UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Can I get your (robot) attention?
Human sensitivity to subtle hints of human-likeness in a humanoid robot's behavior

Permalink

https://escholarship.org/uc/item/9qc821pz

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Ghiglino, David
De Tommaso, David
Willemse, Cesco
et al.

Publication Date

2020

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Can I get your (robot) attention?
Human sensitivity to subtle hints of human-likeness in a humanoid robot's behavior

Ghiglino D. (Davide.Ghiglino@iit.it) [1-2] , De Tommaso D. (Davide.DeTommaso@iit.it) [1], Willemse C. (Cesco.Willemse@iit.it) [1], Marchesi S. (Serena.Marchesi@iit.it) [1], Wykowska A. (Agnieszka.Wykowska@iit.it) [1]

[1] Istituto Italiano di Tecnologia, Via Morego, 30, 16163, Genova, Italy
[2] DIBRIS, Università degli Studi di Genova, Via Opera Pia, 13, 16145, Genova, Italy

Abstract

Designing artificial agents that can closely imitate human behavior, might influence humans in perceiving them as intentional agents. Nonetheless, the factors that are crucial for an artificial agent to be perceived as an animated and anthropomorphic being still need to be addressed. In the current study, we investigated some of the factors that might affect the perception of a robot's behavior as human-like or intentional. To meet this aim, seventy-nine participants were exposed to two different behaviors of a humanoid robot under two different instructions. Before the experiment, participants' biases towards robotics as well as their personality traits were assessed. Our results suggest that participants' sensitivity to human-likeness relies more on their expectations rather than on perceptual cues.

Keywords: Human-robot interaction, humanoid robot, social cognition, intentional stance, mental states, instruction manipulation

Introduction

In everyday life, we are frequently exposed to different smart technologies. From our smartphones to avatars in computer games, and soon perhaps humanoid robots, we are surrounded by artificial agents created to interact with us. Already during the design phase of an artificial agent, engineers often endow it with functions aimed to promote the interaction and engagement with it, ranging from its "communicative" abilities to the movements it produces. The idea that an artificial agent able to behave like a human being would boost the spontaneity and naturalness of interaction is well supported by the literature (Ficocelli, Terao, Nejat, 2015; Mirning et al., 2017; Wiese, Metta & Wykowska, 2017). Providing an artificial agent with human-like behaviors might increase social attunement toward it, and this aspect might be crucial for deploying artificial agents in environments where social interaction with them is desirable (e.g., robot-assisted training for individuals diagnosed with autism; Scassellati, Admondi, Matarić, 2012). In fact, several authors demonstrated the advantages of providing artificial agents with human-like behaviors on the quality of interaction with humans (Hancock et al., 2011; Thepsoonthorn, Ogawa & Miyake, 2018).

Perceiving human-likeness from an artificial agent's behavior appears to be modulated by its behavioral capabilities, ranging from the kinematics of the movement (Gielniak, Liu & Thomaz, 2013) to the agent's responsiveness to external stimuli (Willemse & Wykowska, 2019). Even during the interaction with conspecifics, humans rely partially on motion cues when they need to infer the mental states underpinning behavior. Similar processes might be activated during the interaction with embodied artificial agents, such as humanoid robots. At the same time, a humanoid robot that can faithfully reproduce human-like behavior may undermine the interaction, causing a shift in attribution: from being endearing to being uncanny (Mori, 1970). Furthermore, it is still not clear whether individual biases and prior knowledge related to artificial agents can override perceptual evidence of human-like traits (Hinz, Ciardo & Wykowska, 2019). We hypothesize that human sensitivity to such characteristics varies depending on individual differences and available contextual information. The current study aims to investigate human sensitivity to anthropomorphic characteristics of robot's behavior, based on motion cues, under different conditions of prior knowledge. To meet this aim, we manipulated the human-likeness of the behavior displayed by the robot and the explicitness of instructions provided to the participants. As a secondary aim, we explored some of the individual differences that affect general attitudes towards robots, and the attribution of human-likeness consequently.

Methods

Participants

Seventy-nine participants took part in the experiment (mean age = 24.0, SD = 4.4, 50 females). All participants reported no history of psychiatric or neurological diagnosis, substance abuse, or psychiatric medication. Our experimental protocols followed the ethical standards laid down in the Declaration of Helsinki and were approved by the local Ethics Committee (Comitato Etico Regione Liguria). Each participant provided written informed consent to participate in the experiment. Participants were not informed regarding the purpose of the study before the experiment but were debriefed upon completion.

