Beyond R2D2: Designing Multimodal Interaction Behavior for Robot-specific Morphology

DAPHNE E. KARREMAN, Human Media Interaction, University of Twente, Enschede, The Netherlands
GEKE D. S. LUDDEN, Interaction Design, University of Twente, Enschede, The Netherlands
VANESSA EVERS, Human Media Interaction, University of Twente, Enschede, The Netherlands

Robots are expected to enter the everyday lives of people to entertain, educate, or support them. It is therefore important that people can intuitively understand the behavior of robots. Oftentimes, the behavior of people is used as a model because of its familiarity. However, it is as yet unclear what the best approach is to design interaction behaviors for non-humanoid robots. In this article, we explore two different approaches toward designing behavior for a service robot. The first concerns the commonly used approach of copying human behavior as closely as possible to the robot (*human-translated*). The second approach was inspired by product design methods. The design of the robot's behavior was optimized for the robot's interaction capabilities and hardware modalities (*robot-optimized*). To evaluate people's responses to the two behavior sets for a tour guide robot, an online video study ($N = 204$) and a two-day in-the-wild study ($N > 600$) were performed. Results showed that participants responded slightly more positive to robot-optimized behavior and paid attention to robot-optimized behavior for longer. However, participants remembered more details when the robot showed human-translated behavior. Together, the studies show that it is sometimes better for non-humanoid robots to have robot-optimized behaviors rather than human-translated behaviors.

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**;

Additional Key Words and Phrases: Human-Robot Interaction, Human robot user studies, In-the-wild studies, Social robots, Interaction design, Tour guide robot, Robot appearance, Robot behaviors, Design approaches, Robot design

18

1 INTRODUCTION

R2D2 is a famous robot from the *Star Wars* movies. If we were to encounter R2D2 in the real world, then we would probably not literally understand the beeps and movements it makes, because R2D2 does not closely resemble people in the way it looks, speaks, or behaves. However, we might be able to deduce R2D2's intentions. This implicit understanding may be due to the tendency of people to project human-like behavior on non-human things, such as robots [22, 25], their social responses to computers and technology in general [57, 62], and their inclination to interpret behavior of non-human things as if it were human [32, 41].

Although it is unlikely for us to encounter R2D2 in an everyday situation, it is expected that we will soon encounter similar service robots that offer services or support us in various settings. This is due to the fact that more and more tasks and services are, and increasingly will be, automated. Hence, robots are expected to enter public environments such as museums, hospitals, and shopping malls in the near future. These settings will require robots to interact with people who probably have not encountered robots before.

To have meaningful interactions with people, these robots should detect and process behavior of people as well as react in a normative and social way. Therefore, these robots can be classified as social robots. This description is in line with the definition as formulated by Bartneck and Forlizzi: "*A social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact*" [5].

So far, a large body of work in Human-Robot Interaction has been focused on designing behavior for social robots that closely resemble people (humanoids). However, few researchers have studied the design of behavior for social robots that do not have human features. As a result, little knowledge is available on how to successfully design multimodal behaviors for social robots that lack human-like appearance. In this article, we set out to report and evaluate a systematic approach to develop behavior for social robots that do not closely resemble people, using design practices from the field of Industrial Design and Interaction Design. In this article, we will focus on tour guide robots to understand the influence of robot behavior on people's understanding and experience in context. In this context, multimodal communication and interaction between the guide robots and visitors is important. Also, in this setting the general tendency is to copy human-like behavior.

The article is structured as follows: In the related work, we discuss how robot behavior is predominantly designed currently, and we describe the trends in developing tour guide robots. We then offer our problem statement and describe the two approaches for designing tour guide robot behavior that we compare in this article. Next, we report on two studies, an online video study ($N = 204$) and a two-day in-the-wild study ($N > 600$) in which we compared the two behavior sets in terms of participants' attention and attitude toward the robot. We present the results and we discuss the findings, drawing parallels to previous research in the field of Human-Robot Interaction. We finalize the article with conclusions and directions for future work.

2 RELATED WORK

2.1 Design of Humanoid and Non-human-like Robots and Their Behavior

Humanoid robots are often envisioned to perform tasks in everyday social situations. These humanoids have a similar basic physical structure and kinetic capabilities as people have, and thus are to some degree physically able to behave as people would. For example, these robots are able to walk, use their arms and hands to manipulate things, and swivel their heads. Therefore, these robots seem to be suitable to perform social tasks in everyday environments, such as schools and

museums. Examples of such robots are: NAO (e.g., used in References [8, 29, 51]) or ASIMO (e.g., used in References [56, 67]).

A common strategy to create the behavior for humanoid robots is to copy human behavior. There are several good reasons to do so; people are already familiar with the behaviors [9], people can read the behaviors naturally [61], people seem to expect it from humanoid robots [68], and this way of developing behavior is shown to work for humanoid robots (e.g., References [47, 48]).

However, humanoid appearances and behaviors for robots might also have disadvantages. Developing non-humanoid robots is typically more cost-effective, and they are more reliable, because they have physical forms targeting specific tasks [12]. Moreover, anthropomorphic appearances and behaviors can lead to disadvantages in human-robot (social) interactions. For example, an extremely human-like appearance can in some cases lead to a frightening robot, especially when it starts moving [54]. Another disadvantage can be that a humanoid appearance can raise unrealistic expectations of the abilities of the robot [20, 22]. This can lead to disappointment when the robots are not as skillful, responsive, and advanced as people expect them to be [14, 25].

Furthermore, the possibilities of creating highly anthropomorphic behaviors and appearances are often limited by what is technically possible; see also Reference [12]. For example, in the design of the Snackbot, the width of the head was dictated by the technical components, in this case a Bumblebee stereo camera [49]. Moreover, the robot did not walk, but it was based on an existing wheeled platform, and it had bumpers for safety reasons. As another example, alternatives to copying (human-like) gaze behavior had to be found for the robot Chester, because it did not have a swiveling head. Nevertheless, effective gaze behavior was created when gaze and orientation behavior were designed and controlled jointly [73].

Another design option for interactive service robots is to deliberately create robots that do not resemble people. Creating a robot that does not look like a human does not necessarily limit the capabilities of this robot to interact with people. Several sources (e.g., References [20, 22]) state that anthropomorphism is not the only solution to creating effective social robots. According to DiSalvo, a social robot should leverage robot-ness and product-ness, as well as humanlike-ness, to avoid false expectations of the machine's capabilities and to make people feel comfortable using the product, while the interaction is socially engaging [20]. As such, the robot-ness, product-ness, and humanlike-ness need to be balanced and should be recognizable in appearance and behavior of the robot.

Even though a non-humanoid appearance might be an advantage for robots, there is no validated method available to design behavior for such robots. Currently, the approach is to also copy from human behavior. However, we question whether it is indeed logical to use this human-behavior-translated-to-robot strategy for all kinds of social robots, specifically for robots that do not at all resemble people.

We argue that robot behavior can be optimized for the robot's morphology and modalities. This concept is not new to the field of Human-Robot Interaction. For example, Embgen et al. [23], Bretan, Hoffman, and Weinberg [10], and Cha, Matarić, and Fong [13] already proposed to use abstracted robot-specific behavior consisting of body movement and displaying colored lights, to show a robot's emotion and intentions. Szafir, Mutlu, and Fong explored the design of "natural" and "intuitive" flight motions to improve assistive free-flying robots' abilities to communicate intent while simultaneously accomplishing tasks [71]. Vázquez et al. found that not only a copy of human orientation behavior (attentive orientation) but also an optimized orientation (middle orientation) can be an effective starting point for robot orientation behavior [73]. Sabanovic, Reeder, and Kechavarzi have shown that social behavior, not copied from humans, was preferred over alarm type behavior for minimalistic desk robots [66]. However, guidelines and approaches on

how such behaviors should be designed, to our knowledge are scarce. Ju and Leifer proposed a framework to design implicit interactions for non-living objects, which provides a starting point to create non-verbal behavior [40]. Building on that, we present an approach to develop nonverbal behavior for non-humanoid robots, as an alternative to copying human behaviors.[1]

The motivation for the proposed approach is that the design of appearance, interaction and behavior of social robots finds many parallels in the design of smart interactive products, studied in the fields of product design and interaction design. Similar to social robots, some smart products interact with people in a social and personalized manner. However, to design the appearance and behavior of smart products, industrial designers generally do not rely on copying human behavior or appearance. Instead, they often take a desired user experience as the starting point for design.

Several interaction design and product design researchers provide guidelines on designing the user experience of products, which may be applicable to Human-Robot Interaction. Useful guidelines are for instance given by Forlizzi and Ford, who state that to design a successful product, the focus should be on the user, the product and the context of use [28]. Similarly, Hummels states that not only a product, but a "context for experience" should be designed [38]. The interaction does not have to be as simple as possible, but it should be engaging and the product should encourage to explore the context with all senses to gain rich experiences (e.g., pressing the play button on a cd-player is less engaging than cleaning a LP and putting the needle in the groove to play the music).

In line with these approaches, Hekkert, Mostert, and Stompff used the user experience of "a dance" as the starting point for designing the interaction with a photo copier to stimulate several senses in a surprising way [33]. Desmet, Ortíz Nicolás, and Schoormans successfully created intended interaction experiences, dominant vs. elegant, for two similarly looking devices developed for their study [18]. Because these experiences were not visible in the device, they only were perceived through interaction style, indicating that it is possible to design interaction devices with different product personalities.

In these examples, the designers did not copy human behavior for their product. Instead, they focused on the desired interaction experience that needs to be realized. Because of this, we expect that taking a desired user experience and a robot's functional goals as starting points, is a promising alternative to copying human behavior in the design of socially interactive robots.

So far, the focus was on robot appearance, behavior and user experience in general. In this article, we will focus on behavior for a tour guide robot, therefore, in the next section, we give an overview on the influence of appearance and behavior of tour guide robots on Human-Robot Interaction.

2.2 The Influence of Appearance and Behavior of Tour Guide Robots on Human-Robot Interaction

We focus on non-humanoid robots, because there is no evidence that a guide robot needs to be humanoid in appearance to effectively guide people. Also, a non-humanoid appearance will generally ensure a less complex and challenging development process. Although we focus on non-humanoid

---

[1]While the existence of other robot appearances, such as zoomorphic, is acknowledged, this article mainly focuses on human-like aspects of the robot's appearance and behavior design. Notably, research in human robot interaction is often carried out on early prototypes, that look mechanical or machinelike (e.g., see References [35, 74]). While results from such studies are considered, these robot platforms are not included as examples of either humanoid or non-humanoid robot types for the reason that they are unfinished concepts rather than final designs.

tour guide robots, we expect the findings to be generalizable to other domains of Human-Robot Interaction. In this article, we give a brief overview of the development of tour guide robots. For a more extensive overview of tour guide robots and how their behavior influences Human-Robot Interaction, see Reference [42].

