Inclusive Dialogues: WokeBot Engaging Diversity Dilemmas

Mirjam de Haas
HU University of Applied Sciences
Utrecht, The Netherlands
mirjam.dehaas@hu

Veerle L.N.F. Hobbelink
HU University of Applied Sciences
Utrecht, The Netherlands
veerle.hobbelink@hu.nl

Matthijs Smakman
HU University of Applied Sciences
Utrecht, The Netherlands
matthijs.smakman@hu.nl

ABSTRACT

In today's world, understanding different viewpoints is key for societal cohesion and progress. Robots have the potential to provide aid in discussing tough topics like ethnicity and gender. However, comparably to humans, the appearance of a robot can trigger inherent prejudices. This study delves into the interplay between robot appearance and decision-making in ethical dilemmas. Employing a Furhat robot that can change faces in an instant, we looked at how robot appearance affects decision-making and the perception of the robot itself. Pairs of participants were invited to discuss a dilemma presented by a robot, covering sensitive topics of ethnicity or gender. The robot either adopted a first-person or third-person perspective and altered its appearance accordingly. Following the explanation, participants were encouraged to discuss their choice of action in the dilemma situation. We did not find significant influences of robot appearance or dilemma topic on perceived anthropomorphism, animacy, likeability, or intelligence of the robot, partly in line with previous research. However, several participants hearing the dilemma from a first-person perspective changed their opinion because of the robot's appearance. Future work can expand with different measures such as engagement, in order to shed light on the intricate dynamics of human-robot interaction, emphasizing the need for thoughtful consideration in designing robot appearances to promote unbiased engagement in discussions of societal significance.

CCS CONCEPTS

• **Human-centered computing → User studies**.

KEYWORDS

Inclusive robots, Gender, Ethnicity, human-robot interaction

ACM Reference Format:
Mirjam de Haas, Veerle L.N.F. Hobbelink, and Matthijs Smakman. 2024. Inclusive Dialogues: WokeBot Engaging Diversity Dilemmas. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3610978.3640650

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3640650

1 INTRODUCTION

In today's society, embracing diverse perspectives is recognized as pivotal for societal cohesion and progress. In an ideal world, everyone is treated with respect, irrespective of skin color, gender, religion, or physical abilities. However, entrenched prejudices hinder the realization of such an inclusive society. Embracing diverse perspectives fosters empathy and inclusivity, yet discussing contentious topics like identity—such as ethnicity and gender—poses challenges. Herein lies the potential of technological interventions, notably in the form of facilitative robots to play a role in creating a more empathetic and understanding society.

Using a robot as a facilitator for ethical or gender-posed dilemmas offers unique perspectives that people cannot achieve. The robot has the ability to dynamically change its physical appearance, speech patterns, and language expressions, allowing the robot to portray as a main character in ethical dilemmas. This enables participants to fully empathize with the character, fostering a deeper empathetic connection. For instance, portraying the robot as a woman in a scenario where women are mistreated can enhance people's understanding of the impact on women and possibly result in a deeper discussion about these specific topics. In our study, we used the Furhat robot to guide conversations between participants discussing a social dilemma. This robot went beyond just explaining the issue; it adjusted its appearance and language to embody the role of the central figure in the dilemma. Participants interacted either with the robot portraying the person at the heart of the issue or with another version representing the dilemma from a different perspective.

2 BACKGROUND

In exploring human perceptions of robots, it's evident that factors like gender and ethnicity play pivotal roles. Despite robots' capacity to avoid explicit gender expression, individuals tend to assign gender to robots, even in the absence of gender-stereotypical cues [3, 14]. Given the similarities between social robots and humans, this attribution significantly impacts interactions, leading to the imposition of traditional gender norms onto robots. The gender alignment between a social robot and a human serves as a crucial indicator of our attitudes toward these robots [9, 18]. Gender portrayal in social robots can positively influence user attitudes, fostering favorable views and increased anthropomorphism [9, 17]. Thus, it's evident that gender shapes people's preferences and attitudes towards robots.