Stimuli and Apparatus

In the current study, we sat our participants in a dimly lit sound-attenuated room, in front of an iCub robot (Metta et al. 2008; Natale et al. 2017) that was "playing" a solitaire card

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

game on a laptop located in front of it. We placed a screen connected to a loudspeaker on the right of the iCub robot, on which we played scenes of various movies that were aimed to "distract" the robot from the game. Neither of the screens' displays was visible from the participant's position, but the sound produced by the loudspeaker was audible to everyone in the room (Fig. 1). The setup of the current study was the replica of a previous attentional capture experiment, which involved human participants playing the same solitaire card game while being distracted by the same sequence of movie scenes (see Ghiglino, De Tommaso & Wykowska, 2018 for details).



Figure 1: Experimental setup.

The setup of the current study was the replica of a previous attentional capture experiment, which involved human participants playing the same solitaire card game while being distracted by the same sequence of movie scenes (see Ghiglino, De Tommaso & Wykowska, 2018 for details).

Experimental design and procedure. Prior to the experiment, we asked all participants to complete a brief sociodemographic questionnaire along with the Autism Quotient test (AQ, Baron-Cohen et al., 2001), the Big Five Inventory (BFI, John & Srivastava, 1999) and the Negative Attitude Towards Robots Scale (NARS, Syrdal, et al., 2009). We adopted these questionnaires as they are all freely available, easy to administer, and vastly used to broadly assess individual differences that might affect human-robot interaction (see, for example, Schweinberger, Pohl & Winkler, 2020; Muller & Richert, 2018).

All participants of the present experiment were exposed to two different conditions determined by the behavior displayed by the robot: human-like or machine-like. The order of these conditions was counterbalanced between participants. Each behavior consisted of an 8-minutes sequence of eye- and head-movements.

In the human-like condition, the robot's behavior was derived from the recordings of a human participant's eyes and head movement collected using an eye-tracker (Tobii Pro Glasses 2) and an inertial sensor (Bosch Sensortec BNO055 Intelligent 9-Axis Absolute Orientation Sensor) during the attentional capture experiment mentioned above. Human data recorded in our previous experiment were transferred to the iCub head and eyes using a minimum-jerk controlling algorithm. It is important to point out that the behavior observed in the recordings of the human participant was highly variable: each reaction to a distracting stimulus was different from the others in terms of temporal and spatial kinematics (ranging from minimal and fast to wide and slow movements). The behavior displayed by the robot in the "human-like" condition was aimed to embody the same variability and unpredictability of the behavior recorded from the human.

In contrast, for the machine-like condition, we programmed the robot to display repetitive, predictable, and constant behavior. Thus, the machine-like behavior consisted of only one pattern of neck and eye movements, based on the average temporal and spatial movement dynamics extracted from the human recording of the aforementioned experiment. To maximize the difference between the two conditions, during the machine-like behavior, the robot was programmed to move its eyes from left to right repetitively while "playing" the solitaire card game and to react to each distracting stimulus with exactly the same head turn.

We asked the first forty participants (mean age = 24.1±3.73; mean education = 15.8±2.3; 24 females) to carefully observe the robot's behavior during both conditions without adding any further instruction or information. The remaining thirty-nine participants (mean age =24.3±5.07; mean education = 15.2±2.0; 26 females) were told explicitly, from the beginning of the experiment, that the robot would display two different behaviors, and that their task would be to identify which one was based on a human's recordings.

After each condition, all seventy-nine participants filled out the GodSpeed questionnaire (Bartneck et al., 2009) to assess the tendency to attribute anthropomorphic, animated and likable traits to a robot, and they took part in the InStance test (Marchesi et al., 2019) to investigate the tendency of humans to explain the behavior of a robot using either a mentalistic or a mechanistic vocabulary.

