"Are you sad, Cozmo?" How humans make sense of a home robot's emotion displays

Hannah Pelikan, Mathias Broth and Leelo Keevallik

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-167357>

N.B.: When citing this work, cite the original publication.

Pelikan, H., Broth, M., Keevallik, L., (2020), "Are you sad, Cozmo?" How humans make sense of a home robot's emotion displays, *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI'20)*, , 461-470. <https://doi.org/10.1145/3319502.3374814>

Original publication available at:

<https://doi.org/10.1145/3319502.3374814>

Copyright: ACM Press

<http://www.acm.org/>

© ACM 2020. This is the author's version of the work. It is posted here for your personal use. Not for redistribution.

"Are You Sad, Cozmo?" How Humans Make Sense of a Home Robot's Emotion Displays

Hannah R. M. Pelikan
hannah.pelikan@liu.se
Linköping University
Linköping, Sweden

Mathias Broth
mathias.broth@liu.se
Linköping University
Linköping, Sweden

Leelo Keevallik
leelo.keevallik@liu.se
Linköping University
Linköping, Sweden

ABSTRACT

This paper explores how humans interpret displays of emotion produced by a social robot in real world situated interaction. Taking a multimodal conversation analytic approach, we analyze video data of families interacting with a Cozmo robot in their homes. Focusing on one happy and one sad robot animation, we study, on a turn-by-turn basis, how participants respond to audible and visible robot behavior designed to display emotion. We show how emotion animations are consequential for interactional progressivity: While displays of happiness typically move the interaction forward, displays of sadness regularly lead to a reconsideration of previous actions by humans. Furthermore, in making sense of the robot animations people may move beyond the designer's reported intentions, actually broadening the opportunities for their subsequent engagement. We discuss how sadness functions as an interactional "rewind button" and how the inherent vagueness of emotion displays can be deployed in design.

CCS CONCEPTS

• **Human-centered computing** → **Field studies; Collaborative and social computing design and evaluation methods.**

KEYWORDS

emotion; affect; social robots; robots in the home; conversation analysis; non-lexical sounds; long-term interaction

ACM Reference Format:

Hannah R. M. Pelikan, Mathias Broth, and Leelo Keevallik. 2020. "Are You Sad, Cozmo?" How Humans Make Sense of a Home Robot's Emotion Displays. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI'20)*, March 23–26, 2020, Cambridge, United Kingdom. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3319502.3374814>

1 INTRODUCTION

Emotions have been an important topic in human-robot interaction (HRI) for many years. Both human and robot emotions have been studied, with research investigating the recognition and categorization of human emotions (e.g. [27, 47]) and exploring how emotions

can be expressed in robots [2, 5, 8, 30, 37, 42, 45, 46, 48, 50, 52]. Robot emotions are typically studied by asking participants to evaluate different emotion displays after seeing them in an experimental setting, outside a real world interaction with the robot [8, 42, 45]. Such approaches presuppose the existence of unambiguous emotion displays that are independent of the specific interactional context. Recent work in HRI [12, 25], human-computer interaction [4] and in human-human interaction (HHI) [51, 57] has challenged this perspective. There is mounting evidence that humans fine-tune their consecutive actions to the details of emotion displays in HHI [35, 55] and that they make sense of a robot's emotion displays within locally developing sequential contexts [12]. However, we still know very little about how robot emotion displays are understood and responded to in real world social interaction.

We studied how humans make sense of robotic emotion displays when playing with a robot in their home. We collected and analyzed 19 hours of video data from 4 pairs of German adults and 4 Swedish families interacting with a Cozmo robot (see Figure 1), taking a multimodal conversation analysis (CA) perspective [17, 33]. Approaches rooted in psychology typically focus on individual subjects' interpretation of emotion displays (e.g. [22]). In contrast, our CA approach, rooted in ethnomethodology [13], focuses on the social-interactional achievement of emotion in real world contexts and stays agnostic about the existence of inner feelings [44]. This approach seems particularly suitable for HRI, as it makes the question whether robot emotions are simulated or real [10] irrelevant. From a CA perspective, robotic emotions exist as soon as participants treat robots as having them.

