"I'm Not a Girl": The Myth of Gender Neutrality and the Nonbinary Social Robot

Ruth Mueller[a], Aoi Yasuda[a], and Ryan Blake Jackson[a]

Abstract— We designed an experiment that refers to a robot with singular 'they', as opposed to gendered third person pronouns. This choice may interact with complex human gender norms in unanticipated ways. We pose several questions where the input of queer studies and gender studies scholars would be helpful to HRI research.

I. INTRODUCTION

Philosopher Mark Coeckelbergh has argued that "our talk about and to robots is not a mere representation of an objective robotic or social-interactive reality, but rather interprets and co-shapes" human-robot relations and human mental models of robots [1]. For example, a subject-object ontological distinction is reflected in whether we refer to a robot with the impersonal pronoun 'it' versus personal pronouns like 'they', 'she', or 'he'. However, this choice not only indicates a distinction in one's mental model of the robot (as object vs. subject), but also influences and shapes that mental model and any interaction with the robot. Based on the hypothesis that linguistic pre-construction of the human-robot relation will influence and co-construct the actual relation, Coeckelbergh proposes an experiment that manipulates how the experimenter refers to the robot (both pronouns and name) prior to an interaction to see whether this way of pre-defining the interaction will measurably impact the interaction itself.

However, many interactions with social robots may not include the opportunity for a roboticist to supply a linguistic pre-definition of the robot beforehand. For example, recent research has placed robots in public spaces like museums [2] or shopping malls [3]. Thus, we are interested in how a robot might linguistically (pre-)define its own identity or ontological category by introducing itself in the beginning of an interaction with a human. We parameterize possible robot introductions along variations to the conventionalized human script, "Hi, I'm [NAME]. My pronouns are [PRONOUNS]." Following this script not only allows the robot to explicitly frame itself early in an interaction, but also may prompt people to introduce themselves similarly, which would give the robot information about the human's name and gender[1] for use in future referring expressions. Based on Coeckelbergh's arguments, we expect that a robot specifying personal rather than impersonal pronouns may prompt a more anthropomorphic mental model of the robot and lead people to view the robot as having greater moral status.

To exhaustively explore all variations of our introduction script would require 5 conditions corresponding to 'he', 'she', 'they', and 'it' pronouns and a control condition with no pronouns specified. While he/him/his and she/her/hers are commonly used personal pronouns, we excluded them in an attempt to 1) focus more on the subject-object distinction than on gendered differences, and 2) reduce the number of conditions to make participant recruitment more feasible. Explicitly gendering the robot, while interesting in its own right, could be a confounding variable here; some evidence suggests strong effects of robot gender presentation in, for example, the context of robot noncompliance [4]. While we avoided explicitly gendering our robot and attempted to minimize implicit gender cues, the powerful human tendency to attribute gender to social robots is well documented, and it may not be possible to entirely avoid ascriptions of gender to social robots[5], [6]. Indeed, we will record any participant gendering of our robot, despite using they/them/theirs as the only personal pronoun condition.

II. PROMPTING A LGBTQ+ MENTAL MODEL

The choice to use they/them/theirs as our only personal pronouns adds additional nuance to the interpretation of results because of complex interactions with, and media representations of, nonbinary identities. There is a long history of singular they being used generically in English [7]. However, more recent changes in socio-pragmatic conventions have seen singular 'they' increasingly used for specific referents and also increasingly associated with queer [2] culture and identities [7], [9]. Evidence shows that specific singular 'they' is a viable option for both a gender-neutral and nonbinary pronoun in English [7], [9]. Our concern is that we do not know and cannot control whether it will prompt our robot to be perceived as having a nonbinary humanlike gender, an unknown or unspecified gender, a robot-specific gender [10], or no gender (while remaining a social actor). Which of these mental models the robot's language prompts will influence subsequent perceptions of robot speech and behavior, and will shape the interaction in ways that may not be obvious.

Adding further complexity is the portrayal of nonbinary characters in popular media. There seems to be a pop-cultural awareness of a possible trope in which nonbinary characters are commonly portrayed as nonhuman, especially as robots [11], [12], [13], [14]. Examples include the AI Janet from the TV sitcom *The Good Place* who consistently corrects other's misgendering based on their feminine presentation (stating, e.g., "I'm not a girl") [11], and the bigender robot BMO from the animated series *Adventure Time* [15]. There is, to our knowledge, little formal academic research investigating this trope or how it affects perceptions of robots. Familiarity with these portrayals of nonbinary identities could influence people's interpretations of our robot in complex and unknown ways. We might hypothesize, for example, these representations would prime our participants to think of robots as having nonbinary gender, thus decreasing their likelihood of viewing our robot as having no gender or unknown gender.

