

Embodied Active Listening: How Non-Verbal Backchannel Behaviour Influences Trust and Perception of a Social Robot

Gisela Reyes-Cruz
School of Computer Science
University of Nottingham
Nottingham, UK
gisela.reyesacruz@nottingham.ac.uk

Marta Romeo
School of Mathematical and
Computer Sciences, Heriot-Watt
University
Edinburgh, UK
m.romeo@hw.ac.uk

Dominic Price
School of Computer Science
University of Nottingham
Nottingham, UK
dominic.price@nottingham.ac.uk

Matthew Aylett
School of Mathematical and
Computer Sciences, Heriot-Watt
University
Edinburgh, UK
m.aylett@hw.ac.uk

Chris Greenhalgh
School of Computer Science
University of Nottingham
Nottingham, UK
chris.greenhalgh@nottingham.ac.uk

Joel E. Fischer
School of Computer Science
University of Nottingham
Nottingham, UK
joel.fischer@nottingham.ac.uk



Figure 1: A participant talking to three versions of the Miro robot: a) static, b) responsive, c) random.

Abstract

With the trend to make conversational user interfaces (CUIs) more fluid and humanlike we must consider the impact of anthropomorphism, especially when the CUI is embodied in a device, robot or avatar. Non-verbal backchannel behaviour and active listening is a key research area in modern CUIs and in this paper we explore the impact of non-verbal backchannel behaviour on active listening using a zoomorphic robot, Miro. We carried out a mixed-methods experiment, involving 20 participants interacting with the robot under three behavioural conditions: static, responsive backchannel, and random backchannel. Findings reveal that while participants occasionally misinterpreted the robot's non-verbal cues, there were also notable moments of engagement and perceived connection. In addition, a responsive backchannel increased the perceived trust towards the robot. This suggests that trust in this particular embodied CUI was tied to their perceived social competence, highlighting the role of non-verbal backchannel behaviour in shaping human perceptions of CUI design.

CCS Concepts

• Computer systems organization → Robotics; • Human-centered computing → Empirical studies in HCI.

Keywords

human-robot interaction, HRI, embodied CUI, robots, backchanneling, non-verbal behaviour, trust

ACM Reference Format:

Gisela Reyes-Cruz, Marta Romeo, Dominic Price, Matthew Aylett, Chris Greenhalgh, and Joel E. Fischer. 2026. Embodied Active Listening: How Non-Verbal Backchannel Behaviour Influences Trust and Perception of a Social Robot. In *ACM Conversational User Interfaces 2026 (CUI '26)*, July 21–24, 2026, Bremen, Germany. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3816046.3816228>

1 Introduction

The anthropomorphism of AI-powered robots extends beyond the physical appearance of the robot to their embodied movements [54]. Analysis of more than 70 studies showed that movement positively affects the perceived anthropomorphism of a robot [51]. In particular, non-verbal behaviour is fundamental to human-human and human-robot interactions regardless of whether the robot is humanoid [64]. Non-verbal behaviour opens up possibilities for expressing the pragmatic dimension of language in conversational



This work is licensed under a Creative Commons Attribution 4.0 International License. CUI '26, Bremen, Germany

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2741-2/2026/07
<https://doi.org/10.1145/3816046.3816228>

design [22, 32]. For embodied conversational agents capable of natural language interaction, backchannel cues, i.e., the verbal and non-verbal expression of attention, understanding, or agreement (sometimes referred to as ‘active listening’), influence how the agent is perceived [9]. For CUIs in general, visual feedback to indicate the device is ‘listening’ has been shown to benefit system perception and increase trust [45]. Prior work seems to confirm that non-verbal cues contribute to make conversations with CUIs feel more natural [38] and to anthropomorphism’s influence on trust in human-robot interactions [39]. For instance, participants in previous research have expressed a preference for robots that exhibit backchannel behaviour as companions in stressful situations [8], and have been more engaged with a conversational agent that employs backchannel cues [28].

However, designing effective backchannel cues is challenging, as details such as the precise timing and perceived congruency of the cues matter [45]. Common speech-based dialogue systems that rely on the detection of silences as end-points of utterances are ill-equipped or just ‘too slow’ [3]. Instead, incremental dialogue systems [57] or approaches that predict end points may be needed for effective implementation of backchannels [31].

Furthermore, our research is also situated within a broader debate, which suggests a concern that anthropomorphisation can lead to undesirable consequences such as overtrust or over-reliance [61]. For instance, it has been suggested that “cuteness” in social robots could be exploited as a “dark pattern” to intrude upon people’s privacy at home [30].

To study the effect of non-verbal backchannels during active listening on self-reported trust in and perception of an embodied CUI, we transformed a zoomorphic robot, MiRo-E¹ (hereafter referred to as Miro) into an embodied active listener. In particular, we wanted to explore how non-verbal backchannel behaviours were perceived and understood and whether the system was considered more trustworthy with this ability. To study this we: (1) developed non-verbal backchannel behaviours for Miro inspired by human and digital avatars’ backchannels; (2) set up three versions of Miro, one static throughout the interaction, one employing backchannel behaviours randomly, and one using them responsively and appropriately in a Wizard of Oz (WoZ) experimental set-up; (3) ran an in-person study where 20 participants interacted with the three versions of the embodied CUI and gave us their opinion through a questionnaire, assessing their trust towards the interface in each condition (MDMT [60]) and their general thoughts on the interaction through open-ended questions.

With the work described in this paper, we contribute to the trust and anthropomorphism literature in CUI by unpacking key characteristics and challenges of non-humanoid embodiments and underlining the need to carefully navigate the boundaries between effectiveness and deception in responsible and trustworthy CUIs.

2 Related Work

2.1 Active Listening for CUIs

Being able to equip conversational agents with better and more human-like listening and turn-taking abilities has the potential to

improve their communicative performance and be perceived as more communicative, cooperative and fitting to users’ expectations [19, 65]. Anthropomorphism is not only a physical (surface) phenomenon but can be psychological, simulating social behaviours [52, 54]. Perceiving another as responsive to one’s needs is inherent to the formation of attachment bonds, and it can lead to a higher perceived anthropomorphism of the embodied agent, especially as the technology takes on new roles as a social entity and assistant [13].

In embodied conversational assistants, one of the human-like social behaviours that their embodiments could emulate is active listening [26, 34]. In active listening, many non-verbal behaviours have communicative functions [33]. These behaviours include gestures, head movements, body posture, eye contact, and facial expressions and are mostly referred to as backchannels [9]. Backchannels are a mixture of non-verbal and verbal behaviours such as nodding and sometimes reacting with short phrases like “aha”, “right”, “yeah”, “hmm” that are produced while another person is talking. These backchannel behaviours are typically interpreted as signalling attention, or agreement without disrupting the current speaker’s turn [2, 4]. Also, these signals are often used to encourage the speaker to continue talking or show understanding; however, they can be ambiguous as they are context-dependent [2, 10]. Schroder et al. [55] built human-like conversational listeners for 3D graphical human characters. The system generated verbal and non-verbal backchannel behaviours, which successfully increased user behavioural and felt engagement with the character. Park et al. [44] studied K2 (5-year old) children talking to two Tega soft-toy “monster-like” social robots. Based on observation of child-child dyadic interaction, they created an automated backchannel selection system used by one of the robots, and found that, compared to the other robot producing random backchannels, children perceived the non-random one to be more attentive and directed their storytelling more toward it. Birnbaum et al., examining the effects of a robot’s backchannel responses suggest that humans not only utilise backchannel cues to ascribe social intentions to robots but use them as a source of consolation and security [8].

