BEHAVE: A Set of Measures to Assess Users' Attitudinal and Non-verbal Behavioral Responses to a Robot's Social Behaviors

Michiel Joosse[1], Aziez Sardar[1], and Vanessa Evers[2]

[1] Informatics Institute, Faculty of Science, University of Amsterdam,
PO Box 94323, 1090 GH Amsterdam, The Netherlands
{mjoosse,aziez}@science.uva.nl
[2] Department of Electrical Engineering, Mathematics and Computer Science,
University of Twente.
PO Box 217, NL-7500 AE Enschede, The Netherlands
v.evers@utwente.nl

Abstract. Increasingly, people will be exposed to social robots. In order to inform the design of behaviors for robots that share domestic and public spaces with humans, it is important to know what robot behavior is considered as 'normal' by human users. The work reported in this paper stems from the premise that what would be perceived as socially normative behavior for robots may differ from what is considered socially normative for humans. This paper details the development of a set of measures, BEHAVE, for assessing user responses to a robot's behavior using both attitudinal and physical responses. To test the validity and reliability of the BEHAVE set of measures, a human robot interaction experiment was conducted in which a robot invaded the personal space of a participant. Based on the results from this evaluation, a final set of BEHAVE measures was developed.

Keywords: Human-robot interaction, proxemics, humanlike robots, avoiding behaviors.

1 Introduction

In the future, we can expect that more and more people will encounter robots and may need to interact or work with robots on a regular basis. Where robots were previously confined to industrial environments such as car manufacturing plants, future robots are envisioned to operate in the home [15] and in public spaces. The latter are very different environments for human robot interaction [12]. Industrial environments such as a car assembly line are designed to optimize the robots' operations. In contrast, robots in domestic environments will need to operate in a world specifically designed for humans. As people start sharing their social spaces with robots, the question arises about how they should interact socially. Human social interaction is governed by social norms [12] that dictate what distance to keep, when to acknowledge someone's

B. Mutlu et al. (Eds.): ICSR 2011, LNAI 7072, pp. 84–94, 2011.

presence, engage in eye-contact, approach, start speaking, smile and so on. Social norms differ across cultures and even for people of the same cultural background, interpretation of social behaviors is often fraught with misunderstanding [16].

It is yet unclear whether human social norms are transferable to human robot interaction. Even though there is some initial research that investigates the effect of social normative behavior displayed by robots on human responses to the robots. For instance, [18] investigated US and Chinese responses to explicit and implicit communication, such research has not yielded a complete set of socially normative behaviors for robots that interact with humans.

We aim to contribute to the knowledge in this area by identifying a comprehensive overview of socially normative behaviors for robots. The work reported in this paper concerns the development of a set of measures to evaluate human responses to robot behaviors in order to assess whether they are socially normative. Our aim is to be able to measure people's subjective attitudinal as well as more objective behavioral responses to robot behaviors. With these measures we will then identify which behaviors are most appropriate for human robot interaction. The remainder of this paper will first explain how we developed an initial set of measures. We will then detail how we evaluated the validity and reliability of the measures in an experiment investigating the socially normative behavior of personal space invasion. Finally, we will provide the final set of measures that can be used to measure subject responses to robot social behavior to determine which behaviors are socially normative for human robot interaction.

2 Theoretical Background

2.1 Relevant Measures to Study Human Responses to Social Behaviors of Robots

When evaluating whether a robot adheres to human social norms we have chosen the following approach:

1. From social psychological and behavioral science literature we identify human social behavioral norms (for instance preferred interpersonal space).
2. In an experimental setting, we expose subjects to a human confederate as well as a robot displaying this behavior and either adhering to or violating a specific social norm.
3. Measuring the humans' subjective (attitudinal) and objective (behavioral) response to determine whether adhering to or violating the social norm influences subjects' attitudes and behaviors positively or negatively.

