Effects of Humanlike Conversational Behavior on the Perception of Psychological Anthropomorphism: A Case Study with a Humanoid Robot

Maha Salem

2011

Preprint:

This is an accepted article published in ICSR 2011 : International Conference on Social Robotics 2011. The final authenticated version is available online at: https://doi.org/[DOI not available]

Effects of Humanlike Conversational Behavior on the Perception of Psychological Anthropomorphism: A Case Study with a Humanoid Robot

Maha Salem[1], Friederike Eyssel[2], Katharina Rohlfing[3], Stefan Kopp[4], and Frank Joublin[5]

[1] Research Institute for Cognition and Robotics, Bielefeld University, Germany, msalem@cor-lab.uni-bielefeld.de
[2] Center of Excellence Cognitive Interaction Technology, Bielefeld University, Germany, feyssel@cit-ec.uni-bielefeld.de
[3] Emergentist Semantics Group, Bielefeld University, Germany, kjr@uni-bielefeld.de
[4] Sociable Agents Group, Bielefeld University, Germany, skopp@techfak.uni-bielefeld.de
[5] Honda Research Institute Europe, Offenbach, Germany, frank.joublin@honda-ri.de

Abstract. Previous work has shown that typically human conversational behaviors such as gesturing increase anthropomorphic inferences about artificial communicators such as virtual agents. In an experiment with a humanoid robot, we investigated how far humanlike gesturing behavior would affect anthropomorphic inferences about the robot. Particularly, we examined the effects of the robot's gesturing behaviors on the attribution of typically human traits, experienced psychological warmth with regard to the robot, shared reality and finally, perceived pleasantness of HRI. We hypothesized higher ratings on all dependent measures in the humanlike gesturing (vs. no gesturing) behavior condition. Our results confirm our predictions: when the humanoid robot used humanlike gesturing during interaction, the robot was anthropomorphized more, participants felt more warmth and shared reality with it and experienced the interaction more positively than when the robot gave instructions using no gesturing behavior. These findings show that humanlike behaviors in robotic systems affect both anthropomorphic perceptions and the mental models humans form of a humanoid robot during interaction.

Keywords: Multimodal Interaction and Conversational Skills, Non-verbal Cues and Expressiveness, Anthropomorphism

1 Introduction

Social robotics research is dedicated to designing, developing and evaluating robots that can engage in social environments in a way that is both appealing and intuitive to human interaction partners. Therefore, a social robot's behavior ideally should appear natural, comprehensive and potentially humanlike. For this, an appropriate level of communicative functionality is required which, in turn, strongly depends on the appearance of the robot and attributions thus made to it. Different design approaches can be chosen depending on the social context and use of the robot. Fong et al [7] define four basic categories for social robots with regards to their appearance: anthropomorphic, zoomorphic, caricatured, and functionally designed robots. Anthropomorphic design, i.e. equipping the robot with a head, two arms and two legs, is broadly recommended to support an intuitive and meaningful interaction with humans [3, 4]. It is also considered a useful means to elicit the broad spectrum of responses that humans typically direct toward each other. This phenomenon is referred to as *anthropomorphism*, i.e. the attribution of human qualities to non-living objects. Humanlike body features in a robot increase anthropomorphism, especially when accompanied by "social" movements such as gaze behavior or hand and arm gesture. But to what extent are anthropomorphic inferences determined by the robot's physical appearance and what role, on the other hand, does the robot's social-communicative behavior play with regard to judgments of anthropomorphism?

Given the design of humanoid robots, they are typically expected to exhibit humanlike communicative behaviors, using their bodies for non-verbal expression just as humans do. For instance, co-verbal arm and hand gestures are a key feature of social-communicative behavior, which is frequently used by human speakers during human-human interaction. Crucially, gesture helps the speaker to convey information which cannot be conveyed by means of verbal communication alone. This is particularly true for referential, spatial or iconic information. However, gesture also affects the listener in an interaction. In [8], for example, it was demonstrated that human listeners pay close attention to information conveyed via such non-verbal behaviors. Accordingly, humanoid robots that shall be applied as interaction partners in HRI should generate co-verbal gestures for comprehensible and believable behavior. In addition, providing multiple modalities helps to dissolve ambiguity typical of uni-modal communication and thus increases robustness of communication.