Within this context, our work aims at studying the effect of timely and relevant non-verbal backchannel behaviours on users’ trust towards the system, as well as their general perception of those behaviours.

2.2 Trust in Social Robots

Trust has a fundamental role in technology adoption [7, 29]. Human likeness (resemblance) has been found to influence trust in human-robot interactions [24, 39]. Human-likeness and the anthropomorphisation of robotic technology are contentious. Positively, being able to recognise human behaviours in robots and autonomous systems facilitates the interaction by exploiting behavioural expectations [17]. Negatively, they can be seen as a form of deception [20], perpetuate societal stereotypes [49] and support “dark patterns” potentially used by malicious actors [25, 30].

Social robots are beneficial in a variety of contexts, from health-care [56] to education [43]. In particular, disclosure has been used as a way to improve well-being in vulnerable populations [40], both in dyadic and group interactions [47]. However, looking at the

¹<https://consequentialrobotics.com/miroe>

trust literature, it is clear that giving robots human-like capabilities can also result in overtrust [62], which can have major societal implications. People have shown little restraint in following a robot in an emergency situation, even after it was clear that the robot was defective [50]. Again, after cooperating with a robot to reach a shared goal, people have shown a tendency to build rapport and trust the robot enough to disclose sensitive information [1].

The same challenges, of designing trustworthy and safe interactions that could be of assistance to users, while identifying areas of potential harm, are equally pressing concerns for embodied CUIs.

3 Designing Miro's Backchannel Behaviours

In order to investigate how people perceive and respond to a robot active listener and its non-verbal backchannel behaviours, we used the zoomorphic robot Miro. The current study is not concerned with anthropomorphism in the appearance of the robot. There we chose a zoomorphic (animal-like) robot for the task, with the intention that people interacting with it would focus on its behaviours during the conversation rather than its superficial resemblance (or otherwise) to a human being. Choosing a zoomorphic robot also has the added potential benefit that people may have fewer preconceptions and expectations about its human-like capabilities, which might allow its conversational partners to interact with a more open mind. To create our non-verbal backchannel behaviours, firstly, we explored Miro's capabilities and the different types and range of movements available. These include ears wiggle, eye lids movements, head nod, head tilt, tail shake, body rotation, body moving forward and back. We used the results from the SEMAINE project [6, 55] on backchannels of digital avatars and their different meanings in order to inspire our design of Miro's non-verbal behaviours for the study. However, some humanoid avatar behaviours such as smiling, frowning and raising eyebrows, could not be adapted to Miro due to the design of the robot (i.e. no eyebrows or mouth).

Additionally, the design of behaviours was guided by our professional judgment and expertise in working with social robots, complemented by our individual interpretations of the robot gestures. After several iterations of the process above described (i.e. turning Miro's capabilities into gestures from prior literature or new ones agreed by the team), we defined a final list of "expressions" (Table 1), comprising potential non-verbal signals (gestures and movements) plus short phrases to continue and close the interaction (see Table 1 and Figure 2). No formal experimental validation of the expressions was performed, as the primary goal of the study was to explore participants' individual interpretations in context of their conversations with the robot rather than to assess the degree of correspondence between their interpretations and our own. Gestures were scripted using Miro's Robot Operating System (ROS²) interface and phrases were generated using the CereProc CereVoice toolkit³. A custom web application was built to enable the researchers to trigger the expressions during the user study by pressing a button on the screen (see Procedure in 4.1).

Table 1: Miro expressions including backchannels designed for the study.

Expressions	Description
Head nod	Repeat once, but can be requested multiple times for repeated nodding.
Head shake	Left to right, a couple of times.
Happy	Ear wiggle, tail shake, eyes open 90%, rapid small movements on the pitch motor for two seconds.
Sad	Close eyelids and slightly move head down.
Full sad	Close eyelids and move head down to the lowest position.
Disappointed	Full sad + head shake.
Eyes wide	Fully open eyes, no other movement done.
Surprise	Head up to the highest position and eyes fully open.
Alert	Head up halfway
Listening	Head up halfway + nod
Curious	Slight head tilt
Wink	Left eye lid close and open
Spin	Whole robot moving in circles until next movement is selected.
Wag	Tail moving left to right three times.
Ears and tail	Ear wiggle and tail shake three times.
Sleeping	Head down, eyes closed, idle circular movement.
Active listening	Listening + ongoing nodding
Sad listening	Sad + ongoing nodding
Short phrases	Please continue, Sounds nice, Tell me more, Thank you, That's sad, Unbelievable, Interesting, Yes, Okay No, Oh, Oh no, That's all, Goodbye

4 User Study Methodology

The user study we present in this paper aimed to investigate people's perceptions of an active listener robot, focusing on the effects of the addition of human-like non-verbal backchannel behaviours. Further, the study explored the relationship between those behaviours and people's trust (or lack thereof) in the robot. The study was approved by the School of Computer Science's Ethics Committee at the University of Nottingham.

4.1 Procedure

We conducted a within-subjects experiment with 20 participants. The study lasted approximately one hour and was divided into three stages, described below: pre-study questionnaire; main study; and post-study interview and debrief. The second (main) stage of the study involved three rounds of interaction, each with a nominally different robot (i.e. a different condition). In each round, the interaction was initiated by the robot, with a different but comparable set of questions to ask the participant (see below). After asking each question, the robot assumed the role of active listener with the aim of keeping the conversation going for as long as possible. In some

²<https://www.ros.org/>

³<https://app.cereproc.com/cerevoice-sdk/>



Figure 2: Examples of non-verbal backchannel behaviours defined for the study using the Miro-E robot: a) ongoing nodding (active listening), b) ears and tail wiggle, c) head tilt (curious), d) eyes closed, head down, shaking head left to right (disappointed).

conditions the robot executed non-verbal backchannel behaviours. Occasionally, some verbal feedback was needed to encourage participants to say more or to close the current conversation. The verbal feedback did not change across conditions, employing the same short phrases in each (see the “Short phrases” entry in Table 1). During the study, the robot’s behaviours was not autonomous. A single researcher assumed the role of “wizard”, controlling the robots’ responses through the custom-made app (see Section 3). We employed a Wizard of Oz approach as a first step towards a fully autonomous social robot (see appendix A for full details of Wizard of Oz protocol).

4.1.1 Stage 1. Pre-study information. First, we obtained demographic details and asked our participants to fill in the the Negative Attitudes Towards Robots Scale (NARS) [41]. Participants were then given the list of questions the robots would ask them and were given a few minutes to think about potential answers. This was intended to reduce participant’s anxiety about interacting with the robot. Before moving to the interaction with the robot, the researcher asked a control question (i.e. “Which one is your favourite restaurant in the city?”) to help participants start talking and establish a control case with a human to allow reflection on whether the participants were more comfortable interacting with the researcher or the robot.