The contribution of this paper is threefold. First, we provide a detailed description of how a Cozmo robot's emotion displays are treated in everyday interaction in peoples' homes. Second, we demonstrate how these emotion displays occasion a range of possible interpretations, which may differ from the designers' intentions. We discuss how the relative vagueness of emotion displays can be drawn upon in robot interaction design. Finally, we contribute insights into how robot emotion displays may be used as an interactional resource in HRI [12], with displays of happiness moving the interaction forward and displays of sadness possibly working as an interactional "rewind button". We thereby contribute to the ongoing theoretical debate on robot emotions, demonstrating how they are understood as social actions.

2 EMOTION IN HUMANS AND ROBOTS

When displaying and understanding emotions, humans typically draw on multimodal resources, such as facial expressions, prosody, gestures, posture and spatial movements as well as lexical and syntactic design of verbal utterances [36, 44]. Some of these aspects,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
HRI '20, March 23–26, 2020, Cambridge, United Kingdom

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6746-2/20/03...\$15.00
<https://doi.org/10.1145/3319502.3374814>

such as facial expressions [2, 5, 46, 52] have been picked up in the design of robot emotion displays. The idea to express emotions in robot faces is inspired by Paul Ekman’s work on basic emotions [11], which suggests that some emotions are universal and produced by specific facial muscle groups. HRI work has also moved beyond the face and explored movement [8, 45] and sound [42] to convey emotions. Recently, multimodal expressions of basic emotions have become popular, e.g. combinations of facial expressions and vocal sounds [48] or color, sound and movement/vibration [30, 50].

2.1 Why Should a Robot Display Emotions?

A common view in HRI regards emotion displays as a way to communicate the robot’s internal state and thereby providing users with access to the robot’s intentions (e.g. [5]). This approach has been criticized by work in cognitive science, which suggests that simulating emotions in a robot is not the equivalent to the robot actually having emotions [10]. Another argument for emotion expression in robots is that the robot should align with the users’ emotions (or do the opposite) [20]. Stressing the social function of emotions, recent work in HRI points out that emotion displays need to be produced at appropriate moments in interaction. Jung [25] stresses that an emotion display (e.g. a smile) can be understood differently in different contexts. Fischer and colleagues [12] underline that emotion expression in HHI is strongly dependent on the interactional context and propose that robots should display emotions at those moments where humans would typically expect an emotional response. They point to CA research as providing useful insights on this.

2.2 Emotion as a Public Display

There is a solid amount of research in CA on how humans attend to emotion displays in HHI. They are regularly co-constructed by several people [44]. For instance, humans normally do not laugh alone, but finely coordinate laughing together [6, 15, 24]. Similarly, surprise in story-telling is not so much an involuntary emotional eruption than a collaborative achievement. The teller of a story employs linguistic devices to elicit a surprised response from the listener [56] and previous speakers often project displays of disappointment as the relevant next action [9]. Emotions are not expressed randomly but typically produced in response to a triggering event [18]. Since humans unavoidably make sense of behavior in the context of previous actions [13], the same formal expression can convey a variety of emotions depending on its precise interactional context [43]. Human emotions are sequentially organized and negotiated between specific participants to an interaction, and it is likely that humans also make sense of robot emotions based on the same interactional logic [12]. Read and Belpaeme [41] corroborate the finding that robot emotions may be interpreted differently in different situations in a lab study. This paper sheds light on how humans understand robot emotion displays in the sequential context of real world interaction.

3 METHOD

To study how participants interpret robot emotion displays in the wild, we collected video data of people interacting with a Cozmo robot in their homes. We conducted fieldwork in two countries with

Figure 1: The Cozmo robot with its touch sensitive toy cubes and the smartphone app that controls the robot.

German adults (Study 1) and Swedish families (Study 2). In addition, to include the designer’s reasoning in creating the emotion animations, Cozmo’s audio designer, Ben Gabaldon, was interviewed for 45 minutes.

