

# Making sense of robots together: Examining group interactions with Pepper

Salla Jarske  
Tampere University  
Tampere, Finland  
salla.jarske@tuni.fi

Sanna Raudaskoski  
Tampere University  
Tampere, Finland  
sanna.raudaskoski@tuni.fi

Kirsikka Kaipainen  
Tampere University  
Tampere, Finland  
kirsikka.kaipainen@tuni.fi

Kaisa Väänänen  
Tampere University  
Tampere, Finland  
kaisa.vaananen@tuni.fi

**Abstract**— Human-Robot Interaction (HRI) is traditionally addressed as interaction of one human user with a single robot. In this workshop contribution we reflect upon HRI in group interactions. Our approach follows ethnomethodological conversation analysis, and we present excerpts from a video-recorded study of group interaction with Pepper. We draw attention to how the observations and interpretations of robots’ behaviors are situationally and socially given a meaning by the human participants acting together, and that this inevitably has implications for both design and theory of HRI. We conclude that a robot is configured as social through human interactions, as robots can provide interpretable and observable materials that people ordinarily use in social interactions with each other.

**Keywords**—socially interactive robots, group interaction, HRI scene, ethnomethodology

## I. INTRODUCTION

The field of human-robot interaction (HRI) has predominantly focused on understanding and designing interactions between one user and one robot at a time. While understanding such interactions forms an important basis to the field, it is necessary to understand the variety of situational interactions which often involve multiple people simultaneously.

In this workshop paper we illustrate HRI in group situations with our recently gathered video data on university students interacting with the Pepper [1] robot in groups. We introduce theoretical notions from ethnomethodology to help understand what happens as people together make sense of robot’s behaviors. We argue that to understand group HRI, we should look at HRI as an unfolding activity between human participants, to which the robot is incorporated in a meaningful way.

## II. THEORETICAL BACKGROUND

According to ethnomethodology, social interaction can be understood to have a “public” nature. All the participants interacting are supposed to be able to grasp the meanings in the same way, based on the material that is, available for all members of the interaction [2, 3]. In the same vein, human participants are accountable for their interactions to other people. Even though a single individual was interacting with the robot while others are silent (trying to ask a question, etc), the interactional turns provided by one person become accountable material to the other members of the scene as well, and the meaning associated to these interactions are supposed to be understood in the same way by the participants. In other words, when the person is speaking to a robot (for instance, to ask “What is your name?”) the assumptions of what such a question implies are interpreted the same way by all the participants.

We shall use the term “HRI scene” to refer to the physical space of interaction that consists of all human participants, the robot, and any other material that becomes relevant to the interaction. The point of this term is to highlight that HRI here is not understood only as the specific interactions between a robot and a human, but it is understood broadly to consider the scene in its entirety. For instance, the robot’s gaze direction is observable, shared material for all the participants in the group setting. Interpretations made by individuals are confirmed by others in the scene and negotiated further through the interaction.

## III. THE STUDY

The data in question is recently gathered video data of university students interacting with the Pepper robot. Participants were invited to participate in a video-recorded session during which they could freely interact with our robots. We had two rooms, one with Pepper and one with Nao and Cozmo. The data excerpts in this paper are all taken from Pepper’s room. In the Pepper room, the third author was overseeing technical support for Pepper and interacting with the students, and the first author was moving between the rooms observing the interactions, but not interacting much with the students or the robots.

These were the student’s first encounters with these robots. The participants were familiar with each other and joined as friend groups. They were instructed to choose freely which robot they want to interact with, and they were allowed to switch rooms any time.

## IV. EXCERPTS

The video data excerpts introduced in this paper were analyzed following ethnomethodological conversation analysis [4]. Transcription applied the Jeffersonian transcription system [5]. For an explanation of the symbols used in the transcription, see a list of transcription symbols at the end of this paper.

### A. “Raise your left hand”

Here we illustrate how people together produce understanding of what is going on in the HRI scene. In this situation, the participants have been trying to get Pepper to touch its head with its arm. One participant asked, “can you touch your head?” after which Pepper uttered “of course” slightly raising its right arm in a gesture, but not actually touching its head. The group laughed at the scene and made the interpretation that the robot needs to be told to touch its head directly, and one participant attempted the command “touch your head” multiple times. A while later one participant attempts to get a response from Pepper by stating “raise your arm”. As nothing happens, the participants suggest

a specific arm must be referenced. The group has been unsuccessful in getting responses from Pepper for a while. Then, the interaction proceeds as follows:

01 C: Rai:se your left hand  
 02 P: da ɹdap (.) ↑Thank you (.) da ɹdap  
 ((Pepper's right arm raises slightly))  
 03 D: Thank you:  
 04 A: That's a right one [but okay]  
 05 C: [that command but [But this] is right  
 06 ((joint laughter for 3 seconds))  
 07 P: da ɹdap (.) da ɹdap  
 08 C: Rai:se your right hand  
 09 P: da ɹdap (1.5) da ɹdap  
 10 B: No  
 11 D: It's tra:ined to understand-  
 12 B: She does understand  
 [that command but ]  
 13 A: **[Think she's overwhelmed] (.) ha ha**  
 14 B: =maybe not right now yea:  
 15 E: **((points at Pepper's arms))**  
**seems a bit stressed out**  
**((imitates Pepper's shaky hands))**  
 16 ((joint laughter for 3 seconds))  
 17 C: How many fingers do you have  
 18 ((participants laugh quietly))  
 19 A: It's like [( )]  
 20 E: [Do you have fingers]  
 21 P: da ɹdap (.) da ɹdap  
 22 B: Awwh:  
 23 D: Why is so difficult-  
 24 E: **((leans toward Pepper))**  
**Is is ↑everything oka:y**  
 25 ((joint laughter for 2 seconds))  
 26 B: She understood that (.) but she doesn't  
 wanna answer

Excerpt 1: The participants are trying to get Pepper to raise its arm.

In the excerpt, C utters “raise your left hand” to get Pepper (P) to follow the command (line 01). Pepper makes a beep sound and then utters “thank you” in a cheery voice (line 02), raising its right arm slightly in an animated gesture. Participant A makes the comment “that’s the right arm but okay”, pointing out that the robot raised the wrong arm. Note that it is not in our goals to make inferences about what the participants genuinely believe about the robot’s gestures (i.e., whether the robot misheard the human or not). What we are interested in are the accounts that people provide for these events.

On line 15, E makes the comment “seems a bit stressed out”, while imitating the apparently slightly shaky and awkward hand positions that Pepper has in that moment. (Relatedly, [6] observed that people can find narratives for robot’s embodied aliveness, for example seeing the robot as “tired”.) E’s account is soon followed by another similar one, when E leans slightly toward Pepper and interjects “is everything okay” in a slightly concerned style of voice. Laughter follows, as the utterance is treated humorously by the participants. These accounts provided by E on lines 15 and 24, and by A on line 13 are interesting because they deviate from the interpretations provided by the participants that focused more on the logic behind how Pepper can understand commands.

In this excerpt, A’s and E’s accounts appear during a phase in the interaction where the robot has not been responding to people’s questions and commands correctly, or not at all, which can indicate that these accounts were delivered to explain the robot’s unresponsiveness. However, these accounts in this case do not contribute to understanding why the robot does not actually respond in a factual sense, but it

allows for the participants to create a sense, in this case also with a humorous tone, to how things look like. As the robot brings its own material features to the scene, people can *find* them together and practically account for them based on what any observer can see.

### B. Who Participates in HRI?

Researchers who participate in robot encounters also shape how the robot’s sociality is constituted [e.g., 7]. In the following excerpt, we illustrate how researchers in the scene also participate in giving accounts about the robot. The following HRI scene consists of three participants and one researcher. The previous student group had just left the room, one of the current participants (C) had already been interacting with Pepper (P) with the previous group for a while, and the two new participants (A and B) entered the room for the first time to see Pepper. The researcher (K) moves away from Pepper and stands near the wall observing the situation.

01 P: Select the service you want me to tell more  
 about  
 ((participants gaze at Pepper's touch  
 screen for 10 seconds))  
 00 B: ((presses a button on Pepper's screen))  
 02 P: You can take a look at the visit tampere  
 website on the big touch screen behind me  
 (2.0)  
 ((all three participants take steps to look  
 at the back side of the Pepper robot))  
 03 C: **[Are] you sure about that**  
 04 A: [Oh ]  
 ((the participants step in front of  
 Pepper))  
 05 K: **It's a bit confused**  
 06 A: Ha ha ha  
 07 B: Yea:h

Excerpt 2: Participants are trying out Pepper’s touch screen.

For context, the application that the participants were using was designed to be used at a service center, and the position of the robot would be in front of a touch screen that customers can also use. In the current scene, however, no such touch screen exists. As Pepper makes a reference to the touch screen “on the big touch screen behind me” (line 02), the three participants in front of Pepper immediately start moving behind Pepper, C from the right hand side, and A and B from the left hand side of Pepper. As C has a direct view of Pepper’s back, C states “are you sure about that” immediately returning in front of Pepper. A and B have also moved to their original positions after realizing there is no touch screen behind Pepper.