2.2 Attitudinal Response

In order to measure the attitudinal and behavioral responses, we derived a set of measures based on literature which we will discuss here. Attitudinal measures include perceived human likeness of the robot, attitudes toward robots, trust in the robot, perceived social skills of the robot, and physical and social attraction of the robot. Behavioral measures include body responses such as leaning away, stepping away and facial feature responses such as smiling or looking scared. The origin of each measure is detailed below.

Previous studies have found that humanlike robots elicited different responses than more mechanical looking robots because people saw them as more human or because people had higher expectations of more human-looking robots [15]. It is possible that people expect higher conformity to social norms from more humanlike robots due to the closer human resemblance. One of the measurements to include is therefore the **perceived human-likeness** of the robot. For our set of measures we included a scale devised by Ho & MacDorman [8]. This is a 7-point Likert-scale consisting of six items, for example 'human-made – humanlike'. The only adjustment we made was re-adding the item deleted by Ho and MacDorman "genderless – male or female".

When evaluating people's responses to robots, it is important to consider whether people are predisposed to like or dislike robots in general. Results of a study by Wang et. al. [19] suggested that a more negative attitude toward robots may influence a person's tendency to adhere to a robot's recommendations. In order to measure whether people have a generally negative or positive attitude to robots we included the Negative **Attitude toward Robots** Scale on a 7-point Likert-scale as reported by Nomura. [13]. We have excluded three out of the fifteen original items, these items seamed redundant and somewhat ambiguous. Thus leaving the twelve items as reported in the results section.

Trust is an important factor in the use of social autonomous systems [3]. Because trust conveys a lot about the users' attitudinal response towards the robot, we included a measure of trust: the 7-point Likert- Source Credibility Scale [9] developed by McCroskey, et.al. The 3-item subscale 'competence' was not used. These questions concerned the fulfillment of a task, which is not always relevant for interaction with service robots in domestic/public environments. Because of the same reason the two 'communication items' in the subscale 'extroversion' were not used either. The subscales 'sociability', 'composure' and 'character', with items like "calm-anxious" and "tense-relaxed" were adapted to measure responses to a robot.

Because **likeability and attraction** are found to greatly influence the outcome of human as well as human-robot interaction [11], [14] we included items from the Interpersonal Attraction scale, developed by McCroskey & McCain [10] to measure the likeability of the confederate. This originally is a 15-item 7-point Likert-scale. We used the original 10 items from the 'social-' and 'physical attraction' subscales and adapted the questions to measure responses to a robot, for example: "I think the robot could be a friend of mine".

Perceived social skills are also considered important, because higher social skills could lead to higher conformity to social norms. To measure these skills, we included a five-item scale developed by Wish & Kaplan [20] to measure social skills. This 9-point bipolar scale was modified to a 7-point Likert scale.

2.3 Non-verbal Behavioral Response

In order to measure non-verbal behavioral responses more objective by analyzing video material of people interacting with robots, we developed a tool with which coders are able to evaluate the videos. This tool consists of items concerning: immediacy cues, facial expressions, body movement and overall behavior of the user.

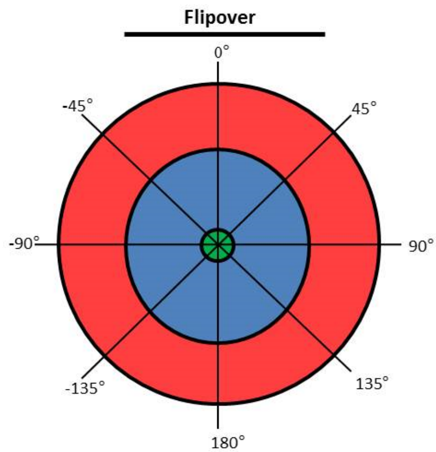

Fig. 1. Step distance and step direction

The immediacy cues consist of step distance and step direction. We developed a “wheel”, in which the step direction could be given in eight different degrees as depicted in figure 1. The distance could be either no step, a step within one’s intimate zone (less than 45 cm) or a step outside the intimate zone.