3.1 The Cozmo Robot

We used Anki’s Cozmo robot (see Figure 1), a small robot inspired by Pixar’s Wall-E and Eve, which is marketed as a toy for children in the age of 8-14. The robot does not have speech recognition and mainly interacts through beeps, movements and animated eyes. Cozmo recognizes human faces, pets and its toy cubes and uses different sensors to track its body orientation. Among many other sounds and animations, Cozmo displays emotions, which according to the sound designer are fitted to specific contexts, such as a particular game or activity (e.g. learning a user’s name and face). The robot is controlled through a smartphone app with different modes such as letting the robot roam freely, teaching it a person’s name and face, playing games or creating simple programs.

3.2 Participants

A total of 28 people (16 adults, 12 children) participated in the study. Study 1 included four pairs of German adults. Study 2 involved four Swedish families with at least one child in the age of 8-14. The entire family was encouraged to play with the robot and the collected data often involve more than one participant interacting with the robot, including siblings and parents. Participants were recruited by convenience sampling through personal contacts and local social media groups. To explore cross-cultural patterns, we included participants with two different native languages and from diverging age groups. The sample size was determined following ethnographic and conversation analytic standards. Compared to quantitative work, our numbers are small. However, our findings are based on detailed analyses of manifest behavior in real time encounters and offer strong empirical evidence of how humans make sense of a particular robot’s design when interacting with it.

3.3 Setting

In both studies, the first author visited participants in their homes and accompanied them during their initial encounter with the robot. The aim of the study was explained and consent was obtained before cameras and robot were turned on. The robot does not come with any indications beyond how to switch it on and the researcher did not give participants further instructions. In Study 1, participants played with the robot for as long as they wanted, which turned out to be between 10-30 minutes. The robot was controlled through the researcher's phone and was roaming freely, a mode in which the robot takes randomized initiatives. In Study 2, families played with the robot for 8-13 days. They controlled the robot through their own smartphones and were free to explore all its functions. In this study, participants were recorded by the first author during introduction and retrieval of the robot whereas they recorded themselves with a simple camcorder on a tripod when interacting with the robot during the days in between. The researcher recorded with two video cameras from two different angles, one focusing on the robot and one on participants. The families were instructed to record not only the robot but also their faces and body and were asked not to stand behind the camera but to join the everyday family activities.

3.4 Ethics

Before the first visit, participants received written information about the study, including a detailed explanation of the videotaping. They were told that their participation was voluntary and that they could withdraw at any time. This information was repeated orally before the start of the study. A briefer and simpler explanation was given to the children. All adult participants signed informed consent to be videotaped and opted-in for different kinds of video usage (e.g. anonymized pictures). Parents signed consent for their children. All participants, including children, were asked for consent before the video cameras and robot were switched on. Since participants largely videotaped themselves in Study 2, they were in control when to videotape and when not to. Along with instructions on how to switch the robot and camera on and off, families received guidance concerning "ongoing consent", stressing that they should always ask all people present whether they agree to being recorded before switching on camera gear. In one case, a child did not like to be videotaped after a few days and the mother then stopped the recordings. Video data are stored on password-protected hard drives, which only the researchers have access to. Participants' real names are replaced by pseudonyms in all transcripts.

3.5 Data Collection and Analytical Method

We collected 19 hours of video data (12.5h recorded by researcher, 6.5h by participants) following CA practices for video recording [31]. Data were transcribed following verbal [23] and multimodal [32] transcription conventions and analyzed from a CA perspective, which focuses on the social achievement of actions and activities in an orderly and sequentially organized manner [19]. Scrutinizing how every action is contingent on a previous one and projects a certain next action, CA studies the sequential organization of verbal and embodied behavior as well as how artefacts are incorporated into sensemaking. The empirical basis of the analysis lies in participants' locally displayed understandings of others' actions.