4.1.2 Stage 2. Robot conditions. Participants were then asked to interact with the robot. The three robot conditions were: (1) static robot, not exhibiting any backchannel behaviours; (2) responsive backchannel robot, controlled by the researcher to display backchannel behaviours at appropriate times; (3) random backchannel robot, controlled by the researcher to display backchannel behaviours randomly throughout the conversation. The same Miro robot was used for the three conditions. However the researcher would leave the room with the robot in between rounds and change a coloured bandana on the robot before bringing it back to the participant to continue the study. This was intended to suggest that this might be a different robot or at least an independent conversation. The set of questions participants were asked by the robot in the three rounds were:

- Round 1: Q1) who are you?, Q2) tell me what you did yesterday, and Q3) tell me a funny memory

- Round 2: Q4) who are you?, Q5) have you read something sad in the news?, and Q6) how is it living in your neighbourhood?
- Round 3: Q7) who are you?, Q8) do you have any plans for tomorrow?, and Q9) how was your high school experience?

The first question was kept the same for the three rounds to establish a baseline of the user subsequent responses during the interaction with the robots and to keep the illusion of interacting with different robots. While the order of the questions remained the same, condition order effects were mitigated by counterbalancing the number of participants across the six sequences (Static–Random–Responsive: 3 participants; Static–Responsive–Random: 3 participants; Random–Static–Responsive: 3 participants; Random–Responsive–Static: 4 participants; Responsive–Static–Random: 4 participants; and Responsive–Random–Static: 3 participants).

After each round participants completed the Multi-Dimensional Measure of Trust (MDMT) Questionnaire v2 [35, 36, 60], which includes 20 items (see appendix B) that assess five dimensions of trust i.e., trust due to the robot being Reliable, Competent, Ethical, Transparent, and/or Benevolent, which are organised into two broader factors of trust: *Capacity Trust* (Reliable, Competent) and *Moral Trust* (Ethical, Transparent, Benevolent).

Participants were also asked four open-ended questions about their experience in that round: 1) do you think the robot was listening? 2) did this robot make you talk more or less compared to the researcher and previous robot (i.e. round)? 3) did you think the robot was understanding you? and 4) do you think this robot has a personality?.

4.1.3 Stage 3. Final questions and debrief. After the last round and its respective questionnaire were completed, participants were asked some final questions, including which robot (i.e. condition) of the three they preferred and why, and their perceptions about trusting (or not) the robots (i.e. conditions) they interacted with. They were then told the purpose of the study, that the robot was controlled by the researcher to show the three intended behaviours.

4.2 Participants Recruitment

Recruitment was done through mailing lists and university channels. Participants were required to be 18 years or older. In total, 20

participants took part in the study: 9 men, 10 women, and 1 non-binary. The age range was 19-69 (3 preferred not to say), mean=34.76, SD=12.99. Participants' occupations included PhD students (9), academics and researchers (5), masters' or undergraduate students (2), retired (2), apps and games developer (1), and development engineer (1). Their region of origin included Europe (9), Central and South America (3), East and South Asia (3), Africa (3), Middle East (1), and preferred not to say (1). Their self-reported familiarity with robots ranged from no familiarity (3), to low (7), mid (5) and high (5). Participants described themselves as talkative (7), not talkative (7) or variable depending on the situation (6). Participants were provided with a £10 Amazon voucher as compensation for their time.

5 Data Collection and Analysis

The data collected during the study included audio and video recordings per participant (approximately 45-50 minutes each) as well as their questionnaire responses.

Questionnaire data were analysed using SPSS Statistics. Likert responses from the MDMT questionnaire (see 4.1.2) were averaged within participants to produce composite subscale scores for capacity and moral trust. Although individual Likert items are ordinal, aggregated multi-item Likert scales are commonly treated as approximately continuous and analysed using parametric methods because such analyses are generally robust to violations of interval scaling assumptions [42].

The score in the NARS questionnaire was used to group subjects into three groups: top and bottom 25th percentile and the mid range. This grouping (*ngrp*) was then used to model a degree of participant variation. Study *conditions* (static, responsive, random) and *trusttype* (moral trust, capacity trust) were also included as factors in a repeated measures ANOVA analysis of self-reported trust. A Mauchley's Test of Sphericity was applied and neither *trusttype* nor *condition* violated the sphericity assumption.

Transcriptions from the recordings were generated using Microsoft's auto transcription service, and were manually checked by one of the researchers. Responses to the open-ended questions after each robot (Stage 2) and the final questions and debrief conversation (Stage 3) were analysed by the lead-author through Braun and Clarke's thematic analysis approach [11] (i.e. data familiarisation, individual inductive coding, grouping codes, and themes generation). Since only one coder was involved, there was no requirement for inter-coder checks or procedures to resolve discrepancies. Initial themes were presented and discussed with the team and collaboratively refined. In addition, the transcriptions were used to identify specific personal details given to the robot during their interactions.

6 Quantitative Results

Repeated measures ANOVA showed a significant effect on self-reported trust for *condition* ($F(2,34)=5.449$, $p < 0.01$) and *trusttype* ($F(1,17)=5.438$, $p < 0.05$). *ngrp* had no significant effect and there was no significant interactions, indicating no significant effect of participants attitude towards robots in the final results. Capacity Trust was higher than Moral Trust across the conditions (mean capacity 3.666 > mean moral 3.3 $p < 0.05$). Post-hoc tests show that the responsive condition was perceived as significantly more

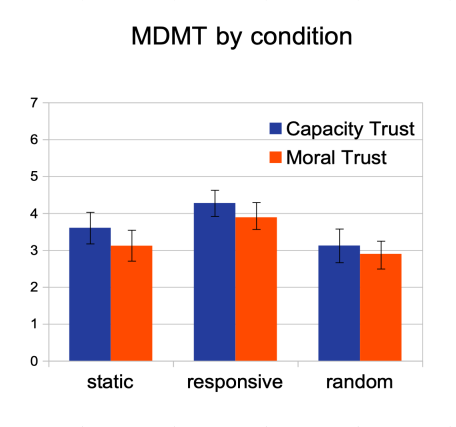


Figure 3: Differences in self-reported trust, by type and across conditions.

trustworthy than either the static condition (mean responsive 4.08 > mean static 3.360, $p < 0.05$) or the random condition (mean responsive 4.08 > mean random 3.3009, $p < 0.01$) (See Figure 3).

7 Qualitative Results

7.1 Interpretation of Robot Expressions

7.1.1 Confusion about robot non-verbal backchannel behaviours. Participants reported that some of the robot's non-verbal behaviours were easily interpreted, such as nodding (e.g. "that felt more natural", P10) and shaking the head left and right. However participants also reported confusion about some gestures, which hindered their interactions with the robot in both the responsive and random conditions (hereafter referred to as the "responsive robot", etc., for brevity). This was especially noticeable with the first moving condition they were exposed to. For example, some participants reported that robot expressions such as closing the eyes, moving the head down, or moving the ears could indicate listening or processing (P1, P2, P7, P9), especially after the participant talked (P9). However twelve participants stated that they could not interpret some of the robot's behaviours (e.g. head down, ears shaking, robot spinning), finding them weird, funny or feeling indifferent to them. This confusion often caused participants to stop talking, as they were trying to understand what was happening with the robot ("I felt like I was agitating it", P8). Even when there was no negative connotation, some participants reported not it did not encourage more conversation. Moreover, three participants noted a mismatch between their expectation of how an animal naturally moves, and the very limited ways in which this animal-like robot behaved (e.g. if the robot was trying to mimic a dog, expressions such as winking and ears shaking do not align with how dogs behave). Familiarity with pets was also a factor influencing participants' interpretations of Miro's expressions; for instance, P2 thought the eye blink was funny because his cat does it, but P10 cited his lack of familiarity with pets as groups for uncertainty understanding the robot's movements.