The first tour guide robots (developed before 2002) were generally basic in appearance. Examples are Rhino [11] and the three entertainment robots in the "Museum für Kommunikation" in Berlin [31]. Some did have more sophisticated humanlike bodies (e.g., Reference [7]) and facial features such as Minerva [72]. What these robots have in common is that technical challenges as well as challenges in navigation and localization were the main focus of the research groups developing and deploying the robots. However, some investigations into people's responses to the robots were carried out. These observations showed people's tendency to "abuse" or "try out" the robots by blocking their path, pressing buttons, and so on [11, 63]. The first tour guide deployments were a novelty [59], and as a result people were unfamiliar with the abilities of robots and the services they provided [11, 31]. This indicates that to guide people through a robust sense making process, and to ensure optimal usage, tour guide robots require careful design.

Even though the tour guide robots developed later than 2002 were equipped with more capabilities than their predecessors, the visitors' experience interacting with the guide robots in the museum did not seem to improve significantly. It is interesting that while robots had increasingly humanlike features and more sophisticated capabilities, these tour guide robots still evoked boundary-searching or "bullying" behaviors from users [15, 69]. This could indicate that while tour guide robots were endowed with more human-like interaction capabilities, insufficient effort has been spent in designing the behaviors to optimize people's understanding, expectations and interaction experience.

Recent studies performed with humanoid tour guide robots focused on the evaluation of singled-out robot behavior, meaning the copy of only one specific behavior to communicate something, such as human-like gaze behavior [77], pointing behavior [6], and orientation behavior [48]. Findings showed that people recognized and responded naturally to the singled-out humanlike behavior the robot performed [6, 36, 48, 77]. This seems to be an argument for identifying effective singled-out behaviors from people and translating these one-to-one to robots. However, as Gehle et al. discussed, when behavior that the robot had to show became more complex (not one singled-out behavior but a combination of behaviors), the interaction between the robot and the visitors became less smooth [29]. We believe this might be due to two reasons. First, singled-out behavior may be effectively translated to a robot, but when combinations of behaviors are needed to display more complex behavior, the combination may not be what people expect and may therefore be more difficult to understand. Second, a one-to-one translation of behavior is not always possible because of the technical limitations of the robot, which leads to unsynchronized combined behaviors.

From our previous studies with tour guide robots, we similarly found that an exact copy of human behavior to robots cannot always be made due to restrictions in the behavior modalities available on the robot (e.g., having no arms for safety reasons, while humans have two), or due to the limitations posed by the robot modalities (e.g., a swiveling head that humans have, but a robot might lack) [45]. This indicates that human-like behavior is usually not optimal for robots that do not closely resemble people. Therefore, we argue that for non-humanoid robots, behavior that is specifically designed for the robot's modalities should be used.

In the remainder of this article, when we refer to robots, we mean robots that do not closely resemble people; we will use the term humanoid robots to indicate robots that closely resemble people.

3 RESEARCH QUESTION

We hypothesize that visitors, in general, will be more positive toward a guide robot with multi-modal behavior specifically designed for the modalities of the robot compared with a robot with behavior copied from people. We expect that "human-translated" behavior will be appropriate for highly anthropomorphic robots, but that this behavior might be perceived as uncanny, inappropriate or difficult to understand for non-humanoid robots. Especially when a one-on-one translation from human to robot is not possible, which is the case when a robot does not closely resemble people. Therefore, we assume that "robot-optimized" behavior that is specifically designed for the morphology of the robot will be perceived as a better fit for robots that do not closely resemble people.

Based on the expectations described above, our main research question is: *"Do people respond more positively to a robot displaying robot-optimized behavior compared with a robot displaying human-translated behavior?"* To answer this question, we focus on people's attention and attitudes toward the robot's behavior to measure its effectivity. To answer the overall research question, we have formulated two hypotheses.

The first hypothesis concerns the way that the robot's behavior will influence the attention of the visitor. We expect that visitors will pay more attention to the robot in the robot-optimized condition. Also, we expect visitors to remember more of the information the robot provided. This leads to the following hypothesis:

H1: Robot behavior optimized for the modalities of a tour guide robot that does not closely resemble people will lead to longer periods of attention and higher recall of the information provided by the robot.

The second hypothesis concerns the way that the robot's behavior will influence the experience and attitudes toward the robot of the visitors. We expect that visitors will have a more positive attitude toward the robot with robot-optimized behavior. This leads to the following hypothesis:

H2: Robot behavior optimized for the modalities of a tour guide robot that does not closely resemble people will lead to more positive attitude toward the robot.

In the following section, we will first explain the two approaches to design the multimodal behavior that we have used, before we describe the methods for evaluation in Section 5 and results in Section 6.

4 TWO APPROACHES TOWARD THE DESIGN OF MULTIMODAL ROBOT BEHAVIOR

To test the assumption outlined above, we developed two sets of behaviors for the robot following two different design approaches. We state that copying human-like behavior to a robot that does not resemble people, limits the options for behavior design, while our newly proposed approach optimizes the use of the robot modalities.

In Figure 1, we present how our newly proposed approach relates to the conventional approach to design multimodal behavior for socially interactive robots. As can be seen from Figure 1, both design approaches start from the observed human tour guide behavior. This is especially valuable when the robot performs social interaction tasks that previously were only performed by people, as is the case for guiding. The right side of Figure 1 shows how the robot morphology may limit the opportunities for behavior when human-like behavior is translated to the robot. This is due to the fact that the robot might not have the modalities or technical features to perform behavior exactly the way that people do. This approach might lead to poor imitation of human behavior or behavior that cannot be performed will be left out. We call this approach *"human-translated."* On the left side, we present the robot-optimized approach. In this approach, the behavior of a
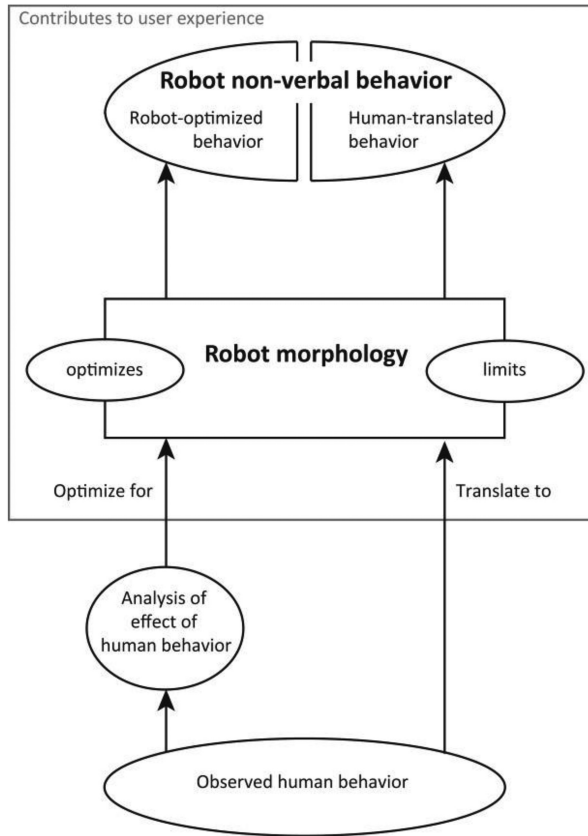
Fig. 1. Schematic view of the two different design approaches—robot-optimized behavior and human-translated behavior—both stem from observed human behavior.

human tour guide is analyzed to understand the interactional outcomes. Interactional outcomes are the effects of multimodal actions that human tour guides aim for with specific guide behaviors during their tour. For each of the multimodal actions a human tour guide performs, we defined the interactional outcome (see Appendix 1). For example, "focus visitor attention on the exhibit" is the interactional outcome of a guide's action "gaze toward the exhibit while pointing toward the exhibit." Subsequently, multimodal behavior that fits the modalities of the robot is designed to accomplish the same interactional outcomes, without necessarily showing the same behavior as a human tour guide would. We call this approach *"robot-optimized."*

The robot we used in our studies is the FROG robot (see Figure 2), and a small model of it. The FROG is developed within the EU FP7 project FROG, which stands for Fun Robotic Outdoor Guide. FROG was developed as a tour guide robot to guide small groups of visitors and engage and educate them at tourist sites. At an earlier stage in the project, we used other robot platforms and the model used for the first study presented in this article to try out and change behaviors and to inform the design of the final FROG robot. FROG has some human-like characteristics. Nevertheless, looking at the full appearance and abilities of the robot, it was clear that FROG does not resemble a person and has a lot of limitations to copy human-like behavior. Therefore, this robot was deemed suitable for the studies we wanted to perform.

Fig. 2. FROG robot.

The design of the two sets of behaviors was done by a small group of designers who together decided how to implement the behavior, of which the process will be explained in the following sections. The implementation of design relates balancing real-world validity and controlled lab studies. For this study, we decided to deal with this balance by using a combination of studies and a robot behavior design that was created and evaluated by designers.

4.1 Approach 1: Human-translated Behavior

The human-translated approach was to translate human behavior as similarly as possible to the robot. As explained previously, this is an approach that is commonly used in the field of Human-Robot Interaction to design social robot behavior.
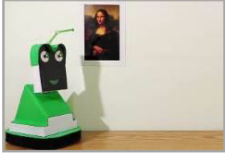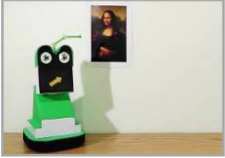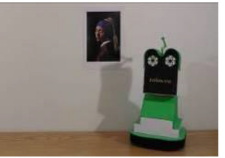
The resulting set of behavior cues mimicked the observed behaviors of human tour guides studied in [43] as much as possible (see Table 1 for an impression, and see Appendix 1 for a full explanation of the interactional outcomes and sets of behaviors for the robot). For example, when a guide would blink with the eyes, an animation in FROG's eyes was added to imitate the blinking, and when a guide would swivel his head to look toward all visitor, the robot would make a full body movement as FROG could not swivel its head to focus on all visitors. The actions of the robot were timed at similar moments as the actions of human tour guides would happen. For example, when human guides want to go to the next exhibit, they were observed to break eye-contact with the group of visitors before they started to move to the next exhibit. We translated this behavior as closely as possible to the robot: When a robot finished providing information about an exhibit, the animation in the eyes of the robot looked downwards, indicating a break of eye-contact before it moved to another exhibit.

4.2 Approach 2: Robot-optimized Behavior

As an alternative to the human-translated approach to designing robot behavior, we propose the robot-optimized approach. This approach is based on principles that are common in product design and interaction design. Starting points for this approach are the robot's functional specifications and the desired user experience of the Human-Robot Interaction.

