Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability

Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, Frank Joublin

2013

Preprint:

This is an accepted article published in Int J Soc Robot. The final authenticated version is available online at: https://doi.org/[DOI not available]

To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability

Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp & Frank Joublin

International Journal of Social Robotics

ISSN 1875-4791

Int J of Soc Robotics
DOI 10.1007/s12369-013-0196-9



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Author's personal copy

Int J Soc Robot
DOI 10.1007/s12369-013-0196-9

To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability

Maha Salem · Friederike Eyssel · Katharina Rohlfing · Stefan Kopp · Frank Joublin

Accepted: 30 April 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Previous work has shown that non-verbal behaviors affect anthropomorphic inferences about artificial communicators such as virtual agents or social robots. In an experiment with a humanoid robot we investigated the effects of the robot's hand and arm gestures on the perception of humanlikeness, likability of the robot, shared reality, and future contact intentions after interacting with the robot. For this purpose, the speech-accompanying non-verbal behaviors of the humanoid robot were manipulated in three experimental conditions: (1) no gesture, (2) congruent co-verbal gesture, and (3) incongruent co-verbal gesture. We hypothesized higher ratings on all dependent measures in the two multimodal (i.e., speech and gesture) conditions compared to the unimodal (i.e., speech only) condition. The results confirm our predictions: when the robot used co-verbal gestures during interaction, it was anthropomorphized more, participants perceived it as more likable, reported greater shared reality with it, and showed increased future contact intentions than when the robot gave instructions without gestures. Surprisingly, this effect was particularly pronounced when the robot's gestures were partly incongruent with speech, although this behavior negatively affected the participants' task-related performance. These findings show that communicative non-verbal behaviors displayed by robotic systems affect anthropomorphic perceptions and the mental models humans form of a humanoid robot during interaction.

Keywords Social human-robot interaction · Multimodal interaction and conversational skills · Non-verbal cues and expressiveness · Anthropomorphism · Robot companions and social robots

M. Salem (✉)
Research Institute for Cognition and Robotics,
Bielefeld University, Bielefeld, Germany
e-mail: msalem@techfak.uni-bielefeld.de

F. Eyssel
Center of Excellence Cognitive Interaction Technology,
Bielefeld University, Bielefeld, Germany
e-mail: feyssel@cit-ec.uni-bielefeld.de

K. Rohlfing
Emergentist Semantics Group, Bielefeld University, Bielefeld, Germany
e-mail: kjr@uni-bielefeld.de

S. Kopp
Sociable Agents Group, Bielefeld University, Bielefeld, Germany
e-mail: skopp@techfak.uni-bielefeld.de

F. Joublin
Honda Research Institute Europe, Offenbach, Germany
e-mail: frank.joublin@honda-ri.de

1 Introduction

Social robotics research is dedicated to designing, developing, and evaluating robots that can engage in social environments in a way that is appealing and intuitive to human interaction partners. However, interaction is often difficult because inexperienced users do not understand the robot's internal states, intentions, and expectations. To facilitate successful interaction, 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.

Anthropomorphic design, i.e., equipping the robot with humanlike body features such as two legs, two arms, and a head, is broadly recommended to support an intuitive

Published online: 31 May 2013

Author's personal copy

and meaningful interaction with humans [3, 5]. It is further considered a useful means to elicit the broad spectrum of responses that humans typically direct toward each other [1]. This phenomenon is referred to as *anthropomorphism*, i.e., the attribution of human qualities to non-living objects [7]. Anthropomorphism is increased when social-communicative behaviors such as gaze or hand and arm gesture are displayed by the robot [5]. 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 non-verbal behavior play regarding judgments of anthropomorphism?

With regard to non-verbal behavior, hand and arm gestures are primary candidates for extending the communicative capabilities of social robots, since they represent a key feature of social-communicative behavior [11]. Human speakers frequently use gesture during interaction, in order to convey information which cannot be expressed by means of verbal communication alone, such as referential, spatial or iconic information [16]. In addition, gesture also affects human listeners in an interaction, as they have been shown to pay close attention to information conveyed via such non-verbal behaviors [8]. Accordingly, humanoid robots that shall be applied as interaction partners should generate co-verbal gestures for comprehensible and believable behavior. Since a large body of research (e.g., [17, 23, 26, 27]) has already focused on the role of robot gaze in human-robot interaction (HRI), our investigations concentrate on hand and arm gesture as a specific subpart of non-verbal communicative behavior.

The present work aims at shedding light on how the implementation of humanlike non-verbal behaviors, specifically hand and arm gestures, affect social perceptions of the robot and HRI. For this purpose, we conducted an experiment using the Honda 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 provide additional communication channels such as gestural behaviors. We addressed this issue in the present experiment by investigating how gesture behavior affects anthropomorphic inferences about the humanoid robot, particularly with regard to perceived humanlikeness, likability, shared reality with the robot and judgments of acceptance, as well as future contact intentions after interacting with the robot.

We first discuss related work in Sect. 2. In Sect. 3, we describe the methodology of our study, including the proposed hypotheses, the experimental design and procedure, the dependent measures, and information on the participants. Results are subsequently presented in Sect. 4. Finally, we discuss the results and conclude by giving an outlook of future work in Sect. 5.

