and the same effect was observed at the beginning of S3 (*M* = 3:01, *SD* = :72) with respect to the end of S2 (*M* = 3:21, *SD* = :83).

5.3.3 Competence. We did not find a significant effect of the game interaction ($F^1 1; 46^o = {:}830; p = {:}367$) nor of repeated interactions ($F^1 2; 36^o = 1{:}807; p = {:}179$) on the perception of competence. What changed across sessions, however, was the correlation between competence and perceived threat. The two constructs were significantly negatively correlated in S2 ($r(47) = -.424, p = .002$), but not in S1 and S3. Moreover, the correlation between competence and discomfort changed over time, with the competence being negatively correlated with discomfort in S2 ($r(47) = -.418, p = .003$) and S3 ($r(38) = -.336, p = .034$), but not in S1. The perception of competence that the robot elicited at the end of S1 and S2 was recalled without significant differences at the beginning of the subsequent sessions (recall at beginning of S2: $F(1, 46) = .137; p = .713$; recall at beginning of S3: $F(1, 37) = 2.568; p = .118$).

5.3.4 Positive Perception (Likability and Warmth). The game interaction had a significant effect on the likability of the robot ($F(1, 46) = 17.458; p < .001$), with the robot being perceived as more likable at the end of the first session ($M = 3.17$, $SD = .77$) than after the first impression ($M = 2.91$, $SD = .73$; Fig. 4 center). No such effect was observed for warmth ($F(1, 46) = 1.243; p = .271$). The likability of the robot ($F(2, 36) = 4.675; p = .016$) also changed across sessions: it increased significantly from S1 ($M = 3.06$, $SD = 0.73$) to S2 ($M = 3.27$, $SD = 0.81$, $p = .019$), and was close-to-significant from S1 to S3 ($M = 3.25$, $SD = 0.72$, $p = .054$). The perceived warmth, instead, remained stable over time ($F(2, 36) = 1.312; p = .282$). In terms of correlations, likability was negatively correlated with perceived threat at S2 ($r(47) = -.368, p = .009$), but not at S1 (albeit a trend was present: $p = .066)$ and S3, and warmth was significantly negatively correlated with discomfort in the same session ($r(47) = -.318, p = .026$). In addition, both likability and warmth were significantly negatively correlated with the score of participants (likability at S3 only, $r(38) = -.358, p = .023$, warmth both at S2, $r(47) = -.361, p = .011$, and S3, $r(38) = -.457, p = .003$).

We did not find any significant difference in the perception of likability and warmth between the end of S1 and the beginning of S2 (likability: $F(1, 46) = .099; p = .755$; warmth: $F(1, 46) = .048; p = .827$), and the end of S2 and the beginning of S3 (likability: $F(1, 37) = 3.221; p = .081$; warmth: $F(1, 37) = 2.172; p = .149$). This indicates that participants recalled their positive perceptions of the robot even after a period of zero exposure.

5.3.5 Negative Perception (Perceived Threat and Discomfort). A significant main effect of game interaction on perceived threat ($F(1, 46) = 8.850; p = .005$) was visible in S1. The perceived threat decreased from the first impression ($M = 2.03$, $SD = .68$) to the end of the first session ($M = 1.86$, $SD = .66$; Fig. 4 right). The same significant effect was observed for discomfort ($F(1, 46) = 27.010; p < .001$) with the discomfort after the first impression ($M = 2.26$, $SD = .88$) decreasing after the first session of interaction ($M = 1.94$, $SD = .79$). Perceived threat and discomfort were also significantly affected by repeated interactions (threat: $F(2, 36) = 4.558; p = .017$; discomfort: $F(2, 36) = 10.332; p < .001$) with perceived threat decreasing from S1 ($M = 1.90$, $SD = 0.67$) to S3 ($M = 1.67$, $SD = 0.52$, $p = .023$) and from S2 ($M = 1.84$, $SD = 0.65$) to S3 ($p = .010$), and discomfort decreasing only from S1 ($M = 2.02$, $SD = 0.79$) to S3 ($M = 1.62$, $SD = 0.55$, $p < .001$). The perceived threat was negatively correlated with participants' score at S1 ($r(47) = -.386, p = .006$) and S2 ($r(47) = -.376, p = .008$), but not at S3. It was also negatively correlated with involvement with the robot and involvement with the game at S2 and S3, but not at S1 (see section 5.2.2). While there was no significant difference in the perceived threat recalled by participants in the beginning of S2 ($F(1, 46) = 2.182; p = .146$) and S3 ($F(1, 37) = 1.143; p = .292$), they recalled the robot as eliciting more discomfort ($M = 2.09$, $SD = .90$) than it actually did ($M = 1.94$, $SD = .79$) in the beginning of S2 ($F(1, 46) = 5.807; p = .020$) but not at the beginning of S3 ($F(1, 37) = .385; p = .539$).