After the completion of both experimental sessions and the questionnaires, all participants were asked if they noticed any difference between the two behaviors displayed by the robot. In case of a positive answer, participants were asked to elaborate on their answer, explicating which one of the two behaviors they thought it was more similar to human behavior and why. We expected that participants who noticed the difference between the two conditions would be unanimous on the "correct" attribution of human-likeness. However, we received unexpected human-likeness attributions toward the machine-like condition that we kept into consideration during the data analysis. Eventually, this final explicit question allowed us to differentiate people in terms of sensitivity to the behavioral manipulation and in terms of correctly attributed/misattributed human-likeness.

Data Analysis

To explore the effects of our experimental manipulation, several mixed effect general linear models (GLM) were applied in R studio. In each model, we considered the responses in the GodSpeed questionnaire and in the InStance

test as separate dependent variables and each participant's intercept as a random factor. We included instruction manipulation (Explicit vs No Instructions) and the behavior displayed by the robot (HumanLike vs MachineLike) as fixed factors. This family of models allowed us to explore the main effects of the single factors and the interaction between the two.

Additionally, we aimed at exploring the effect of participants' individual attribution of human-likeness on the InStance and the GodSpeed ratings. Thus, we further grouped our participants based on their sensitivity to the subtle differences between the robot's behaviors and on the explicit attribution of human-likeness (provided at the end of the experiment). To avoid confounding effects and/or overfitting of the data, we analyzed participants that received explicit instructions separately from participants that received no instructions. This decision was made also taking into consideration the way participants distributed themselves in the three response groups across the two instructions conditions (under no instructions: 14 correctly attributed human-likeness, 9 misattributed human-likeness, 17 no attribution; under explicit instructions: 31 correctly attributed human-likeness, 8 misattributed human-likeness, 0 no attribution). This between-groups difference was tested using a chi-squared test. For all the mixed models, pairwise post-hoc comparisons were estimated using the Tukey method. Due to the way linear mixed models partition variance, and the lack of consensus on the calculation of effect sizes for individual model terms (Rights and Sterba, 2019), we estimated standardized effect sizes only in post-hoc analyses.

To investigate individual differences that affect human sensitivity to subtle hints of human-likeness in a humanoid robot's behavior, we calculated Pearson's correlation coefficients between the AQ, BFI, NARS, sociodemographic information, GodSpeed questionnaire, and Instance tests. Since we were interested in assessing individual differences that might play a role in the general attitude towards robots, for each participant we used the averages of the GodSpeed subscales and InStance scores as input variables of the correlation matrix.

Results

Instruction Manipulation and Robot Behavior

Instance ratings. We did not find any significant effects on the InStance scores due to the instructions manipulation ($F(1, 77)=0.41$, $p=.522$), of the behavior displayed by the robot ($F(1, 77)=2.16$, $p=.146$) or of the interaction between the two ($F(1, 77)=0.57$, $p=.455$) (Fig. 2). **GodSpeed ratings.** We found a significant interaction effect on the Anthropomorphism scores between instructions manipulation and behavior displayed by the iCub ($F(1, 77)=5.64$, $p=.020$), paralleled by a main effect of the behavior ($F(1, 77)=11.11$, $p=.001$). A null effect of instructions manipulation emerged from the data on this subscale ($F(1, 77)=0.05$, $p=.82$). Under explicit instructions, planned comparisons revealed a significant difference in Anthropomorphism scores: participants tended to attribute higher anthropomorphism to the human-like behavior than to the machine-like behavior ($t(77)=4.01$, $p<.001$). The same pattern was found on the Animacy subscale scores, highlighting an interaction between the instructions and the behavior ($F(1, 77)= 9.33$, $p=.003$), a main effect of the behavior ($F(1, 77)= 9.08$, $p=.004$) and a non-significant effect of the instructions ($F(1, 77)=0.20$, $p=.654$). Planned comparisons pointed out a significant difference in the Animacy scores between the human-like and the machine-like behaviors in the group that received explicit instructions ($t(77)=4.26$, $p<.001$). Interestingly, for the Likeability subscale scores, we found a single main effect of the instruction manipulation ($F(1, 77)=12.14$, $p<.001$), but neither a significant effect of behavior ($F(1, 77)=3.50$, $p=.065$) nor of interaction ($F(1, 77)=2.03$, $p=.158$). Post-hoc comparisons revealed a significant difference between the two instructions provided to the participant on the perceived likeability of the robot both after the human-like ($t(77)=-2.71$, $p=.038$) and after the machine-like ($t(77)=-3.93$, $p<.001$) behaviors (see Fig. 2 for details).