The interactional outcomes of observed human tour guide behavior (see Reference [43] for details on guide behaviors) were used to design robot-optimized behavior. Moreover, we have studied the visitor experiences of following a human guided tour. By designing behavior for the robot tour

Table 1. Example of Robot Behaviors Derived from Interactional Outcomes

Interactional outcome	Human behavior cues	Human-translated Direct copy of human tour guide behavior	Robot-optimized Optimal use of modalities to achieve interactional outcomes
Keep visitors engaged and get implicit feedback	Gaze at one person and change to the others sometimes Stand close to exhibit, but not in front of it (in front of pillar etc.) Orientation toward visitors	Keep gazing at one person, and by turning on the spot alternate to others (not able to swivel head only) Stand close to exhibit but not in front of it Orientation toward visitors	Show supporting terms/visuals on screen Scan visitors every once in a while with pointer Stand close to exhibit but not in front of it Orientation toward visitors
Direct attention to the exhibit	Point into exhibit, supporting while saying "there" etc. Gaze to exhibit when pointing	Slightly turn toward exhibit Point into exhibit Gaze into exhibit	Point into exhibit with arrow and pointer
Finish at exhibit	Break eye-contact Look in the direction of next exhibit Point toward next exhibit Start moving in direction of next exhibit	Animation of eyes 'look downwards' Turn slightly on the spot, as if looking at all visitors, then find small gap in between visitors to drive through Point toward next exhibit with pointer Start moving in direction of next exhibit	Eyes start pulsing Pointer turns from visitors toward next direction Start moving in direction of next exhibit Keep pointing to next exhibit On screen "please follow me"

guide, we emphasized the parts of behavior that evoked positive experiences and limited the negative experiences. To do so, we selected the interactional outcomes to design multimodal behavior for based on expected positive visitor experiences. For example, visitors did not like too long explanations, thus the explanations would be kept short. For details on visitor experiences in tourist sites, see Reference [44].

In this article, we will mainly focus on the functional specifications of the robot as a starting point for the design of robot's behavior. To explore different options for the robot-optimized behavior set, we used a morphological chart [39]. This is a method taken from Industrial Design to systematically come to design options. By using this method, several ideas on how to use a modality to communicate the intention of the robot (interactional outcome) were generated.

Figure 3 is an example of one of the morphological charts we created. In this example, we generated different options for the behavior set for "finishing at an exhibit." For each interactional outcome, a morphological chart was created. For each modality of the robot, several ideas for behavior were generated. Together the designers evaluated which combination of behaviors of the different modalities would be the best option for the behavior set for each of the interactional outcomes. The resulting behavior sets were implemented in the robot. Examples of robot-optimized behavior are: The robot scans for visitors with its pointer camera, instead of making a full body movement; the robot uses two different types of animations for the eyes, one attention-attracting for the drive-mode and a more subtle one for the explain-mode; the screen of the robot did not continuously show a face but also showed additional information, instructions or intentions.

Using a morphological chart leads to one specific design for a set of behaviors while other combinations might also be effective and successful. The behavior set we have chosen and evaluated with several designers for the tour guide robot is one of many possibilities. In this article, we explore whether this alternative approach to designing behavior for social robots might be useful and how it influences the effectivity and understandability of social robots in interaction with people when compared to behavior designed following a human-translated approach. Independent of the specific set of behaviors chosen for the robot, it is important to create behavior that is consistent over time during the complete interaction between people and the robot.

5 METHODS FOR EVALUATION

To check whether participants responded differently toward the sets of behaviors, we prepared an online video study and an in-the-wild interaction study. The online video study allowed us to collect a more controlled response to the two sets of behavior of the robot and the in-the-wild observations were less controlled by nature but offer richer insights into people's experiences when encountering a service robot in a semi-public place, as we will explain below. Furthermore, Ju and Takayama suggested that the use of different experimental methods can address or mitigate potential limitations inherent in each method [41]. In this section, the methodology of each study will be reported separately, while the results of the two studies will be jointly reported in the results section, because both studies offer insight in different aspects of the hypotheses.

Video studies in Human-Robot Interaction research are an efficient way to collect data on visitors' perception of the robot, understandability of the robot behavior sets and visitor attitudes toward the robot [24, 50, 53, 76]. A video of isolated tour guide behaviors was not expected to provide an experience similar to experiencing an actual robot in an actual cultural site. However, by showing videos of the different sets of robot behaviors, we were able to find whether and which differences participants observed between the sets of robot behavior and how participants perceived and evaluated the different sets of behaviors in terms of (product) characteristics.
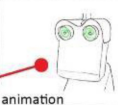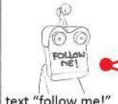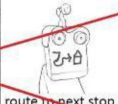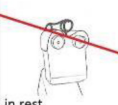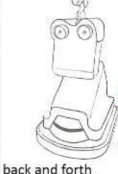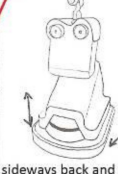
Fig. 3. Example of morphological chart.

In-the-wild studies in Human-Robot Interaction elicit valuable insights into how people will interact with robots in unstructured settings [65]. In our study, we observed reactions from visitors that encountered and joined the robot for a guided tour. Furthermore, some of these visitors were randomly invited to participate in a short interview, in which they were asked for their experiences. As a result, the in-the-wild study gave us the opportunity to find out how spontaneous visitors reacted to the robot and how they experienced a tour of FROG.

5.1   Manipulation of the Robot's Behavior in Both Studies

We developed human-translated behavior and robot-optimized behavior for the FROG robot. To control for the number of different behaviors, we developed the two sets of behaviors for the same interactional outcomes. In Table 1, we present two examples of interactional outcomes, how human tour guides perform them, and how we translated and optimized them for the robot. Note that these images only give an impression of the differences in the behavior sets, because the behaviors of the guide and the robot are very dynamic and difficult to grasp in stills. Table 1 presents the behaviors using images from the video study only, the behaviors in the in-the-wild study were similar. See also Appendix 1 for an exhaustive list of behaviors used in both studies, and see Appendix 2 for an overview of the different eye-animations of the robot.

5.2   Method of the Online Video Study

The video study had a between-subjects design with robot behavior design as the independent variable. For this purpose, we developed two videos, one with the human-translated robot behavior and the other with the robot-optimized behavior set.

5.2.1   The Robot. The robot that we used as a tour guide in both videos was a small 3D model (approximately 30cm in height) of the actual robot developed in the EU FP7 project FROG. We made this model to closely resemble the design of the robot used in the project, since we specifically wanted to understand the effect of behavior rather than that of appearance. The robot model

was made of cardboard, foam, colored paper, and plastic sheets. This model allowed us to use the following modalities of the robot simulate different behaviors in the videos: animated eyes, pointer on top, visuals on "screen," whole body movement and driving.

The robot itself would not talk, but in the video a voice-over was used for the explanations. The voice-over was a computer voice (female) with speech generated by a text to speech engine (NaturalReader with US Crystal voice). See Table 2 for an overview of the main differences between the human-translated and the robot-optimized behavior used for the robot in the online video study.
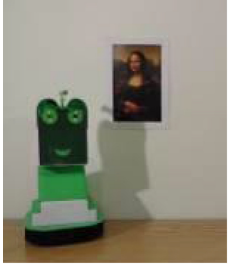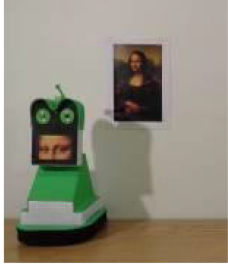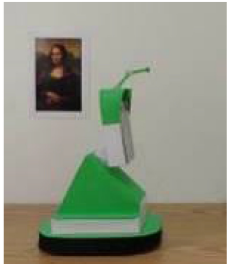
5.2.2 Participants. A total of 204 participants (132 males, 58 females, 14 preferred not to disclose, aged 15–58, mean age 31.6 years) evaluated the stimuli. All participants were recruited through Crowdflower, a crowdsourcing site, which allowed and stimulated the involvement of a wide range of participants. The participants came from different continents (most from Europe 42% and Asia 31%). Of the participants, 69.1% had no previous experience with social robots, 15.7% had little previous experience with social robots, 7.4% had much previous experience with social robots, 1% answered "other," and 6.9% did not answer the question. Of all participants, 146 participants (71.6%) rated their level of English "good" to "very good," 40 participants (19.6%) rated their level of English as "average," only 5 participants (2,5%) rated their level of English as "bad" to "very bad," and 6.5% of the participants did not answer the question.

5.2.3 Measures. In line with our research questions and hypotheses, we prepared a questionnaire that focused on participant's attitudes toward the robot behavior and on participant's attention. Measures that focused on participant's attitudes included constructs of the Godspeed questionnaire [4] and the Source Credibility Scale [64] as well as four items that were added to mask the intention of the questionnaire (obviousness, novelty, being qualified and reliability of the robot). The constructs of the Godspeed that were used were Anthropomorphism, Likeability, Animacy and Perceived Intelligence. The constructs of the Source Credibility Scale that we used were Sociability, Extraversion, Competence and Character. Three items ("artificial-lifelike," "unfriendly-friendly," and "unintelligent-intelligent") were asked once in the questionnaire, but used in two different constructs during analysis, because they appeared in constructs of the Godspeed and the Source Credibility Scale. Furthermore, the participants were asked to rate the behavior of the robot on the product personality scale [30, 55].

Questions that focused on participant's attention included nine questions that evaluated whether participants had understood and/or were distracted by the behavior of the robot and four questions to measure the recollection of details. We used information related to two paintings, which were the *Mona Lisa* by da Vinci and the *Girl with the Pearl Earring* by Vermeer. These paintings were chosen because many people are familiar with these paintings but would be interested to learn details about them that are not widely known. This way, the information provided would be new and interesting. Also, it allowed to test for recall of details afterwards. The actual recollection of details was measured by presenting the participants four multiple choice questions to test whether people had remembered the details of the story that the robot told (four options, of which one was always "I can't remember"). The story and questions were based on the story used for the study presented in Reference [45].

Participants who observed both videos were asked to answer four questions. One was a question on preference for one of both robot behavior sets, the other three were open questions examining why participants preferred a specific set of behavior and what differences they found between the behaviors. Participants who observed both videos were not invited to fill out the same questionnaire again for the second video, as participants would have been biased after seeing the first

Table 2. Overview of Differences Between Human-translated and Robot-optimized Behavior Sets

Main characteristics of human-translated behaviors set			
The robot shows a smiling mouth on the screen and looks toward the visitors.	When the robot points at the painting, it uses its pointer, turns in that direction and the eyes indicate that they look to the painting.	Sometimes the robot shows a picture on the screen, that it shows to all visitors by turning slightly on the spot.	When the robot drives to the next point, the pointer is in rest, and the screen shows a smiling mouth.
Main characteristics of robot-optimized behaviors set			
The robot is always oriented toward one point (the middle of a group of visitors) and scans for visitors with its pointer.	The robot points by using its pointer and an arrow on the screen. It keeps its original orientation (does not turn to all visitors).	The robot often shows details of objects and other pictures on the screen. It keeps its original orientation (does not turn to all visitors).	While driving, the eyes of the robot are animated in another way than during the explanations. Also, the pointer points at the next exhibit.

video, and they would already know the story, making questions about recall, for example, unreliable. However, remarks of these participants about whether they observed differences and what the differences were according to them gave us valuable information about the quality of our manipulations. Last, all participants were asked some demographical details, such as age, gender, education level, profession, and experience with social robots.