2 Related Work

Much research (e.g., [2, 4, 13, 18]) has evaluated complex gesture models for the animation of virtual characters. Several studies 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 or by utilizing speech alone without any accompanying gestures. Participants subsequently rated Max's emotional state and 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 using speech as the only modality.

Despite the relevant implications of such studies, findings from virtual agents cannot be easily transferred to robot platforms. First, the presence of real physical constraints in a robot may influence the perceived level of realism. Second, given a greater degree of embodiment and physicality, interaction with a robot is potentially richer than with a virtual agent. As humans share the same interaction space, they can walk around or even touch the robot in an interaction study. Consequently, the interaction experience is different, which is likely to affect the outcome of the results.

In the area of human-robot interaction, a large body of work (e.g., [17, 23, 26, 27]) has studied the effect of robot gaze as an important aspect of non-verbal behavior. In contrast, not much research has focused on hand and arm gestures in particular and the evaluation of their effects in HRI studies (see [19] for an overview of related work). Therefore, our work centers on speech-accompanying arm movements. However, given the substantial correlation between gaze and hand gesture behavior in human communication, the interplay between these two non-verbal communication modalities needs to be further investigated in the future, as already done, for example, by Iio et al. [10].

Our present approach is theoretically based on social psychological research on the (de-)humanization of social groups [9, 14]. To illustrate, Haslam et al. [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 the human nature dimension is typically used to measure 'mechanistic dehumanization',[1] we conversely employ this measure to assess the extent to which a robot is perceived as humanlike. We further assess the degree of anthropomorphism

[1] According to Haslam et al. [9], when people are denied human nature, they are implicitly or explicitly objectified or likened to machines rather than to animals or humans.

Author's personal copy

attributed to the humanoid robot by measuring participants' perceptions of the robot's likability, shared reality with the robot, and future contact intentions.

By adapting measures of anthropomorphism from social psychological research on human nature traits [9, 14], 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 contribute to a deeper understanding of determinants and consequences of anthropomorphism.

In the following, we will present an experiment that tested the effects of unimodal vs. multimodal communication behavior on perceived anthropomorphism and likability, experienced shared reality, and contact intentions with regard to a humanoid robot.

3 Method

To gain a deeper understanding of how communicative robot gesture may impact and shape user experience and evaluation of human-robot interaction, we conducted a between-subjects experimental study using the Honda humanoid robot. For this, an appropriate scenario for gesture-based HRI 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. The main motivation for choosing a task-based interaction was to realize a largely controllable yet meaningful interaction which would allow for a measurable comparison of participants' reported experiences. In the given task, the robot referred to various objects by utilizing either unimodal (speech only) or multimodal (speech and either congruent or incongruent accompanying gesture) 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, whereas the task-related performance of each participant was derived from the error rate, i.e., the number of objects that were not correctly placed during the experimental task.

3.1 Hypotheses

Based on findings from gesture research in human-human as well as in human-agent interaction we developed the following hypotheses for gesture-based HRI:

1. Participants who receive multimodal instructions from the robot (either congruent or incongruent) will evaluate the robot more positively and anthropomorphize it more than those who receive unimodal information (i.e., speech-only).
2. A robot that occasionally performs non-matching (i.e., incongruent) gestures will be preferred over one that performs no gestures at all.
3. Participants who are presented with incongruent multimodal instructions by the robot will perform worse at the task than those who are presented with unimodal or congruent multimodal information by the robot.

3.2 Materials

Participants interacted with the Honda humanoid robot (year 2000 model) [15]. 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 multimodal robot behavior [21]. The framework combines conceptual representation and planning with motor control primitives for speech and arm movements of a physical robot body. To ensure minimal variability in the experimental procedure, the robot was partly controlled using a Wizard-Of-Oz technique [24] during the study. The experimenter initiated the robot's interaction behavior from a fixed sequence of predetermined 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 using the implemented action generation framework for speech and gesture synthesis [21]. The ordering of the utterance sequence remained identical across conditions and experimental runs. The robot's speech was identical across conditions and was generated using the text-to-speech system *MARY* (Modular Architecture for Research on speech sYnthesis) [22] set to a neutral voice. 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 Design

The experiment was set in a simulated kitchen environment in a robot lab (see Fig. 1). The humanoid robot served as a household assistant. Participants were told that their task was to help a friend who was moving house. They were asked to unpack a cardboard box containing nine kitchen items and to put these into the cupboard that was part of the kitchen set-up. Specifically, the objects comprised a thermos flask, a sieve, a ladle, a vase, an eggcup, two differently shaped chopping boards and two differently sized bowls. Participants were allowed to move freely in the area in front of the robot. Given the participant's non-familiarity with the friend's kitchen environment, the robot was made to assist with the task by providing information on where to put the

