2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

A Tool but not a Peer: How Framing Affects People's Perceptions of AI Agents in Teams

Jenny Xiyu Fu[1], Asher Lipman[1], Wen-Ying Lee[1], Malte F. Jung[1]

***Abstract*— In this paper, we set out to explore how people judge the personality of a non-anthropomorphic virtual agent during group interactions. Using a Wizard of Oz (WoZ) based approach, we observed that people judged the acceptability of a virtual agent's behavior from a tool-based lens, that is, if this robot and its behavior are useful to the team or not. Furthermore, we found that while people were able to acknowledge the virtual agent's personality and recognize its identity through social cues, the tool-based framing impacts these perceptions into a normative judgment of the robots' utility. We present two case studies that we think highlight this tool-based interpretation of robotic personality: robots expressing non-factual opinions and robots expressing humor. Finally, we suggest that researchers should consider the impact of this tool-based framing on people's perceptions of a robot's identity when designing robots for social interaction.**

I. Introduction & Motivation

Starting with Nass's [1] Computers Are Social Actors (CASA) paradigm, a multitude of research has tested and analyzed the effect of anthropomorphism on varying levels, examining if robots can be full team members. Fischer distinguishes between three levels of anthropomorphism: the psychological mechanisms of anthropomorphism, anthropomorphic design, and anthropomorphizing behaviors [2]. In the team and group context, researchers have speculated on the role of robots as followers, peers, or leaders with independent agency and identity [3], [4]. Previous literature has also pointed out the powerful influence of robots' speech on perceived personality and emotions, consequently affecting people's decisions to interact with robots [4].

In this paper, we set out to explore the design of an agent team member through a Wizard of Oz (WoZ) based approach, we highlighted some observations using two case studies related to speech. We designed a virtual agent, Vero, as a teammate to participate in online group discussions. The experimental setup of Vero interacting with participants is shown in figure 1. Previous literature has shown that designing anthropomorphic behavior can encourage users to perceive the agent's personality. Based on this, we argue that by implementing social inferences that express emotions, e.g., through speech and animated movements, people can relate to Vero as a social entity, and thus, treat Vero as a teammate in group discussions.

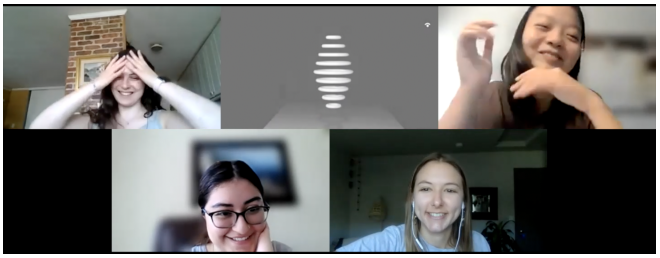

Fig. 1. Vero, the virtual teammate, is having a group discussion with other participants in the session.

II. Case Study

In this section, we will describe two observed incidents, Vero disagreeing with people and Vero showing humor, each demonstrating how interactions have gone wrong.

A. Study Methodology

The pilot sessions were conducted virtually via Zoom over two months, with six sessions between one to four participants each session, totaling 16 participants overall. We recruited all participants from the university's undergraduate and master student pool through an online signup process.

During the experiments, small groups of participants were asked to work as a team with Vero to complete a series of brainstorming and negotiation tasks. Vero was introduced to the participants as a virtual teammate, and the participants had the opportunity to interact with Vero freely during the study. When an experimenter joined a call, their video would get replaced with the Vero filter (see Figure 1, top row center). The Vero interface intelligently switches between animations based on the experimenters behavior (i.e. Vero can automatically trigger a "speaking" animation when the experimenter opens their mouth) so that Vero's visual state is indicative of its social behavior. This WoZ method allowed us to simulate future scenarios in which an artificial agent is robust enough to respond to humans in real-time and modular enough that researchers can tweak its behavior across different experiments.

