Understanding Anthropomorphism: Anthropomorphism is not a Reverse Process of Dehumanization

Jakub Złotowski[1,2](✉), Hidenobu Sumioka[2], Christoph Bartneck[1], Shuichi Nishio[2], and Hiroshi Ishiguro[2,3]

[1] HIT Lab NZ, University of Canterbury, Christchurch, New Zealand
jaz18@uclive.ac.nz
[2] Hiroshi Ishiguro Laboratory, Advanced Telecommunications Research Institute International, Kyoto, Japan
[3] Department of System Innovation, Graduate School of Engineering Science, Osaka University, Osaka, Japan

Abstract. Anthropomorphism plays an important role in human interaction with robots. However, our understanding of this phenomenon is still limited. In the previous research, we proposed to look at the work on dehumanization in order to understand what factors can affect a robot's anthropomorphism. Moreover, considering that there are two distinct dimensions of humanness, a two-dimensional model of anthropomorphism was proposed. In this paper we present a study in which we manipulated the perceived intentionality of a robot and appearance (Robovie R2 vs Geminoid HI-2), and measured how they affected the anthropomorphization of the robots on two dimensions of humanness. We did not find statistically significant differences in attribution of human traits and mind along two dimensions of humanness. However, after dividing the traits based on their valence, we found that Geminoid HI-2 was attributed significantly more negative human traits than Robovie R2. These results do not support the proposed two-dimensional model of anthropomorphism.

Keywords: Anthropomorphism · Dehumanization · Human-Robot Interaction · Moral agency

1 Introduction

People have a strong tendency to attribute human-like qualities and characteristics to non-human agents. This process is known as anthropomorphization and has been heavily investigated in the field of Human-Robot Interaction (HRI). Previous research proposed several factors that affect the extent to which a robot is anthropomorphized: appearance [3], movement [15], verbal communication [14], emotions [2] or gestures [12]. These studies regarded anthropomorphism as a uni-dimensional space between a machine and a human. However,

A. Kheddar et al. (Eds.): ICSR 2017, LNAI 10652, pp. 618–627, 2017.
https://doi.org/10.1007/978-3-319-70022-9_61

the uni-dimensionality of this phenomenon has never been proved. In this paper, we argue that it is not possible to understand anthropomorphism without understanding what is humanness.

1.1 Humanness

We can learn what does it mean to be a human by looking at a process of depriving objectified humans of characteristics regarded as crucial in order to be perceived and treated as a human. This process is known as dehumanization. The work by [7] reveals that humanness is a two-dimensional concept. The characteristics that distinguish us from the related category of animals are called Uniquely Human (UH). The second dimension is called Human Nature (HN) and it represents characteristics that are typical of or central to humans. These "core" characteristics may differ from those that distinguish us from other species, e.g. being cultured distinguishes us from other animals, but is not a core characteristic of humans. On the other hand, emotionality is a fundamental characteristic of humans, but does not distinguish us from animals. Haslam [7] in his model of dehumanization proposed that UH characteristics are developed in later stages of people lives and reflect socialization, such as intelligence, secondary emotions and intentionality. On the other hand he argues that HN attributes are inborn biological dispositions, such as cognitive openness, primary emotions and warmth.

Further evidence supporting existence of two dimensions of humanness comes from the work on mind attribution. Gray et al. [4] showed that people differentiate between two types of mind attribution: experience (the capability for feelings and sensations) and agency (the capability to act and intend). The research shows that the concepts of dehumanization and mind attribution overlap to a great extent [6]. Agency corresponds to UH and experience to HN dimensions.

Understanding the processes of dehumanization and mind attribution can help to identify the key characteristics for robots to affect their anthropomorphism. Furthermore, in order to fully understand anthropomorphism it may be beneficial to think about it as a two-dimensional space. This new perspective could help us to better understand some of the results of previous studies that were hard to interpret, e.g. the uncanny valley phenomenon seems to be caused by the experience dimension rather than agency [5].