The present work aims at shedding light on how the implementation of humanlike behaviors, such as gestures, affect social perceptions of the robot and HRI. For this purpose, we conducted an experiment in which we employed a humanoid robot as an interaction partner. Since this robot prototype lacks visible facial features that could potentially enrich the interaction with human users (e.g. by conveying emotional states of the system), this emphasizes the necessity to rely on additional communication channels, e.g. gesturing behaviors. Therefore, we examined this issue in the current experiment by investigating how humanlike gesturing behavior would affect anthropomorphic inferences about the humanoid robot, particularly with regard to the attribution of typically human traits, liking and shared reality with the robot and finally, judgments of acceptance and pleasantness of the HRI experience.

2 Related Work

A large body of work has evaluated complex gesture models for the animation of virtual characters (e.g., [13], [2]). Several recent studies have investigated the human attribution of naturalness to virtual agents. In one such study [13], the conversational agent Max communicated by either utilizing a set of co-verbal gestures alongside speech, typically by self touching or movement of the eyebrows, or by utilising speech alone without any such accompanying gestures. Participants subsequently rated Max' current emotional state and its personality, e.g. by indicating the extent to which Max appeared aggressive or lively. The results of the study showed that virtual agents are perceived in a more positive light when they produce co-verbal gestures rather than acting in a speech-only modality. In [2] Bergmann et al. modelled the gestures of Max based on real humans' non-verbal behavior and subsequently set out to question the communicative quality of these models via human participation. The main finding was that Max was perceived as more likeable, competent and humanlike when gesture models based on individual speakers as opposed to a collection of speakers or when no gestures at all were applied.

Despite the relevant implications of these studies, it is difficult to transfer the findings from virtual targets to robot platforms. Firstly, the presence of real physical constraints may influence the perceived level of realism. Secondly, given a greater degree of embodiment, interaction with a robot potentially appears richer. Since humans share the same interaction space with the robot, they can walk around or even touch a real robot in an interaction study. As a result, the interaction experience is different, which is expected to affect the outcome of the results.

To measure the degree of anthropomorphism attributed to the humanoid robot, we assessed participants' attribution of essentially human traits made to the robot: on the one hand, we measured perceptions of the robot's interpersonal warmth, a core dimension of human social cognition [6]. On the other hand, we assessed anthropomorphic inferences of our participants by asking them to rate the robot interaction partner with regard to uniquely human traits ([9], [10], [14]). Our approach is theoretically based on social psychological research on the dehumanization of social groups [9]. To illustrate, [9] have proposed two distinct senses of humanness at the trait-level. Specifically, they differentiate "uniquely human" and "human nature" traits. While "uniquely human" traits imply higher cognition, civility and refinement, traits indicating "human nature" involve emotionality, warmth, desire and openness. Since we already use a warmth measure to tap the robots emotionality, we only assess perceptions of Haslam's [9] human uniqueness dimension in this experiment.

By adapting these measures of anthropomorphism from social psychological research on uniquely human traits ([9], [6]), we complement existing work on the issue of measurement of anthropomorphism in social robotics (see [1] for a review). Thus, by presenting a social psychological perspective on anthropomorphism and new possible ways of measurement to the HRI community, we aim to contribute to a deeper understanding of determinants and consequences of anthropomorphism. In the following, we will present an experiment that tested the effects of uni-modal vs. multi-modal communication behavior on perceptions of warmth, perceived anthropomorphism, experienced shared reality and contact intentions with the robot.

3 Method

To gain a deeper understanding of how communicative robot gesture might impact and shape user experience and evaluation of HRI, we conducted a between-subjects experimental study using a humanoid robot. For this, an appropriate scenario for gesture-based human-robot interaction was designed and benchmarks for the evaluation were identified.

The study scenario comprised a joint task that was to be performed by a human participant in collaboration with the humanoid robot. In the given task, the robot referred to various objects by utilizing either multi-modal (speech and gesture) or uni-modal (speech only) utterances, based on which the participant was expected to perceive, interpret and perform an according action. Data documenting the participant's experience was collected after task completion using a questionnaire.