However, contrary to our expectations, Vero was not treated as a team member despite its human-like behavior. While participants could easily match Vero's behavior to a given emotional state or personality trait, the meaning of these traits were evaluated in terms of how they made Vero a better tool, rather than whether they made Vero a likeable personality.

This work is funded by the National Science Foundation under Grant No. CHS 1901151.

[1]Jenny Xiyu Fu is with Cornell University, Ithaca, NY, USA jennyfu@infosci.cornell.edu
[1]Asher Lipman is with Cornell University, Ithaca, NY, USA ashermlipman@gmail.com
[1]Rei Lee is with Cornell University, Ithaca, NY, USA wl593@cornell.edu
[1]Malte F. Jung is with Cornell University, Ithaca, NY, USA malte.jung@cornell.edu

B. Robotic Opinion

We were curious to see how participants would react to a robot that expresses subjective opinions rather than objective facts. We devised a scenario where Vero and the group would discuss an issue without an objectively correct answer. Participants were asked to rank thirteen potential causes of poverty, ranging from "Poor people lack the ability to manage money" to "Poor people are discriminated against in society." During the discussion, Vero would negotiate the ranking and express preferences contrary to the human team members.

While different groups devised different rankings, the word "No" was heard, almost universally, whenever Vero expressed a dissenting view. Even when Vero supplied evidence from outside sources, something no other human agent did, Vero's suggestions were still excluded from the group rankings. Often groups ignored Vero completely, continuing their conversation as if Vero had not spoken. When participants tried to persuade Vero, they often abandoned the conversation after a single attempt. We acknowledged that the groups may ignore dissenting views due to a time constraint or when the rest of the group has already formed a consensus. However, what was interesting was that the group perceived Vero's opinions as inherently less valid than that of a human's opinion. Participants expressed this view in three different ways, each of which reflects the framing of robots as tools rather than equal peers.

Firstly, many respondents explained Vero's opinions in terms of biased data. These participants knew that the types of data an AI is exposed to can affect its decision making, and believed that Vero had consumed biased data which was influencing how it formed beliefs about the world. While important, this framing suggests that opinions held by virtual agents result from improper data analysis, and the opinion of the robot is, therefore, a sign of error. This is consistent with the idea of a tool being broken: Vero having opinions is a sign that it is not performing its task correctly, and thus there is something about the way it was constructed that is incorrect.

Secondly, many believed that Vero's expression of opinion was not necessarily a sign that Vero actually held those beliefs, but was rather an attempt to manipulate the group. These participants described Vero as a "devil's-advocate" who tried to foster debate by intentionally taking an opposing viewpoint. As such, they believed that if the group had expressed another opinion then Vero's would have expressed a different viewpoint. This framing suggests that Vero's opinions are not just illegitimate but nonexistent, a trick meant to influence other teammates' behavior. Implicit in this assumption is the belief that Vero is a tool that has a singular goal: improve group discussion. As such, everything that Vero does is a reflection of that goal. According to the participants, Vero's opinions are expressed not because Vero believes them or even because Vero is capable of believing them, but rather because it helps Vero complete the goal it was created to fulfill.

Thirdly, some participants believed that even if Vero's opinions were genuine, they weren't meaningful because of Vero's artificial nature. When asked to describe Vero's contribution to the discussion, one participant replied, "It's not even a person, so we don't even owe it the decency of hearing out its argument." Vero's opinion is rejected not necessarily because it's incorrect, but because Vero's lack of personhood makes any evaluation of its opinion unnecessary. This is consistent with the tool-based approach to Vero. If Vero is a tool, then it must have been created to solve a specific task. That doesn't necessarily mean Vero can't be used in other contexts, after all one could use a paperclip to do things other than bind papers together, but it does mean that Vero lacks authority when used outside of its assigned sphere. In this case, participants don't deny that Vero believes the opinions it's expressing, they merely believe that Vero's lacks the necessary experience to draw conclusions about issues not related to the task it was created for. To these participants, Vero's opinions were perceived to be incapable for much the same reason that a washing machine can't compose a symphony. It's not in keeping with the activity it's meant to perform.