5.2.4 Procedure. Participants were asked to watch a video and to fill out a questionnaire about the behavior of a robot that gave a tour in a small art gallery. They were informed that the questions would be about the behavior of the robot and the information the robot gave.

Two short stop-motion videos (1:37min and 1:39min) were prepared; one for the human-translated condition and one for the robot-optimized condition. Both videos showed the robot presenting two artworks in a museum-like setting. For both videos, the story line was kept the same and the same sound file (a voice offering explanations about the paintings in English) was used. Only the multimodal behavior sets were different. Furthermore, the videos were controlled for robot activity to ensure that the robot was not more active in one of both conditions, as this seemed to be the cause of the found differences in a previous study [45].

The videos and online survey were distributed through a link on Crowdflower that led to Survey-monkey. On the first page of the questionnaire, the participants were asked for their consent. After that, the online questionnaire started with a video of either the robot in the human-translated condition or of the robot in the robot-optimized condition. The participants were randomly assigned to one of the two conditions. After seeing the video, the participants were asked to evaluate the behavior of the robot and how they experienced the behavior of the robot. The items in the questionnaire were presented on 5-point Likert scales or 5-point semantic differentials, unless stated otherwise. The items of the constructs and other questions in the questionnaire were randomized per page, unless order of the questions was important.

After the participants had finished the evaluation, they could decide to see the video of the other condition as well. Participants who decided to see the second video were asked to indicate which of the two sets of robot behaviors they preferred, and they answered three open questions about why they preferred a specific behavior, what the main differences between the robot's behaviors in the two videos were and whether they had any suggestions for the behavior of the robot. After seeing the second video and responding to the 4 items, participants were directed to the last page. When participants chose not to see the other video, they were redirected to the last page directly.

On the last page of the questionnaire, all participants were asked demographical information. The questionnaire contained 83 questions in total (participants had to fill in 79 questions when they did chose to not see the other condition). After participants completed the questionnaire they received a code to obtain a small payment for their participation.

5.2.5 Data Preparation. After preparing the data, a total of 204 participants were left (from the original 251 who started the questionnaire). The participants originally were evenly divided over the two conditions. However, some participants were removed from analysis, because they either did not complete the questionnaire or they responded with the same answer to all items, for example they always chose 5. Hence, due to the fact that 47 participants were removed from analysis, we had 114 cases for the human-translated condition, and 90 cases for the robot-optimized condition. Out of the remaining 204 participants, 166 (81% of all participants) compared both sets of robot behaviors.

5.2.6 Reliability of Measures. We used Cronbach's Alpha to check the reliability of our scales. A Cronbach's Alpha above 0.7 is considered reliable [60]. Therefore, we only were able to use the combined scales for the constructs Likeability, Perceived Intelligence, Sociability, and Understandability of robot behavior in the analyses. See Table 3 for an overview of the found Cronbach's Alphas for each of the constructs. Furthermore, we analyzed results for all individual items of the questionnaire.

5.2.7 Manipulation Check. As a manipulation check for differences between both behavior sets, one item asked the subset of participants who observed both sets of robot behavior, whether they

Table 3. Cronbach's Alpha for Constructs

Measurement	# items	Found Alpha α
Godspeed – Anthropomorphism	5	0.69
Godspeed – AnimacyMechanical – Organic was left out to increase the Cronbach's Alpha.	5	0.69
Godspeed – Likeability	5	0.83
Godspeed – Perceived Intelligence	5	0.70
Source Credibility Scale – Sociability	3	0.72
Source Credibility Scale – Extraversion	3	0.44
Source Credibility Scale – Competence	3	0.67
Source Credibility Scale – Character	3	0.67
Understandability of robot behavior	7	0.85

found any differences between both robot behaviors and if so, then which. In the comparison of both sets of robot behaviors, 37 did not answer the questions or gave answers unrelated to the questions and were therefore removed from the sample. Of the remaining 129 participants, 95 participants (73.6% of the participants who saw both videos) stated they saw differences in the robot behavior after seeing both videos. Thirty-four participants (26.4%) stated that they did not see any differences. The condition, age or nationality was not found to influence these results. Therefore, we concluded that the manipulation was merely successful.

In the following section, we describe the method of the in-the-wild study, before we report on the combined results of both studies.

5.3 Method of the In-the-wild Study

In the in-the-wild study the robot gave tours to spontaneous visitors in the Royal Alcázar in Seville, Spain.

5.3.1 The Robot. The robot used for the in-the-wild study was the FROG platform; see Figure 4. The different sets of behaviors were programmed on this platform. To drive around FROG was controlled remotely. The controller of the robot was the same for all sessions performed during the study. FROG had enough battery power to give tours for approximately 90min. Afterwards the robot needed to recharge for several hours.

The robot was designed to attract and keep the attention of the visitors using its body movements, the eyes (sequences played with the LED-lights in the eyes), the pointer on top of the robot (pointing at things and sequences played with the LED-lights in the pointer), a touchscreen that was used to show pictures and movies, and prerecorded synthesized speech. In this experiment, the visitors were not asked to use the touchscreen. Furthermore, the robot would not understand the visitors if they should talk to it. See Table 4 for an overview of the differences between the human-translated and the robot-optimized behavior used for the robot in the in-the-wild study.

5.3.2 Participants. All visitors of the Royal Alcázar were potential participants for the in-the-wild study, because they were all free to join or to ignore the robot. Based on chance, visitors participated in the study.

We estimated that in each condition about 300 participants joined the robot at some point during the tours (approximately 15 tours in human-translated condition, and approximately 14 tours in robot-optimized condition). We did this by counting participants who joined the FROG tour and interacted with the robot. Interactions with the robot could range from joining the robot for the

Fig. 4. FROG in the Hall of Festivities (in human-translated condition).

explanation of one point of interest to an entire tour of explanations at five points of interest or longer. Only visitors who joined the tour for at least one explanation were counted in this number. People who stood close to the robot, but also people who followed the tour from a distance, were counted.

5.3.3 Procedure. One short tour was prepared for the study, while two different sets of behaviors were prepared. In the first condition the robot communicated its intentions with human-translated behaviors, in which human behaviors were mimicked as closely as possible. In the second condition the robot showed robot-optimized behavior, in which the behavior was optimized for the modalities of the robot. The sets of behaviors were largely similar to the sets of behaviors used in the online video study. However, the behaviors were now performed by a functional robot and used in a different heritage setting. The multimodal behaviors were therefore performed at other moments in time than in the video study. The two sets of behaviors were controlled for activity of the robot to ensure that the robot was not more active in one of both conditions.

The tour consisted of five explanations at five different points of interest, which took about 20s. The total duration of the tour was 3–5min. Deviations in duration of the tours occurred, because the robot had to wait for visitors to step aside or to navigate around visitors present in the room to reach the next point of interest. See Figure 5 for a layout of the tour.

Table 4 shows an example of how an explanation was given at a point of interest in the tour, in the two conditions. In this example, in the human-translated condition the robot moves its body toward the visitors to attract attention, while in the robot-optimized condition, it uses its pointer to obtain a similar effect. The camera images clearly show visitor reactions (and therefore allowed us to observe differences in reactions between conditions), while the differences in robot eye and screen behavior are less well observable.

The robot was brought into the room at three timeslots on two different days. Before the robot started the guided tours, the researchers set up the camera and placed signs to provide visitors with information about the study and about consent. Visitors who joined the tour of the robot were not further informed about the study before the tour started. If people decided to enter anyway, then they automatically gave their consent for the use of the obtained video data for academic research.

Table 4. Overview of the Differences in one Explanation at One Point of Interest of the Tour Between Human-translated and Robot-optimized Behavior in the Wild

Example of one explanation at one point of interest using the human-translated behavior set
The robot turned its whole body slightly toward the different visitors while giving information. On the screen a mouth was visible and one explaining picture was shown for a few seconds. When the robot pointed at the exhibit, the pointer and eyes pointed there and the robot turned its whole body slightly in that direction.
Example of one explanation at one point of interest using the robot-optimized behavior set
The robot stood in a position and used the pointer to search for people. Before the explanation started, the screen showed "come closer." During the explanation, the screen was mainly black, only one explaining picture was shown for a few seconds. When the robot pointed at the exhibit, the pointer pointed in that direction, but the orientation of the robot did not change. After the explanation, the screen showed "follow me."

It was stated that this included use of the material for publications. If visitors did not wish to be recorded, then they were advised to contact one of the researchers who could be identified by their FROG project badge.

According to protocol, if this happened, then the camera would be stopped and recordings would be destroyed. However, none of the visitors made such a request.

A camera (that was within visitors' sight and not hidden) was used to record the visitors' reactions. The camera was placed at a height of 2m.

The main aim of the in-the-wild study was to observe visitors' reactions and interactions with the robot. Additionally, to improve our understanding of the observations, several visitors were interviewed about their experiences of following a robot guided tour. For the interviews, we randomly chose from the pool of visitors who joined the robot tour for at least one full explanation at one point of interest. The interviews were held in groups of one to three visitors who visited the Royal Alcázar together and joined the robot together. In the human-translated condition, 6 interviews (10 people) were recorded. In the robot-optimized condition, 11 interviews (16 people) were recorded.
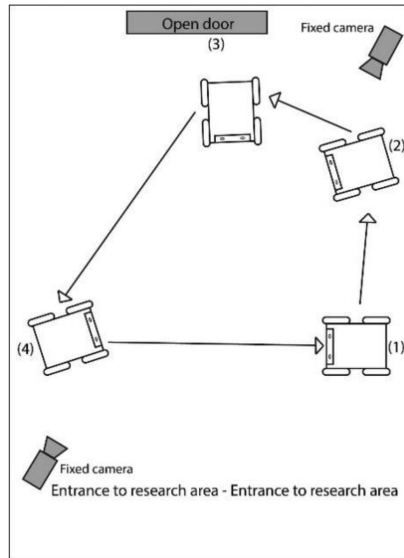
Fig. 5. Layout of the tour.

The interviews had a semi-structured set-up. Questions included: "Which aspect of the robot or the tour got your attention first?" "Can you describe your experience of following the robot guided tour, similar to how you would tell it to your friends and family at home?" "Can you describe the robot in three words?" "In what way is the robot guided tour different from information boards, audio guides or tour guides?" "Imagine the Royal Alcázar would decide to use these kinds of robots in other areas of the site as well, what would you think of that?" "Do you have suggestions for the robot or other remarks?" The interviews took approximately 2–5min. The interviews were recorded using a voice recorder when people gave consent to record, otherwise only notes were taken, which was the case for only one interview. The interviews were performed in English or occasionally in Dutch.