To learn more about how users make sense of the robot's emotion displays, we therefore focus on moments when several people are gathered around the robot, as they would negotiate meanings among them. Furthermore, the multiparty nature of our data is highly relevant for understanding how a robot functions in an actual home, as it may lead to greater interactional complexity (such as talking in overlap with the robot and with each other). Cozmo was designed for such everyday environments where more than one person is regularly present, and some of the games presume multiple players. Analytically we draw on the understanding of robot actions that participants' display publicly. By studying how humans hearably and visibly deal with various cues in real time we obtain access to sensemaking in the wild. This analytic procedure goes beyond experimentally induced settings such as think-aloud protocols and interviews/questionnaires after an interaction has terminated, which rely on participants' (re)construction of the actual event for scientific purposes. In contrast, the CA method identifies salient action patterns through transcribing and analyzing video recordings of real world interaction in detail. Identifying patterns that recur across instances, CA allows generalization from data without losing empirical grounding in local specificity. Targeting interactional sequences, ethnomethodological and CA studies have yielded insights into interaction with robots [34, 59], voice interfaces [40] and copying machines [53].

4 DISPLAYS OF HAPPINESS

There is a variety of animations in Cozmo's SDK that are designed to express happiness. They consist of a combination of sounds, "smiling" eyes and movement of the robot's forklift arms and head, and they are different in different contexts (e.g., finishing a trick, winning a game or learning a person's face). The particular animation that we will be scrutinizing here (see Figure 2) occurs after Cozmo has learned a new face, represented by 26 instances in our data.

Figure 2: Transcription, spectrogram and visuals during the "happy to see you" animation. Cozmo lifts its head and displays "smiley" eyes. It then produces a long sound sequence during which it lifts its forklift arms and moves them in the air. After a short silence, the robot produces another sound with rising intonation, while slightly nodding its head.

We selected this animation as it is produced in a context in which humans are directly collaborating with the robot (e.g. positioning their face in front of the robot's camera) rather than only watching the robot accomplish a "trick", such as singing or lifting its toy cubes. As Cozmo's sound designer pointed out, the process of scanning a new face is "difficult because we [the animation team] needed to try to create this whole closed system of 'I'm going to scan a face, I'm in the process of scanning, very clearly identifying that I successfully did.' And it was a very conscious decision as a group that we needed to build that robot-human relationship by making sure that Cozmo always reacted very happy. He is happy to see you. [...] We needed a very positive successful moment there."

In 16 cases people responded to the animation. In 8 out of these they laughed or smiled (see Excerpt 1). In 7 cases they reacted with an embodied response beyond laughter (see Excerpt 2) and in 1 case the participants reinterpreted the animation (see Excerpt 3). In the remaining 10 cases that will not be discussed here, the animation was ignored, typically because participants had already moved on before the happy animation was completed. The excerpts reflect the variety of observed instances.

Excerpt 1 illustrates a typical sequential context in which the "happy to see you" animation is triggered. This is one of the 8 cases in which participants respond with laughter or smiles. Husband (DAD) and wife (MOM) have played with Cozmo (COZ) for a while. When their adult son Jonas (SON) comes home, they call him to have the robot learn his name. The researcher types his name into the app interface and the robot then scans Jonas' face for several seconds, for which he has to stay idle. After finishing the learning process, the robot is now reading out the newly learned name. Please see appendix A for transcription conventions. Focus lines are highlighted using blue for Cozmo's emotion display and green for human responses.

Excerpt 1. E18-12-30 [19:31-19:45]

```
01 COZ jona::s+
    son +smiles-->11
02 MOM j↑A:::[: jo]-
    yes jo-
03 COZ [a↑::]:o::w
04 (0.4)
05 MOM ↑a::[:o]
06 COZ [jon]as:
07 (0.5)
08 MOM (h)&e(h)e[(h)a% ]
09 COZ [>dudedu] & dudidu< dudel↑a=
    coz &lifts and moves arms&
    mom %smiles-->11
10 SON =(h)a#(h)a
    im #Image1
11 (1.5)%+(0.5)
    mom -->%
    son -->+
12 DAD du wie hieß nochma der andere roboter,
    you what's the other robot called again?
```

The robot reads out Jonas name (line 01), to which his mother responds with a "yes", followed by the first syllable of Jonas name (l. 02). She stops her talk when the robot produces another sound (l. 03). The mother responds with a similar sound (l. 05), which again is overlapped by the robot starting to talk, repeating Jonas name (l. 06). After a short silence (l. 07), the mother starts to laugh (l. 08). Once more, she turns silent after the robot starts producing the "happy to see you" animation (l. 09). As soon as the sound of the animation finishes, Jonas responds to it with laughter as well (l. 10, Image 1). After this, mother and son stay silent for 2 seconds, and also stop smiling (l. 11). The father then starts a new sequence, asking about another robot (l. 12).