In our previous work, we explored the suitability of a two-dimensional model of anthropomorphism in HRI [16]. In our model we proposed that a robot can be independently anthropomorphized on HN and UH dimensions, depending on what characteristics it is perceived as possessing. That is if a robot is perceived as intelligent or intentional, it will be anthropomorphized on UH dimension as suggested by the work on dehumanization [7]. On the other hand, a robot perceived as emotional or warm will be anthropomorphized on HN dimension. Since these dimensions are based on folks conceptions of humanness, the subjective perception of humans is more important than the objective properties of a robot, i.e. it is sufficient that people perceive a robot as possessing these characteristics irrespective of whether the robot possesses them or is even able to posses them.

Supporting the idea of a two-dimensional model of anthropomorphism, perceived emotionality of a robot affected only attribution of HN traits. However, the other factor manipulated in out previous study, perceived intelligence, did not affect neither UH nor HN. Considering that in the model of dehumanization proposed by [7], intelligence is a characteristic of UH dimension we expected that in the context of HRI increased perceived intelligence of a robot, will increase attribution of UH traits to it. The lack of the effect of perceived intelligence may indicate a limited impact of it on a robot's anthropomorphism or that UH dimension is less relevant in general in the context of HRI. The second interpretation would also mean that anthropomoprhism and dehumanization are not reverse processes since their dimensions differ. In the current work, we wanted to shed additional light on that possibility and investigated the impact of a robot's perceived intentionality and appearance on anthropomorphism. The research on dehumanization [7] suggests that perceived intentionality should be associated with UH and appearance with experience [5] dimensions. Since the concepts of dehumanization and mind attribution overlap [6] we formulated the following two hypotheses:

– *H1*: A robot perceived as intentional will be attributed more UH traits and agency than an unintentional robot.
– *H2*: A more humanlike looking robot will be attributed more HN traits and experience than a machine-like robot.

In the context of the relationship between anthropomorphism and dehumanization, if our H1 is supported, the results would suggest that the lack of the effect of perceived intelligence on UH in the previous study [16] was due to the limited impact of it on anthropomorphism. On the other hand, if our H1 is not supported, it would mean that two different factors (perceived intelligence and intentionality) that are supposed to be associated with UH do not affect that dimension in the context of anthropomorphism. In that case, two independent studies would suggest that anthropomorphism and dehumanization are not reverse processes since their dimensions differ (the lack of relevance of UH for anthropomorphism, while being a dimension of dehumanization).

1.2 Moral Agency

Apart from investigating the potential existence of two dimensions of anthropomorphism, it is also important to show how differentiating between the dimensions can impact HRI. In the recent years we can observe increased interest in moral responsibility of robots for their actions [10]. The work on dehumanization and mind attribution suggests that depending on what anthropomorphic characteristics a robot will have, may affect whether it will be perceived as responsible for its actions. The work on human perception of moral agency suggests that people perceived high in agency [6] and UH [1] are perceived as blameworthy for their wrongdoings. Based on that, we formulated the third hypothesis as:

– *H3*: A robot perceived as intentional will be attributed more responsibility for its immoral actions than an unintentional robot.

2 Method

We conducted an experiment with 2×2 between-subjects design with perceived internationality of robots and their appearance as factors. The perceived intentionality of robots was manipulated by their cheating (intentional) or non-cheating (unintentional) behavior and appearance by using a machine-like and human-like robot.

2.1 Participants

Fifty-two (34 male and 18 female) native Japanese speakers were recruited for this study. They were undergraduate students of various universities and departments located in Kansai region. They received ¥2000 as time compensation. Their age range was between 18 and 27 years ($M = 21.54, SD = 1.93$).

2.2 Materials and Apparatus

In this study participants interacted either with Geminoid HI-2 or Robovie R2. Geminoid HI-2 is a highly humanlike robot that is a copy of a real person. On the other hand, Robovie R2 is a machine-like robot that has some humanlike features, such as arms or head. Both robots spoke with the same synthesized voice. In order to ensure correct reactions from the robots we implemented this experiment as a Wizard-of-Oz study, where the robots' actions were controlled by a researcher who sat in another room. Since both robots use different operating systems, we implemented their reactions in a way that kept the response delay the same for both platforms.