5.3.4   Data Analysis. We analyzed the video data by using DREAM (Data Reduction Event Analysis Method), a method we developed to annotate rich video data in a focused and fast manner and to anticipate for too early interpretations. The method is based on "thin slices of behavior" [1], "the grounded theory method" [16], and coding with multiple coders. For the analysis of the video data of in-the-wild Human-Robot Interaction, only "thin slices," sequences of three stills of the video of each of the robot's explanation at a point of interest, were used to analyze the data. Going from video data to sequences of stills implies a large reduction of data, which speeds up the data analysis. Nevertheless, the outcomes still are reliable.[2]

The advantage of using this method is that only the moments of interaction are analyzed, by the use of three images per action. This led to a series of sequences of three images to be annotated. The time that the robot was going from one point to the next was not taken into account in this analysis.

To analyze the sequences, two researchers (#1 and #2) discussed the definition of the codes and created an affinity diagram [17], on which the final code scheme was based. The found codes (53) were clustered into 10 categories, of which each category contained several codes. For example,

---

[2]The full description about the development and evaluation of DREAM can be found in Reference [42].

the category FORMATIONS contained 8 codes: 1 person as close as possible, 2 people as close as possible, 2 different groups of 2 or more people, group resembles a large guide group with more than 11 people, people stand far away, formation is a line, formation is a semi-circle, formation of group is unstructured (note: each of these codes should be used once when visitors perform these actions, and more than one code of this category can be applied to a sequence).

Atlas.ti [3] was used to code the data and the inter-rater reliability was calculated with CAT [52]. The main researcher (#1) annotated the dataset with the codes from the code scheme. Afterwards a third researcher (#3) was asked to annotate 25% of the sequences using the same codes. The annotations of researchers #1 and #3 were compared to calculate the inter-rater reliability. The inter-rater reliability was $\kappa = 0.61$ (Fleiss Kappa), which indicates that the codes applied were considerably consistent between researchers.

Overall, the robot gave a similar number of explanations at points of interests in both conditions (73 explanations at point of interests in human-translated condition and 71 explanations at points of interests in the robot-optimized condition), so, we could compare the numbers found with the annotation of the data. For the current study, we used the cumulative numbers of annotations of all explanations at points of interests taken together, because we wanted to find the general differences in visitor actions and reactions between the robot behavior sets.

Because of the small number of interviews, the interview data of the in-the-wild study was mainly used to illustrate the observations with experiences from real visitors. Nevertheless, analysis of the number of similar remarks made provided some insights in the reason why people chose to join or leave the robot.

5.3.5 Manipulation Check. To check for successful manipulation of the behavior sets, we searched for observable differences in the behavior of the visitors between conditions that we will present in the results section. As we found differences between the behavior sets in the video study, and the behavior sets for the video study and the in-the-wild study were similar of set up, we assumed visitors would observe these differences as well. Moreover, we do not have indications that these observed differences in visitors behaviors were introduced due other factors than the two sets of robot behavior. Therefore, based on the results of the online video study and the in-the-wild study, we state that we were largely successful in creating two different types of robot behavior.

6 RESULTS

To answer the main research question, we focused on two aspects. These two aspects are the *participant's attention* during the explanation of the robot at an exhibit and *participant's attitudes toward the robot* during the explanations at points of interests of the FROG guided tour. We analyzed these aspects through the questionnaire data of the online video study and observations as well as interviews of the in-the-wild study. In this section, we present the combined results of both studies.

6.1 Attention

H1: Robot behavior optimized for the modalities of a tour guide robot that does not closely resemble people will lead to longer periods of attention and higher recall of the information provided by the robot.

First, to gain insight into how visitor attention was influenced by the different behavior sets, we analyzed the quantitative data of the online video study. To check whether participants felt distracted by one of both behavior sets, we performed Mann-Whitney tests with the items of the attention scale as dependent variables and the condition as independent variable. We found no

significant differences in self-reports on how participants understood the robot's stories or participants' distraction or attention to the artworks between conditions.

Furthermore, we analyzed to what extent participants remembered correctly the details provided by the robot. We did this by summing up correct answers per participant, incorrect answers per participant, and "I can't remember" answers per participant. For each of the categories participants could score between 1 and 4 as there were 4 questions. Note that the sum of the three categories for each participant also was 4 (unless they did not fill out all questions). We used these items in Mann-Whitney tests to check for differences between conditions. We found that participants who observed the human-translated behavior set ($M = 2.18$, $SD = 1.44$) gave significantly more correct answers than participants in the robot-optimized condition ($M = 1.78$, $SD = 1.34$) $U = 4280$, $Z = -1.977$, $p = 0.048$. And vice versa, we found that participants in the robot-optimized condition ($M = 1.32$, $SD = 1.25$) gave significantly more incorrect answers than participants who observed the robot with the human-translated behavior set ($M = 0.91$, $SD = 1.03$) $U = 4178.5$, $Z = -2.291$, $p = 0.022$.

Second, we analyzed the actions of the visitors based on the coded video data of the in-the-wild study. To check which of both robot behavior sets kept the user attention for a longer time span we counted the number of explanations at points of interests that visitors joined the robot in each condition. In the human-translated condition, we observed more often that people left the tour after one explanation at one point of interest (23 times), compared with the robot-optimized condition (16 times). Contrarily, in the human-translated condition, we observed less often that people left the tour after two explanations at two points of interest (8 times), compared with the robot-optimized condition (15 times). Furthermore, in the robot-optimized condition, we observed visitors following the robot for more than five explanations at points of interest, which was more than one full tour (5 times), where we only found this one time in the human-translated condition. No differences between conditions were found for joining the robot for 3–5 explanations at different points of interest.

To check whether one of both behavior sets was more understandable to visitors, we counted in which condition visitors complied better with the requests of the robot to come closer. Those requests were verbally asked by the robot, but in the robot-optimized condition the request was also visible on the screen of the robot. We evaluated the compliance of the visitors by checking how visitors reacted to the robot's request to come closer. We found that when one visitor followed the robot, this visitor in the robot-optimized condition complied better to the request "please come closer" (23 times of 71 explanations at points of interest in this condition) than single visitors in the human-translated condition (6 times of 73 explanations at points of interest in this condition). We did not find differences for compliance of groups of 2 or more people to come closer to the robot.

Third, we analyzed the interview data of the in-the-wild study. To check whether the different behavior sets had influence on the understanding of the visitors, we analyzed whether visitors reacted differently to the question "did you understand the behavior of the robot?" We found that in both conditions visitors stated that they understood the behavior. However, in 5 interviews out of 11 in the robot-optimized condition, visitors explicitly added to their answers what they understood the robot's intentions and instructions were. In contrast, visitors in human-translated condition never added such remarks.

As a result, H1 is partly supported. The robot in robot-optimized condition was better able to keep the visitors attention for a longer time span, people seemed to comply more to the robot with robot-optimized behavior and people seemed to understand the robot-optimized behavior better, meaning that the first part of H1 is supported. However, visitors recalled less details of the story of the robot correctly when the robot used robot-optimized behavior, thus the second part of H1

was not supported. Overall, the robot-optimized and human-translated behaviors both score well on different aspects, indicating that both behaviors are suitable to use for a robot that does not resemble people.

6.2 Attitude Toward the Robot

H2: Robot behavior optimized for the modalities of a tour guide robot that does not closely resemble people will lead to more positive attitude toward the robot.

To gain insight into the aspects that influence visitors' attitudes toward the robot, we again started with the analysis of the online questionnaire data. To test whether one of both conditions was evaluated more positively, we performed Mann-Whitney tests with condition as independent variable and the reliable constructs of the Godspeed and Source Credibility scales as dependent variables. We did not find any statistical differences for the constructs of the Godspeed and the Source Credibility scale.

We also performed Mann-Whitney tests with condition as independent variable and the individual items of the Product Personality scale, and all individual items of the Godspeed and the Source Credibility scales as dependent variables. We found that the robot with the human-translated behaviors ($Mdn = 4$, $SD = 0.98$) scored significantly higher on the item 'serious' from the product personality scale than the robot with the robot-optimized behaviors ($Mdn = 3$, $SD = 1.04$) $U = 4220.5$, $Z = -2.283$, $p = 0.022$. Moreover, we found statistical differences on the item Mechanical-Organic from the Godspeed scale $U = 2789.5$, $Z = -5.8$, $p < 0.001$. Participants rated the robot with the human-translated behavior ($Mdn = 2$, $SD = 0.95$) as more mechanical and the robot with the robot-optimized behavior ($Mdn = 3$, $SD = 1.14$) as significantly more organic.

Furthermore, to check which robot behavior set participants preferred, we compared the answers to the corresponding question. We found that of the participants who observed both behavior sets, 76 participants (45.8%) preferred the robot with the human-translated behaviors, while 90 participants (54.2%) preferred the robot with robot-optimized behaviors. We did not find that age, gender, nationality, or education level had influence on these outcomes.

Moreover, to check what factors play a role in the preference for one of both behavior sets, we analyzed the answers to the open questions of the questionnaire data. We found no clear distinction for participants' preferences based on the answers to the open questions.

Second, we used the coded video data of the in-the-wild study to analyze the visitors' attitudes toward the robot behaviors. To check whether one of both behavior sets elicited visitors to behave more attracted to the robot, we counted the times that visitors took pictures of the robot. We found that in robot-optimized condition, visitors took a picture 20 times (of 71 explanations at points of interest in this condition), while in human-translated condition, visitors took a picture of the robot only 6 times (of 73 explanations at points of interest in this condition).

Finally, we analyzed the interviews to gain insight into the attitude of visitors toward the robot. To check whether the robot attracted the same kind of visitors between conditions, we asked them if they would prefer a human tour guide over the robot tour guide. We found that more frequently during interviews in the human-translated condition (2 out of 6 interviews), than during interviews in the robot-optimized condition (1 out of 11 interviews) visitors stated they would prefer a human tour guide. Even though we did not specifically ask for, in 5 out of 11 interviews, visitors in the robot-optimized condition mentioned that they really did not like human tour guides and the large groups, against no remarks like this during the interviews out of 6 with the visitors in the human-translated condition. Also, we found that in 4 out of 11 interviews, visitors in the robot-optimized condition mentioned that they liked the fact that they could leave the robot when they wanted to

go somewhere else or when they wanted to focus on something else, without offending the guide, while in none of 6 interviews did visitors in the human-translated condition make this remark.

Thus, H2 is partially supported, as we found only small differences between people's attitudes toward the different robot behaviors. The robot-optimized behaviors seemed to fit the robot better, and people liked a robot tour guide with robot-optimized behavior better, but when people observed both behavior sets for the robot, preferences for one of both behavior sets were evenly divided. Therefore, we argue that the robot-optimized behavior was slightly more preferred over human-translated behavior but did not lead to a significantly more positive attitude toward the robot.

7 DISCUSSION

In this section, we will first discuss the results of both studies, we then go on to discuss the implications of using the robot-optimized design approach to design behavior for robots that do not look like people, and finally we will discuss the use of mixed methods for the evaluation of robot behavior.