Author's personal copy

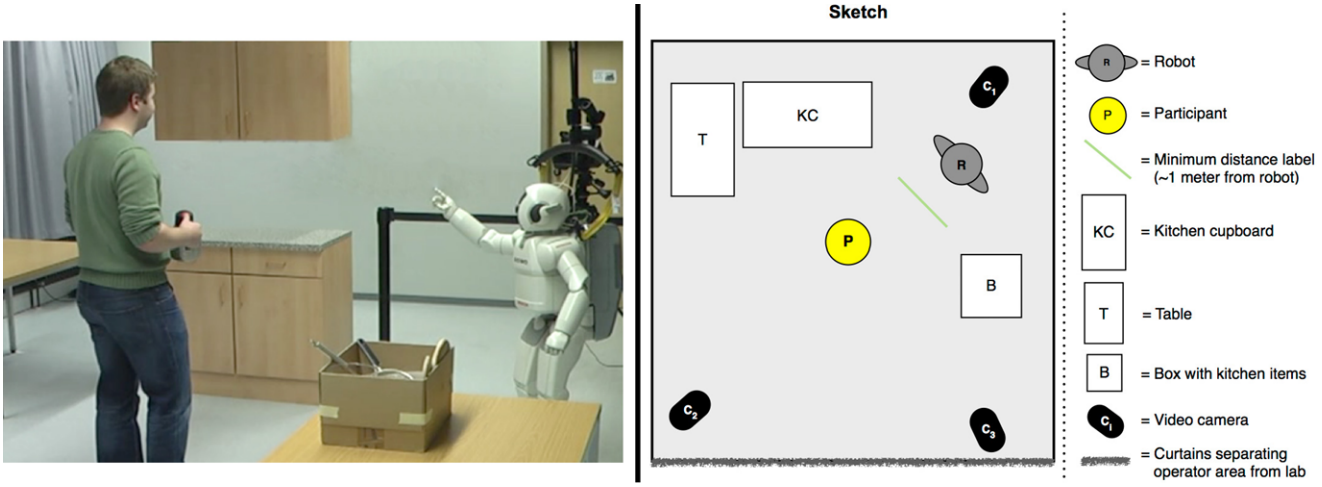

Fig. 1 The experimental setting: the robot provides the participant with information about the storage location of the object (*left*); sketch of the experimental lab (*right*). Reprinted from [20]

respective kitchenware. In case the participant did not understand where the item had to be stored, a table situated beside the kitchen cupboard was provided for alternative placement. Participants were instructed not to guess the location of the item in such case, so that their performance could be correctly evaluated afterwards. The experimental setting is illustrated in Fig. 1.

Conditions We manipulated the non-verbal behaviors that were displayed by the humanoid robot in three experimental conditions:

1. In the *unimodal (speech-only)* condition, the robot presented the participant with a set of nine 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.
2. In the *congruent multimodal (speech-gesture)* condition, the robot presented the participant with the identical set of nine verbal instructions used in condition 1. In addition, they were accompanied by a total of 21 corresponding gestures explaining where each object should be placed. Speech and gesture were semantically matching, e.g., the robot said "put it up there" and pointed up. Simple gaze behavior supporting hand and arm gestures (e.g., looking right when pointing right) was displayed during the interaction.
3. In the *incongruent multimodal (speech-gesture)* condition, the robot presented the participant with the identical set of nine verbal instructions used in condition 1. In addition, they were accompanied by a total of 21 gestures, out of which ten gestures (47.6 %) semantically matched the verbal instruction, while the remaining eleven gestures (52.4 %) were semantically non-matching, e.g., the robot occasionally said "put it up there" but pointed downwards. Simple gaze behavior supporting hand and arm gestures (e.g., looking right when pointing right) was displayed during the interaction.

The *incongruent multimodal* condition was designed to decrease the reliability and task-related usefulness of the robot's gestures. In other words, participants in this group were unlikely to evaluate the use of the additional gesture modality solely based on its helpfulness in solving the given task. In addition, this experimental condition decreased the predictability of the robot's actions, which may lead to increased uncertainty associated with the robot during interaction. According to what Epley et al. [7] refer to as *effectance motivation*, the anxiety caused by such uncertainty along with the desire to predict an agent's behavior can affect people's tendency to anthropomorphize non-human agents. Finally, the choice to combine semantically matching gestures with non-matching ones in this condition was made to avoid a complete loss of the robot's credibility after a few utterances.

Verbal Utterance In order to keep the task solvable in all three experimental conditions, spoken utterances were designed in a self-sufficient way, i.e., gestures used in the multimodal conditions were supplementary to 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 multimodal 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].>`

Author's personal copy

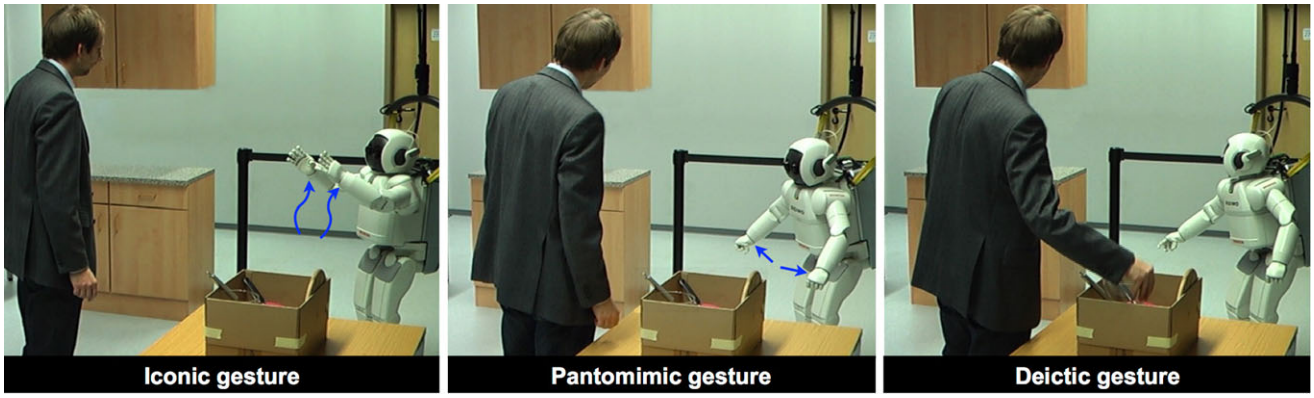

Fig. 2 Examples of the three gesture types performed by the robot in the multimodal conditions (adapted from [21]; *arrows* indicate the movement trajectory and direction of dynamic gestures): *iconic gesture* illustrating the shape of the vase; *pantomimic gesture* conveying the act of opening the lower cupboard; *deictic gesture* pointing at the designated location