3.1 Participants and Design

A total of 41 participants (21 female, 20 male) took part in the experiment, ranging in age from 20 to 61 years ($M = 30.68$ years, $SD = 10.00$). All subjects were German native speakers and were recruited at Bielefeld University, Germany. Based on five-point Likert scale ratings, participants were identified as having negligible experience with robots ($M = 1.32$, $SD = 0.61$), whereas they reported moderate skills regarding technology and computer use ($M = 3.78$, $SD = 1.04$). Participants were randomly assigned to one of two experimental conditions that manipulated non-verbal robot behaviors (uni-modal vs. multi-modal communication).

3.2 Materials

Participants interacted with the Honda humanoid robot (year 2000 model)[11]. Its upper body comprises a torso with two 5DOF arms and 1DOF hands, as well as a 2DOF head. To control the robot, we used a previously implemented speech-gesture generation model which allows for a real-time production and synchronization of multi-modal robot behavior [15]. The framework combines conceptual representation and planning with motor control primitives for speech and arm movements of a physical robot body.

During the study, the robot was partly controlled using a Wizard-Of-Oz technique to ensure minimal variability in the experimental procedure. The robot's speech was identical across conditions and was generated using the text-to-speech system ***M**odular **A**rchitecture for **R**esearch on speech s**Y**nthesis* (MARY)[16] set to a neutral voice.

The experimenter initiated the robots interaction behavior from a fixed sequence of pre-determined utterances, each of which was triggered when the participant stood in front of the robot. Once triggered, a given utterance was generated autonomously at run-time. The ordering of the utterance sequence remained identical across conditions and experimental runs.

The entire interaction was filmed by three video cameras from different angles, while the experimenter observed and controlled the interaction from the adjacent room.

3.3 Experimental Setting

The experiment was set in a simulated kitchen environment in a robot lab (see Fig. 1). The humanoid played the role of a household robot. Participants were told that their task was to help a friend who was in the midst of moving his household. To do so, they were asked to unpack a cardboard box full of kitchen appliances and to put these into the cupboard that was part of the kitchen set-up. Specifically, the box contained nine kitchen items. It was unknown to participants, where they were supposed to put these items. Importantly, however, participants were informed that the humanoid robot would help them to solve the task by telling them where to put the respective kitchenware. The experimental setting is illustrated in Figure 1.

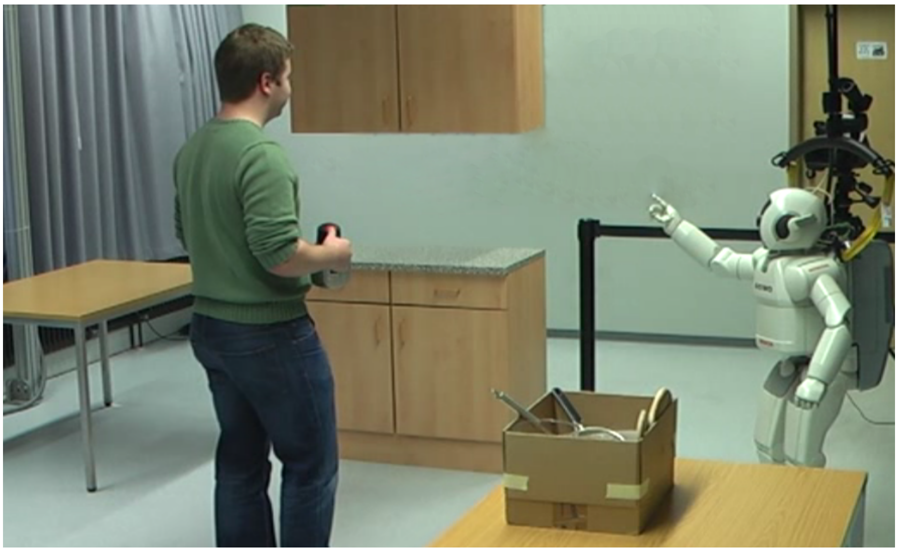

Fig. 1. The experimental setting; the robot provides the participant with information about the designated storage location of the object.