Black & Yacoob [2] defined motion cues for each of the six universal emotions happiness, sadness, surprise, fear, disgust and anger. We included the items as 5-point Likert-scaled questions in order to check if these emotions appeared, for example “raising mouth corners”. We hypothesized that gaze, staring eye contact would be important, as it has several communication functions, like establishment and recognition of a social relationship [1]. This includes smiling, by raising one’s mouth corners. Therefore, we included eight items, for example: “participant looked away”, “participant made eye-contact” and “mouth corners raised”.

Body posture is also indicates someone’s interest. Guerrero defined leaning and touching as immediacy / involvement cues [5]. We included three items, based on the work of Guerrero, these were “the participant leaned away from the confederate/ robot”, “participant leaned towards / from the confederate/robot” and “participant was touching him/herself”.

The overall emotions were coded by the items defined by Guerrero [5]. We included five 5-point Likert-type items, with which we used the coders’ ability of determining the mental state of mind of the user. Examples of these items are: “anxious – calm”, “distracted – focused” and “restless – still”.

3 An HRI Experiment to Test the Validity and Reliability of the Measures

In order to test the validity and reliability of our measures we administered them in a human-robot interaction experiment. The experiment concerned the socially normative behavior of maintaining personal space, an specific example of social normative behavior is interpersonal distance. People keep what they see as an appropriate distance from each other. This leads them to stand further apart when standing in a big, open room than in a crowded elevator. This distancing behavior is called proxemics, first defined by Edward T. Hall as "the physical and psychological distancing from others"[6]. The personal space cannot be invaded without causing some sense of discomfort [6, 7]. Four different personal space zones (See table 1) have been defined which tend to hold true for most people, although there are factors which influence proxemics, such as gender and social (cultural) norms [1].

Table 1. Personal space zones as defined by Hall. [6]

Personal space zone	Range
Intimate zone	0 - 0.45 m.
Personal zone	0.45 - 1.2 m.
Social zone	1.2 - 3.6m.
Public zone	3.6m +

Recently, Mumm & Multu [11] carried out a study where subjects were instructed to retrieve something from behind the robot that was positioned in the centre of a room. The robot either followed the subject with its gaze (turned its head to follow the movement of the subject as they walked past the robot to retrieve the item at the back of the robot) or kept its gaze fixed to a specific point in the room. They found that subjects kept a larger distance from the robot in the condition where the robot followed the subject with its gaze. The findings suggest that a robot that shows more humanlike behaviors or that seems aware of the humans around it is given more personal space. Even though this is an extrapolation from their conclusions, it seemed as if the subjects in Mumm and Mutlu's study were adhering to a social norm to give more room to a being that was aware of their presence. Other research found that women maintained a larger distance from robots than men when the robot was looking at their face [17]. Similarly, a robot's appearance was found to influence proxemic behaviors. In research by Syrdal et.al. [15] people allowed a mechanical robot to approach them closer (57 cm.) than a humanoid robot (52 cm.) with more anthropomorphic attributes. This could be due to the fact that humanlike robots lead to higher expectations of conformity to social norms. No difference between men and women was reported.

Another factor that could influence the (dis)comfort of the personal space invasion is approach speed. In earlier work from Dautenhahn & Walters a speed of 0.4 m/s was considered a good speed, if perhaps a little too slow [4] when approaching a seated person. On the contrary, a speed of 1.0 m/s was considered as being too fast [18], when approaching a standing person.

For the experiment where the BEHAVE measures were administered, we expected that people would have higher expectations of humans than robots concerning adhering to social norms. We therefore thought that when a human invaded a subject's personal space compared with when a robot would invade the personal space, people would have more negative attitudes toward the human and would display more avoiding behaviors (negative body and facial behaviors). We also expected this effect would be stronger in the case that the approach was faster than what is considered a socially normative approach speed.