Example: *Please take the vase and place it on the left side of the lower 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 multimodal conditions, 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 the shape or size of objects
- *Pantomimic gestures*, e.g., hand movement performed when opening cupboard doors

Examples of the three gesture types are illustrated in Fig. 2.

3.4 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.*"). 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 instruction and, ideally, placed each item into its correct 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 unimodal (speech-only) condition, all utterances including the greeting and farewell were presented verbally; in the multimodal (speech-gesture) conditions, 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.

3.5 Dependent Measures

Participants were asked to report their interaction experience with the robot and rate their perception of its behavior based on a post-experimental questionnaire. Video data was analyzed to evaluate participants' performance during the task. Data analysis focused on two main aspects, namely *perception of the robot* and *task-related performance of participants*.

With regard to participants' *perception of the robot*, the degree of *humanlikeness* attributed to the robot was assessed using Haslam et al.'s list of human nature traits which comprises ten characteristics and represents a measure from social psychological research on the (de-)humanization of social groups [9]. The questionnaire items forming this index

Author's personal copy

Table 1 Dependent measures, respective questionnaire items and scales used to evaluate the *perception of the robot*

Index/measure	Questionnaire items	Scale
Humanlikeness	Curious, friendly, fun-loving, sociable, trusting, aggressive, distractible, impatient, jealous, nervous	1 = strongly disagree, 5 = strongly agree
Likability	Polite, sympathetic, humble	1 = strongly disagree, 5 = strongly agree
Shared reality	"How close do you feel to the robot?" "How pleasant was the interaction with the robot for you?" "How much fun did you have interacting with the robot?"	1 = not at all, 5 = very much
Future contact intentions	"Would you like to live with the robot?"	1 = not at all, 5 = very much

Table 2 Dependent measure, questionnaire item and scale used to evaluate the *task-related performance of participants*

Measure	Questionnaire item	Scale
Competence self-rating	"How competent were you in solving the task?"	1 = not at all, 5 = very

are presented in Table 1 together with other measures used to evaluate participants' perception of the robot. In particular, participants' perception of the robot's *likability* was assessed using three questionnaire items. Their degree of *shared reality* with the robot was evaluated based on three further items which tap perceptions of similarity and experienced psychological closeness to the robot [6]. The shared reality index also covers aspects of human-robot acceptance, as participants had to indicate how much they enjoyed the interaction with the robot. Finally, participants' *future contact intentions* with regard to the robot were measured using a single item. The dependent measures evaluating participants' perception of the robot were used to test the first and the second hypotheses.

Task-related performance of participants was measured in two ways: first, *subjective assessment* was measured using a questionnaire item asking participants to assess their own performance (see Table 2); second, *objective assessment* was derived from the task-related error rate, i.e., the number of objects that were not correctly placed during the experimental HRI task. This objective measure was used to test the third hypothesis.

3.6 Participants

A total of 62 participants (32 female, 30 male) took part in the experiment, ranging in age from 20 to 61 years ($M = 30.90$ years, $SD = 9.82$). All participants were German native speakers recruited at Bielefeld University, Germany. Based on five-point Likert scale ratings, participants were identified as having negligible experience with robots ($M = 1.35$, $SD = 0.66$) and moderate skills regarding technology and computer use ($M = 3.74$, $SD = 0.97$). Participants were randomly assigned to one of the three experimental conditions that manipulated the robot's non-verbal behaviors, while maintaining gender- and age-balanced distributions.

4 Results

4.1 Perception of the Robot

First, reliability analyses (Cronbach's α) were conducted to assess the internal consistencies of the dependent measures where applicable. The indices proved sufficiently reliable, given a Cronbach's α of 0.78 for the index reflecting *humanlikeness*, a Cronbach's α of 0.73 for the *likability* index, and a Cronbach's α of 0.78 for the *shared reality* index respectively. Consequently, participants' responses to the respective items were averaged to form the three outlined indices. To test the effect of experimental conditions on the dependent measures, analyses of variance (ANOVA) and Tukey's post-hoc tests were conducted with a 95 % confidence interval for pairwise comparisons between condition means. Mean values and standard deviations are summarized in Table 3 and are visualized in Fig. 3.

Results show a significant effect of condition on all dependent measures. Specifically, they confirm that the manipulation of the robot's gestural behavior had a significant effect on participants' ratings of the *humanlikeness* index which reflects their attribution of human nature traits to the robot ($F(2,58) = 4.63$, $p = 0.01$). It also had a significant effect on their assessment of the robot's *likability* ($F(2, 59) = 3.65$, $p = 0.03$). Furthermore, analyses indicate that the manipulation of the robot's non-verbal behavior had a significant effect on participants' ratings of the *shared reality* measure ($F(2, 59) = 4.06$, $p = 0.02$) as well as on their *future contact intentions* ($F(2, 58) = 5.43$, $p < 0.01$).