6 DISCUSSION

The Game Interaction Effect (RQ1)

We found that *allowing participants to interact with the robot in a fun and engaging collaborative game positively influenced their impressions of it.* Participants judged the robot to be more humanlike and less threatening after interacting with it, they liked it better and felt more comfortable with it. This result is in line with Złotowski et al. as it confirms that "a positively toned interaction" can increase a robot's likability [60]. Unlike Bergmann et al., we did not find the perceived warmth of the robot to decrease after the game interaction [10], but to remain stable. This difference might be due to the fact that, unlike us, Bergmann and colleagues measured first impressions after 15 s. Despite the robot displaying much more advanced language understanding and interaction capabilities in the game (a reason for potentially improved competence perception) and its limited geographic knowledge (a reason for a potential decrease in perception), *we did not find the robot's perceived competence to be influenced by the game interaction.* This finding suggest that *a robot's competence is judged more on its initial appearance and general interaction modalities and less on the content of the interaction.*

The Effect of Repeated Interactions (RQ2)

While some researchers have investigated the persistence of first impressions within the same interaction [10][44][60], our experiment is novel in that it focuses on the effect of well-controlled repeated interactions occurring with multiple days of zero exposure in between. As highlighted in Figure 5, *our findings indicate that the mental image of a robot is built over time as the perceptual dimensions that contribute to it stabilize at different moments.* The perceived competence of the robot was determined in the first two minutes of chat with the robot and, after that, remained stable across sessions. The perceived anthropomorphism increased from the first impression to the first game interaction, and then remained unchanged. Likability increased, but only up to the second session despite the continuously positive interaction [60]. Perceived threat and discomfort were the most fluctuating dimensions and kept decreasing until the last session of the experiment, a trend we also observed in the changing correlations involving the two.

The Consistency of People's Memory (RQ3)

Overall, we found *participants to correctly recall the perception of the robot from the previous interaction session.* Perceived anthropomorphism and discomfort were the only exceptions. At the beginning of both S2 and S3, participants *recalled the robot to be less humanlike than they had perceived it to be after the previous interaction.* This is especially interesting because participants' perception of the robot's anthropomorphism was otherwise stable across all three sessions. We saw a similar effect with *perceived discomfort, which participants recalled as higher at the beginning of S2 than it actually*

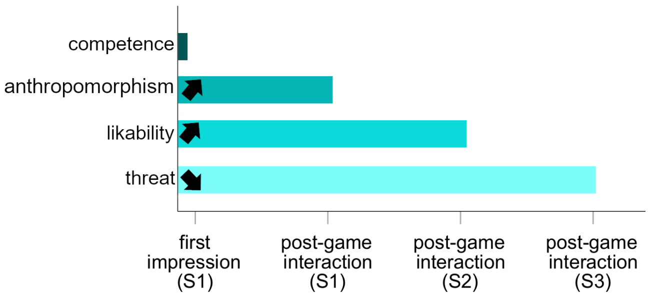

Figure 5: Time frame required for perceptual dimensions to stabilize. Arrows indicate increasing/ decreasing trend.

was at the end of the previous session. However, this effect vanished at the beginning of S3. We hypothesize that the widespread conception of a robot as a mechanical device and "bad" fictional character might have distorted the mental image of the robot [26].

The Curse of the Morph Embodiment (RQ4/5)

We did not find any significant interaction effect between session and embodiment, which suggests that the *effects of repeated interaction with a robot generalize across all levels of anthropomorphism that we investigated.* This implies, however, that *initial biases against a robot do not wear out over time.* Participants judged the humanlike robot as more competent, warmer and talked with it as with another person only based on its humanlike facial appearance, and these differences in judgement persisted even after three exactly identical interaction sessions. Even though our robot had very advanced remote-controlled interaction capabilities, it could not recover from its initial unappealing appearance. These findings extend the perceptual mismatch theory [28] by suggesting that *a perceptual mismatch negatively influences more dimensions than the robot's likability alone.*

Relevance to the HRI Community

Our findings are important for the HRI community in several ways: (1) by highlighting the durability of the effect of appearance on a robot's perception, we strengthen the importance of a robot's visual design; (2) by demonstrating that, even though perceived threat and discomfort can decrease over time, differences between robot embodiments remain constant, we shed new light on the persistence of the uncanny valley effect; (3) by analyzing the changes in the perception of a robot over time and the accuracy of people's recall of such perception, we show how pre-existing expectations can influence people's situated perception of a social robot, and (4) by assessing the perception of a robot at different moments, we find that perceptual dimensions evolve differently over time.