Moreover, ethnicity emerges as a significant element alongside gender. Biases arising from predominantly white-designed robots, advocating for diverse representations in design [16]. Recent studies indicate that individuals attribute race to robots, akin to how they assign gender and species [5, 8]. Participants showed having 'shooter-bias' towards robots characterized as black [5], with

interventions using intermediate colors, like brown, effectively mitigating these biases [1]. Studies also indicate a trend of presumed discrimination against non-white robots, extending to assumptions about their behavior [12]. Moreover, robot design affects user experience, which makes it important to create an inclusive experience for all end users [15]. These insights prompt a shift in Human-Computer Interaction research, pushing beyond superficial appearances to foster discussions and change biases, aligned with Critical Race Theory [13].

The studies described earlier analyze how participants perceive the appearance of robots and how these appearances influence the conversation and attitudes with the robot. We explore the potential of manipulating robot appearances to facilitate concurrent interactions among participants engaged with the robot. Rather than solely examining the influence of appearances on conversations and attitudes, our emphasis lies in changing the robot's appearance as a catalyst for multi-participant engagement with the robot and the influence of this on participants' perception.

3 METHODS

3.1 Participants

We conducted an experiment with 61 adults at a science festival (Betweter Festival) in the Netherlands. Participants were assigned to a dilemma and condition (Table 1). During the science festival, participants were invited to sit with the robot to discuss an ethical dilemma in groups of two (except one group of three). This experiment was approved by the ethical committee of the HU University of Applied Sciences.

3.2 Experiment

The experiment included two dilemmas (Gender, Ethnicity) and two conditions (congruent, incongruent). We selected two dilemmas from a dilemma game developed by Universiteit Utrecht [6]. The dilemmas were selected from the themes Gender & Genderidentity and Cultural, Ethnic & Religious (further referred to as Gender and Ethnicity). We altered the dilemmas to be more general and to create a first-person scenario. In the congruent conditions, the robot explained a dilemma from a first-person perspective. The robot told the story as if a situation had occurred wherein it was discriminated against. Its appearance was altered to fit the dilemma (for instance, "I am Kione . In front of others, a white colleague tells a racist joke about me. You are one of the bystanders, what would you do?"). During the Gender dilemma, the robot had a (white) female voice, face, and name. For the Ethnicity dilemma, the robot had the voice, face, and name of a black male. In the incongruent condition for both dilemmas, the robot told the story from a third-person perspective and was given a white male voice, face, and name (for instance, "I am Daan. In front of others, a white colleague tells a racist joke about a colleague, who has just walked away. You are one of the bystanders, what would you do?").

3.3 Materials

The robot used was a Furhat robot (Furhat robotics, [2]). We manipulated the robot's genderedness and ethnicity through voice, name, and facial features (see [14]). We selected one white male face character from the Furhat library (Furhat character: Jamie), one white female face (Furhat character: Isabel), and one black male face (Furhat character: Kione). See for all robot characters Figure 2c. For the male (black and white) and female characters, we selected Dutch male and female voices from the Furhat library (male: Daan22k_HQ, female: Jasmijn22k_HQ). The robot was further equipped with a woolen hat and pin to increase humanness (Figure 1a). We programmed the robot using the online Blockly programming tool.

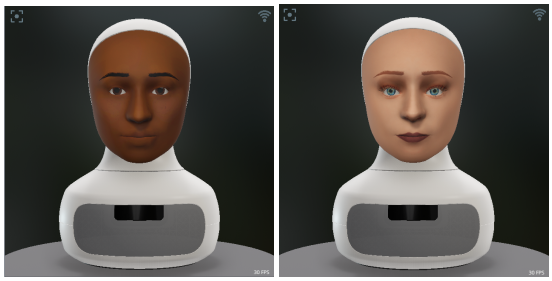
(a)

(b)

Figure 1: Set-up during the experiment. a: The Furhat robot with the white male character face and woolen hat. b: Two participants engaging in discussion after the robot explained the dilemma.