In this excerpt Jonas, the person whose name the robot just learned, reacts to the happy animation. He produces a short laugh while looking at the robot. Participants treat the name learning sequence as successfully finished and initiate a new topic.

A similar instance can be found in Excerpt 2, in which the robot has just learnt the first family member's name, that of sister (SIS) Anna, who is playing with the robot together with her brothers (BR1 and BR2) as well as their parents. As in the previous excerpt, the robot reads out Anna's name twice before producing the "happy to see you" animation, to which Anna responds with an embodied response beyond laughter (one of 7 similar cases). Also in this case, the robot's sounds overlap with other participants' turns-at-talk.

Excerpt 2. FAM4_19-09-03_cam2_1 [04:58-05:13]

```
01 COZ an|na
02 (0.4)
03 SIS oh han sa ANNA ((MOM and DAD? laugh))
    oh he said anna
04 COZ anna
05 (0.2)
06 BR1 tar mitt &n- hm &[mitt namn nu] &
    take my n- hm my name now
07 COZ [>dudedu] dudi+du< dudel↑a
    coz &.....&lifts and moves arms&
    sis +.....-->
08 (0.4)+(0.2)#(0.3)
    sis .....+pets COZ-->
    im #Image2
((1. 09-13 omitted. BR2 asks about button on COZ's back))
14 MOM nu ska vi ta en ny person
    now we will take a new person
```

Cozmo reads out Anna's name (l. 01), which she acknowledges by saying "oh he said Anna" (l. 03). Following the pre-programmed sequence for learning a new name, the robot repeats her name (l.

04). Anna's older brother (BR1) then requests that Cozmo learn his name next (l. 06). While he is making his request, Cozmo starts playing the happy animation (l. 07). Anna responds to the animation by petting Cozmo on the back (l. 08, Image 2). A short side-sequence evolves when Anna's twin brother (BR2) inspects the button on Cozmo's back (l. 09-13). The face learning activity is then resumed by their mother who announces "now we will take a new person" (l. 14).

In sequential terms, the "happy to see you" animation occurs "late", as participants have already moved on to the learning of the next person. Although the robot's emotion display overlaps with participants' talk, it is still treated as relevant in a non-verbal way by Anna, who pets the robot. Like Jonas in Excerpt 1, Anna, the person whose name was learnt, is positioning herself as the addressee of this happy animation. As in Excerpt 1, participants move on to next matters after the animation, indicating that this constitutes a successful and unproblematic completion of an interactional sequence.

In Excerpts 1 and 2, we have seen that humans respond to the robot sound in ways that are very much in line with what was intended by the designers. The robot learns a new face and name, and then plays a happy animation, which participants also respond to in the specific context of the name learning activity. However, as Excerpt 3 illustrates, participants do not always act according to the designers' intended sequence. While other cases could also be counted into this category, this is one clear and arguably incontestable manifestation of a reinterpretation of the intended design in our data. This case demonstrates that participants make sense of a design in its specific context of occurrence. The following sequence was recorded before the beginning of a dinner party, where a husband (HUS) and wife (WIF) had arrived ahead of other guests and enjoyed playing with the robot while the hosts were making dinner preparations. In this excerpt, the robot is learning husband Ulrich's name. Ulrich does not wait until the robot displays that the face learning sequence has ended, but initiates a new sequence, unrelated to the face learning activity.