Limitations and Future work

Our study was carried out in the lab to keep environmental factors fixed and isolate the effect of repeated interactions on people's perception of social robots. Moreover, it mostly involved participants coming from a Computer Science background, who have an unusual positive attitude towards robots [7]. Since our findings suggest a lasting effect of negative feelings towards our robot even within this potentially positively biased sample group, we believe that the participants' background did not have a negative impact on our findings. However, in the future, it would be interesting to bring the same scenarios in the wild and include participants with a wider background and a more gender-balanced distribution. In our work, we limited the number of sessions to three, so we can only speculate how our findings generalize to further sessions. Since the mere exposure to other humans has a peak in effect between 10 and 20 exposures [11], it would be interesting to extend the present work to more than 10 sessions.

The blended embodiment that we chose for our study was limited to a head and the overall perceived discomfort and threat that it elicited was already low at the first impression for all three levels of humanlikeness. To investigate how our findings extend to robots eliciting overwhelming uncanny feelings during first impressions, we suggest to use different platforms like the Geminoid robot.

In this study, the robot was remote-controlled by an operator to avoid overshadowing the game interaction with errors in language understanding. The operator might have given participants an overly positive impression of the robot's abilities, but it unlikely affected their perception as participants were unaware of it until debriefing. The presence of the robot while participants answered the questionnaires could have led them to rate it favorably. Yet, the robot was in idle mode and was not looking at participants while they filled out the questionnaires. Moreover, the iPad's screen was not facing the robot and was initially introduced as containing information hidden from it. Hence, we have no reason to believe people considered the robot as having access to their responses.

7 CONCLUSION

In this paper, we presented the results of a long-term study investigating the effect of repeated interactions on the perception of a social robot. Participants engaged in a collaborative geography game with the blended robot head Furhat in three game sessions occurring with three to ten days of zero exposure in between. Furhat's anthropomorphism was varied between participants to study the influence of embodiment on the perception of a robot over time. We found the perceptual dimensions composing participants' mental image of the robot to stabilize within different time frames. While perceived competence was judged quickly and remained stable after only two minutes of social chat, playing a game with the robot improved participants' impressions of its anthropomorphism and likability, which kept increasing up to the second session. Perceived threat and discomfort, instead, kept fluctuating until the last session. Interestingly, the robot we designed to look unappealing could not recover from its initial negative perceptions. These findings can contribute to the field of HRI as they highlight the relevance of a robot's visual design and shed new light on the importance of allowing participants time to interact with a robot before measuring their perception of it.

ACKNOWLEDGEMENTS

This work was supported by the COIN project (RIT15-0133) funded by the Swedish Foundation for Strategic Research. Thanks to S. Walkötter for Figure 2, R. Kessler for assisting with the technical setup and R. Manuvinakurike for his work on the game dynamics.

REFERENCES

[1] Muneeb Imtiaz Ahmad, Omar Mubin, and Joanne Orlando. 2017. Adaptive social robot for sustaining social engagement during long-term children–robot interaction. *International Journal of Human–Computer Interaction* 33, 12 (2017), 943–962.

[2] Muneeb Imtiaz Ahmad, Omar Mubin, Suleman Shahid, and Joanne Orlando. 2019. Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: a long-term child-robot interaction study. *Adaptive Behavior* (2019), 243–266.

[3] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: A back-projected human-like robot head for multiparty human-machine interaction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7403 LNCS (2012), 114–130.

[4] Linda Albright, David A. Kenny, and Thomas E. Malloy. 1988. Consensus in Personality Judgments at Zero Acquaintance. *Journal of Personality and Social Psychology* 55, 3 (1988), 387–395.

[5] Nalini Ambady, Frank J Bernieri, and Jennifer A Richeson. 2000. Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In *Advances in experimental social psychology*. Vol. 32. Elsevier, 201–271.

[6] Moshe Bar, Maital Neta, and Heather Linz. 2006. Very first impressions. *Emotion* 6, 2 (2006), 269–278.

[7] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic. 2019. Human-Robot Interaction-An Introduction. (2019).