3.4 Procedure

After they were seated, participants were given headphones to reduce background noise created by the festival (see 2b for the setup of the experiment). The robot started the experiment by introducing itself and explaining the procedure. It gave a practice dilemma that was not related to appearance, selected from the dilemma game from the category "Gender & Genderidentity", about homosexuality [6]. Then, it explained the experimental dilemma (Gender or Ethnicity) and gave four options on possible ways to act in the scenario. After the explanation, participants were asked to first write down their answers and then discuss this with the other participant. After this discussion, participants were asked to fill out a questionnaire. Since participants were allowed to discuss freely and however long they liked, the interaction lasted anywhere between 7 and 15 minutes.



(a) (b)

(c)

Figure 2: a: Black male character (Furhat Kione). b: White female character (Furhat Isabel). c: White male character (Furhat Jamie)

We asked whether participants changed their dilemma answers based on the discussion, or based on the robot's appearance, and if so, how. Additionally, we selected one item from four Godspeed sub-scales (Anthropomorphism: fake-natural; Animacy: artificial-lifelike; Likeability: unfriendly-friendly; Intelligence: incompetent-competent; [4]). Participants could rate the four items on a five-point Likert scale.

4 RESULTS

The majority of participants (85%) reported that their opinions remained unchanged following discussions with other participants. Furthermore, 88% of the participants indicated that the robot's appearance did not impact their opinions. Among the participants who felt that the robot's appearance influenced or possibly influenced their opinions (N = 7), the majority (N = 6) interacted with the robot that altered its appearance to fit the dilemma (congruent).

Table 1 displays participants' average scores on the four Godspeed items (Anthropomorphism, Animacy, Likeability, Intelligence) for each condition. A two-way MANOVA was conducted, with the four Godspeed items as dependent variables, and dilemma type (gender, ethnicity) and robot appearance (congruent, incongruent) as independent variables. There was no main effect ($F(8, 112) = 0.79$), Pillai's trace = 0.11, $p = .62$, eta = .05).

4.1 Observations

Throughout the experiment, our observations of the participants revealed several distinct interaction patterns with the robot and among themselves. All participants followed instructions given by the robot and engaged in a discussion after interacting with the robot. While the majority of participants both looked at and listened to the robot, some individuals focused solely on the robot's voice due to background noise. Consequently, they forgot to visually engage with the robot during the experiment. On the other hand, some participants exclusively directed their attention toward the robot, possibly focusing on technical or visual details, unintentionally overlooking the robot's verbal communication.

In addition, during the discussions, we noted that, although most participants shared the same opinion and chose the same option provided by the robot, some participants deliberated over the alternative options presented for the dilemmas. Despite sharing the same opinion, these participant pairs engaged in debates and provided justifications for their viewpoints. Furthermore, we observed that participants did not merely compare the four options and articulate their own opinions; at times, they also discussed alternative options or scenarios to illustrate when this dilemma might or might not occur.

		Ethnicity		Gender	
		Incongruent	Congruent	Incongruent	Congruent
	N	14	14	17	16
Anthropomorphism	M (sd)	2.6 (1.1)	2.4 (1.3)	2.6 (0.9)	2.2 (1.0)
Animacy	M (sd)	2.6 (1.2)	1.6 (0.7)	2.2 (0.9)	2.06 (1.2)
Likeability	M (sd)	4.4 (0.7)	4.4 (0.9)	4.3 (0.8)	4.1 (0.9)
Intelligence	M (sd)	3.8 (0.8)	3.4 (1.2)	3.8 (1.0)	3.5 (1.2)

Table 1: Means, standard deviations, and counts for measurements (Godspeed, questionnaire items) of the two dilemmas (Ethnicity, Gender) and conditions (congruent, incongruent).