If the robot is indeed perceived as having nonbinary gender, and regardless of whether that perception is tied to representations in media, the unfortunate pervasiveness of dehumanizing rhetoric directed towards nonbinary people as a means of discriminatory gender policing (see [16]) might also impact our results. In participants whose attitudes have been shaped by (and reflected in) such language, those attitudes might counteract any humanizing effects or tendency towards moral status from framing the robot in a personal (subject) rather than impersonal (object) way.

III. CONCLUSION

We have discussed an experiment that may unintentionally intersect with gender nonbinariness in ways that, though not central to our research questions, are important to consider when designing robots in the complex ecosystem of human gender norms. We emphasize that an absence or inconsistency of gender cues does not necessarily yield a "gender neutral" robot. Rather, we should study the possibility of robots being attributed nonbinary and nonconforming genders, and leverage the expertise of scholars in queer studies and gender studies.

[a]Department of Computer Science, Harvey Mudd College, Claremont, CA, USA {rmueller, ayasuda, bjackson}@hmc.edu

[1]Useful for languages with gendered third person pronouns like English.

[2]We use the term 'queer' in the academic sense that emphasizes non-heteronormativity and contests oppressive social constructions of sexual orientation and gender [8] in opposition to its historic use as an epithet.

REFERENCES

[1] M. Coeckelbergh, "You, robot: on the linguistic construction of artificial others," *AI & society*, vol. 26, no. 1, pp. 61–69, 2011.

[2] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive robotics: The influence of robot gender on human behavior," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 2563–2568.

[3] T. Nomura, T. Uratani, T. Kanda, K. Matsumoto, H. Kidokoro, Y. Suehiro, and S. Yamada, "Why do children abuse robots?" in *Proceedings of HRI Extended Abstracts*, ser. HRI'15 Extended Abstracts, 2015.

[4] R. B. Jackson, T. Williams, and N. Smith, "Exploring the role of gender in perceptions of robotic noncompliance," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 559–567.

[5] C. Nass, Y. Moon, and N. Green, "Are machines gender neutral? gender-stereotypic responses to computers with voices," *Journal of applied social psychology*, vol. 27, no. 10, pp. 864–876, 1997.

[6] K. Seaborn and P. Pennefather, "Neither" hear" nor" their": Interrogating gender neutrality in robots," *arXiv preprint arXiv:2205.00183*, 2022.

[7] A. C. Saguy and J. A. Williams, "A little word that means a lot: A reassessment of singular they in a new era of gender politics," *Gender & Society*, vol. 36, no. 1, pp. 5–31, 2022.

[8] D. L. Zosky and R. Alberts, "What's in a name? exploring use of the word queer as a term of identification within the college-aged lgbt community," *Journal of Human Behavior in the Social Environment*, vol. 26, no. 7-8, pp. 597–607, 2016.

[9] E. D. Bradley, J. Salkind, A. Moore, and S. Teitsort, "Singular'they'and novel pronouns: gender-neutral, nonbinary, or both?" *Proceedings of the Linguistic Society of America*, vol. 4, no. 1, pp. 36–1, 2019.

[10] K. Seaborn, P. Pennefather, and H. Kotani, "Exploring gender-expansive categorization options for robots," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–6.

[11] C. Prevas, "We need more non-binary characters who aren't aliens, robots, or monsters," https://electricliterature.com/we-need-more-non-binary-characters-who-arent-aliens-robots-or-monsters/.

[12] S. Taskas, "Non-binary characters in games - player option, robot, alien, or mystery?" https://medium.com/@staskas/non-binary-characters-in-games-player-option-robot-alien-or-mystery-c7662d32711b.

[13] T. Ågotnes, W. Van Der Hoek, J. A. Rodríguez-Aguilar, C. Sierra, and M. Wooldridge, "On the logic of normative systems," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[14] "Non-human non-binary," https://tvtropes.org/pmwiki/pmwiki.php/Main/NonHumanNonBinary, accessed: June 23rd 2022.

[15] C. McNabb, *Nonbinary gender identities: History, culture, resources*. Rowman & Littlefield, 2017.

[16] I. Hatch and A. Randolph, "Digital shame, dehumanization, and dismissal: An analysis of gender policing," *Pursue: Undergraduate Research Journal*, vol. 4, no. 1, p. 4, 2021.