Robot's Behavior and Participants' attribution

The frequencies of participants' human-likeness attribution were different between the two instructions we provided them with ($\chi^2(2)= 23.47$, $p<.001$). Thus, we ran the subsequent analyses separately for the two instructions groups.

No instructions group. No main effect of the robot's behavior was found on the InStance ratings ($F(1, 37)=0.06$, $p=.805$), nor on the Anthropomorphism ($F(1, 37)= 0.35$, $p=.558$), Animacy ($F(1, 37)= 0.59$, $p=.448$) and Likeability ($F(1, 37)=0.23$, $p=.638$) subscales of the GodSpeed.

Similarly, participant's human-likeness attribution did not affect neither InStance ratings ($F(2, 37)=1.04$, $p=.363$), nor Anthropomorphism ($F(2, 37)=1.95$, $p=.157$), Animacy ($F(2, 37)= 0.52$, $p=.602$) or Likeability scores ($F(2, 37)=0.42$, $p=.662$).

No interaction between robot's behavior and participants' attribution was found on the InStance ($F(2, 37)=1.76$, $p=.186$) and Anthropomorphism ($F(2, 37)=2.21$, $p=.124$). An interaction between behavior and attribution was found on Animacy ($F(2, 37)=6.16$, $p=.004$) and Likeability ($F(2, 37)=7.65$, $p=.002$) scores. The effect on the Animacy scores did not survive post-hoc comparisons. Post-hoc analysis revealed that participants misattributing human-likeness tended to attribute higher likeability to the robot displaying the machine-like behavior, compared to the human-like behavior ($t(37)=3.77$, $p=.035$, $d=1.24$).

Explicit instructions group. No main effect of the robot's behavior was found on the InStance ratings ($F(1, 37)=0.01$, $p=.940$), or on the Anthropomorphism ($F(1, 37)=0.51$, $p=.477$), Animacy ($F(1, 37)=0.19$, $p=.669$) and Likeability ($F(1, 37)=0.22$, $p=.639$) subscales of the GodSpeed.

No main effect of participant's attribution was found on the InStance ratings ($F(1, 37)=0.40$, $p=.529$), or on the Anthropomorphism ($F(1, 37)= 1.45$, $p= .235$), Animacy ($F(1,$

37)=0.00, p=1.00) and Likeability (F(1, 37)=0.45, p=.504) subscales of the GodSpeed.

A significant interaction between the robot's behavior and participant's human-likeness attribution was found on the InStance ratings (F(1, 37)=4.31, p=.045), and on the Anthropomorphism (F(1, 37)= 16.89, p<.001), Animacy (F(1, 37)= 25.40, p<.001) and Likeability (F(1, 37)=4.36, p=.044) scores (Fig. 3). The interaction effect on the InStance scores did not survive post-hoc comparisons. Planned comparisons revealed that participants making the correct human-likeness attribution provided higher ratings on Anthropomorphism (t(37)=5.33, p<.001, d=-1.75), Animacy (t(37)=6.04, p<.001, d=-1.98) and Likeability (t(37)=2.82, p=.036, d=-0.92) subscales after seeing the human-like

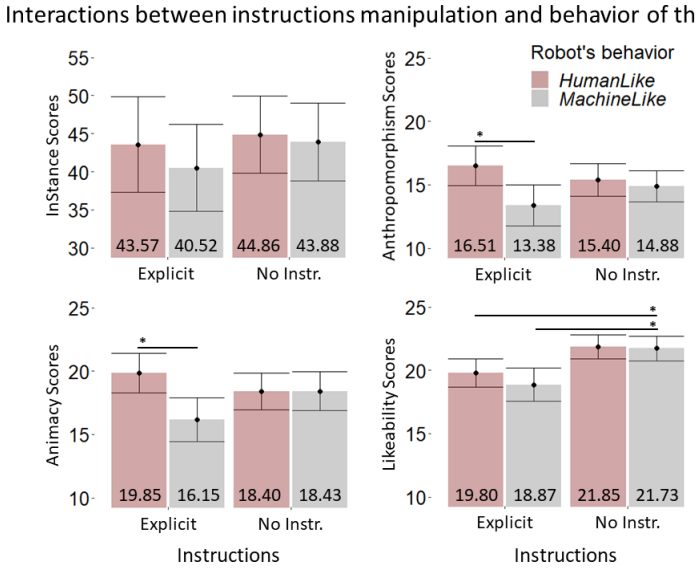


Figure 2: Bar charts showing the fixed effect on questionnaire scores due to the interaction between the instructions provided to the participants and the behavior displayed by the robot (GLM). Error bars: +/- 1. SE. Asterisks denote significant comparisons. Numbers indicate the mean value of each cell.