5 DISCUSSION AND CONCLUSION

This study delved into how participants perceived a robot acting as a conversation starter for ethical dilemmas. The robot presented these dilemmas either from an external viewpoint or embodied specific perspectives, like a female or black viewpoint. As expected, the robot effectively facilitated discussions among participants, leading to active engagement with the dilemmas. Notably, in five instances, the robot's appearance played a role in decision-making. Of those, four came from the group where the robot embodied the female or black viewpoint. This hints at the potential for further exploration into how robot appearances can enhance participant interactions during conversations.

We did not observe any differences in the aspects measured by the Godspeed concepts: anthropomorphism, animacy, likability, or perceived intelligence. This lack of significant differences aligns with a study examining robots of different colors [1], which explored anthropomorphism regardless of gender. However, this contrasts with prior studies indicating that gender can influence participants' perceptions of anthropomorphism [9, 17]. These results may be attributed to the robot maintaining a consistent overall appearance despite changes in facial features, resulting in no differences in the conditions. We did not investigate how people engaged

with the robot and the story due to its appearance. Therefore, exploring how a robot's embodiment influences engagement would be an interesting question to explore.

This study did not delve into the thematic discussions among participants or conduct a detailed analysis of their individual opinions and conversations. Our observations were limited to a distant perspective, only allowing for high-level insights into the interactions. It could be interesting to specifically examine the topics people discussed in the different conditions. In cases where participants noticed or altered their opinions based on the robot's appearance, they might have mentioned this during the discussion. Additionally, we did not explore whether people felt more engaged in the story due to the robot's appearance, we only asked them whether their opinion changed because of this appearance.

Throughout the experiment, we observed that the festival setting likely hindered participants from fully engaging with the robot. The background noise demanded their complete attention to properly listen to the robot through a headphone, often resulting in less focus on visually observing the robot. Consequently, many participants failed to notice the facial changes; only five participants indicated that the robot's appearance influenced their opinion. This highlights the need for future studies to explore methods that specifically attract attention to the robot's face, ensuring that its appearance is noticed. This is aimed at exploring in more detail whether people's biases influence their opinions or if they can recognize being more open to alternative viewpoints [13].

Conducting research in this constrained setting brought forth further challenges. Due to time constraints during the experiment, we opted to select a limited number of Godspeed [4] items. However, to ensure more robust measurements, future research should consider including all items. Because participants' time was limited, and privacy was considered important, we refrained from collecting personal data, such as age, gender, or race. Nonetheless, it is acknowledged that personal differences may potentially influence viewpoints and outcomes. While we observed that many participants arrived at the experiment together and therefore probably already knew each other, this information should be collected in future work. Alternatively, existing relationships between participants could be avoided altogether. Therefore, future research should incorporate the collection of such participant information for a more comprehensive understanding of potential influences on the study's results.

In addressing the ethical considerations surrounding robots with gender or race expressions, it is crucial to take into account diverse perspectives on robot design. The avoidance of perpetuating societal biases through technological implementations, such as robots, and the promotion of inclusivity and fluid identity can only be achieved by careful consideration of robot design and ethical practices [10]. Furthermore, ensuring a diverse participant group is essential to account for the varied experiences individuals have in their interactions with these robots [11]. It's worth noting that in both this research and broader societal contexts, there is a tendency to consider white males as the default group, while other race and gender groups are often labeled as 'others', thereby activating power dynamics between these groups [7]. Future work should aim to enhance inclusivity by increasing the representation of characters of different genders and races, thereby fostering a more inclusive environment.

Our study revealed participants' willingness to engage in discussing ethical dilemmas when facilitated by a robot. This underscores the potential for robots to effectively stimulate discussions on complex ethical issues, fostering a deeper sense of connection and engagement among participants. However, our findings highlight that more research should be done on the manner in which the robot tells a story to explore the impact of employing congruent facial features with the main character in ethical dilemmas.