2.3 Procedure

Participants were asked to play "Acchi muite hoi", a traditional Japanese game, with a robot called "Robo". They were told that a prototype of a robot was prepared for a tournament and they will test it. The robot has to win at least 80% of the games (which is 8 rounds out of 10 they were asked to play) or otherwise it will be destroyed. During the game, one participant moves a finger sideways in front of an opponent and at the same time the opponent moves her head sideways. Both of them do it after "hoi" in a phrase "Acchi muite hoi" that is said by one of the players. If a player moving his head does it in the same direction as the other player moving her finger, he loses the game. If they move in different directions then a player moving his head is the winner.

After explaining the rules of the game to the participants, the researcher played three example games with a robot. During a pilot study it became apparent that participants tried to cheat a robot by delaying their responses. Therefore, during the third example game, the researcher explained that timing is important and if a participant's response will be delayed, the robot will request to repeat the game. After confirming that participants understood the rules of the game, the researcher left the room leaving a participant alone with a robot. After all ten rounds were played, the experimenter returned to the room and asked a robot about the outcome of the game. Finally, participants were requested to fill out questionnaires on a computer and had a brief interview with a researcher who asked whether they noticed something unusual during the game. Participants were free to answer this question the way they wanted.

In the current setup a robot was a player moving its head and saying “Acchi muite hoi”, and a participant moved her finger. Previous research [13] shows that cheating behavior of a robot during a game increases perceived intentionality of it. In the unintentional (control) condition, a robot played all 10 rounds fairly and at the end it admitted not winning enough many games after it was asked about the outcome by the experimenter. In the intentional condition, during rounds 3, 6 and 9 a robot “cheated” by waiting until a participant moves her finger before making its own move or by changing the direction. Furthermore, when asked by the experimenter at the end of the game about the outcome, it claimed to have won more than 80% of games. The popularity of the game by people of all ages ensures that participants had good understanding of the rules and should understand when a robot makes an illegal move. Since some participants might have opposed the outcome announced by the robot, the researcher in both conditions informed them that the game was recorded and the video will be used to verify that everything worked correctly.

2.4 Measurements

As a manipulation check we measured perceived intentionality of a robot using three questions (“This robot is capable of doing things on purpose.”, “This robot is capable of planned actions.” and “This robot has goals.”) on a 7-point scale from strongly disagree (1) to strongly agree (7).

Perceived humanlikeness of a robot was measured on a scale derived from [9] that was back-translated to Japanese. It consists of 6 items that are rated on a 5-point semantic differential scale (e.g. “Please rate your impression of the robot on these scales: 1-inanimate, 5-living”).

Anthropomorphism was measured by the attribution of human traits belonging to UH and HN dimensions ([8], Japanese translation [11]). Participants were asked to rate the extent to which a robot possesses them on a 7-point Likert scale (e.g. “The robot is... sociable”) from not at all (1) to very much (7).

In addition, we measured attribution of mind to a robot using its 2 dimensions: experience and agency ([4], Japanese version [11]) – e.g. “To what extent do you think that Robo is capable of... hunger” on a 5-point Likert scale.

Moral agency of a robot was measured on a 7-point Likert scale using 2 items: “Robo deserves blame for its wrong-doings” and “Robo is ‘morally responsible’ for performing immoral behaviours”.

3 Results

Due to the brevity, the exact results of statistical tests are reported only for statistically significant differences at the level $p < .05$.

3.1 Perceived Intentionality

The scale created to measure perceived intentionality had weak internal consistency ($\alpha = .62$). Since the scale was not previously validated it was deemed as not sufficiently reliable to use as a manipulation check. Instead, we used descriptive statistics based on the post-study interviews. In the unintentional condition, none of the participants indicated any suspicious behaviour of a robot. In the intentional condition 20 participants said that a robot tried to cheat during the game, 4 indicated that it had some technical problems and 2 did not report any unusual behavior of a robot.