[8] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. My robotic doppelgänger-A critical look at the uncanny valley. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 269–276.

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

[10] Kirsten Bergmann, Friederike Eyssel, and Stefan Kopp. 2012. A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *International Conference on Intelligent Virtual Agents*. Springer, 126–138.

[11] Robert F Bornstein. 1989. Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychological bulletin* 106, 2 (1989), 265–289.

[12] Dana R Carney, C Randall Colvin, and Judith A Hall. 2007. A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality* 41, 5 (2007), 1054–1072.

[13] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The robotic social attributes scale (rosas): Development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*. ACM, 254–262.

[14] Jacob Cohen. 1969. *Statistical power analysis for the behavioral sciences*. Academic Press.

[15] Kerstin Dautenhahn. 2004. Robots we like to live with?!-a developmental perspective on a personalized, life-long robot companion. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, 17–22.

[16] Maartje De Graaf, Somaya Ben Allouch, and Tineke Klamer. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in human behavior* 43 (2015), 1–14.

[17] Maartje de Graaf, Somaya Ben Allouch, and Jan AGM van Dijk. 2016. Long-term evaluation of a social robot in real homes. *Interaction studies* 17, 3 (2016), 462–491.

[18] Maartje De Graaf, Somaya Ben Allouch, and Jan Van Dijk. 2017. Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 224–233.

[19] Ap Dijksterhuis and Pamela K. Smith. 2002. Affective Habituation: Subliminal Exposure to Extreme Stimuli Decreases Their Extremity. *Emotion* 2, 3 (2002), 203–214.

[20] Friederike Eyssel, Ricarda Wullenkord, and Verena Nitsch. 2017. The role of self-disclosure in human-robot interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 922–927.

[21] Eli J Finkel, Michael I Norton, Harry T Reis, Dan Ariely, Peter A Caprariello, Paul W Eastwick, Jeana H Frost, and Michael R Maniaci. 2015. When does familiarity promote versus undermine interpersonal attraction? A proposed integrative model from erstwhile adversaries. *Perspectives on Psychological Science* 10, 1 (2015), 3–19.

[22] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C Schultz, et al. 2005. Designing robots for long-term social interaction. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1338–1343.

[23] Kerstin Sophie Haring, Katsumi Watanabe, David Silvera-Tawil, Mari Velonaki, and Tomotaka Takahashi. 2015. Changes in perception of a small humanoid robot. In *2015 6th International Conference on Automation, Robotics and Applications (ICARA)*. IEEE, 83–89.

[24] Kerstin S Haring, Katsumi Watanabe, Mari Velonaki, Chad C Tossell, and Victor Finomore. 2018. FFABâĂŤThe Form Function Attribution Bias in Human–Robot Interaction. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (2018), 843–851.

[25] Joachim R. Höflich and Afifa El Bayed. 2015. *Perception, Acceptance, and the Social Construction of Robots—Exploratory Studies*. Springer International Publishing, Cham, 39–51.

[26] Aike C. Horstmann and Nicole C. KrĂďmer. 2019. Great Expectations? Relation of Previous Experiences With Social Robots in Real Life or in the Media and Expectancies Based on Qualitative and Quantitative Assessment. *Frontiers in Psychology* 10 (2019), 939.

[27] Lauren J Human, Gillian M Sandstrom, Jeremy C Biesanz, and Elizabeth W Dunn. 2013. Accurate first impressions leave a lasting impression: The long-term effects of distinctive self-other agreement on relationship development. *Social Psychological and Personality Science* 4, 4 (2013), 395–402.

[28] Jari Kätsyri, Klaus Förger, Meeri Mäkäräinen, and Tapio Takala. 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 (2015), 390.

[29] Csaba Kertész and Markku Turunen. 2017. What Can We Learn from the Long-Term Users of a Social Robot?. In *International Conference on Social Robotics*. Springer, 657–665.

[30] Angela Y. Lee. 2001. The mere exposure effect: An uncertainty reduction explanation revisited. *Personality and Social Psychology Bulletin* 27, 10 (oct 2001), 1255–1266.

[31] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. 2012. Personalization in HRI: A longitudinal field experiment. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 319–326.

[32] Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. Empathic robots for long-term interaction. *International Journal of Social Robotics* 6, 3 (2014), 329–341.

[33] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics* 5, 2 (2013), 291–308.

[34] Manja Lohse, Frank Hegel, and Britta Wrede. 2008. Domestic applications for social robots: An online survey on the influence of appearance and capabilities. (2008), 21–32.