7.1.2 Robot (not) paying attention as a salient moment. Twelve participants interpreted the robot moving a lot and looking away during their interactions as indications of not paying attention to them when talking, especially the case with the random robot condition (nine participants), there the robot was interpreted as “*doing its own thing*” (P1, P5, P7). As these behaviours were often distracting, noisy, and interruptive, participants felt that the robot was not really listening or losing interest in their responses, which often provoked emotional reactions, such as annoyance and frustration or even feeling disrespected (P17), making them not want to keep talking (“*It’s not giving me any attention, so why should I bother?*”, P5). One participant, who during the interaction told the robot off for not listening, reflected on the awkwardness of the situation afterwards (“*It’s a robot, for goodness’s sake! Get crossed because he’s not listening to me, how silly is that?*”).

Participants further reflected on the notion of paying attention or not in human-human interaction, for instance P1 reflected on how some neurodivergent people act in similar ways (i.e. looking away) and that it is not necessarily bad but a non-normative way of being, which perhaps could be applied to robots (“*Maybe he’s the kind of robot that likes to move*”, P1). Likewise, some participants stated that talking to the random robot was reminiscent of talking to children in their lives, who did not show the usual social cues to indicate listening, (“*I was like, why is it looking that way? But then again, I’m a father of a three-year-old. Sometimes I’m speaking to her, and she’s listening and then out of nowhere she’s just going to stare at the wall... So it wasn’t completely weird to me*”, P6).

7.2 Perceptions About Interactions With The Robot

7.2.1 Moments of immersion and connection during interaction. There was a general sentiment that robots cannot understand in the same way that humans do (e.g. they are only parsing information, their movements are meant to evoke emotions, but these are not real), and it was evidenced by the fixed and generic nature of the questions and utterances, the mismatch between verbal and non-verbal behaviours (particularly with the random robot), and the lack of understanding of meaning (“*I think it understood that I was saying something, but not necessarily what it was*”, P11).

However, about half of the participants expressed that there were moments, if sometimes brief, in which they *felt* the robot was listening to some extent. As ten participants explained, it felt that the responsive robot (and sometimes random robot) reacted in appropriate ways through its expressions, such as turning its head towards them, nodding when asking a question, or reacting accordingly to the content of what they were saying (e.g. sad or funny story). Nine participants expressed that verbal utterances and follow-up questions by the robot caused them to believe the robot was listening; this was specifically the case for the static robot, which was found to be more timely and quick in its verbal delivery. Nonetheless, redundancy of behaviours and utterances were noted, that in conjunction with expressions not fitting the situation, and sequential prompting of questions, led to no conclusive evidence that the robot was listening. Participants’ understanding and sense-making of the technical capabilities and workings of the

robot shaped their perceptions about whether the robot was understanding them or not. For instance, fourteen participants expressed that there was *some form of understanding* showcased by the robot in the form of follow-up questions and utterances, recognition of meaning or emotion, and fitting verbal and non-verbal responses. However, nine of them explicitly stated that it could be functioning through picking keywords and/or tone of voice, using a Large Language Model (LLM) to generate responses, being pre-programmed to say and react in specific ways during specific questions, or by being remotely controlled.

Despite this awareness of the robot limitations or perceptions of how it was working, thirteen participants indicated that, surprisingly, there were moments of immersion and connection during their interactions with the robot (“*And the [responsive] one, I know it didn’t understand me, but maybe during the interaction there were moments where I was immersed enough to behave as if it understood me*”, P1). Interestingly, for two participants who were convinced the robot was being controlled or pre-programmed (and who reported high familiarity with robots), the most enjoyable interaction occurred with the random robot, which was unpredictable and not appropriate, but for that very same reason more fun and entertaining to interact with (P3, P15). Five participants mentioned that talking to the robot felt like talking to a pet, a baby or a child, that even when you know they may not be really understanding you, you still talk to them. As P3 put it, their responses to the robot were carefully crafted to fit its perceived capabilities (e.g. slowing down talk, using simpler words, controlling accent, and pronouncing in a clearer way). Three participants remarked on the robot’s childish voice, which in some cases influenced them to perceive and treat the robot as a child (P3, P17) or on the other hand, made them find the whole encounter “*creepy*” (P2).

7.2.2 Robot preference and robot personalities. Overall, participants’ preferences for robot type varied. Nine participants preferred the responsive robot, six participants preferred the static robot, four participants preferred the random robot, and one stated there was no marked preference (P13) as they felt no effect or emotion, although it was noticeable the robot behaved in different ways. The responsive robot was fun and lively (P1), reacted in a more human way with appropriate gestures that added to the interaction (P6, P7, P10, P14, P18, P19, P20). The static robot was perceived as a better listener as it paid attention in comparison to the other robots (P5). Even if still and quiet, its responses were seemingly fitting and timely (P8, P11, P16, P17), even humorous (P12). The random robot’s unpredictability made it feel more playful (P3), and funny and entertaining (P15). Others also preferred the random robot as they found the responsive robot too noisy and reacting too much (P4), or felt too pressured to talk by it (P9). Interestingly, these two participants indicated that talking less to the robot was desirable as they consider themselves introverted and non-talkative, and that being prompted to talk more, was stressful.

When talking about robot personality, participants were leaning more towards the fact that robots cannot have a personality, but it is easier to ascribe one when comparing their different behaviours. While the static robot was perceived as quiet and calm, the responsive robot was perceived as bubbly, happy, cheerful, extroverted, lively or sassy, and the random robot was perceived as care-free,

quirky, hyperactive, rude, and a jerk. Interestingly, four participants (P1, P2, P4, P14) gendered the robot conditions, perceiving the static one as a boy or a man (e.g. *“This is a white straight man way of listening, just complete stern... no signals that they’re listening, just there”*, P2), and the other conditions as girls (*“it shared some reactions... really try to connect with you”*, P4). However, the colour of the bandanas could have been a confounding factor (i.e. blue for static robot).

7.3 Would People Trust a Robot Active Listener?

7.3.1 Mistrusting the robot’s ability to interact. When asked to reflect on whether they trusted the robot differently in the different conditions and what personal information they shared, participants talked about trust in the robot’s technical and social capabilities. For instance, not trusting that the robot understands precludes communication (P14). Likewise, talking to a robot is less influenced by trust in relation to data protection and more about if it is fun or engaging to interact with it (P13). Willingness to share any information decreased if the conversation with the robot felt disconnected (P6), or when they felt the random robot was being disruptive or rude (P5). Participants reported talking less to the robot due to their limited ability to engage in conversation (whether perceived, confirmed, or unknown). Even when the participants’ task was to talk and the robot’s purpose was to listen, some participants purposely asked questions to it, as conversation was seen as a two-way activity (P2, P8, P15). Some participants expressed that they would be willing to open up more to the robot if it showed signs of empathy or interest in what they had to say (P6, P7, P9), indicating that although the responsive robot had some of that, it was not enough for them to fully ease into the interactions. Nonetheless, participants mostly reflected that they talked more to the researcher than to the robot, as they felt more at ease or felt that the researcher could relate more to their answers (i.e. about their favourite restaurant). While some of them talked at length to the robot (especially those self-described as talkative), they emphasised that the interaction felt more like an interview in which they merely provided information rather than a normal, human, conversation.