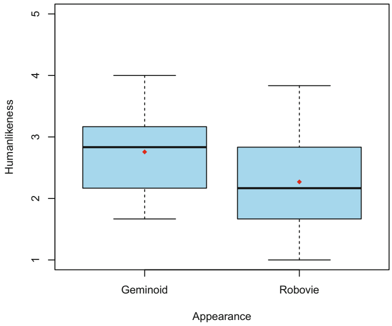
3.2 Perceived Humanlikeness

The scale of perceived humanlikeness had reliability at the level of Cronbach's $\alpha = .6$. However, as the scale was previously validated in the context of HRI and used in several papers, all items were kept. A two-way ANOVA with intentionality and appearance as factors showed a significant main effect of appearance on perceived humanlikeness ($F(1, 48) = 7.47, p = .009$, $\eta^2_G = .14$). Geminoid HI-2 ($M = 2.76, SD = 0.63$) was perceived as more humanlike than Robovie R2 ($M = 2.27, SD = 0.68$), see Fig. 1a. No other main or interaction effects were found.

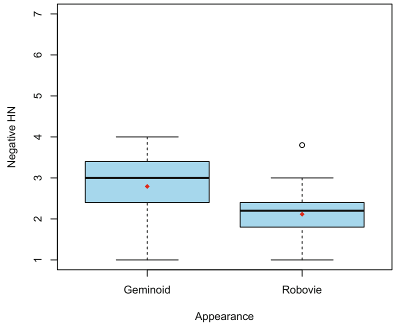
3.3 Human Trait Attribution

We looked at the attribution of human traits along two dimensions (UH and HN) as indicators of anthropomorphism. Although the reliability of HN and UH scales was unsatisfactory, $\alpha = .58$ and $\alpha = .57$ respectively, these scales are well validated in psychology and HRI, and all the items were kept. A two-way ANOVA with intentionality and appearance as factors did not show any statistically significant main or interaction effects on either of the scales. Considering the work of [11] that showed that Japanese differentiate between positive and negative dimensions of HN and UH scales, we calculated a composite measures of positive HN, negative HN, positive UH and negative UH using the items indicated in [11].

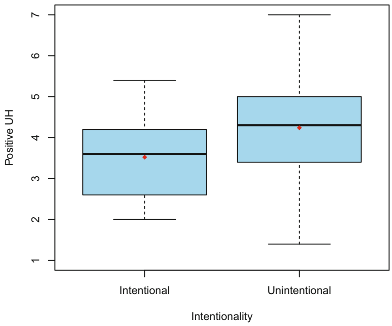
A two-way ANOVA with intentionality and appearance as factors did not show any statistically significant main or interaction effects on positive HN. However, there was a statistically significant main effect of appearance on attribution of negative HN traits, $F(1, 48) = 10.87, p = .002$, $\eta^2_G = .19$. Geminoid HI-2 ($M = 2.79, SD = 0.84$) was attributed more negative HN traits than Robovie R2 ($M = 2.12, SD = 0.63$), see Fig. 1b.



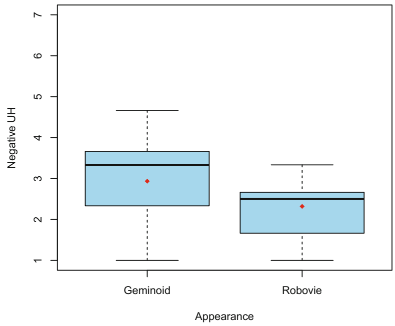
(a) Humanlikeness of robots based on their appearance.



(b) Attribution of negative Human Nature traits to robots based on their appearance.



(c) Attribution of positive Uniquely Human traits to robots based on their perceived intentionality.



(d) Attribution of negative Uniquely Human traits to robots based on their appearance.

Fig. 1. The effects of appearance and intentionality on anthropomorphism measures. Mean values are indicated by red dots. (Color figure online)

A two-way ANOVA showed a statistically significant main effect of intentionality on positive UH, $F(1, 48) = 5.23, p = .03$, $\eta^2_G = .1$. An unintentional robot ($M = 4.24, SD = 1.19$) was attributed more positive UH traits than an intentional robot ($M = 3.52, SD = 1.02$), see Fig. 1c.

A two-way ANOVA with intentionality and appearance as factors revealed a statistically significant main effect of appearance on negative UH, $F(1, 48) = 6.20, p = .02$, $\eta^2_G = .11$. Geminoid HI-2 ($M = 2.94, SD = 1.02$) was attributed more negative UH traits than Robovie R2 ($M = 2.32, SD = 0.73$), see Fig. 1d.