3.1 Participants

A total of 92 participants (52 men and 40 women) participated in this study, aged between 18 and 70. (M=24.8, SD = 9.5). 85% of the participants were from the Netherlands, 7% from other European countries and 8% from countries outside Europe. 79% of the participants indicated they (previously) owned a pet. Of the 47 participants who interacted with the robot, 42% of the participants indicated that they had no prior experience with robots and 12% had built robots themselves.

3.2 Methods

We conducted a controlled 2 (human confederate vs. robot confederate) x 2 (comfortable approach vs. uncomfortable approach) between-group laboratory experiment. Depending on the condition, either a robot or a human approached a subject who was viewing a poster and subsequently, the agent invaded the personal space of the subject. The subjects' behaviors were videotaped and later analyzed. The subjects' attitudes were assessed by a post session questionnaire. Only the participants who interacted with the robot were given the human likeness [8] and attitude towards robots questions [16].

Fig. 2. Modified Nomad Scout

The robot used was a modified Nomad robot (see figure 2), with a height of 140 cm. and a diameter of 40 cm. The robot was operated with a joystick from an adjacent room. Depending on the gender of the participant, the confederate was either a men or woman, of average size and posture and wore a white shirt. The approach speed was either comfortable, normal (0.4 m/s), or uncomfortable, fast (1.0 m/s). Behavioral responses were collected using two video cameras: one positioned above the flipover with poster and one in the corner of the experiment room. See figure 3 for the experiment lab layout.

3.3 Experimental Procedure

After being welcomed, informed of the procedure and having signed a consent form, the participants were led to the experiment room. In the corridor they were told that they should not mind the technical equipment (robot condition), or the "other participant" (human condition). Invasion of personal space was not mentioned to the participants. During the experiment, the participants were instructed to search for figures on a poster. After a minute, the human or robot confederate would move towards the participant, and invade the participant's personal space zone. After entering the intimate personal space zone (45 cm.), the agent would turn its "head" towards the participant's poster. The behavior of the participant was observed, after which a debriefing consisting of a questionnaire and brief interview questions took place.

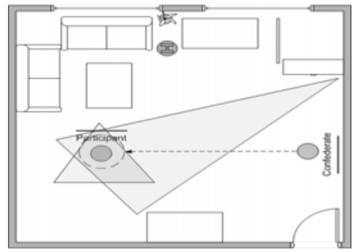

Fig. 3. Experiment lab layout

The video data from our two cameras (see figure 3), the behavioral measurements, were coded by the categories facial expression, intimacy cues and immediacy. Each pair of videos was coded by three coders. There were a total of 30 different coders. None of the coders knew the participants, the coding was done anonymously. The Intraclass Correlation Coëfficient (ICC) was used to calculate intercoder reliability.

4 Results

We will report the reliability of the measures and the final scales, results of the experiment itself will be published elsewhere. Together, the final scales as reported in table 2 and table 3 were included in the final BEHAVE measurement tool.

Table 2. Attitudinal measures: final items and their reliability

Attitude towards robots - NARS [13]	Chronbach's
Subscale Interaction:	α = .708
I would feel uneasy if I was given a job where I had to use robots.	
I would hate the idea that robots or artificial intelligences were making judgments about things.	
I would feel very nervous just standing in front of a robot.	
Subscale Social:	
I would feel uneasy if robots really had emotions.	
Something bad might happen if robots developed into living beings.	
I feel that if I depend on robots too much, something bad might happen.	
I am concerned that robots would be a bad influence on children.	
I feel that in the future society will be dominated by robots.	
I feel that in the future, robots will be commonplace in society.	
Subscale Emotion:	
I would feel relaxed talking with robots.	
If robots had emotions, I would be able to make friends with them.	
I feel comfortable being with robots.	
Attractiveness of the robot- Interpersonal Attraction Scale (physical) [10]	Chronbach's
I think the robot is quite handsome	α = .820
The robot is very sexy looking	
I find the robot very attractive physically	
I don't like the way the robot looks	
The robot is somewhat ugly	