7.3.2 Trust concerns are shaped by personal views on sharing personal data. While a few participants expressed concerns about data potentially being recorded by the robot (P2) especially if connected to the Internet (P4), being curious about who has access to the data (P9), and overall not feeling ready to trust machines (P8), others reported no concerns due to generally being open people (P10, P17), or not feeling that the questions by the robot was leading them to share private information (P9, P11, P16). Some participants reported no concerns about personal data in general, as *“everything is online”* (P12) or feeling they are *“a boring person”* (P18). Likewise, it was remarked that *“personal means different things to different people”* (P15). Looking at what kind of information participants were sharing with the robot, they did not differ much between conditions. Our participants were confident sharing their name, country of origin, occupation and employer, and even identifiable information about family members. Being part of a ethically approved study from a higher education institution and having read a participant information sheet potentially put them at ease and confident that no information would be leaked or retained.

Although numbers do not differ much, the majority of participants shared their name (9 to the static robot, 15 to the responsive robot, 15 to the random robot); their occupation (19-static, 20-responsive, 19-random); their employer (12-static, 15-responsive, 14-random). Additional personal information was shared to a minor degree, such as their country of origin (4-static, 5-responsive, 4-random); details about family members, partners and living arrangements (5-static, 8-responsive, 4-random); and the neighbourhood they live in (6-static, 2-responsive, 3-random).

8 Discussion

First, we sum up the findings from the study in relation to trust. Second, we recap how people interpreted the backchannel behaviours and reflect on the implication for the design of non-humanoid embodied agents/CUIs. Third, we also discuss the fine line between backchannels being interpreted as active listening vs. being perceived as interruptions and the challenges this raises for the design of backchannels for CUIs. Finally, we relate our findings to the discussion on disclosure and deception in CUI studies and design.

8.1 Trust

In terms of trust, the responsive condition increased both capacity trust and moral trust towards the robot. In all conditions, self-reported moral trust tracks a little below capacity trust, perhaps reflecting that there is little evidence available to directly support moral assessment so it is largely inferred from capacity, but with some additional reservation [66]. The responsive condition also led to more moments of subjectively meaningful engagement with the robot (13/21), which seemed to arise primarily when the robot’s response was both timely and understood as being appropriate and proportionate. In this condition the robot was also felt to be more human, although, at least in our experiment, participants remained generally unconvinced that the robot could or did really understand them. The random condition evoked the lowest trust levels (although not significantly lower than the static version). This reflects the way that its behaviours were often perceived as disruptive in themselves (noisy, inappropriate) and/or as an indication of attention away from the listener, through the robot’s orientation. In other words they were not received as acknowledgements or signal to continue but rather as signals to stop or sources of confusion. By using a zoomorphic robot, we have tried to highlight the importance of the platform behaviour beside its appearance [27], and this result demonstrates that human-like backchannel behaviours and psychological anthropomorphism are fundamental elements in designing conversational robots that can listen [54]. Moreover, our findings reflect the complex and multifaceted nature of trust in the context of CUIs [18].

8.2 Interpretation of backchannel behaviours

For both the random and responsive conditions, some of the robot behaviours were found to be confusing. Behaviours with common human analogues (e.g. nodding meaning agreement or listening, or shaking head left to right meaning no) were the least problematic. The non-human zoomorphic behaviours, such as rotating the ears or spinning round, were a common source of confusion and uncertainty, which in itself disrupted the interaction. While choosing an

ambiguous zoomorphic robot was intended to mitigate the impact of its visual human-likeness (or otherwise), it clearly created other expectations with respect to its behaviour [23]. It is likely that the research team, working with the robot over time, have collectively established a more-or-less common orientation towards the robot, drawing on a range of external analogies (including a mixture of cat and dog behaviours). However the participants lacked this common understanding and found it much harder to interpret these novel behaviours. It may have been less disruptive to use a restricted, common, set of behaviours. Alternatively, perhaps a more explicit identification between the robot and a common animal stereotype would have reduced confusion.

Interestingly, comments on the human-likeness of the robot also referenced the voice of the robot, considered similar to that of a child and contributing to the anthropomorphisation of the Miro robot. Developing an appropriate voice for conversational interface has become a priority for the community only in recent years [16, 37], as its influence on trust becomes more and more clear [53, 59] as well as its influence on interface anthropomorphism related to gender assignment [14, 15]. Gender assignment also happened during our experiment, probably due to the colours of the bandanas used to signal the three different conditions. These considerations signal that anthropomorphism can happen in many different ways and that the full range of characteristics, both physical and psychological, that are at the core of the phenomenon are not yet fully understood. Our results highlight once more the fundamental role of voice in conversational interface design and the need to balance every possible anthropomorphism clue very carefully.

8.3 Active listening vs interruption

The timing of backchannel cues is crucial to achieve its intended purpose (e.g. show attention and encourage speaker to continue). In some occasions the timing was deemed successful by participants, but there were cases in which the behaviours were disruptive, potentially due to the Wizard of Oz approach employed. While embodied backchannel feedback by humans tends to be non-disruptive [2], embodied CUIs struggle to replicate that fluidity. The robot's servos and motors were also relatively loud in operation, and the motion profiles were preset, which may have added to a sense of interruption or awkwardness in the robot's actions. At least in part these issues may have contributed to the static robot being well received by some people, i.e. being seen as attentive and listening well. In the static condition, participants can focus on the robot's verbal responses. This, together with the fact that the random condition scored the lowest in terms of trust, would seem to indicate that it is better to have no backchannels than random or disruptive backchannels, at least in terms of eliciting trust.

8.4 Disclosure and deception

Ultimately, the type of information disclosed to the three conditions did not differ significantly. This could have been due to the safe environment in which the experiment was run. However, we see clear indications of heightened engagement, perceived familiarity and perceived trust towards the robot showing responsive backchannel behaviours. Prior CUI literature has found that backchanneling translates into more emotional disclosure, which can be particularly

positive for people who are in need of emotional support [12]. However, this can also have undesired effects, especially in prolonged and long-term human-embodied agent interactions, from creating emotional dependency to privacy issues related to the amount of information that users will be more and more willing to share with AI companions [63].

With our work, we have shown yet another way on how anthropomorphism can bring “dark patterns” (more appropriately called deceptive or hidden patterns) into conversational interfaces and underlined the importance of thinking about clear ethical design guidelines to outline what are anthropomorphic characteristics and how they play a role in shaping embodied CUI interactions and intention to use [21].

8.5 Limitations and Future Work

The choice of a zoomorphic robot (and consequently its design and available range of movements) limited both the definition of the non-verbal backchannel behaviours employed in this study, and the expectations that participants had about them (e.g., if it looks like a dog it should probably move like a dog). Although some of our findings can inform other forms of design and interaction with embodied CUIs and social robots, broad generalisations cannot be made as it is unknown to what extent they apply to other embodiments, whether humanoid, zoomorphic, digital or robotic. Our work highlights the need for more structured approaches for designing and developing backchannels [5].