Author's personal copy

Table 3 Mean values of the dependent measures reflecting participants' *perception of the robot* (standard deviations in parentheses)

Measure	Condition		
	Unimodal	Congr. multimodal	Incongr. multimodal
Humanlikeness	1.98 (0.58)	2.15 (0.58)	2.55 (0.68)
Likability	3.69 (0.97)	3.92 (0.81)	4.36 (0.59)
Shared reality	3.23 (0.93)	3.75 (0.76)	3.92 (0.70)
Future contact intentions	2.63 (1.30)	2.95 (1.40)	3.90 (1.14)

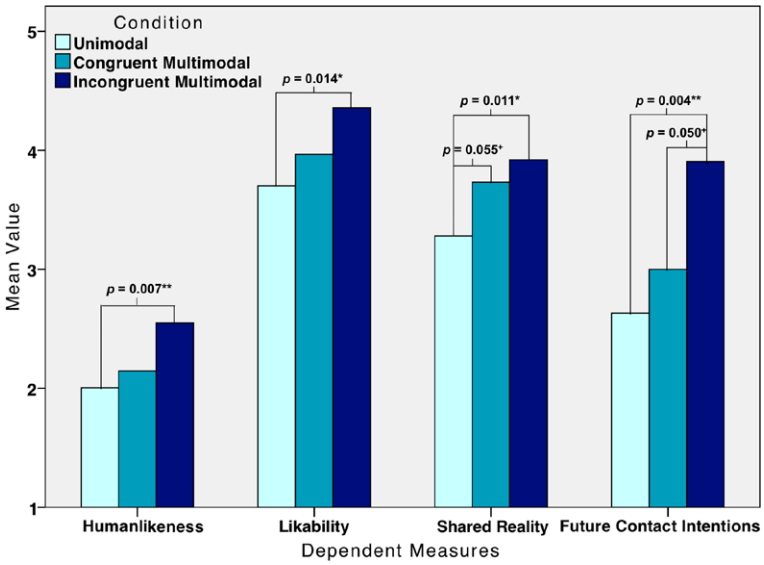

Fig. 3 *Bar chart* visualizing the mean ratings and significant effects for the dependent variables measuring participants' *perception of the robot*; $^{+} = p < 0.10$, $^{*} = p < 0.05$, $^{**} = p < 0.01$

Tukey post-hoc comparisons of the three groups indicate that participants in the incongruent multimodal condition ($M = 2.55$, $SD = 0.68$) rated the perceived *humanlikeness* of the robot significantly higher than participants in the unimodal condition ($M = 1.98$, $SD = 0.58$), $p < 0.01$. That is, when the robot performed gestures that were to some extent incongruent with speech, participants anthropomorphized it significantly more than when it did not gesture at all.

Moreover, participants reported significantly greater perceived *likability* when interacting with a robot whose verbal utterances were occasionally accompanied by non-matching gestures in the incongruent multimodal condition ($M = 4.36$, $SD = 0.59$) than when it was only using speech ($M = 3.69$, $SD = 0.97$), $p = 0.01$.

Participants also experienced greater *shared reality* with the robot when it used either congruent ($M = 3.75$, $SD = 0.76$) or incongruent ($M = 3.92$, $SD = 0.70$) multimodal behaviors than when it relied on unimodal communication only ($M = 3.23$, $SD = 0.93$); this effect was approaching significance for the comparison of unimodal versus congruent multimodal behavior, $p = 0.06$, and was significant when comparing the unimodal with the incongruent multimodal condition, $p = 0.01$.

Finally, participants' assessment of *future contact intentions* with regard to the robot was also significantly higher in the condition with partially incongruent speech-accompanying gesture behavior ($M = 3.90$, $SD = 1.14$) than in the unimodal condition ($M = 2.63$, $SD = 1.30$), $p < 0.01$. Remarkably, average ratings of whether participants would like to live with the robot were much higher in the incongruent multimodal condition than in the congruent multimodal condition group ($M = 2.95$, $SD = 1.40$), $p = 0.05$.

Although comparisons between the unimodal and the congruent multimodal condition were not statistically significant at the 5 % level, they indicate a trend towards higher mean ratings for all dependent measures in the congruent multimodal condition. Similarly, comparisons between the congruent multimodal and the incongruent multimodal groups were not statistically significant at $p < 0.05$, however, the results throughout indicate a trend towards higher mean ratings in favor of the incongruent multimodal group.

These results and observed trends with regard to participants' *perception of the robot* support both Hypotheses 1 and 2 which predicted higher ratings on all dependent measures in the two multimodal groups when compared to the unimodal group.