REFERENCES

[1] Arifah Addison, Christoph Bartneck, and Kumar Yogeeswaran. 2019. Robots can be more than black and white: examining racial bias towards robots. In *Proceedings of the 2019 AAAI/ACM Conf. on AI, Ethics, and Society*. 493–498.

[2] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems: COST 2102 Int'l Training School, Dresden, Germany, February 21-26, 2011, Revised Selected Papers*. Springer, 114–130. https://doi.org/10.1007/978-3-642-34584-5_9

[3] Gaye Aşkın, İmge Saltık, Tuğçe Elver Boz, and Burcu A Urgen. 2023. Gendered actions with a genderless robot: Gender attribution to humanoid robots in action. *International Journal of Social Robotics* (2023), 1–17.

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1 (2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[5] Christoph Bartneck, Kumar Yogeeswaran, Qi Min Ser, Graeme Woodward, Robert Sparrow, Siheng Wang, and Friederike Eyssel. 2018. Robots And Racism. In *Proceedings of the 2018 ACM/IEEE Int'l Conf. on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. ACM, New York, NY, USA, 196–204. https://doi.org/10.1145/3171221.3171260

[6] Caspar Bok, Linas Cepinskas, Frederike Groothoff, Kim van Hees, Hai Le, Merel Nissen, Zoulikha El Ouakili-Saadi, Lina Senen, Gary Sheikkariem, and Hylko Tas. [n. d.]. Het Dilemmaspel. Equality, Diversity & Inclusion (EDI) Office. https://www.uu.nl/organisatie/equality-diversity-inclusion/edi-voor-jou/activiteiten/dilemmaspel

[7] Patricia Hill Collins. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge, New York, NY.

[8] Friederike Eyssel and Dieta Kuchenbrandt. 2012. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology* 51, 4 (2012), 724–731.

[9] Friederike Eyssel, Dieta Kuchenbrandt, Simon Bobinger, Laura De Ruiter, and Frank Hegel. 2012. 'If you sound like me, you must be more human' on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Proceedings of the seventh annual ACM/IEEE int'l conf. on Human-Robot Interaction*. 125–126.

[10] Donna J. Haraway. 1985. *A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century*. Routledge.

[11] Sandra Harding. 1986. *The Science Question in Feminism*. Cornell University Press, Ithaca, NY.

[12] Jessica K. Barfield. 2021. Discrimination and Stereotypical Responses to Robots as a Function of Robot Colorization. In *Adjunct Proceedings of the 29th ACM Conf. on User Modeling, Adaptation and Personalization* (Utrecht, Netherlands) *(UMAP '21)*. ACM, New York, NY, USA, 109–114. https://doi.org/10.1145/3450614.3463411

[13] Ihudiya Finda Ogbonnaya-Ogburu, Angela D. R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. *Proceedings of the 2020 CHI conf. on Human Factors in Computing Systems* (2020). https://api.semanticscholar.org/CorpusID:218483077

[14] Giulia Perugia and Dominika Lisy. 2023. Robot's gendering trouble: a scoping review of gendering humanoid robots and its effects on HRI. *International Journal of Social Robotics* (2023), 1–29.

[15] Yolanda A Rankin and Jakita O Thomas. 2019. Straighten up and fly right: Rethinking intersectionality in HCI research. *Interactions* 26, 6 (2019), 64–68. https://doi.org/10.1145/3363033

[16] Robert Sparrow. 2020. Robotics Has a Race Problem. *Science, Technology, & Human Values* 45, 3 (2020), 538–560. https://doi.org/10.1177/0162243919862862

[17] Benedict Tay, Younbo Jung, and Taezoon Park. 2014. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior* 38 (2014), 75–84.

[18] Ellis Van den Hende and Ruth Mugge. 2012. The role of gender congruity for anthropomorphized product perception. *ACR North American Advances* (2012).