There were some instances of participants gendering the robot based on its static and non-static behaviours (which is often not the case for zoomorphic robots [58]), as well as different use of pronouns when referring to it (i.e. he, she, it). These represent initial insights outside our primary focus. Further research is required to elucidate the role of robot backchanneling in gender bias [46].

The use of the Wizard of Oz approach introduced certain limitations, including delays in responding to participant inputs that occasionally overlapped with subsequent responses, accidental errors, and misjudgement of the timeliness or appropriateness of non-verbal backchannel behaviours. Lastly, this paper presents insights from mixed-methods responses (questionnaires, post-interaction interviews, and some information provided when talking to the robots); further and more in-depth analysis of the video and audio recordings, including use of language, reactions to specific robot's behaviours and how the interactions unfolded moment-by-moment will be explored in future work, as well as expanding the participant pool to reach a broader spectrum of technology consumers.

9 Conclusion

In this work, we explored the role of non-verbal backchannel behaviours in shaping trust and engagement during human-embodied conversational interactions. By implementing three behavioural conditions (static, responsive, and random) on a zoomorphic robot, we examined how timely and contextually appropriate non-verbal cues influence users' perceptions of a robot's active listening capabilities and trustworthiness. Our findings suggest that responsive backchannel behaviours significantly enhance self-reported trust, and can foster moments of perceived connection. Participants often interpreted these behaviours as signals of attentiveness and social

competence, even when aware of the robot's technical limitations. Conversely, random behaviours often disrupted the interaction, leading to confusion, frustration, and reduced willingness to engage. Interestingly, while some participants preferred the static condition for its perceived attentiveness and minimal distraction, others found the random condition entertaining, highlighting the diversity of user expectations and preferences. The study also revealed that anthropomorphic behaviours, even in the non-humanoid embodiment Miro-E, can evoke emotional responses and influence how the conversation unfolds, raising important ethical considerations around overtrust and data privacy.

The results underscore the importance of designing socially competent CUIs that balance effectiveness with transparency. While anthropomorphism can enhance engagement, it must be carefully calibrated to avoid misleading users or fostering undue reliance. Using non-human embodiments for CUIs also has its own particular characteristics and challenges. Future work should investigate how these findings generalise across different embodiment morphologies and contexts, and explore autonomous implementations of responsive behaviours. Ultimately, this research contributes to the growing discourse on responsible and trustworthy social interfaces and embodied CUIs, advocating for design strategies that support meaningful, trustworthy and interpretable interaction without compromising user agency or safety.

Acknowledgments

This work was supported by Engineering and Physical Sciences Research Council [grant numbers EP/V00784X/1, EP/Y009800/1], through funding from Responsible AI UK.

References

- [1] Alexander Mois Aroyo, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. 2018. Trust and Social Engineering in Human Robot Interaction: Will a Robot Make You Disclose Sensitive Information, Conform to Its Recommendations or Gamble? *IEEE Robotics and Automation Letters* 3, 4 (2018), 3701–3708. doi:10.1109/LRA.2018.2856272
- [2] Agnes Axelsson, Hendrik Buschmeier, and Gabriel Skantze. 2022. Modeling feedback in interaction with conversational agents—a review. *Frontiers in Computer Science* 4 (2022), 744574.
- [3] Matthew Peter Aylett, Eva Szekeley, Donald Mcmillan, Gabriel Skantze, Marta Romeo, Joel Fischer, and Gisela Reyes-Cruz. 2023. Why is my Agent so Slow? Deploying Human-Like Conversational Turn-Taking. In *Proceedings of the 11th International Conference on Human-Agent Interaction* (Gothenburg, Sweden) (HAI '23). Association for Computing Machinery, New York, NY, USA, 490–492. doi:10.1145/3623809.3623976
- [4] Adrian Bangerter and Herbert H. Clark. 2003. Navigating joint projects with dialogue. *Cognitive Science* 27, 2 (2003), 195–225. doi:10.1207/s15516709cog2702_3
- [5] Samuel R. M. Barrett and James E. Young. 2025. Toward Designing a Toolkit for Intuitive Design of Backchanneling Behaviour in Social Robots. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction* (Melbourne, Australia) (HRI '25). IEEE Press, 1236–1239.
- [6] Elisabetta Bevacqua, Sathish Pammi, Sylwia Julia Hyniewska, Marc Schröder, and Catherine Pelachaud. 2010. Multimodal backchannels for embodied conversational agents. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents* (Philadelphia, PA) (IVA'10). Springer-Verlag, Berlin, Heidelberg, 194–200.
- [7] Raden Agoeng Bhimasta and Pei-Yi Kuo. 2019. What causes the adoption failure of service robots? a case of Henn-na hotel in japan. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) (UbiComp/ISWC '19 Adjunct). Association for Computing Machinery, New York, NY, USA, 1107–1112. doi:10.1145/3341162.3350843
- [8] Gurit E. Birnbaum, Moran Mizrahi, Guy Hoffman, Harry T. Reis, Eli J. Finkel, and Omri Sass. 2016. What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior* 63 (2016), 416–423. doi:10.1016/j.chb.2016.05.064
- [9] Peter Blomsma, Gabriel Skantze, and Marc Swerts. 2022. Backchannel behavior influences the perceived personality of human and artificial communication partners. *Frontiers in Artificial Intelligence* 5 (2022), 835298.
- [10] Peter Blomsma, Julija Vaitonytė, Gabriel Skantze, and Marc Swerts. 2024. Backchannel behavior is idiosyncratic. *Language and Cognition* 16, 4 (2024), 1158–1181. doi:10.1017/langcog.2024.1
- [11] Virginia Braun and Victoria Clarke. 2021. *Thematic analysis: A practical guide*. SAGE publications Ltd.
- [12] Eugene Cho, Nasim Motalebi, S. Shyam Sundar, and Saeed Abdullah. 2022. Alexa as an Active Listener: How Backchanneling Can Elicit Self-Disclosure and Promote User Experience. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 273 (Nov. 2022), 23 pages. doi:10.1145/3555164
- [13] Sara Cooper, Alessandro Di Fava, Oscar Villacañas, Tatiana Silva, Victor Fernández-Carbajales, Luis Unzueta, Manex Serras, Luca Marchionni, and Francesco Ferro. 2021. Social robotic application to support active and healthy ageing. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 1074–1080.
- [14] Martina De Cet, Moperceivedhammad Obaid, and Ilaria Torre. 2025. Breaking the Binary: A Systematic Review of Gender-Ambiguous Voices in Human-Computer Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [15] Martina De Cet, Miriam Sturdee, Mohammad Obaid, and Ilaria Torre. 2025. Sketching Robots: Exploring the Influence of Gender-Ambiguous Voices on Robot Perception. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 103–112.
- [16] Balandino Di Donato, Marta Romeo, and Matthew Aylett. 2025. Sound design for HRI: a case study exploring appropriateness, naturalness, and sociality of a zoomorphic robot. In *AM'25: Proceedings of the 20th International Audio Mostly Conference*. Association for Computing Machinery (ACM).
- [17] Brian R. Duffy. 2003. Anthropomorphism and the social robot. *Robotics and autonomous systems* 42, 3-4 (2003), 177–190.
- [18] Justin Edwards and Elaheh Sanoubari. 2019. A need for trust in conversational interface research. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) (CUI '19). Association for Computing Machinery, New York, NY, USA, Article 21, 3 pages. doi:10.1145/3342775.3342809
- [19] Jonathan Ehret, Andrea Bönsch, Patrick Nossol, Cosima A. Ermert, Chinthusa Mohanathanasan, Sabine J. Schlittmeier, Janina Fels, and Torsten W. Kuhlen. 2023. Who's next? Integrating Non-Verbal Turn-Taking Cues for Embodied Conversational Agents. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents* (Würzburg, Germany) (IVA '23). Association for Computing Machinery, New York, NY, USA, Article 27, 8 pages. doi:10.1145/3570945.3607312
- [20] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review* 114, 4 (2007), 864.
- [21] Reza Etemad-Sajadi, Antonin Soussan, and Théo Schöpfer. 2022. How Ethical Issues Raised by Human-Robot Interaction can Impact the Intention to use the Robot? *International Journal of Social Robotics* 14 (06 2022). doi:10.1007/s12369-021-00857-8
- [22] Mary Ellen Foster. 2019. Face-to-face conversation: why embodiment matters for conversational user interfaces. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) (CUI '19). Association for Computing Machinery, New York, NY, USA, Article 13, 3 pages. doi:10.1145/3342775.3342810
- [23] Moojan Ghafurian, Gabriella Lakatos, and Kerstin Dautenhahn. 2022. The zoomorphic MiRo robot's affective expression design and perceived appearance. *International journal of social robotics* 14, 4 (2022), 945–962.
- [24] Peter A. Hancock, Theresa T. Kessler, Alexandra D. Kaplan, John C. Brill, and James L. Szalma. 2021. Evolving trust in robots: specification through sequential and comparative meta-analyses. *Human factors* 63, 7 (2021), 1196–1229.
- [25] Martin Hannig, Ruth Stock-Homburg, and Merlind Knof. 2025. The privacy paradox in interactions with service robots in the workplace. *The International Journal of Human Resource Management* 36, 3 (2025), 385–413. arXiv:https://doi.org/10.1080/09585192.2024.2445135 doi:10.1080/09585192.2024.2445135
- [26] Hung-Hsuan Huang, Natsumi Konishi, Sayumi Shibusawa, and Kyoji Kawagoe. 2015. Can a Virtual Listener Replace a Human Listener in Active Listening Conversation?. In *Proceedings of the International Workshop on Emotion Representations and Modelling for Companion Technologies*. 33–39.
- [27] Zuzanna Janeczko and Mary Ellen Foster. 2022. A Study on Human Interactions With Robots Based on Their Appearance and Behaviour. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 33, 6 pages. doi:10.1145/3543829.3544523
- [28] Jin Yea Jang, Saim Shin, and Gahgene Gweon. 2024. Minimal yet big impact: How ai agent back-channeling enhances conversational engagement through conversation persistence and context richness. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. 14509–14521.