Table 2. (*Continued*)

Likeability of the robot - Interpersonal Attraction Scale (social) [10]	Chronbach's
I think the robot could be a friend of mine	α = .777
It would be difficult to meet and talk with the robot	
The robot just wouldn't fit into my circle of friends	
The robot and me could never establish a personal friendship with each other	
I would like to have a friendly chat with the robot	
Human likeness - Ho & MacDorman [8]	Chronbach's
Artificial – Natural	α = .821
Human-made – Humanlike	
Without definite Lifespan – Mortal	
Inanimate – Living	
Mechanical Movement – Biological movement	
Synthetic – Real	
Genderless – Male or Female	
Trust in the robot- Social Credibility Scale [9]	Chronbach's
Good natured - irritable	α = .789
Cheerful-gloomy	
Poised-Nervous	
Tense-Relaxed	
Calm-Anxious	
Dishonest-Honest	
Unsympathetic-Sympathetic	
Good-Bad	

The internal reliability of the attitudinal measures was high, except for one, the social credibility scale (Chronbach's α = .48). After removing the items "friendly-unfriendly" and "timid-bold", the reliability became acceptable (Chronbach's α = .789). The final items and their internal reliability are included in table 3. The internal reliability of the social skills measure was very low (α < 0.5); therefore this measure was excluded from BEHAVE.

For the behavioral variables to measure avoidance behavior, included in table 3, intercoder reliability was low for many of the items. The items with a high α (indicating that the coders may have interpreted the responses roughly the same) were combined into one independent variable, called "avoiding behavior" (Chronbach's α = 0.836); a variable from 1 to 5, with 1 being "no avoiding behavior" and 5 "a lot of avoiding behavior". The ICC of the step direction variable was high (.829, α = .936), as well as that of the step distance (.878, α = .956).

Table 3. Avoidance behavior measure: final items and reliability

Item	
Participant made eye-contact	α = .836
Did the participant say anything the moment that PSI occurred?	
Participant laughed out loud	
Participant was distracted from his/her task	
Mouth corners raised	
Participant leaned away from the confederate/robot	
Eyebrows raised	
Eyes open wide (to expose more white)	

5 Discussion and Conclusion

In this paper, we presented a tool for measuring socially normative robot behavior. The variables human likeness, attraction, likeability and attitude towards robots are all included to be able to explain differences in response to a robot's behavior in an experimental setting. The variables trust and avoidance/engaging behavior are the dependent variables of interest in order to measure whether certain robot behaviors are considered socially normative. We believe the set of measures as well as the video-coding method will prove to be an extremely valuable resource in the near future as it allows researchers to assess the quality of their robot's behavior in an experimental setting. This indicates, in our opinion the extent to which the behavior is experienced as socially normative by the users.

The scales showed high internal reliability and the question that remains is whether the scales are valid and indeed measure perception of socially normative behavior. Because the trust in robots scale is based on the previously validated McCroskey source credibility scale, its validity is more easily argued for. The behavior scale is more difficult to assess the validity of. When looking at the final scales, the questions seem to address (positive) engagement or (negative) avoidance behavior. Even though this is the first study to assess the BEHAVE set of measures, we think that engagement is a valid indication of users' behavioral responses to robots and we expect that future research will confirm this. Even though we believe these results are sound, replication of this experiment would prove useful to determine the external reliability.

We adopted both attitudinal and behavioral measures. Of the behavioral measures the immediacy cues evaluated by the coders had the highest intercoder reliability. We believe that it is very difficult for coders to assess subtle facial feature movements and possibly technological advances in facial recognition can be used in the future to objectively detect emotional responses.

The attitudinal measures (NARS, attraction, likeability, human likeness and trust) were very reliable. We tried to measure the social skills but it resulted in an incorrigible low Alpha. The cause of this was probably because the questions were multi-interpretable. In spite of this, we still think it is imperative to measure the believed social skills of the social robot. Future work will focus on developing a set of items to measure the perceived social skills of the robot as it will be an indication of how socially able users found the robot and possibly, how socially normative the robot is.