7.1 Discussion of the Results

The combined results of our study show that even though we only found small differences between both conditions, participants seemed to be slightly more positive toward the tour guide robot with robot-optimized behavior. Therefore, we argue that using a robot-optimized approach to develop behavior for robots that do not closely resemble people in appearance is promising next to using a human-translated approach.

People recalled more details correctly when the robot showed human-translated behavior. This might be because they recognized the human-translated behavior and were distracted by the robot-optimized behavior. Also, people were attracted for longer by the robot-optimized behavior, probably because it was new and surprising. This novelty effect might have influenced the results and should be further studied.

Nevertheless, we also found that the robot-optimized behavior was perceived as more organic for the robot. This may indicate that robot-optimized behavior better fits the morphology of a non-humanoid robot than human-translated behavior. Moreover, the robot-optimized behavior was perceived less serious compared to human-translated behavior and preferred by people who do not like human tour guides. This might explain why people who joined the robot guided tour reacted slightly more positive toward the robot with robot-optimized behavior.

In the studies, we did not find major differences on effectiveness of either the human-translated behavior set or the robot-optimized behavior set in terms of participant's attention and attitudes toward the robots. However, we had expected to find clear differences between the two behavior sets. This could be explained by the degree of humanlike-ness of FROG. Possibly, the modalities of FROG were not perceived as very distinct from human modalities (e.g., because FROG has dynamic eyes, and a pointer that could be interpreted as one arm). Hence, FROG's characteristics might have led to behavior sets that were perceived as being quite similar.

Based on this explanation, we expect that perceived differences between the behavior sets we applied may be more extreme when they are applied to robots that look much less human-like and use modalities that are very distinct from human modalities. For example, the use of light, sound, and projection might give more options to create a distinct design for robot-optimized behavior. Furthermore, the use of modalities that are not a replacement for human modalities might lead to more creative behavior design for robots. Using different robot platforms in a study similar to the studies presented here would further strengthen insights in the differences between human-translated behavior and robot-optimized behavior, and in the effectiveness of human-translated

behavior for robots that do not closely resemble people. Compared to other studies in which robot behavior was developed and evaluated, the differences we found in responses toward the two sets of behavior may seem marginal. However, there are two main differences between our study and previously performed studies.

First, multimodal behavior strongly influences Human-Robot Interaction. Vázquez et al., for example, show in their paper that gaze and orientation should be designed jointly [73]. Even though some researchers have implemented multimodal behaviors in their robots (e.g., References [37, 73]), knowledge on how to effectively design a set of intuitive and unambiguous multimodal behaviors for social robots is limited. In previous research, behavior cues have been mostly studied in isolation. For instance, it was studied how gaze [2], gestures [19], or body movement [48] separately influence Human-Robot Interaction. This has led to valuable insights, however, the interplay between modalities has not been a topic in these studies. When more modalities are used at the same time, which is almost always the case in "natural" behavior, the effectiveness of specific behaviors might be different than suggested in these kinds of studies. However, this manner of studying effectiveness of behavior also introduces a limitation as it becomes more difficult to understand what causes the differences influenced by the different behavior sets. Thus, a combination of both approaches might be necessary to fully understand the effect of robot behavior on Human-Robot Interaction.

Second, with this study, we venture beyond previous studies on behavior design for social robots (e.g., References [45, 48, 70]), because in these studies human-like robot functional interaction behavior is compared to random or no interaction behavior (control condition). As a result, the developed functional interaction behavior was in all situations perceived better than the lack of behavior or randomly performed behavior. Therefore, the only conclusion that can be drawn from these studies is that designing behavior for a robot is better than not designing behavior for a robot. In contrast, in this article, we explored behavioral cues in multimodal form and evaluated two carefully and functionally designed alternatives to find which set of behaviors would be preferred. Based on our findings, we concluded that both designed sets of behaviors (the human-translated as well as the robot-optimized behavior set) work for the robot. However the robot-optimized behaviors seemed to fit the non-humanoid robot FROG slightly better than human-translated behavior.

7.2   Design Approach to Develop Robot Behavior

The field of Human-Robot Interaction is a relatively young field. Because of this, methods and knowledge from other fields are used extensively. Human Robot Interaction strongly relies on knowledge from Social Sciences, Behavioral Sciences, Psychology, and Human-Computer Interaction. So far, the fields of product and interaction design have had a modest influence on Human-Robot Interaction, although there are some good examples of the contribution of industrial design on Human-Robot Interaction. Examples are the development and evaluation of the Hug [21], the development of the Snackbot [49] and a study in which new behavior for a Roomba vacuum cleaner was designed starting from a defined user experience [34]. These studies were all performed by product design or interaction design researchers. By presenting our alternative design approach to design robot behavior, we contribute to transferring knowledge from product and interaction design to Human-Robot Interaction and show what opportunities this field offers.

During our studies of state-of-the-art literature, we found that usually no methodological approach is used to create behavior for social robots. As a result, the multimodal behavior of the robot might be inconsistent and unpredictable. This is a problem, because several researchers state that robot behavior should be consistent over time [22, 46, 58, 74, 75]. Currently, creating consistent robot behavior is usually done by adding a human-like personality to the robot. This personality profile informs how behaviors should be performed (e.g., the use of a large amplitude for gestures

for an extravert robot). Despite the limitations, we used an imitation of human-like behavior to create human-translated behavior for the robot.

In the robot-optimized approach, we used the morphological chart to systematically develop behavior for the different modalities of the robot. The morphological chart makes the design choices visible which makes it easier for designers to discuss the best options. Additionally, the use of images instead of text only to describe behavior of the robot, helps designers to agree on what the behavior of the robot will look like and it will make early interpretation on how the behavior will be experienced more feasible. Furthermore, this way of working ensures the documentation of design choices throughout the project and more systematic improvements for the behavior. For these reasons, we found that the creation of the morphological chart and all possible options for robot behaviors was worth the effort, and we would use this method again when developing behaviors for a robot that does not closely resemble people or that lacks human-like modalities.

The behavior we developed for FROG was multimodal in both cases, meaning that the robot used more than one modality at the same time to communicate. Cha et al. argue that it is difficult to create multimodal redundant behavior, as non-humanoid robots only have a few communication channels [12]. Nevertheless, in human-human communication, but also in human-machine interaction (e.g., traffic lights changing position and color), often redundant signals are used to make sure everyone can understand the signal given. Therefore, we have chosen to create multimodal behavior that is redundant, so that people interacting with the robot can understand the behavior, even if they might miss some cues.

We argue that the added value of this article is to start designing robot behavior from the point of view of realizing interactional outcomes rather than copying humanlike behavior to a robot (as we showed in Figure 1). Actually, not the specific design for robot-optimized behavior as presented in this article is of interest to the field of Human-Robot Interaction, but the approach to determine the robot behavior that follows the morphology of the robot and starts with understanding interactional outcomes in human-human interaction. Hence, following the robot-optimized approach will lead to more flexibility in robot design and robot behavior design. Even though other designers/researchers might not come up with the exact behaviors as presented in the study, this alternative design approach for robot behaviors should lead to behavior that fits the robot better (e.g., perceived as more organic) and follows the morphology of the robot more naturally. Therefore, we conclude that robot-optimized behavior is a promising approach to design nonverbal behavior for robots that do not closely resemble people in appearance.

7.3   Use of Mixed Methods for Evaluation

To be able to draw results on the value of our alternative design approach to design behavior for robots, we used a combination of study methods. We are aware that both studies introduced some limitations.

First, the videos in the video study only took about 1:40min. Therefore, people may not have had enough time to interpret and understand certain behaviors. This might explain why people recalled more details correctly in the robot with human-like behavior condition, because people were more familiar with this behavior.

A limitation of the in-the-wild study is that the sessions of this study were performed in different timeslots during two days. In the robot-optimized condition the experiment was stopped due to technical issues and was continued for 45 minutes in the afternoon of the same day. This might have had some influence on the visitors' reactions toward the robot. However, as we found that the visitors' opinions about the robot behaviors did not change over sessions, we decided to use the interviews of all sessions in the analysis.

Additionally, with this article, we show how in-the-wild Human-Robot Interaction studies can be set up.

8   CONCLUSION AND FUTURE WORK

To conclude, we answer the question: "*Do people respond more positively to a robot displaying robot-optimized behavior compared with a robot displaying human-translated behavior?*" Based on the results of both studies, we found that the both methods to design robot behavior lead to understandable tour guide behavior for FROG. Nevertheless, the proposed robot-optimized behavior set for FROG follows the robot's morphology better and leads to slightly more positive attitudes and responses from visitors toward the robot. Therefore, we argue that this approach is promising to design behavior for robots and further exploration of the robot-optimized design approach seems valuable.

Our alternative approach to develop robot behavior was based on methods used in product design and interaction design. While this field is developing methods to design smart interactive products that people can intuitively interact with and establish relationships with, the field of Human-Robot Interaction is increasingly focusing on similar aims in the development of social robots. Therefore, we argue that these fields can mutually learn from one another.

As a result, the difference between functionally designed smart products, which include robots such as FROG, and biologically inspired social robots as was proposed by Fong et al. [26] is narrowing. This is because knowledge and approaches to develop robots and smart products become more widely accepted, understood and used. On the one hand, consumer products become more and more robotic, in a sense that they become more autonomous, take over tasks from people and communicate in a human-like manner, for example with speech. Consumer products that have become more robotic are robotic vacuum cleaners such as the Roomba, which people might perceive as social [27] and autonomous cars, which are originally designed in the fields of product design and interaction design, but now need more and more knowledge from the field of Human-Robot Interaction. On the other hand, a growing number of robots that were developed with knowledge from the field of Human-Robot Interaction become interactive consumer products. Think, for example, of the introduction of Aldebaran's Pepper[3] or Breazeal's Jibo[4], robots that are designed to assist, engage, and entertain people in their homes as a day-to-day companion. Consequently, this wider view enriches the development of social robots and interactive smart products.

We expect that more specific modalities, such as light, sound, and movement, will lead to better distinguishable robot optimized behavior. Future studies should focus on these aspects. Moreover, a focus on developing more robot-specific modalities will lead to consumer robots that not all closely resemble people. Therefore, we argue that developing robot-optimized behavior specifically for the modalities of a robot that does not closely resemble people in appearance is the future of robot behavior design.

---

[3]For more information, see https://www.aldebaran.com/en/cool-robots/pepper.
[4]For more information, see www.jibo.com.