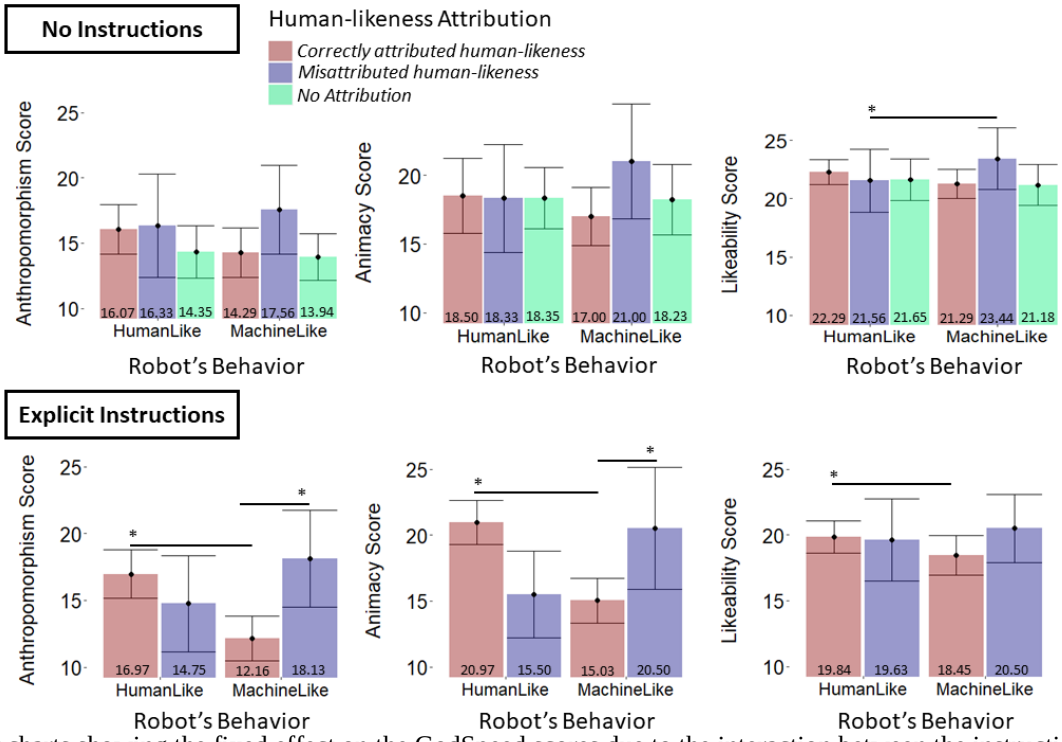


Figure 3: Bar charts showing the fixed effect on the GodSpeed scores due to the interaction between the instructions provided to the participants and the behavior displayed by the robot (GLM). Error bars: +/- 1. SE. Asterisks denote significant comparisons. Numbers indicate the mean value of each cell.

behavior. Additionally, participants providing the unexpected attribution rated higher than participants providing the expected attribution after seeing the machine-like behavior on the Anthropomorphism ($t(37)=-3.23$, $p=.010$, $d=-0.30$) and on the Animacy ($t(37)=-2.97$ $p=.021$, $d=-0.72$) subscales.

Individual differences

Our analyses showed small negative correlations between the InStance score and years of education ($r(77)=-.257$, $p=.022$) and AQ scores ($r(77)=-.246$, $p=.029$). We also found small negative correlations between the Likeability subscale of the GodSpeed questionnaire and the three subscales of the NARS ("Situations of Interaction": $r(77)=-.27$, $p=.018$; "Social Influence": $r(77)=-.23$, $p=.040$; "Emotions in Interaction" $r(77)=-.29$, $p=.008$). Additionally, our results pointed out a systematic, although small, set of correlations between BFI and NARS subscales. Specifically, the Conscientiousness subscale of the BFI correlates negatively with the subscales "Situations of Interaction" ($r(77)=-.32$, $p=.004$) and "Social Influence" ($r(77)=-.28$, $p=.012$) of the NARS. In addition, the Neuroticism subscale of the BFI correlates positively with the subscales "Situations of Interaction" ($r(77)=.24$, $p=.036$) and "Emotions in Interaction" ($r(77)=.28$, $p=.011$) of the NARS.