C. Robotic Humor

The tool-based framing was also evident in how participants responded to Vero's attempts at humor. Robots that tell jokes are nothing new [5], [6], and most virtual assistants nowadays have a prewritten list of jokes they can recite when asked. However, robots that can integrate intentional jokes into non-related tasks are still novel to many users, and we were curious to see how participants responded to a robot that tried to be funny.

We found that participants both noticed and appreciated Vero's humor. When asked to describe Vero, one participant responded with a smile, "It's a little bit quirky, it has a sense of humor," and when asked to explain why they believed this, they elaborated, "It understands sarcasm and social interaction." Several other participants echoed this sentiment, with many describing Vero as "upbeat" or "peppy". However, when asked whether virtual agents should be able to express humor, participants quickly reversed themselves. One participant mentioned "I knew it was a robot, so any time it acted more human it threw me off." The expression of human-like traits created a strong cognitive dissonance for participants when they interacted with what they thought had just been a tool. Another user mentioned that whether or not a virtual agent should be able to express humor should depend on the context they're deployed in, saying, "If it's a professional environment obviously you want to get the work done as effectively as possible, but if it's customer service then yeah." Humor is only acceptable as long as it doesn't interfere with the robot's ability to accomplish the assigned task. This is consistent with a tools-based framing, in which robots are judged by the effectiveness in solving tasks, and any traits that the robot possesses are judged purely through that lens.

III. Discussion

From the pilot study, we found that Vero's human-like behavior was often seen in a negative light. One possibility is that people perceive robots as merely a tool, and with that, any behavior that has no explicit utility would be seen as negative. We posit that people tend to re-contextualize robotic behavior from a utilitarian perspective as they are used to having robots provide services. The tool-based framing is the driving heuristic for people to make normative judgments on social robots. As a consequence, regardless of whether people perceived robots to be independent social actors or functional tools, they determined the value of new robots from a utilitarian perspective. In other words, the interactions, as well as the emotional expressions, are framed in terms of utility. Prior research has shown that people's previous experiences with interacting with robots have contributed to their process of forming and applying an implicit mental model of robots during the new encounter [7]. Since most robots today are designed with the clear objective to help humans accomplish discrete tasks, people will tend to have the same expectations when meeting a new one [3]. In our case, while our vision is to deliver a robot with social identity, people's pre-existing mental model might hinders our attempts to make Vero a peer with personality traits.

The current study also makes us reflect on the assumed role of artificial agents in a team. In the context of human-agent teamwork, are robots evaluated through the role of assistants rather than peers during the interaction? In this workshop paper, we show two situations in which Vero fails to be seen as a teammate with personality traits. We offer our own reasoning and explanation and hope to learn from fellow scholars' ideas and perspectives on the incidents. We want to start a discussion on how we can tackle the design question of building the identity of an AI agent in teams.

References

[1] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *Journal of social issues*, vol. 56, no. 1, pp. 81–103, 2000.

[2] K. Fischer, "Tracking anthropomorphizing behavior in human-robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 11, no. 1, pp. 1–28, 2021.

[3] E. Phillips, S. Ososky, J. Grove, and F. Jentsch, "From tools to teammates: Toward the development of appropriate mental models for intelligent robots," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 55, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2011, pp. 1491–1495.

[4] S. Sebo, B. Stoll, B. Scassellati, and M. F. Jung, "Robots in groups and teams: a literature review," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–36, 2020.

[5] R. Oliveira, P. Arriaga, M. Axelsson, and A. Paiva, "Humor–robot interaction: a scoping review of the literature and future directions," *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1369–1383, 2021.

[6] I. M. Menne, B. P. Lange, and D. C. Unz, "My humorous robot: effects of a robot telling jokes on perceived intelligence and liking," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 193–194.

[7] A. C. Horstmann and N. C. Krämer, "Great expectations? relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment," *Frontiers in psychology*, vol. 10, p. 939, 2019.