4.2 Task-related Performance of Participants

Results of participants' *subjective assessment* ratings are shown in Table 4. Participants generally rated their own

Author's personal copy

Table 4 Mean values of the measure indicating participants' *subjective assessment* (standard deviations in parentheses)

Measure	Condition		
	Unimodal	Congr. multimodal	Incongr. multimodal
Competence self-rating	4.60 (0.50)	4.67 (0.58)	4.05 (0.81)

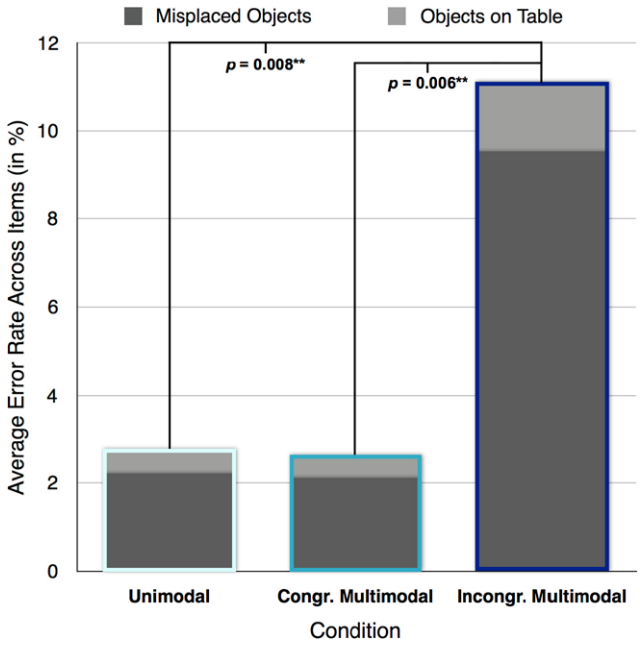

Fig. 4 Bar chart visualizing participants' *objective assessment* based on the average error rate per group across all nine objects handled in the experimental task

competence as high, with mean values between 4.05 and 4.60 in all three groups. However, a one-way ANOVA indicated a significant effect of experimental condition on participants' self-ratings regarding their task-related competence, $F(2, 59) = 5.83$, $p < 0.01$. Pairwise comparisons with Tukey's post-hoc test further revealed that participants in the incongruent multimodal group ($M = 4.05$, $SD = 0.81$) rated their own performance significantly worse than participants both in the unimodal group ($M = 4.60$, $SD = 0.50$), $p = 0.02$, and in the congruent multimodal group ($M = 4.67$, $SD = 0.58$), $p < 0.01$.

Results of *objective assessment* ratings are illustrated in Fig. 4. The average error rate across all nine kitchen objects handled in the experiment was found to be highest for the incongruent multimodal condition with a total average error rate of 11.12 %. This comprises an error rate of 9.53 % with regard to misplaced objects and an additional mean error of 1.59 % with regard to objects that were placed on the adjacent table, indicating that the participant had failed to understand the robot's instruction. In comparison, average error rates were much lower in the other two conditions with a combined error rate of 2.78 % (2.22 % misplaced objects) in the unimodal condition and a combined error rate of 2.65 % (2.12 % misplaced objects) in the congruent multimodal condition. In fact, a Kruskal-Wallis test showed a significant effect of condition, $\chi^2(2) = 9.06$, $p = 0.01$. Mann-Whitney tests were conducted with a Bonferroni correction to follow up this finding, yielding a significant difference both between the unimodal and the incongruent multimodal groups ($U = 127.50$, $p < 0.01$, $r = -0.38$) and between the congruent and incongruent multimodal groups ($U = 132.00$, $p < 0.01$, $r = -0.39$). That is, in accordance with Hypothesis 3, participants who received partly incongruent instructions from the robot performed significantly worse than participants who received either unimodal or congruent multimodal instructions from the robot.

Finally, Spearman's correlation analysis showed a significant negative correlation between *objective* and *subjective* assessment measures ($r = -0.41$, $p = 0.001$). That is, participants' self-ratings were generally in line with their objective assessments: the more mistakes participants made in the experimental task, the lower they actually rated their own competence afterwards.

5 Discussion and Conclusion

We conducted an experiment to investigate how hand and arm gestures affect anthropomorphic perceptions and the mental models humans form of a humanoid robot. For this, we manipulated the non-verbal behaviors of the humanoid robot in three experimental conditions: (1) no gesture, (2) congruent gesture, and (3) incongruent gesture. We particularly focused on participants' attribution of typically human traits to the robot, likability, shared reality, as well as future contact intentions. By 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 theory-driven approach is characterized by the application of social psychological theories of (de-) humanization [9, 14] to HRI. By adapting these measures of anthropomorphism from research on human nature 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. We hypothesized a positive effect of robot gesture on dependent variables measuring participants' *perception of the robot* (H1 and H2). In addition, we predicted a negative effect of incongruent gesture with regard to participants' *task-related performance* (H3).

Author's personal copy

The results support Hypotheses 1 and 2 by showing that the robot's gestural behavior tends to result in a more positive subsequent evaluation of all dependent measures by the human participants. Intriguingly though, this observation was only statistically significant at the 5 % level when the incongruent multimodal condition was compared to the unimodal condition. That is, when the robot performed partly non-matching gestures, it was perceived and rated more positively than when it only used speech or when it performed congruent multimodal behavior. Specifically, this means that partly incongruent multimodal behavior resulted in greater perceived humanlikeness, likability, shared reality, and future contact intentions with regard to the robot.

In this way, the results actually exceed the hypothetical expectations, especially those expressed by Hypothesis 2: not only do they indicate that a robot with occasionally incorrect gestures is more favorable than a non-gesturing robot; they surprisingly suggest that human interaction partners even favor such partly incongruent multimodal behavior over completely matching multimodal behavior. At first, this finding appears counterintuitive—how can it therefore be interpreted?