APPENDIX
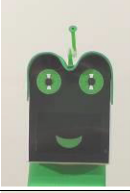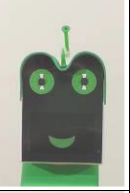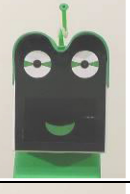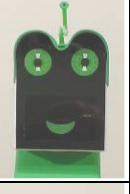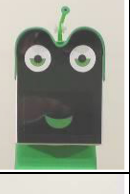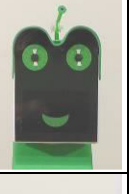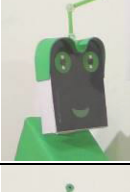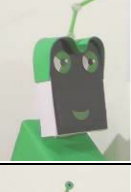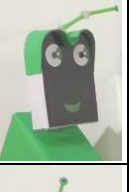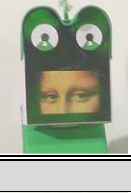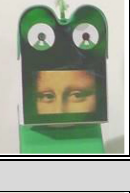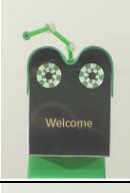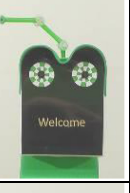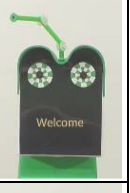
A.1   List of Behaviors in Both Studies

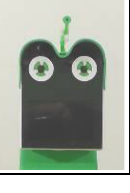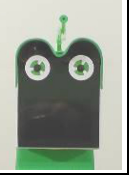| Interactional outcome	Human tour guide behavior [43]	Human-translated behavior	Robot-optimized behavior
Catch the attention of the visitors	Check for busy and empty spots near exhibit Make eye contact with all of the visitors Start with a louder voice to introduce themselves. Finish at first spot by breaking eye contact	Eyes blink once in a while Screen shows a mouth Turn on the spot to 'look' at all visitors Start with introduction of itself Pointer points at artworks when robot mentions them Finish at first spot with eyes look down	Eyes in "drive mode" (attention drawing eyes, busily animated) Pointer scans for visitors Screen shows "Welcome" Start with introduction of itself Pointer and arrow on screen point at artworks when robot mentions them Finish at first spot by showing "Follow me" on the screen
Start tour at first exhibit	Check if visitors follow the tour Start with trivial information Pick one visitor at a time to address	Eyes blink once in a while Screen shows a mouth Turn front toward visitors Start with trivial information through a prerecorded voice Check visitors by turning on the spot Pick one visitor to look at	Eyes in "exhibit mode" (understated animation) Scan visitors with pointer Start with trivial information through a prerecorded voice Turn front toward place visitors should position themselves for optimal viewing the exhibit
Offer information about exhibit Keep visitors engaged and check them to get implicit feedback	Gaze at one person and move gaze to the others intermittently	Eyes blink once in a while Screen shows a mouth Direct gaze at one person, and by turning on the spot alternate gaze-focus toward others	Eyes in "exhibit mode" Screes shows supporting visuals Scan visitors every once in a while by moving pointer
Direct the visitor's attention to the exhibit/ Keep the visitor's attention to the story	Point to exhibit, while using directing language ("there you can see") Gaze toward exhibit while pointing Show brought along visuals to support story (photos or objects) Depict story with gestures Stand close to the exhibit, but not in front of it Orientation of body toward visitors	Point to exhibit with pointer Gaze toward exhibit while pointing by turning on the spot and having eyes directed to artwork Shortly show additional supporting visual on screen Stand close to exhibit but not in front of it Orientation of body toward visitors	Point to exhibit with pointer and arrow on screen Show additional supporting visual on screen Stand close to exhibit but not in front of it Orientation of body toward visitors

Interactional outcome	Human tour guide behavior [43]	Human-translated behavior	Robot-optimized behavior
Prepare group to move attention Indicate direction of next exhibit Finish at exhibit Start moving to next exhibit	Change subject Tell visitors which exhibit they will move to next Point in direction of next exhibit Look in the direction of next exhibit Break eye-contact Start moving in direction of next exhibit	Eyes blink once in a while Screen shows a mouth Explain what is the next exhibit Shortly point in that direction with pointer Turn on the spot, as if looking at all visitors, then find small gap in between visitors to drive through Eyes look down Start moving in next direction	Eyes in "exhibit mode" Explain what is next exhibit Point in that direction with pointer Turn to next direction Pointer points to next exhibit and sometimes scans for visitors Eyes change to "driving mode" Screen shows "Follow me"
Guide visitors to next exhibit	Walk in front of group. Talk to visitors who are in front Only go for short distances	Eyes blink every now and then Screen shows mouth Move to next location	Eyes in "driving mode" Screen shows "Follow me" Move to next location Pointer keeps pointing and sometimes scans for visitors
Indicate having arrived at next exhibit	Stop waking Turn body toward visitors Make eye contact with all of the visitors	Stop driving Turn with front toward visitors Turn on the spot to look at all visitors	Stop driving Turn toward visitors Eyes in "exhibit mode" Pointer scans for visitors
End tour	Tell visitors tour is over Break eye contact	Tell visitors tour is over Eyes look down	Tell visitor tour is over Eyes calm down Pointer moves back to rest position Screen shows "Goodbye"

A.2   Eye Animations for Both Conditions

All pictures have a duration of 0.25s in the animations

Human-translated behavior						
General						
Blink						
Break eye contact						
Look to painting						
Look down						
Robot-optimized behavior						
While driving						
During explanations						

ACKNOWLEDGMENTS

REFERENCES

[1] Nalini Ambady and Robert Rosenthal. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychol. Bull* 111, 2 (1992), 256–274. DOI:https://doi.org/10.1037//0033-2909.111.2.256

[2] Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. 2014. Conversational gaze aversion for humanlike robots. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction.* 25–32. DOI:https://doi.org/10.1145/2559636.2559666

[3] Atlas.ti. 1993. Atlas.ti. Retrieved from www.atlasti.com.

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2008. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot* 1, 1 (2008), 71–81. DOI:https://doi.org/10.1007/s12369-008-0001-3

[5] Christophe Bartneck and Jodi Forlizzi. 2004. A design-centred framework for social human-robot interaction. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'04).* 591–594. DOI:https://doi.org/10.1109/ROMAN.2004.1374827

[6] Maren Bennewitz, Felix Faber, Dominik Joho, and Sven Behnke. 2007. Fritz, a humanoid communication robot. In *Proceedings of the 16th IEEE International Conference on Robot & Human Interactive Communication.* 1072–1077.

[7] Rainer Bischoff and Volker Graefe. 2002. Demonstrating the humanoid robot HERMES at an exhibition: A long-term dependability test. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems; Workshop on Robots at Exhibitions.*

[8] Dan Bohus, Chit W. Saw, and Eric Horvitz. 2014. Directions robot : In-the-wild experiences and lessons learned. In *Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems.* 637–644.

[9] Cynthia Breazeal. 2003. Toward sociable robots. *Rob. Auton. Syst* 42, 3–4 (2003), 167–175. DOI:https://doi.org/10.1016/S0921-8890(02)00373-1

[10] Mason Bretan, Guy Hoffman, and Gil Weinberg. 2015. Emotionally expressive dynamic physical behaviors in robots. *Int. J. Hum. Comput. Stud.* 78 (2015), 1–16. DOI:https://doi.org/10.1016/j.ijhcs.2015.01.006

[11] Wolfram Burgard, A. B. Cremers, Dieter Fox, Dirk Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. 1999. Experiences with an interactive museum tour-guide robot. *Artif. Intell.* 114, 1 (1999), 3–55.

[12] Elizabeth Cha, Yunkyung Kim, Terrence Fong, and Maya J. Mataric. 2018. A survey of nonverbal signaling methods for non-humanoid robots. *Found. Trends Robot.* 6, 4 (2018), 211–323. DOI:https://doi.org/10.1561/2300000057

[13] Elizabeth Cha, Maja Matarić, and Terrence Fong. 2016. Nonverbal signaling for non-humanoid robots during human-robot collaboration. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI'16).* 601–602.

[14] Jeong-gun Choi and Myungsuk Kim. 2009. The usage and evaluation of anthropomorphic form in robot design. In *Proceedings of the Design Research Society Conference: Undisciplined!* 331/1-331/14.

[15] Aurélie Clodic, Sara Fleury, Rachid Alami, Raja Chatila, Gérard Bailly, Ludovic Brèthes, Maxime Cottret, Patrick Danès, Xavier Dollat, Frédéric Eliseï, Isabelle Ferrané, Matthieu Herrb, Guillaume Infantes, Christian Lemaire, Frédéric Lerasle, Jérôme Manhes, Patrick Marcoul, Paulo Menezes, and Vincent Montreuil. 2006. Rackham: An interactive robot-guide. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'06).* IEEE, 502–509.

[16] Juliet M. Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qual. Sociol.* 13, 1 (1990), 3–21. DOI:https://doi.org/10.1007/BF00988593

[17] Catherine Courage and Kathy Baxter. 2005. *Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques.* Morgan Kaufmann. DOI:https://doi.org/10.1038/vital339

[18] Pieter M. A. Desmet, Juan Carlos Ortíz Nicolás, and Jan P. Schoormans. 2008. Product personality in physical interaction. *Des. Stud.* 29, 5 (2008), 458–477. DOI:https://doi.org/10.1016/j.destud.2008.06.003

[19] Elisabeth T. van Dijk, Elena Torta, and Raymond H. Cuijpers. 2013. Effects of eye contact and iconic gestures on message retention in human-robot interaction. *Int. J. Soc. Robot.* 5, 4 (2013), 491–501. DOI:https://doi.org/10.1007/s12369-013-0214-y

[20] Carl F. Disalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. 2002. All robots are not created equal : The design and perception of humanoid robot heads. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS'02).* 321–326.

[21] Carl F. DiSalvo, Francine Gemperle, Jodi Forlizzi, and Elliott Montgomery. 2003. The hug: An exploration of robotic form for intimate communication. In *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'03).* 403–408. DOI:https://doi.org/10.1109/ROMAN.2003.1251879

[22] Brian R. Duffy. 2003. Anthropomorphism and the social robot. *Rob. Auton. Syst.* 42, 3-4 (2003), 177–190. DOI: https://doi.org/10.1016/S0921-8890(02)00374-3

[23] Stephanie Embgen, Matthias Luber, Christian Becker-Asano, Marco Ragni, Vanessa Evers, and Kai O. Arras. 2012. Robot-specific social cues in emotional body language. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'12).* 1019–1025. DOI: https://doi.org/10.1109/ROMAN.2012.6343883

[24] Vanessa Evers, Talia Brodecki, and Pamela Hinds. 2008. Relational vs. Group self-construal: Untangling the role of national culture in HRI. In *Proceedings of the 3rd ACM/IEEE International Conference on Human robot Interaction (HRI'08).* 255–262.

[25] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *Proceedings of 4th International Conference on Social Robotics.* 199–208. DOI: https://doi.org/10.1007/978-3-642-34103-8_20

[26] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Rob. Auton. Syst.* 42, 3–4 (2003), 143–166. DOI: https://doi.org/10.1016/S0921-8890(02)00372-X

[27] Jodi Forlizzi and Carl F. Disalvo. 2006. Service robots in the domestic environment: A study of the roomba vacuum in the home. In *Design.* 258–265.