[35] Nikita Mattar and Ipke Wachsmuth. 2014. Let's Get Personal. In *International Conference on Human-Computer Interaction*. Springer, 450–461.

[36] Phil McAleer, Alexander Todorov, and Pascal Belin. 2014. How do you say 'Hello'? Personality impressions from brief novel voices. *PloS one* 9, 3 (2014), e90779:1–e90779:9.

[37] Rachel McDonnell, Martin Breidty, and Heinrich H. Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics* 31, 4 (2012), 91:1–91:10.

[38] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[39] Lilia Moshkina and Ronald C Arkin. 2005. Human perspective on affective robotic behavior: A longitudinal study. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1444–1451.

[40] Laura P Naumann, Simine Vazire, Peter J Rentfrow, and Samuel D Gosling. 2009. Personality judgments based on physical appearance. *Personality and social psychology bulletin* 35, 12 (2009), 1661–1671.

[41] Radosław Niewiadomski, Virginie Demeure, and Catherine Pelachaud. 2010. Warmth, competence, believability and virtual agents. In *International Conference on Intelligent Virtual Agents*. Springer, 272–285.

[42] Tatsuya Nomura, Takayuki Kanda, and Tomohiro Suzuki. 2006. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *Ai & Society* 20, 2 (2006), 138–150.

[43] Heather L O'Brien and Elaine G Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69.

[44] Maike Paetzel and Ginevra Castellano. 2019. Let Me Get To Know You Better: Can Interactions Help to Overcome Uncanny Feelings?. In *Proceedings of the 7th International Conference on Human-Agent Interaction*. ACM, 59–67.

[45] Maike Paetzel and Ramesh Manuvinakurike. 2019. "Can you say more about the location?" The Development of a Pedagogical Reference Resolution Agent. In *Dialog for Good - Workshop on Speech and Language Technology Serving Society (DiGo)*.

[46] Maike Paetzel, Ramesh Manuvinakurike, and David DeVault. 2015. âĂIJSo, which one is it?âĂİ The effect of alternative incremental architectures in a high-performance game-playing agent. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 77–86.

[47] Aaron Powers and Sara Kiesler. 2006. The advisor robot: tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 218–225.

[48] Beatrice Rammstedt and Oliver P John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality* 41, 1 (2007), 203–212.

[49] Rolf Reber, Piotr Winkielman, and Norbert Schwarz. 1998. Effects of perceptual fluency on affective judgments. *Psychological Science* 9, 1 (1998), 45–48.

[50] Harry T. Reis, Michael R. Maniaci, Peter A. Caprariello, Paul W. Eastwick, and Eli J. Finkel. 2011. Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology* 101, 3 (2011), 557–570.

[51] Laurel D Riek. 2012. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012), 119–136.

[52] Emmy Rintjema, Rianne Van Den Berghe, Anne Kessels, Jan De Wit, and Paul Vogt. 2018. A Robot Teaching Young Children a Second Language: The Effect of Multiple Interactions on Engagement and Performance. *ACM/IEEE International Conference on Human-Robot Interaction* March (2018), 219–220.

[53] Astrid M Rosenthal-Von Der Pütten and Nicole C Krämer. 2014. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior* 36 (2014), 422–439.

[54] Catharina Vesterager Smedegaard. 2019. Reframing the Role of Novelty within Social HRI: from Noise to Information. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 411–420.

[55] Kristen Stubbs, Debra Bernstein, Kevin Crowley, and Illah Nourbakhsh. 2005. Long-Term Human-Robot Interaction: The Personal Exploration Rover and Museum Docents. *Proceeding of the 2005 conference on artificial intelligence in education* (2005), 621–628.

[56] Michael Sunnafrank and Artemio Ramirez. 2004. At first sight: Persistent relational effects of get-acquainted conversations. *Journal of Social and Personal Relationships* 21, 3 (2004), 361–379.

[57] Janine Willis and Alexander Todorov. 2006. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science* 17, 7 (2006), 592–598.

[58] Timothy J Wood. 2014. Exploring the role of first impressions in rater-based assessments. *Advances in Health Sciences Education* 19, 3 (2014), 409–427.

[59] Robert Zajonc. 1968. Attitudinal effects of mere exposure. *Journal of personality and social psychology* 9, 2 (1968).

[60] Jakub A. Złotowski, Hidenobu Sumioka, Shuichi Nishio, Dylan F. Glas, Christoph Bartneck, and Hiroshi Ishiguro. 2015. Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology* 6 (2015), 1–13.