The present analyses particularly focused on participants' attribution of typically human traits to the robot and resulting anthropomorphic inferences. The results may be better understood if the robot's partly incongruent co-verbal gestures are not just considered as non-matching utterances, but as unpredictable behavior. From this perspective, the present findings are actually in line with previous research on anthropomorphism and social robots suggesting that implementing some form of unpredictability in a robot's behavior can create an illusion of the robot being 'alive' [5]. Thus, participants in this group may have attributed intentions to the robot based on its unpredictable behavior, e.g., by assuming that it deliberately tried to confuse them; indeed, several participants in the incongruent multimodal group approached the experimenter after the interaction, reporting that the robot was "cheeky" or was "trying to fool" them. A similar interpretation is based on the concept of *effectance* [25], i.e., the motivation to interact effectively in one's environment or, applied to anthropomorphism, with non-human agents. In this sense, the attribution of human characteristics to non-human agents serves to increase a person's ability to make sense of an agent's actions, to reduce uncertainty associated with the agent, and to increase confidence in future predictions regarding this agent [7].

An alternative interpretation of the results is that participants perceived the robot's incongruent gestures as errors or 'imperfections' which made the robot appear more humanlike and less machinelike, and as a result, generally more likable. These interpretations of the results suggest that a certain level of unpredictability or 'imperfection' in a humanoid robot, as given in the incongruent gesture condition, can actually lead to a greater attribution of human traits to the robot and a more positive HRI experience. Although this observation certainly depends on the given context and task, e.g., whether or not the correctness and reliability of the robot's behavior are vital, as well as on the length and frequency of interaction, it could potentially lead to a paradigm shift in the design of the 'perfect' social robot or artificial companion. Thus, it should be further elucidated in future HRI research.

The analysis of participants' task-related performance based on the number of objects that were not correctly placed at their designated locations revealed a significant effect of condition supporting Hypothesis 3. The results suggest that the partly non-matching gestures affected the participants' perception of the robot's instructions and had a negative impact on their performance. This interpretation is further supported by the fact that participants in the incongruent multimodal group rated their own competence at solving the task significantly lower than participants in the other two groups. The observed correlation between subjective and objective assessment measures thus indicates good self-assessment on the part of the participants.

In view of this finding, it appears even more surprising that the mean ratings of the dependent variables measuring participants' perception of the robot were highest in the incongruent condition. That is, although the robot's behavior negatively affected the participants' task-related performance, they still rated the robot as being more likable, reported greater shared reality with the robot, and expressed a greater desire to live with it than participants in the other groups. These findings therefore emphasize the positive impact of the incongruent gesture condition on participants' evaluation and anthropomorphization of the robot and should be systematically investigated in future studies.

Future research should also address the generalizability of our findings regarding anthropomorphic inferences and incongruent modalities with other robotic platforms, e.g., with non-humanoid robots. Moreover, 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 the extent to which anthropomorphic inferences, likability, shared reality and future contact intentions are determined by the robot's arm gestures versus gaze alone. Ideally, since this was not considered in our current experiment, the robot's behavior should also be parameterized to adapt to the human participant's feedback in future studies. Finally, the results from this study apply to the specific population of novices interacting with the Honda humanoid robot for a short period of time; it is indeed possible that multiple interaction sessions with the robot may result in different findings. Therefore, such long-term effects should be further investigated using similar measures as in the present study.

Author's personal copy

For the time being, however, the present results emphasize the importance of displaying gestural behaviors in social robots as significant factors that contribute to smooth and pleasant HRI. Finally, by revealing the positive impact of the incongruent gesture condition on participants' evaluation of the robot, our findings contribute to an advancement in HRI and give new insights into human perception and understanding of gestural machine behaviors.

Acknowledgements The work described was supported by the Honda Research Institute Europe and the Center of Excellence 'Cognitive Interaction Technology'.