- [29] Bing Cai Kok and Harold Soh. 2020. Trust in robots: Challenges and opportunities. *Current robotics reports* 1, 4 (2020), 297–309.
- [30] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'Dark Pattern' in Home Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 374–381. doi:10.1109/HRI.2019.8673274
- [31] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. 2017. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. 127–136.
- [32] Minha Lee. 2020. Speech acts redux: Beyond request-response interactions. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–10.
- [33] LibreTexts. 2013. Communication in the Real World - An Introduction to Communication Studies. <https://commons.libretexts.org/book/socialsci-18424>. [Online; accessed 2025-09-26].
- [34] Zvi Lotker. 2023. The Art of Active Listening. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*. 470–474.
- [35] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction*. Elsevier, 3–25.
- [36] Bertram F Malle and Daniel Ullman. 2023. Measuring human-robot trust with the mdmt (multi-dimensional measure of trust). *arXiv preprint arXiv:2311.14887* (2023).
- [37] Conor McGinn and Ilaria Torre. 2019. Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots. In *2019 14th ACM/IEEE international Conference on human-robot interaction (HRI)*. IEEE, 211–221.
- [38] Mohammad (Matt) Namvarpour and Afsaneh Razi. 2025. The Art of Talking Machines: A Comprehensive Literature Review of Conversational User Interfaces. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 45, 18 pages. doi:10.1145/3719160.3736621
- [39] Manisha Natarajan and Matthew Gombolay. 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) (HRI '20). Association for Computing Machinery, New York, NY, USA, 33–42. doi:10.1145/3319502.3374839
- [40] Yohei Noguchi, Hiroko Kamide, and Fumihide Tanaka. 2020. Personality Traits for a Social Mediator Robot Encouraging Elderly Self-Disclosure on Loss Experiences. *J. Hum.-Robot Interact.* 9, 3, Article 17 (May 2020), 24 pages. doi:10.1145/3377342
- [41] Tatsuya Nomura. 2014. Influences of experiences of robots into Negative Attitudes toward Robots. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (2014), 460–464. <https://api.semanticscholar.org/CorpusID:6019985>
- [42] Geoff Norman. 2010. Likert scales, levels of measurement and the “laws” of statistics. *Advances in health sciences education* 15, 5 (2010), 625–632.
- [43] Irena Papadopoulos, Runa Lazzarino, Syed Miah, Tim Weaver, Bernadette Thomas, and Christina Kouloughlioti. 2020. A systematic review of the literature regarding socially assistive robots in pre-tertiary education. *Computers & Education* 155 (2020), 103924. doi:10.1016/j.compedu.2020.103924
- [44] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017. Telling Stories to Robots: The Effect of Backchanneling on a Child's Storytelling. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) (HRI '17). Association for Computing Machinery, New York, NY, USA, 100–108. doi:10.1145/2909824.3020245
- [45] Yoonvin Park, Doha Kim, and Hayeon Song. 2025. I Know You're Listening: Designing Visual Backchannels for Voice User Interfaces. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces*. 56–59.
- [46] Giulia Perugia and Dominika Lisý. 2023. Robot's gendering trouble: a scoping review of gendering humanoid robots and its effects on HRI. *International Journal of Social Robotics* 15, 11 (2023), 1725–1753.
- [47] Simon Riches, Lisa Azevedo, Alkesh Vora, Ina Kaleva, Lawson Taylor, Peipei Guan, Priyanga Jeyarajaguru, Harley McIntosh, Constantina Petrou, Sara Pisani, et al. 2022. Therapeutic engagement in robot-assisted psychological interventions: A systematic review. *Clinical Psychology & Psychotherapy* 29, 3 (2022), 857–873.
- [48] Laurel D Riek. 2012. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012), 119–136.
- [49] Daan Robben, Eriko Fukuda, and Mirjam De Haas. 2023. The Effect of Gender on Perceived Anthropomorphism and Intentional Acceptance of a Storytelling Robot. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) (HRI '23). Association for Computing Machinery, New York, NY, USA, 495–499. doi:10.1145/3568294.3580134
- [50] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 101–108. doi:10.1109/HRI.2016.7451740
- [51] Eileen Roesler, Dietrich Manzey, and Linda Onnasch. 2021. A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science robotics* 6, 58 (2021), eabj5425.
- [52] Marta Romeo, Peter E. McKenna, David A. Robb, Gnanathusharan Rajendran, Birthe Nessel, Angelo Cangelosi, and Helen Hastie. 2022. Exploring Theory of Mind for Human-Robot Collaboration. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 461–468. doi:10.1109/RO-MAN53752.2022.9900550
- [53] Marta Romeo, Ilaria Torre, Sébastien Le Maguer, Alexander Sleat, Angelo Cangelosi, and Iolanda Leite. 2025. The Effect of Voice and Repair Strategy on Trust Formation and Repair in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 14, 2 (2025), 1–22.
- [54] Peter A. M. Ruijten, Diane H. L. Bouten, Dana C. J. Rouschop, Jaap Ham, and Cees J. H. Midden. 2014. Introducing a Rasch-Type Anthropomorphism Scale. *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2014), 280–281. <https://api.semanticscholar.org/CorpusID:15539742>
- [55] Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Bjorn Schuller, Etienne de Sevin, Michel Valstar, and Martin Wollmer. 2012. Building Autonomous Sensitive Artificial Listeners. *IEEE Transactions on Affective Computing* 3, 2 (2012), 165–183. doi:10.1109/T-AFFC.2011.34
- [56] Arielle AJ Scogio, Erin D Reilly, Jay A Gorman, and Charles E Drebing. 2019. Use of social robots in mental health and well-being research: systematic review. *Journal of medical Internet research* 21, 7 (2019), e13322.
- [57] Gabriel Skantze and Anna Hjalmarsson. 2010. Towards incremental speech generation in dialogue systems. In *Proceedings of the SIGDIAL 2010 Conference*. 1–8.
- [58] Roger Andre Søraa. 2017. Mechanical genders: how do humans gender robots? *Gender, Technology and Development* 21, 1-2 (2017), 99–115.
- [59] Ilaria Torre, Jeremy Goslin, Laurence White, and Debora Zanutto. 2018. Trust in artificial voices: A “congruency effect” of first impressions and behavioural experience. In *Proceedings of the technology, mind, and society*. 1–6.
- [60] Daniel Ullman and Bertram F. Malle. 2019. Measuring Gains and Losses in Human-Robot Trust: Evidence for Differentiable Components of Trust. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 618–619. doi:10.1109/HRI.2019.8673154
- [61] Daniel Ullrich, Andreas Butz, and Sarah Diefenbach. 2021. The development of overtrust: An empirical simulation and psychological analysis in the context of human-robot interaction. *Frontiers in Robotics and AI* 8 (2021), 554578.
- [62] Daniel Ullrich, Andreas Butz, and Sarah Diefenbach. 2021. The development of overtrust: An empirical simulation and psychological analysis in the context of human-robot interaction. *Frontiers in Robotics and AI* 8 (2021), 554578.
- [63] Anouk Van Maris, Nancy Zook, Sanja Dogramadzi, Matthew Studley, Alan Winfield, and Praminda Caleb-Solly. 2021. A new perspective on robot ethics through investigating human-robot interactions with older adults. *Applied Sciences* 11, 21 (2021), 10136.
- [64] Nungduk Yun and Seiji Yamada. 2025. Investigation of Factors That Influence Human Presence and Robot Anthropomorphism in Telepresence Robot. *IEEE Access* 13 (2025), 4996–5008. doi:10.1109/ACCESS.2024.3516782
- [65] Mateusz Zarkowski. 2019. Multi-party turn-taking in repeated human-robot interactions: an interdisciplinary evaluation. *International Journal of Social Robotics* 11, 5 (2019), 693–707.
- [66] Ke Zhou, Luo Lan, and Zhiqiang Yan. 2023. Human's moral judgements towards different social actors: A cross-sectional study. *British Journal of Developmental Psychology* 41, 4 (2023), 343–357.