The statement “are you sure about that” delivered by C on line 03 is accounted by K on line 05 “It’s a bit confused”. In here, as with the previous excerpt, the account is delivered after something has been unclear or problematic to the participants about the robot. Another way to manage such an error as this could have been to explain the context accurately, but the account “it’s a bit confused” served as an explanation that was enough for the participants to move on with the interaction.

This interaction is somewhat different from the previous one, because participant K was not directly engaging in interaction with the robot the same way as A, B and C were. K was standing further away and due to her role overseeing the participants’ interaction and the robot’s technical functioning. However, this excerpt illustrates that participants that are not directly engaging with the robot are still practically

part of the HRI scene. K participated in giving a social account of the robot by stating that the robot is “a bit confused”, and this was confirmed by participant A’s utterance “yeah”.

The point of illustrating this with an excerpt was to show that the boundaries of what is HRI are somewhat elusive. Should we understand HRI as the exchange between the robot and a human? Is a certain proximity required? Does the interaction have to be directly toward the robot? These are interesting questions, because it appears that the HRI in this case could also be considered as the interactions between human participants orienting toward a robot in the HRI scene. This definition would also incorporate K as part of the HRI, because even though she is not directly engaging with the robot, she is engaging in the unfolding scene that is shared by the human participants. In this regard, human-robot interaction in groups could be viewed as “HRI scenes” that consist of all the participants, the robot, and other material that becomes relevant.

### C. Lessons Learned for Interaction Design

In our research project, we are interested youth civic contexts, and often our concepts relate to engaging youth in civic topics or activities with the use of social robot technology. Sometimes these concepts involve the idea that youth would approach or engage with the robot in groups, and that they would ideally gain important knowledge from the robot. However, how such encounters would take place in the real world is something we are just beginning to study. To make more appropriate designs, we need to study how people orient towards certain interactional attempts by the robot in real situations. Moreover, it should be recognized that much of what ends up happening in these situations depends on the human interactions that people engage in to construct and maintain certain understandings about the robot’s behaviors and the overall HRI scene. It would be helpful if designers could identify aspects they have control over, and the aspects that are out of their control, so to speak.

To illustrate this further, we shall use an example from one of our earliest design projects. As part of a course project in user experience design, we created an exploratory prototype application about plastic recycling for the Pepper robot. The robot’s purpose was to talk about plastic recycling and engage people in a dialogue to persuade towards better environmental behaviors. The dialogue was originally designed for one-to-one interaction with the robot, and we did not consider group interaction much during the design phases. In our recording session, the university students briefly tried the application.

The script created for the interaction followed mostly the tablet interface and certain possible responses that people can provide to move onwards, which means that as designers we were working on the assumption that the interaction would go, step by step, a certain way. However, as it turns out, people can take interaction in directions that were not predicted by the designers.

```
01 P: Microplastics are very small particles of
      plastics that can end up in sea food (.)
      do you think this is a problem (.) da ɹdap
02 (2.0)
03 B: ɹye:s
04 P: da ɹdap (.) I agree with you:
05 B: ɹaha
06 P: did you know (.) fabric made of ɹsynthetic
      material can release microplastics when
      it's washed (.) do you own synthetic
      clothes made of lycra or spandex (.)
```

```
da ɹdap
07 B: ((crosses arms and leans toward Pepper))
      is that a trick question
      (2.0)
08 C: are you made of plastic
09 P: Great job
10 B: Mm: ha ha
11 P: Now that you know more (.) tell me (.)
      which concerns you most (.) about
      microplastics
12 B: ((leans toward Pepper))
      are you made of plastic
13 P: da ɹdap da ɹdap
14 B: ((points at Pepper's screen))
      when it's in this mode it's
15 C: yeah
```

Excerpt 3: Participants orienting to Pepper’s question.

Participant B that was “gazed” by the robot did not respond to the question “do you own any synthetic clothing...” treating it as a question having implications (line 07). Participant C brought attention to how Pepper as a robot consisting of plastic parts is talking about plastics as a serious environmental issue (line 08).

## V. IMPLICATIONS

In this paper we wanted to draw attention to how people orient to the robot’s movements and utterances by actively working to constitute a sense for the interactions. We observed that people attempted to produce understanding of the robot’s capabilities and programming but also produced accounts of the robot that had a more social nature to them (“she’s overwhelmed”, “is everything okay”, “it’s a bit confused”). The robot was treated both as a machine and as something more agential. This phenomenon has been previously observed by Morana Alač [8].