Discussion

The main aim of the current study was to assess whether the information available prior to the interaction with an artificial agent modulates human sensitivity to subtle hints of an agent's human-likeness. Our data showed that prior knowledge related to the behaviors that we implemented in the robot affected the sensitivity to behavioral manipulation. When we provided no *a-priori* information related to the nature of the behaviors implemented in the robot, participants overlooked the details of the behaviors. Consequently, nearly half of the sample provided with no instructions was not able to recognize any difference between the human-like and the machine-like behaviors. Furthermore, even those participants who spotted the differences between the behaviors often misattributed human-likeness. In addition, we could not find any significant differences in their InStance and GodSpeed scores between conditions. In contrast, all the participants who received the explicit instructions detected a difference between the two behaviors, and this was reflected in the anthropomorphism, animacy, and likeability attributed to the robot.

When we prompted our participants' attention to notice hints of human-likeness in the behaviors of the robot, they tended to differentiate more their answers in the GodSpeed questionnaire between conditions, as if their perception of anthropomorphism, animacy, and likeability depended mainly on their belief of what a human-like movement should look like. This suggests that subtle evidence of behavioral human-likeness might be too weak of a signal during tasks merely involving the observation of artificial agents' behavior. This might be related to the fact that in natural interactions with humans, we are usually not monitoring (or not being asked to monitor) the human-likeness of the counterpart's behavior. Thus, human-likeness might be an implicit feature of human behavior, which we derive only if needed to explain the behavior of a non-human agent. Therefore, during everyday life, our sensitivity to such subtle hints might be low, as we more likely perceive "gestalt" relations between behavioral and contextual elements rather than pure and distinct behavioral features (Spelke, 1990; Hamlyn, 2017).

Our results suggest that the concept of human-likeness itself varies across individuals, overriding perceptual evidence – our participants tended to confirm their own biases and modulated their responses on the GodSpeed questionnaire based on their own perception of human-like behavior, rather than the actual human-like behavior of the robot. This casts a shadow on the idea that having artificial agents able to behave exactly like human beings would, improve social interaction with them, as people appear to have very different priors related to the concept of human-likeness. Indeed, participants perceived the robot as more likeable when they received no information related to its behaviors, regardless of its human-likeness. In other words, perceived, but not actual, human-likeness influenced the likeability of the robot. Thus, the attractiveness of interacting with a humanoid robot might be independent of the subtle behaviors it displays, but might rather depend on the users' attitudes toward it. This further suggests that the less an individual knows about the process of implementation of behavior in a robot, the more they enjoy the interaction with it and perceives it as more engaging. Taken together, these results suggest that the differences in knowledge between participants override perceptual evidence and tweak individual sensitivity to behavioral cues.

The presence of individual differences that affect the way humans interact with artificial agents is further supported by the correlation between BFI and NARS subscales. Our results showed that certain personality traits, such as neuroticism and conscientiousness, influenced participants' attitudes towards robots. High neuroticism scores are often associated with the tendency to experience negative emotions during social interaction (Kaplan et al., 2015). The positive correlation between neurotic traits and NARS scores supports previous literature, suggesting that neurotic people might experience discomfort during the interaction with artificial agents, similarly how they feel in interactions with other humans (Müller & Richert, 2018). On the other hand, high conscientiousness often relates to better self-regulation and emotional stability, which positively affect social interaction (Smith, Barstead, Rubin, 2017), and might as well ease the interaction with artificial agents. The negative correlation between InStance and AQ scores further support the idea that social abilities affect humans' general attitude towards artificial agents. Indeed, people with higher autistic traits appeared to have difficulties with explaining the behavior of a robot in terms of the underpinning mental states, relying more on mechanistic terms rather than on mentalistic