References

1. Argyle, M., Dean, J.: Eye-Contact, Distance and Affiliation. Sociometry 28(3), 289–304 (1965)
2. Black, M.J., Yacoob, Y.: Tracking and Recognizing Rigid and Non-Rigid Facial Motions using Local Parametric Models of Image Motions. In: Fifth International Conference on Computer Vision, pp. 374–381 (1995)
3. Cramer, H., Evers, V., van Someren, M., Ramlal, S., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B.: The effects of transparency on trust and acceptance in interaction with a content-based art recommender. User Modeling and User-Adapted Interaction 5(18), 455–496 (2008)
4. Dautenhahn, K., Walters, M., Woods, S., Koay, K.L., Nehaniv, C.L., Sisbot, E.A., Alami, R., Siméon, T.: How I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, pp. 172–179. ACM Press, New York (2006)
5. Guerrero, L.K.: Observer Ratings of Nonverbal Involvement and Immediacy. In: Manusov, V.L. (ed.) The Sourcebook of Nonverbal Measures. Psychology Press, Philadelphia (2004)
6. Hall, E.T.: The Hidden Dimension. Doubleday, New York (1966)
7. Hayduck, L.: Personal Space: Where We Stand Now. Psychological Bulletin 94(2), 293–335 (1983)
8. Ho, C.C., MacDorman, K.F.: Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. Computers in Human Behavior 26(6), 1508–1518 (2010)
9. McCroskey, J.C., Jensen, T., Valencia, C.: Measurement of the credibility of peers and spouses. Papers presented to the International Communication Association, Montreal (1973)
10. McCroskey, J.C., McCain, T.A.: The Measurement of Interpersonal Attraction. Speech Monographs 41, 261–266 (1974)
11. Mumm, J., Mutlu, B.: Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In: Proceedings of the 6th International Conference on Human-robot Interaction, pp. 331–338 (2011)
12. Nakauchi, Y., Simmons, R.: A Social Robot that Stands in Line. Autonomous Robots 12, 313–324 (2002)
13. Nomura, T.: Measurement of Negative Attitudes Towards Robots. Interaction Studies 7(3), 437–454 (2006)
14. Patterson, M.L.: An arousal model of interpersonal intimacy. Psychology Review 83(3), 235–245 (1976)
15. Syrdal, D., Dautenhahn, K., Walters, M.L., Koay, K.L.: Sharing Spaces with Robots in a Home Scenario – Anthropomorphic Attributions and their Effect on Proxemic Expectations and Evaluations in a Live HRI Trial. In: Proceedings of the 2008 AAAI Fall Symposium, pp. 116–123 (2008)

16. Taylor, O.L.: Cross Cultural Communication: An Essential Dimension of Effective Education. The Mid-Atlantic Equity Center, Maryland (1987)
17. Takayama, L., Pantofaru, C.: Influences on proxemic behaviors in human-robot interaction. In: Proceedings of the International Conference on Intelligent Robotic Systems: IROS 2009, St. Louis, MO, pp. 5495–5502 (2009)
18. Wang, L., Rau, P., Evers, V., Robinson, B., Hinds, P.: Responsiveness to Robots: Effects of Ingroup Orientation & Communication Style on HRI in China. In: Proceedings of Human Robot interaction Conference HRI 2009, San Diego, March 11-13, pp. 247–248. ACM Press, New York (2009)
19. Wang, L., Rau, P., Evers, V., Robinson, B., Hinds, P.: When in Rome: the role of culture & context in adherence to robot recommendations. In: Proc. of the 5th ACM/IEEE International Conference on Human-Robot Interaction, pp. 359–366. ACM Press, New York (2010)
20. Wish, M., Kaplan, S.T.: Toward an Implicit Theory of Interpersonal Communication. Sociometry 40(3), 234–246 (1977)