Crucially, we manipulated the non-verbal behaviors that were displayed by the humanoid robot:

- In the **uni-modal** *(speech-only)* condition, the robot presented the participant solely with a set of verbal instructions to explain where each object should be placed. The robot did not move its body during the whole interaction; no gesture or gaze behaviors were performed.
- In the **multi-modal** *(speech-gesture)* condition, the robot presented the participant with the identical set of verbal instructions used in condition 1, however, accompanied by gestures to supplement the spoken utterances. Simple gaze behavior supporting hand and arm gestures (e.g. look right when pointing right) was displayed during interaction.

Verbal Utterance: In order to keep the task solvable under both conditions, spoken utterances were designed in a self-sufficient way, i.e. gestures used in the multi-modal condition contained information that was also addressed by speech. Each instruction presented by the robot typically consisted of two or three so-called *utterance chunks*. Based on the definition provided in [12], each *chunk* refers to a single idea unit represented by an intonation phrase and, optionally in a multi-modal utterance, by an additional co-expressive gesture phrase. The verbal utterance chunks were based on the following syntax:

- **Two-chunk utterance:**
  `<Please take the [object]> <and place it [position+location].>`
  Example: *Please take the thermos flask and place it on the right side of the upper cupboard.*
- **Three-chunk utterance:**
  `<Please take the [object],> <then open the [location]>`
  `<and place it [position].>`
  Example: *Please take the eggcup, then open the right drawer and place it inside.*

Gestures: In the multi-modal condition, the robot used three different types of gesture along with speech to indicate the designated placement of each item:

- **Deictic gestures**, e.g. to indicate positions and locations
- **Iconic gestures**, e.g. to illustrate shape/size of objects
- **Miming gestures**, e.g. hand movement performed when opening cupboard doors or using a ladle

A total of 20 speech-accompanying gestures were generated as part of the instructions during each trial. Out of these, 55% semantically matched the verbal instruction, while the remaining 45% of gestures were semantically non-matching, e.g., the robot occasionally said "put it up there" but pointed downwards. This behavior was chosen to decrease the reliability and task-related usefulness of the gesture, so that participants did not evaluate the use of gestures solely based on the helpfulness of the additional modality in solving the task.

3.4 Hypothesis

As the main hypothesis, we predicted that participants who received multi-modal instructions from the robot would anthropomorphize the robot more than those who are presented uni-modal information by the robot (using only speech).

3.5 Experimental Procedure

Participants were tested individually. First, they received experimental instructions in written form. Subsequently, they entered the robot lab, where the experimenter orally provided the task instructions. They were then given the opportunity to ask any clarifying questions before the experimenter left the participant to begin the interaction with the robot. At the beginning of the experiment, the robot greeted the participant and introduced the task before commencing with the actual instruction part. The robot then presented the participant with individual utterances as described in the experimental design. Each utterance was delivered in two parts: the first part referred to the object (e.g. "*Please take the thermos flask*"); the second part comprised the item's designated position and location (e.g. "*...and place it on the right side of the upper cupboard.*").[1]

Whenever the participant resumed a standing position in front of the robot in order to signal readiness to proceed with the next instruction, the experimenter sitting at a control terminal triggered the robot's subsequent behavior. The participant then followed the uttered instructions and placed each item into its correct location. As explained in the briefing prior to commencing the experimental task, participants were requested to place the object on a table adjacent to the kitchen cupboard if unsure about where the item should be placed, rather than trying to guess its location. At the end of the interaction, the robot thanked the participant for helping and bid them farewell. Participants interacted with the robot for approximately five minutes. In the uni-modal (speech-only) condition all utterances including the greeting and farewell were presented verbally; in the multi-modal (speech-gesture) condition, all utterances including the greeting and farewell were accompanied by co-verbal gestures.

After interacting with the robot, participants were led out of the lab to complete a post-experiment questionnaire to evaluate the robot and the interaction experience. Upon completion of the questionnaire, participants were carefully debriefed about the purpose of the experiment and received a chocolate bar as a thank-you before being dismissed.

3.6 Dependent Measures

We asked participants to report the degree to which they anthropomorphized the robot by using various dimensions:

[1] The delivery of utterances was split as a result of a pilot test which revealed that participants frequently turned away from the robot to grab the object right after it was named and would then pay no more attention to the robot. As a consequence, participants subsequently had difficulty to assess the robot's non-verbal behavior after completing the task.