A Wizard of Oz Protocol

Responses to the guidelines for reporting Wizard of Oz studies by Riek [48].

Robot:

- *How many robots were used?* One robot, three conditions or rounds.
- *What kind(s) of robot(s)?* Zoomorphic.
- *What level(s) of autonomy?* All components were controlled by the wizard
- *What were the robot's capabilities?* Non-verbal backchannel behaviours and short verbal phrases (see Table 1).
- *What hypotheses did the researcher have for the robot?* Whether non-verbal backchannel behaviours during active listening have an effect on self-reported trust in the robot.

Users:

- *How many users participated in total, and per experimental trial?* 20 participants, 3 robot conditions each.

- *What were the user demographics, sampling procedure, etc.?* See participant recruitment section (4.2).
- *What instructions were provided to the users?* They were asked to respond to the robot's questions, which were provided in full before the robot rounds started.
- *What behavioural hypothesis does the researcher have about the users?* Whether different non-verbal backchannel behaviours in embodied active listening could be conducive of increased self-reported trust in the robot.
- *Was the simulation convincing to the users?* Some participants believed the robot was moving autonomously whereas others guessed it was being controlled by someone.
- *What expectations did the user have about the robot, before and after the experiment?* Attitudes towards the robot before the study were collected using the NARS questionnaire [41]—see results in section 6. Interpretations and expectations after the study are extensively reported in section 7.

Wizard:

- *How many wizards were used?* One.
- *What were the wizard demographics (e.g., the researcher, lab mates, naive)?* An HRI researcher, co-author of this paper.
- *Did the wizard know the behavioral hypothesis of the experiment?* Yes.
- *What were the wizard production variables and how were they controlled for?* Short verbal phrases aimed to be minimal and consistent across the three conditions. Non-verbal backchannel expressions not to be used for the static condition. Timely and appropriate non-verbal backchannel behaviours to be used for the responsive condition based on the wizard criterion (e.g. nodding, displaying happy or sad expressions based on participant conversation). Triggering a different non-verbal behaviour every five seconds, aided by a counter integrated on the WoZ interface displaying seconds and milliseconds since previous action triggered.
- *What were the wizard recognition variables and how were they controlled for?* There was no constraint for what the wizard could recognise.
- *How did the experimenter control for wizard error (deliberate and accidental)?* No deliberate errors were included in the study. Accidental errors occasionally occurred but were not measured or identified.
- *How much and what sort of training did wizards receive prior to starting the experiment?* The wizard conducted two different pilot testing sessions before the study was conducted, as well as a series of demos that showcased the designed gestures and movements for the robot, which all collectively helped to obtain an understanding of what it meant to produce timely and appropriate backchannel behaviours.

General:

- *Where did the experiment take place?* Research lab.
- *What were the environmental constants and how were they controlled?* The same room and space configuration was used for all participants.

- *What scenarios did the researchers employ?* Three conditions: static, responsive and random non-verbal backchannel behaviours. Each condition contained three different questions asked by the robot (see section 4.1.2).
- *Was this experiment part of an iterative design process?* No.
- *Does the paper discuss the limitations of WoZ?* See limitations section (8.5).

B Multi-Dimensional Measure of Trust v2 (MDMT) Questions

The MDMT [35, 36, 60] consists of 20 items that are rated using the scale from 0 (Not at all) to 7 (Very), with the option "Does not fit".

Table 2: MDMT v2 questions[illegible]