3.4 Mind Attribution

The internal consistency of both measures of mind attribution was high, experience $\alpha = .85$ and agency $\alpha = .71$. Two-way ANOVAs with appearance and intentionality as between subjects factors did not show any statistically significant main or interaction effects for attribution of either experience or agency. Based on the work by [11] with Japanese subjects, the experience scale was divided into positive and negative dimensions. Two-way ANOVAs showed that neither of these scales is affected by either appearance or intentionality.

3.5 Moral Agency

A two-way ANOVA with appearance and intentionality as between subjects factors did not reveal any statistically significant main or interaction effects on perception of a robot as a moral agent.

4 Discussion and Conclusions

The result of the post-study interviews show that participants verbally described a robot's behavior in intentional terms after it changed its responses during a game. This is in line with the work by [13] and shows that the manipulation in our study worked as planned.

The *H1*, a robot perceived as intentional will be attributed more UH traits and agency than an unintentional robot, was not supported by the results. Neither the attribution of UH traits nor agency to a robot was affected by its perceived intentionality. The work on dehumanization suggests that intentionality is related with UH [7]. In our previous work [16], we did not find an effect of another factor (intelligence) on attribution of UH traits to a robot. Put together, these results suggest that UH dimension has smaller impact on anthropomorphism than HN.

The previously proposed by us two-dimensional model of anthropomorphism [16] is not supported by our results. This indicates that anthropomorphism and dehumanization are not reverse processes since factors related with UH dimension of dehumanization do not affect attribution of UH traits in the context of anthropomorphism. Therefore, not all factors that lead to dehumanization of humans, can make a robot more anthropomorphic. Robotic platform designers who want to create an anthropomorphic robot should focus on factors affecting HN, such as emotionality or ability to feel, rather than UH.

The *H2*, a more humanlike looking robot will be attributed more HN traits and experience than a machine-like robot, was not supported by the results. Although, Geminoid HI-2 was perceived as more humanlike than Robovie R2, there was no statistically significant difference in attribution of HN traits or perceived capacity to experience. This result is not consistent with the work by [5,11] who suggested that more humanlike robots are attributed more capabilities to experience.

There are two primary differences between these studies and our work, which can explain different outcomes. Gray and Wegner [5] and Kamide et al. [11] used images and videos of robots. Therefore, participants in their studies had to base their judgments only on a robot's appearance. On the other hand, participants in our study were involved in an actual interaction with a robot. Even if they made an initial ascription of HN traits or capability to experience to a robot, during an interaction they were able to verify whether a robot really possessed these traits. The second difference between the work by [5] and ours is that their machine-like robot had no face. It is possible that a face is required for an agent to be attributed a capacity to experience, which produced a statistically non-significant result in our study since both robots had faces.

Although, there were no statistically significant differences in the attribution of UH and HN traits, the analysis of their positive and negative subscales shows an important cultural difference for studies with Japanese participants. Kamide et al. [11] suggested that Japanese differentiate not only between UH and HN traits, but also their positive and negative dimensions. This is supported by our data as participants differently attributed negative traits to Geminoid HI-2 and Robovie R2. It is possible that another factor than human-likeness is responsible for this result. For example, the stern appearance of Geminoid HI-2 compared with a child-like and friendly appearance of Robovie R2 may be responsible for that difference. Future studies may consider using a more friendly looking android, such as Erica, to investigate this possibility.

The *H3*, a robot perceived as intentional will be attributed more responsibility for its immoral actions than an unintentional robot, was not supported by the results. The results show that a robot which behaviour was perceived as more intentional, did not affect its perceived responsibility for wrongdoings. Furthermore, more humanlike appearance of the robot did not result in higher perceived moral agency. In other words, people do not perceive highly humanlike robots to be more responsible for their actions than robots that have machine-like features. The hypothesized effect of intentionality on perceived moral agency was based on a premise that agency/UH traits will moderate the effect of intentionality on moral agency. However, as the robots perceived as intentional were not attributed more UH traits or agency than the unintentional robots, our manipulation did not affect also their perceived moral agency.