First, we measured perceived humanlikeness of the robot based on Haslam's [9] list of ten uniquely human traits (broadminded, humble, organized, polite, thorough, cold, conservative, hard-hearted, rude, shallow).

Second, participants were asked to report how interpersonally warm they rated the robot prototype [6]. Warmth was assessed using four traits (polite, friendly, sympathetic, sociable).

We further administered three items ("How close do you feel to the humanoid robot", "How pleasant was the interaction with the robot for you?", "How much fun did you have interacting with the robot?") to assess participants degree of shared reality with the robot [5]. The shared reality index taps perceptions of similarity and experienced psychological closeness to the robot. Moreover, it covers aspects of human-robot acceptance, because participants had to indicate how much they enjoyed the interaction with the robot.

Finally, using a single item to measure participants' future contact intentions, we asked participants to indicate to what extent they would be willing to live with the robot. All responses were given on 5-point Likert scales, with endpoints 1, *not at all*, and 5, *very much*.

4 Results

First, reliability analyses (Cronbach's α) were conducted to assess the internal consistency where applicable. For the dependent measures, the indices proved highly reliable, given a Cronbach's α of .78 for the index reflecting uniquely human traits, a Cronbach's α of .86 for the 'perceived warmth' index, and a Cronbach's α of .81 for the 'shared reality' index respectively. Consequently, participants' responses to the respective items were averaged to form indices of anthropomorphism, perceived warmth, and shared reality. We then conducted t-tests on the dependent measures to test the hypothesis, namely that the humanoid robot's use of gesture would result in higher ratings on all dependent variables than when using speech only.

Results show a significant effect of condition on all dependent measures: Participants attributed more uniquely human traits to the robot in the multi-modal condition (M = 2.55, SD = 0.68) than in the uni-modal condition (M = 1.98, SD = 0.58), $t(39)$ = -2.88, p = 0.007. Moreover, participants reported greater perceived warmth when interacting with the robot whose verbal utterances were accompanied by gestures (M = 4.56, SD = 0.37) than when it was speaking only (M = 3.63, SD = 0.95), $t(39)$ = -4.12, $p < 0.001$. Participants also experienced greater shared reality with the robot when it used multi-modal behaviors (M = 3.92, SD = 0.70) than when it relied on uni-modal communication only (M = 3.23, SD = 0.93), $t(39)$ = -2.68, p = 0.01. Finally, participants' assessment of whether they would like to live with the robot was also higher in the condition with speech-accompanying gesture behavior (M = 3.90, SD = 1.14) than in the one without (M = 2.63, SD = 1.30), $t(38)$ = -3.31, p = 0.002. Fig. 2 illustrates how the pattern of means was in the predicted direction for all dependent variables.

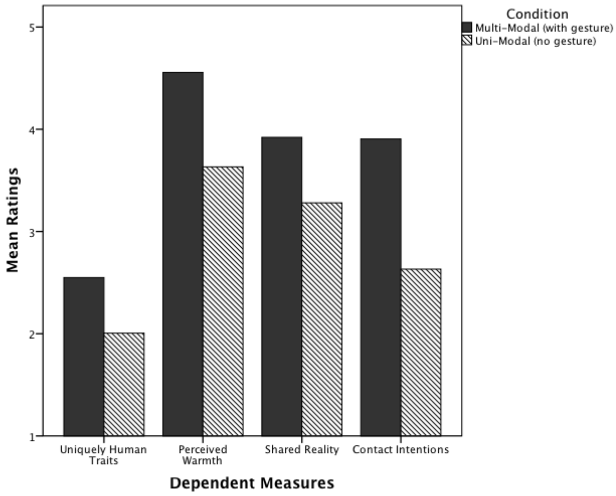

Fig. 2. Mean ratings of dependent measures as a function of experimental condition.