[28] Jodi Forlizzi and Shannon Ford. 2000. The building blocks of experience: An early framework for interaction designers. In *Proceedings of the 3rd Conference on Designing Designing Interactive Systems: Processes, Practices, Methods, and Techniques.* 419–423. DOI: https://doi.org/10.1145/347642.347800

[29] Raphaela Gehle, Karola Pitsch, and Sebastian Wrede. 2014. Signaling trouble in robot-to-group interaction. emerging visitor dynamics with a museum guide robot. In *Proceedings of the 2nd International Conference on Human-agent Interaction.* 361–368.

[30] Pascalle C. M. Govers. 2004. Product personality. Delft University of Technology.

[31] Birgit Graf and Oliver Barth. 2002. Entertainment robotics: Examples, key technologies and perspectives. *Safety* 6, 7 (2002), 8–12.

[32] Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *Am. J. Psychol.* 57, 2 (1944), 243–259.

[33] Paul Hekkert, Marc Mostert, and Guido Stompff. 2003. Dancing with a machine: A case of experience-driven design. In *Proceedings of the International Conference on Designing Pleasurable Products and Interfaces.* 114–119. DOI: https://doi.org/10.1145/782896.782925

[34] Bram Hendriks, Bernt Meerbeek, Stella Boess, Steffen Pauws, and Marieke Sonneveld. 2011. Robot vacuum cleaner personality and behavior. *Int. J. Soc. Robot.* 3, 2 (2011), 187–195. DOI: https://doi.org/10.1007/s12369-010-0084-5

[35] Pamela J Hinds, Teresa L Roberts, and Hank Jones. 2004. Whose job is it anyway? a study of human–robot interaction in a collaborative task. *Human-Computer Interact. Interact.* 19, 1 (2004), 151–181.

[36] Chien-Ming Huang, Takamasa Iio, Satoru Satake, and Takayuki Kanda. 2014. Modeling and controlling friendliness for an interactive museum robot. In *Proceedings of the Conference on Robotics: Science and Systems.*

[37] Chien-Ming Huang and Bilge Mutlu. 2014. Learning-based modeling of multimodal behaviors for humanlike robots. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction.* 57–64. DOI: https://doi.org/10.1145/2559636.2559668

[38] Caroline Hummels. 1999. Engaging contexts to evoke experiences. In *Proceedings of the 1st International Conference on Design and Emotion.* 39–45.

[39] John Chris Jones. 1992. *Design Methods.* John Wiley & Sons.

[40] Wendy Ju and Larry Leifer. 2008. The design of implicit interactions: Making interactive systems less obnoxious. 24, 3 (2008), 72–84.

[41] Wendy Ju and Leila Takayama. 2009. Approachability: How people interpret automatic door movement as gesture. *Int. J. Des.* 3, 2 (2009).

[42] Daphne E. Karreman. 2016. Beyond R2D2—The design of nonverbal interaction behavior optimized for robot-specific morphologies. Ph.D. Dissertation. University of Twente.

[43] Daphne E. Karreman, Betsy E. M. A. G. van Dijk, and Vanessa Evers. 2012. Contextual analysis of human non-verbal guide behaviors to inform the development of FROG, the fun robotic outdoor guide. *Hum. Behav. Underst.* LNCS 7559 (2012), 113–124. DOI: https://doi.org/10.1007/978-3-642-34014-7_10

[44] Daphne E. Karreman, E. M. A. G. van Dijk, and Vanessa Evers. 2012. Using the visitor experiences for mapping the possibilities of implementing a robotic guide in outdoor sites. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'12).* 1059–1065.

[45] Daphne E. Karreman, Gilberto U. Sepúlveda Bradford, Betsy E. M. A. G. van Dijk, Manja Lohse, and Vanessa Evers. 2013. Picking favorites: The influence of robot eye-gaze on interactions with multiple users. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems.* 123–128.

[46] Heeyoung Kim, Sonya S. Kwak, and Myungsuk Kim. 2008. Personality design of sociable robots by control of gesture design factors. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication.* 494–499. DOI : https://doi.org/10.1109/ROMAN.2008.4600715

[47] Yoshinori Kuno, Kazuhisa Sadazuka, Michie Kawashima, Keiichi Yamazaki, Akiko Yamazaki, and Hideaki Kuzuoka. 2007. Museum guide robot based on sociological interaction analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07).* 1191–1194.

[48] Hideaki Kuzuoka, Yuya Suzuki, Jun Yamashita, and Keiichi Yamazaki. 2010. Reconfiguring spatial formation arrangement by robot body orientation. In *Proceeding of the 5th ACM/IEEE International Conference on Human-robot Interaction (HRI'10).* 285–292. DOI : https://doi.org/10.1145/1734454.1734557

[49] Min Kyung Lee, Jodi Forlizzi, Paul E. Rybski, Frederick Crabbe, Wayne Chung, Josh Finkle, Eric Glaser, and Sara Kiesler. 2009. The snackbot: Documenting the design of a robot for long-term human-robot interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction (HRI'09).* 7–14.

[50] Manja Lohse, Frank Hegel, Agnes Swadzba, Katharina Rohlfing, Sven Wachsmuth, and Britta Wrede. 2007. What can I do for you ? Appearance and application of robots. In *Proceedings of the Workshop on the Reign of Catz and Dogz? The Role of Virtual Creatures in a Computerised Society.*

[51] Manja Lohse, Reinier Rothuis, Jorge Gallego Peréz, Daphne E. Karreman, and Vanessa Evers. 2014. Robot gestures make difficult tasks easier: The impact of gestures on perceived workload and task performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1459–1466. DOI : https://doi.org/10.1145/2556288.2557274

[52] Chi-Jung Lu and Stuart W. Shulman. 2008. CAT (Coding Analysis Toolkit). Retrieved from http://cat.ucsur.pitt.edu/.

[53] Karl F. MacDorman. 2006. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *Proceedings of the Long Symposium: Toward Social Mechanisms of Android Science.* 26–29.

[54] Masahiro Mori. 1970. The uncanny valley. *Energy* 7, 4 (1970), 33–35.

[55] Ruth Mugge, Pascalle C. M. Govers, and Jan P. L. Schoormans. 2009. The development and testing of a product personality scale. *Des. Stud.* 30, 3 (2009), 287–302. DOI : https://doi.org/10.1016/j.destud.2008.10.002

[56] Bilge Mutlu, Jodi Forlizzi, and Jessica Hodgins. 2006. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proceedings of the IEEE-RAS Conference on Humanoid Robots (Humanoids'06).* 518–523.

[57] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *J. Soc. Issues.* 56, 1 (2000), 81–103. DOI : https://doi.org/10.1111/0022-4537.00153

[58] Donald A. Norman. 2004. Emotional design: Why we love (or hate) everyday things. Basic Books.

[59] Illah R. Nourbakhsh, Clay Kunz, and Thomas Willeke. 2003. The mobot museum robot installations: A five year experiment. In *Proceedings of the IEEE/RJS International Conference on Intelligent Robots and Systems.* 3636–3641.

[60] Jum C. Nunnally and Ira H. Bernstein. 1967. *Psychometric Theory.* McGraw, New York.

[61] Yusuke Okuno, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Providing route directions: design of robot's utterance, gesture, and timing. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI'09).* 53–60. DOI : https://doi.org/10.1145/1514095.1514108

[62] Byron Reeves and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, New York.

[63] Diego Rodriguez-Losada, Fernando Matia, Ramón Galan, and Agustín Jimenez. 2002. Blacky, an interactive mobile robot at a trade fair. In *Proceedings IEEE International Conference on Robotics and Automation.* 3930–3935. DOI : https://doi.org/10.1109/ROBOT.2002.1014341

[64] Rebecca B. Rubin, Philip Palmgreen, and Holward E. Sypher. 2004. *Communication Research Measures; A Sourcebook.* Lawrence Erlbaum Association, Publishers, London.

[65] Selma Sabanovic, Marek P. Michalowski, and Reid Simmons. 2006. Robots in the wild: Observing human-robot social interaction outside the lab. In *Proceedings of the 9th IEEE International Workshop on Advanced Motion Control.* 576–581.

[66] Selma Sabanovic, Sarah Reeder, and Bobak Kechavarzi. 2014. Designing robots in the wild: *In situ* prototype evaluation for a break management robot. *J. Hum-Robot Interact.* 3, 1 (2014), 70–88. DOI : https://doi.org/10.5898/JHRI.3.1.Sabanovic

[67] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *Int. J. Soc. Robot.* 5, 3 (2013), 313–323. DOI : https://doi.org/10.1007/s12369-013-0196-9

[68] Maha Salem, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2011. A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication.* 247–252. DOI : https://doi.org/10.1109/ROMAN.2011.6005285

[69] Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2006. Interactive humanoid robots for a science museum. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction.* 305–312.

[70] Candace L Sidner, Cory D Kidd, Christopher Lee, and Neal Lesh. 2004. Where to look: A study of human-robot engagement. In *Proceedings of the 9th International Conference on Intelligent User Interfaces*. 78–84. DOI:https://doi.org/10.1145/964442.964458

[71] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2014. Communication of intent in assistive free flyers. In *Proceedings of the 9th ACM/IEEE International Conference on Human-robot Interaction (HRI'14)*. 358–365.

[72] Sebastian Thrun, Maren Bennewitz, Wolfram Burgard, , Frank Dellaert, Dieter Fox, Dirk Hahnel, Charles Rosenberg, Nicholas Roy, Jamie Schulte, and Dirk Schulz. 1999. MINERVA: A second-generation museum tour-guide robot. In *Proceedings of the IEEE International Conference on Robotics and Automation*. DOI:https://doi.org/10.1109/ROBOT.1999.770401

[73] Marynel Vázquez, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E. Hudson. 2017. Toward robot autonomy in group conversations: Understanding the effects of body orientation and gaze. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 42–52.

[74] Michael L. Walters, Dag S. Syrdal, Kerstin Dautenhahn, René te Boekhorst, and Kheng Lee Koay. 2008. Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots*. 24, 2 (2008), 159–178. DOI:https://doi.org/10.1007/s10514-007-9058-3

[75] Sarah N. Woods, Kerstin Dautenhahn, Christina Kaouri, René te Boekhorst, and Kheng Lee Koay. 2005. Is this robot like me? links between human and robot personality traits. In *Proceeding of the 5th IEEE-RAS International Conference on Humanoid Robots (Humanoids'05)*. 375–380.

[76] Sarah N. Woods, Michael L. Walters, Kheng Lee Koay, and Kerstin Dautenhahn. 2006. Methodological issues in HRI: A comparison of live and video-based methods in robot to human approach direction trials. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*. 51–58.

[77] Akiko Yamazaki, Keiichi Yamazaki, Yoshinori Kuno, Matthew Burdelski, Michie Kawashima, and Hideaki Kuzuoka. 2008. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 131–139.