In group HRI, people are not only interacting with the robot but also with other humans. Based on our observations, the group-robot situation appears as a multi-party interaction that orients toward a socially interactive robot. While there are several instances where it appears that only one person is interacting with the robot at a time, the other participants were not irrelevant, but meaningful to the unfolding scene. When groups interacted with Pepper, they worked together to make sense of the robot’s functions and produced accounts of the robot’s behaviors as social and agential.

We are hesitant to draw the conclusion that these social accounts were evidence of people’s perceptions or understandings of robot’s sociality in some psychological sense. Rather, we would argue that these are instances where the robot is configured as social through human interactions, as robots can provide interpretable, observable materials that people ordinarily use in social interactions with each other, such as gaze, gestures, and speech (see [9] for a more detailed discussion).

Studying more of these types of real encounters with a focus on human interactions could produce more understanding about how interactions from the robot can be treated, contributing to better interaction design and theory. Additionally, this would aid in identifying better what roles robots could take in ordinary public human environments. As an example, if the robot’s role were to be a facilitator for group interactions, we would have to consider how people together operate to give the robot such an interactional role and status in practice.

## VI. CONCLUSION

In this workshop paper we have provided three different excerpts from our recently gathered video data to illustrate how people together reconfigure and shape the unfolding interaction with a social robot by accounting for robot's behaviors. We conclude that for the interaction design of group-robot interactions in public settings, more studies are needed to examine how different interactions from the robot can be treated by human participants in real situations. We also note that in group interactions, people's interactions with each other play a crucial role in how the HRI unfolds, implying that the definition of what counts as HRI may need to be more inclusive to what people say and do around the robot, and not focus only on what happens directly between a robot and a human.

## TRANSCRIPTION SYMBOLS

The selected transcription symbols were adapted from [5]. Note that we used the double brackets to describe multimodal behaviors, but in [5] the double brackets are not limited to this purpose. The transcription of "beep" sounds produced by Pepper have been adapted from [10].

(( ))	transcriber's descriptions
(0.2)	a timed pause
(.)	a micropause
[ ]	overlapping talk
wo:rd	stretched sound
<u>word</u>	emphasis
↑word	rise in intonation
↓word	drop in intonation
ha ha	laughter
( )	uncertain part
da dup	Pepper's "beep" sound

## ACKNOWLEDGMENT

We thank Vera Diaz and Vilma Pohjonen for their role in the design and implementation of "PlasticBot" prototype during the course project with the first author.

## REFERENCES

- [1] SoftBank Robotics. <https://www.softbankrobotics.com/emea/en/pepper> (retrieved 23.1.2022)
- [2] H. Garfinkel, "A conception of, and experiments with, trust as a condition of stable concerted actions," In: *Motiv. Soc. Interact. Cogn. Approaches*, O. J. Harvey, Ed. New York: Ronald Press, 1963, pp. 187–238.
- [3] J. Heritage. *Harold Garfinkel and Ethnomethodology*. Polity Press, Cambridge, 1984.
- [4] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn-Taking for Conversation," *Language*, vol. 50, no. 4, p. 696–735, 1974.
- [5] G. Jefferson. "Glossary of transcript symbols with an introduction," In: *Conversation Analysis: Studies from the first generation*, G. H. Lerner, Ed. Amsterdam: John Benjamins, 2004, pp. 13–31, <https://doi.org/10.1075/pbns.125.02jef>
- [6] J. Parviainen, L. van Aerschoot, T. Särkikoski, S. Pekkarinen, H. Melkas, and L. Hennala, "Motions with emotions? A phenomenological approach to understanding the simulated aliveness of a robot body," *Techne Res. Philos. Technol.*, vol. 23, no. 3, pp. 318–341, 2019.
- [7] M. Alač, J. Movellan, and F. Tanaka, "When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics," in *Soc. Stud. Sci.*, vol. 41, no. 6, pp. 893–926, 2011.
- [8] M. Alač, "Social robots: Things or agents?" in *AI Soc*, vol 31, pp. 519–535, 2016, <https://doi.org/10.1007/s00146-015-0631-6>.
- [9] S. Jarske, S. Raudaskoski, and K. Kaipainen, "The Social of the Socially Interactive Robot: Rethinking human-robot interaction through ethnomethodology," in *Proceedings of Robophilosophy*, J. Seibt, M. Nørskov, and O. S. Quick, Eds. 2020, pp. 97–105, <https://doi.org/10.3233/FAIA200915>
- [10] H. R. M. Pelikan, and M. Broth, "Why That Nao? How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk," in *CHI'16*, pp. 4921–4932, 2016.