5 Discussion and Conclusion

We conducted an experiment to investigate how humanlike gesturing behavior affects anthropomorphic perceptions and the mental models humans form of a humanoid robot. We particularly focussed on participants' attribution of typically human traits to the robot, perceived warmth, shared reality, as well as future contact intentions with regard to the robot. Specifically, applying a wide range of dependent variables, we examined to what extent anthropomorphic inferences on the human's side are attributed to the design, and to what extent to the behavior of the robot. Our findings show that the robot's gesturing behavior significantly affected subsequent evaluation by the human participants. Firstly, multi-modal (speech-gesture) behavior displayed by the robot resulted in greater anthropomorphic inference than in the uni-modal (speech-only) condition. Secondly, perception and evaluation of the robot was more positive when the humanoid displayed non-verbal behaviors in the form of co-verbal gestures. Interestingly, this is also true for hand and arm gestures that do not always semantically match the information conveyed via speech. These findings suggest that even when a robot occasionally makes an "inappropriate" gesture, it is still more favorable over a robot that does not perform any gestures at all.

Our theory-driven approach is characterized by the application of social psychological theories of (de-)humanization [9, 10] to HRI. By adapting these measures of anthropomorphism from research on uniquely human traits, we contribute to existing work on the issue of measurement of anthropomorphism in social robotics, and thus to a deeper understanding of determinants and consequences of anthropomorphism.

Future research should investigate the generalizability of our findings regarding anthropomorphic inferences with other robotic platforms, for instance with non-humanoid robots. Furthermore, it should systematically examine the impact of gaze behavior displayed by the robot in an isolated experimental set-up without hand and arm gesture. This way we can investigate to what extent anthropomorphic inferences are determined by the robot's arm gestures versus gaze behavior alone. For the time being, the present findings emphasize the importance of displaying humanlike gesturing behaviors in social robots as significant factors that contribute to smooth and pleasant human-robot interaction.

Acknowledgement. The work described is supported by the Honda Research Institute Europe.

References

1. C. Bartneck, E. Croft, and D. Kulic. Measuring the anthropomorphism, animacy, likeability, perceived intelligence and safety of robots. In *Proceedings of the Metrics of Human-Robot Interaction Workshop, Technical Report 471*, pages 37–41, 2008.
2. K. Bergmann, S. Kopp, and F. Eyssel. Individualized gesturing outperforms average gesturing – evaluating gesture production in virtual humans. In *Proceedings of the 10th Conference on Intelligent Virtual Agents*, pages 104–117. Springer, 2010.
3. C. Breazeal. Designing sociable robots. pages 1–263, 2002.
4. B. R. Duffy. Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3-4):177–190, 2003.
5. G. Echterhoff, E.T. Higgins, and J.M. Levine. Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science*, 4:496–521, 2009.
6. S.T. Fiske, J.C. Cuddy, and P. Glick. Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Science*, 11(2):77–82, 2006.
7. Terrence Fong, Illah R. Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4):143–166, 2003.
8. S. Goldin-Meadow. The role of gesture in communication and thinking. *Trends in Cognitive Science*, 3:419–429, 1999.
9. N. Haslam, P. Bain, S. Loughnan, and Y. Kashima. Attributing and denying humanness to others. *European Review of Social Psychology*, 19:55–85, 2008.
10. N. Haslam, Y. Kashima, and S. Loughnan. Subhuman, inhuman and superhuman: Contrasting humans with nonhumans in three cultures. *Social Cognition*, 26:248–258, 2008.
11. Ltd. Honda Motor Co. The Honda Humanoid Robot Asimo, year 2000 model. http://world.honda.com/ASIMO/technology/spec.html. 2000.
12. S. Kopp and I. Wachsmuth. Synthesizing Multimodal Utterances for Conversational Agents. *Computer Animation and Virtual Worlds*, 15(1):39–52, 2004.
13. N. Krämer, N. Simons, and S. Kopp. The effects of an embodied conversational agents nonverbal behavior on users evaluation and behavioral mimicry. In *Proceedings of Intelligent Virtual Agents (IVA 2007)*, pages 238–251. Springer, 2007.
14. S. Loughnan and N. Haslam. Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science*, 18:116–121, 2007.
15. M. Salem, S. Kopp, I. Wachsmuth, and F. Joublin. Towards an integrated model of speech and gesture production for multi-modal robot behavior. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pages 649–654, 2010.
16. M. Schröder and Jürgen Trouvain. The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. 6:365–377, 2003.