Excerpt 3. E18-12-31 [02:51-03:16]

```

01 HUS  hm m "prob"- machmas ma [anders]
        hm m tr(y)- let's do it differently
02 COZ                                     [wa:::]:::
03      (0.9)
04 HUS  ma[gst du magst du giesin]ger bier?=  
        do you like do you like Giesinger beer?  
        ((beer from local brewery))
05 COZ  [ulrich ]
06 WIF  =(h)a
07 RES  e(h)[h]
08 COZ  [w]a::o::w
09 HUS  A:[H ]
        oh
10 RES  [ha]ha]ha[haha]
11 HUS  [haha] [haha]haha]ha ha ha he]
12 WIF  [haha (h)(h)(h) (h)(h)(h)]
13 COZ  [°° rich°° ]
14 HUS  &[.hhh [hehe [ha]heha .h] (h)](h)]
15 WIF  [(h) (h)(h)(h) (h) (h)]
16 RES  [(h)(h) (h)(h) (h) (h)]
17 BOS  [hi: ]ha]
18 COZ  [>dudedu&: dud]idu< dudel[a=  
        coz &lifts and moves arms&
19 HUS  =ahJA=  
        oh yes

```

((l. 20-23 omitted. WIF, HUS, HOS, RES laughing))

24 HUS volle zustimmung
full approval

((l. 25-28 omitted. WIF, RES, HUS, HOS laughing))

29 WIF [da überschlägt sich ja glei](h)i
and ((as a result)) he is jumping for joy

30 HUS [ha ha ha ha ha ha ha ha ha ha]

31 ((more joint laughter))

32 HUS willst du 'n schluck trinken?
would you like to drink a sip?

Ulrich has been trying to get Cozmo to do things for a while. He now says that he will try something different (l. 01) and proceeds to ask Cozmo whether it likes the local beer he is drinking (l. 04). This is typical of the beginning of human encounters where locally available materials are often commented on at guest arrival [38]. While Ulrich is formulating this question, the robot reads out his name (l. 05). Ulrich's question attracts laughter from his wife and the researcher (l. 06-07). The robot then produces a sound that can be heard as an enthusiastic "wow" (l. 08). Ulrich responds to it with a loud "ah" (l. 09), a German interjection that has been associated with displaying surprise [16], similar to "oh" in English. Since previous attempts to make the robot produce a relevant response had been unsuccessful, the robot sound that is produced (seemingly) in response to the question might come unexpected to Ulrich (and the other participants). Again, this is responded to by laughter from the researcher (l. 10, 16), Ulrich (l. 11, 14), his wife (l. 12, 15) and the host (l. 17), who started watching the scene from the kitchen door. Cozmo repeats the husband's name (l. 13), but it is unclear whether participants actually hear it. When the robot plays the "happy to see you" animation (l. 18), Ulrich stops laughing. Right after the sound of the animation stops, he says "ahja" (l. 19), which has been characterized as an information receipt in German [3]. Ulrich laughs and then proceeds to say "full approval" (l. 24), thereby treating the animation as a relevant turn and formulating his understanding of the robot's behavior as a positive response to his question. His wife reinforces this by formulating the same in her own terms, saying that the robot is jumping for joy (l. 29). Like in the previous excerpts, participants move on after the robot's display of happiness, and Ulrich continues, asking the robot whether it wants a sip of his beer (l. 32).

While the display of happiness was designed for completing a successful scanning of a face, the excerpt shows that humans attend to current local contingencies that may place such a pre-designed animation into the temporal trajectory of a different action sequence. Here, the question "do you like beer?" (l. 04) forms the first pair part of an adjacency pair, which should be followed by a second pair part, a response to the question [49]. The display of happiness (l. 18) accordingly gets interpreted as a positive response to the question, leading participants to the conclusion that Cozmo likes beer. By continuing the sequence and asking the robot whether it wants a sip of his beer, Ulrich recontextualizes the previous question and robot response as a pre-sequence [49], which serves as a preparation for his subsequent offer. The participant has thus constructed the interactional sequence in his terms, recontextualizing the robot sounds in a way that fits the temporal emergence of that sequence.

In this excerpt, the happy animation is treated differently than intended by the designers. However, similar to Excerpts 1 and 2,

the robot's emotion display is again dealt with as a relevant turn. While participants in Excerpt 3 offer more extensive responses to the emotion display than in earlier cases, they still treat the happy animation as unproblematic by moving on. As we will see in the next section, this is different from the animation designed as sad.