Acknowledgements. The authors would like to thank Kaiko Kuwamura, Daisuke Nakamichi, Junya Nakanishi, Masataka Okubo and Kurima Sakai for their help with data collection. This work was partially supported by JST CREST (Core Research for Evolutional Science and Technology) research promotion program "Creation of Human-Harmonized Information Technology for Convivial Society" Research Area, ERATO and ISHIGURO symbiotic Human-Robot Interaction Project.

References

1. Bastian, B., Laham, S.M., Wilson, S., Haslam, N., Koval, P.: Blaming, praising, and protecting our humanity: the implications of everyday dehumanization for judgments of moral status. Br. J. Soc. Psychol. **50**(3), 469–483 (2011)

2. Eyssel, F., Hegel, F., Horstmann, G., Wagner, C.: Anthropomorphic inferences from emotional nonverbal cues: a case study. In: Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Viareggio, Italy, pp. 646–651 (2010)
3. Fischer, K., Lohan, K.S., Foth, K.: Levels of embodiment: linguistic analyses of factors influencing HRI. In: Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI 2012), pp. 463–470 (2012)
4. Gray, H.M., Gray, K., Wegner, D.M.: Dimensions of mind perception. Science **315**(5812), 619 (2007)
5. Gray, K., Wegner, D.: Feeling robots and human zombies: mind perception and the uncanny valley. Cognition **125**(1), 125–130 (2012)
6. Gray, K., Waytz, A., Young, L.: The moral dyad: a fundamental template unifying moral judgment. Psychol. Inq. **23**(2), 206–215 (2012)
7. Haslam, N.: Dehumanization: an integrative review. Pers. Soc. Psychol. Rev. **10**(3), 252–264 (2006)
8. Haslam, N., Loughnan, S., Kashima, Y., Bain, P.: Attributing and denying humanness to others. Eur. Rev. Soc. Psychol. **19**(1), 55–85 (2009)
9. Ho, C., MacDorman, K.: Revisiting the uncanny valley theory: developing and validating an alternative to the godspeed indices. Comput. Hum. Behav. **26**(6), 1508–1518 (2010)
10. Kahn Jr., P.H., Kanda, T., Ishiguro, H., Gill, B.T., Ruckert, J.H., Shen, S., Gary, H.E., Reichert, A.L., Freier, N.G., Severson, R.L.: Do people hold a humanoid robot morally accountable for the harm it causes? In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 33–40. ACM (2012)
11. Kamide, H., Eyssel, F., Arai, T.: Psychological anthropomorphism of robots: measuring mind perception and humanity in Japanese context. In: Herrmann, G., Pearson, M.J., Lenz, A., Bremner, P., Spiers, A., Leonards, U. (eds.) ICSR 2013. LNCS (LNAI), vol. 8239, pp. 199–208. Springer, Cham (2013). doi:10.1007/978-3-319-02675-6_20
12. Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., Joublin, F.: Effects of gesture on the perception of psychological anthropomorphism: a case study with a humanoid robot. In: Mutlu, B., Bartneck, C., Ham, J., Evers, V., Kanda, T. (eds.) ICSR 2011. LNCS (LNAI), vol. 7072, pp. 31–41. Springer, Heidelberg (2011). doi:10.1007/978-3-642-25504-5_4
13. Short, E., Hart, J., Vu, M., Scassellati, B.: No fair!! an interaction with a cheating robot. In: 5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010, Osaka, Japan, pp. 219–226 (2010)
14. Sims, V.K., Chin, M.G., Lum, H.C., Upham-Ellis, L., Ballion, T., Lagattuta, N.C.: Robots' auditory cues are subject to anthropomorphism. In: Proceedings of the Human Factors and Ergonomics Society, vol. 3, pp. 1418–1421 (2009)
15. Wang, E., Lignos, C., Vatsal, A., Scassellati, B.: Effects of head movement on perceptions of humanold robot behavior. In: Proceedings of the 2006 ACM Conference on Human-Robot Interaction (HRI 2006), vol. 2006, pp. 180–185 (2006)
16. Złotowski, J., Strasser, E., Bartneck, C.: Dimensions of anthropomorphism: from humanness to humanlikeness. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, pp. 66–73. ACM (2014)