vocabulary. This might be due to the familiarity that a person has regarding a certain vocabulary when interpreting behaviors in general. In addition, we also found a negative correlation between the Instance test score and the participants' education. We speculate that participants with a higher level of education might be more familiar with the design and functionality of technology in general. This prior knowledge might bias them to explain our robot's behavior relying more on its mechanical apparatus rather than its "desires" and "intentions". We postulate that personality traits and attitudes that play a role in the interaction between humans translate into different approaches towards artificial agents as well. This hypothesis is further supported by the negative correlation between NARS subscales and the perceived Likeability of the robot, indicating that participants' attitudes towards robots affect their engagement during the interaction. Future studies should further explore individual differences that affect participants' behavior and attitudes toward robots to understand whether they play a similar role during human-human and human-robot interactions.

In conclusion, our study suggests that individual knowledge, beliefs and biases play a major role in modulating human perception of an artificial agent's behavior. These influences seem to be even stronger than perceptual evidence during observational scenarios and need to be taken into consideration in future studies.

Acknowledgments

This work has received support from the European Research Council under the European Union's Horizon 2020 research and innovation programme, ERC Starting grant ERC-2016-StG-715058, awarded to Agnieszka Wykowska.The content of this paper is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.

References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders*.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*.

Ficocelli, M., Terao, J., & Nejat, G. (2015). Promoting interactions between humans and robots using robotic emotional behavior. *In Proceedings of IEEE transactions on cybernetics*.

Ghiglino, D., De Tommaso, D., & Wykowska, A. (2018). Attributing human-likeness to an avatar: the role of time and space in the perception of biological motion. *In Proceedings of the 10th International Conference on Social Robotics*. Springer, Cham.

Gielniak, M. J., Liu, C. K., & Thomaz, A. L. (2013). Generating human-like motion for robots. *The International Journal of Robotics Research*.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*.

Hinz, N. A., Ciardo, F., & Wykowska, A. (2019). Individual differences in attitude toward robots predict behavior in human-robot interaction. *In International Conference on Social Robotics*. Springer, Cham.

John, O. P., and Srivastava, S. (1999). "The Big Five trait taxonomy: History, measurement and theoretical perspectives". *In Handbook of personality: Theory and research*. New York: Guilford.

Kaplan, S. C., Levinson, C. A., Rodebaugh, T. L., Menatti, A., & Weeks, J. W. (2015). Social anxiety and the Big Five personality traits: The interactive relationship of trust and openness. *Cognitive behaviour therapy*.

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the Intentional Stance towards humanoid robots? *Frontiers in psychology*.

Metta, G., Sandini, G., Vernon, D., Natale, L., & Nori, F. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. *In Proceedings of the 8th workshop on performance metrics for intelligent systems*. ACM.

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*.

Mori, M. (1970). The uncanny valley. *Energy*.

Müller, S. L., & Richert, A. (2018). The Big-Five Personality Dimensions and Attitudes to-wards Robots: A Cross-Sectional Study. *In Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*.

Natale, L., Bartolozzi, C., Pucci, D., Wykowska, A., Metta, G. (2017), The not-yet-finished story of building a robot child, *Science Robotics*.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological methods*.

RStudio Team (2015), RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, URL: http://www.rstudio.com/.

Scassellati, B., Admoni, H., & Matarić, M. (2012). Robots for use in autism research. *Annual review of biomedical engineering*.

Schweinberger, S. R., Pohl, M., & Winkler, P. (2020). Autistic traits, personality, and evaluations of humanoid robots by young and older adults. *Computers in Human Behavior*.

Smith, K. A., Barstead, M. G., & Rubin, K. H. (2017). Neuroticism and conscientiousness as moderators of the relation between social withdrawal and internalizing problems in adolescence. *Journal of youth and adolescence*.

Syrdal, D. S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2009). The negative attitudes towards robots scale and reactions to robot behavior in a live human-robot interaction study. *Adaptive and emergent behaviour and complex systems*.

Thepsoonthorn, C., Ogawa, K. I., & Miyake, Y. (2018). The relationship between robot's nonverbal behaviour and human's likability based on human's personality. *Scientific reports*.

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology*.

Willemse, C., & Wykowska, A. (2019). In natural interaction with embodied robots, we prefer it when they follow our gaze: a gaze-contingent mobile eye-tracking study. *Philosophical Transactions of the Royal Society*.