References

1. Bartneck C, Croft E, Kulic D (2008) Measuring the anthropomorphism, animacy, likeability, perceived intelligence and safety of robots. In: Proceedings of the metrics of human-robot interaction workshop, technical report 471, pp 37–41
2. Bergmann K, Kopp S, Eyssel F (2010) Individualized gesturing outperforms average gesturing—evaluating gesture production in virtual humans. In: Proceedings of the 10th conference on intelligent virtual agents. Springer, Berlin, pp 104–117
3. Breazeal C (2002) Designing sociable robots. AAAI Press, Menlo Park
4. Buisine S, Abrilian S, Martin J-C (2004) Evaluation of multimodal behaviour of embodied agents—cooperation between speech and gestures. In: From brows to trust, vol. 7, pp 217–238
5. Duffy BR (2003) Anthropomorphism and the social robot. Robot Auton Syst 42(3–4):177–190
6. Echterhoff G, Higgins ET, Levine JM (2009) Shared reality: experiencing commonality with others' inner states about the world. Perspect Psychol Sci 4:496–521
7. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 114(4):864–886
8. Goldin-Meadow S (1999) The role of gesture in communication and thinking. Trends Cogn Sci 3:419–429
9. Haslam N, Bain P, Loughnan S, Kashima Y (2008) Attributing and denying humanness to others. Eur Rev Soc Psychol 19:55–85
10. Iio T, Shiomi M, Shinozawa K, Akimoto T, Shimohara K, Hagita N (2011) Investigating entrainment of people's pointing gestures by robot's gestures using a WOZ method. Int J Soc Robot 3(4):405–414
11. Kendon A (1986) Some reasons for studying gesture. Semiotica 62:1–28
12. Kopp S, Wachsmuth I (2004) Synthesizing multimodal utterances for conversational agents. Comput Animat Virtual Worlds 15(1):39–52
13. Krämer N, Simons N, Kopp S (2007) The effects of an embodied conversational agent's nonverbal behavior on user's evaluation and behavioral mimicry. In: Proceedings of the 7th conference on intelligent virtual agents, vol 4722. Springer, Berlin, pp 238–251
14. Loughnan S, Haslam N (2007) Animals and androids: implicit associations between social categories and nonhumans. Psychol Sci 18:116–121
15. Ltd Honda Motor Co (2000) The Honda Humanoid Robot Asimo, year 2000 model
16. McNeill D (1992) Hand and mind: what gestures reveal about thought. University of Chicago Press, Chicago
17. Mutlu B, Shiwa T, Kanda T, Ishiguro H, Hagita N (2009) Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In: Proceedings of the 4th ACM/IEEE international conference on human-robot interaction, pp 61–68
18. Neff M, Wang Y, Abbott R, Walker M (2010) Evaluating the effect of gesture and language on personality perception in conversational agents. In: Proceedings of the 10th international conference on intelligent virtual agents. Springer, Berlin, pp 222–235
19. Salem M (2012) Conceptual motorics—generation and evaluation of communicative robot gesture. Logos Verlag, Berlin. PhD Dissertation
20. Salem M, Eyssel F, Rohlfing K, Kopp S, Joublin F (2011) Effects of gesture on the perception of psychological anthropomorphism: a case study with a humanoid robot. In: Proceedings of the third international conference on social robotics. Lecture notes in artificial intelligence, vol 7072. Springer, Berlin, pp 31–41
21. Salem M, Kopp S, Wachsmuth I, Rohlfing K, Joublin F (2012) Generation and evaluation of communicative robot gesture. Int J Soc Robot 4(2):201–217. Special issue on expectations, intentions, and actions
22. Schröder M, Trouvain J (2003) The German text-to-speech synthesis system MARY: a tool for research, development and teaching. Int. J. Speech Technol., 365–377. doi:10.1023/A:1025708916924
23. Sidner CL, Kidd CD, Lee C, Lesh N (2004) Where to look: a study of human-robot engagement. In: Proceedings of the 9th international conference on intelligent user interfaces, pp 78–84
24. Steinfeld A, Jenkins OC, Scassellati B (2009) The Oz of Wizard: simulating the human for interaction research. In: ACM/IEEE international conference on human-robot interaction (HRI), pp 101–108
25. White RW (1959) Motivation reconsidered: the concept of competence. Psychol Rev 66:297–331
26. Yamazaki A, Yamazaki K, Kuno Y, Burdelski M, Kawashima M, Kuzuoka H (2008) Precision timing in human-robot interaction: coordination of head movement and utterance. In: Proceedings of the ACM/SIGCHI conference on human factors in computing systems, pp 131–140
27. Yoshikawa Y, Shinozawa K, Ishiguro H, Hagita N, Miyamoto T (2006) Responsive robot gaze to interaction partner. In: Proceedings of the robotics: science and systems conference, pp 37–43

Maha Salem received her Ph.D. in Computer Science with a focus on Social Robotics from Bielefeld University, Germany, in 2012. She subsequently worked as a Postdoctoral Research Associate at Carnegie Mellon University in Qatar and is currently a member of the Adaptive Systems Research Group at the University of Hertfordshire, UK. Her research interests lie in the field of social human-robot interaction with special focus on multimodal interaction and non-verbal expressiveness. She is further interested in cross-cultural HRI as well as in safety aspects for the design of robotic assistants.

Friederike Eyssel is an Assistant Professor and head of the group "Gender and Emotion in Cognitive Interaction Technology" at the Center of Excellence 'Cognitive Interaction Technology' (CITEC) in Bielefeld, Germany. She earned her Masters Degree in Psychology from University of Heidelberg, Germany, in 2004 and received her PhD in Psychology from Bielefeld University, Germany, in 2007. Friederike Eyssel is interested in various research topics ranging from social robotics and intelligent social agents to attitudes, sexual violence, sexism, dehumanization, and prejudice reduction.

Katharina Rohlfing received her Ph.D. in Linguistics from Bielefeld University, Germany, in 2002. Since 2008, she is head of the Emergentist Semantics Group within the Center of Excellence 'Cognitive Interaction Technology' (CITEC), Bielefeld University, Germany. She is interested in multimodal learning processes.

Stefan Kopp is a Professor and research group leader at the Center of Excellence 'Cognitive Interaction Technology' (CITEC) and principle investigator in the SFB 673 'Alignment in Communication' at

Author's personal copy

Bielefeld University. His research interests center around enabling intelligent systems, including robots and virtual characters, to engage in human-like conversational interaction and cooperation, for which he combines empirical studies with cognitive modeling and machine learning techniques. Stefan Kopp is the current president of the German Cognitive Science Society (GK).

Frank Joublin received a European Ph.D. degree in Neurosciences from the University of Rouen, France, in 1993. He is now Principal Scientist at the Honda Research Institute Europe, Germany, and since 2008 board member of the CoR-Lab Graduate School, Bielefeld University, Germany. His research interests include auditory scene analysis, language and semantic acquisition and developmental robotics.