5 DISPLAYS OF SADNESS

The variety of Cozmo's animations for sadness is smaller than those designed for happiness. The sound designer stated that "sad is an easy emotion to communicate". This seems in line with research findings that negative emotions are easier to convey in a robot [50] and that sadness can be conveyed through one modality (sound) alone [30]. The animation that we will be using here (see Figure 3) occurs most often as an indication of failure in a "fist bump" activity that Cozmo randomly initiates and which is explained in the app. Engaging in a fist bump, Cozmo holds its forklift arms up and waits for a user to "bump" their fist against the robot's arms. If no user engagement is detected, Cozmo plays the sad animation.

Figure 3: Transcription, spectrogram and visuals during the sad animation. Cozmo first lifts its head and forklift arms. It then produces a sound with falling intonation, moving arms and head down. After about a second of silence, the robot produces a longer sound with falling intonation and turns away. The robot displays "sad" eyes during the animation.

Out of 19 cases in which Cozmo attempted to do a fist bump, 11 resulted in the sad animation being played. Only in 4 out of the 11 cases participants actually recognized that the robot was trying to get them to engage in a fist bump. When they did not understand the activity, the animation was ignored (2 cases) or only minimally acknowledged (2 cases). In 3 cases participants reinterpreted the sad animation within their current alternative action trajectory, as we will show in Excerpt 6.

Excerpt 4 illustrates a typical response to Cozmo's display of sadness in cases where the participants had recognized that the robot was engaging in the fist bump activity. Son, mother and father are playing with the robot, which is holding its forklift arms out for a fist bump.

While the son is holding his fist against Cozmo's arms, his mother says "fist bump" (l. 01). After a short silence with no apparent reaction from Cozmo (l. 02), she asks whether one has to say something

Excerpt 4. FAM3_19-08-07_P3 [03:19-03:33]

01	MOM	fistbump	
	son	>>holding fist against COZs forklift arms-->	
02		(0.3)	
03	MOM	m*åste man säga nåt sånt	
		does one have to say something like that?	
	son	-->*	
04		(0.3)	
05	COZ	&wa: &[o]	
	coz	&arms down&	
06	DAD	[ne]j	
		no	
07	DAD	[>jag vet inte<]	
		I don't know	
08	MOM	[mj a:: är du] led[sen?]	
		are you sad?	
09	COZ	[&wa]]w&a wa wao:	
	coz	&turns2MOM&	
10	MOM	m .pt nej cozmo::	
		no cozmo	
11		(0.9) & (0.2) & (0.3)	
	coz	&turns further&	
12	COZ	wao	
13		(0.6)	
14	MOM	ja? =	
		yes?	
15	COZ	=owa	
16		(0.2)	
17	MOM	>vill du göra de?< ((moves fist forward))	
		do you want to do it?	

like this (while doing the activity) (l. 03). During her turn, the son moves his fist away from the robot. After another short silence (l. 04), Cozmo starts producing the sad animation, playing a sound with falling intonation and moving its forklift down (l. 05). While the father produces an answer to his partner's question (l. 06-07), she instead responds to the robot's sound, producing a high-pitched "mjaa" vocalization, which can be heard as comforting. She then asks the robot "are you sad?" (l. 08). Formulating her understanding as a question, she requests Cozmo to confirm her assessment. This confirmation is provided in the continuation of the sad animation and as Cozmo turns away from son and dad, towards mom. As the second part of the animation is longer and combined with a turning away (l. 09), it can be seen as an upgraded version of the earlier action. The mother says "no cozmo" (l. 10) with the last syllable lengthened and sounding sad herself. Both the "mjaa" and the "no" align with the negative display of the robot. In the silence that follows (l. 11), Cozmo turns further towards the mother and produces another sound (l. 12), which she responds to with a "yes?" (l. 14). After another robot sound (l. 15), the mother asks whether Cozmo wants to do it and holds her fist out in front of the robot's face (l. 17). The mother thereby seizes the very first opportunity to launch a remedial action, thus re-considering the previous actions instead of moving on to next matters.

As this excerpt demonstrates, the sad animation is treated quite differently from what we saw with the happy animation. The mother produces a comforting sound and a verbal account of her interpretation of the robot's behavior, which gets confirmed through the robot's subsequent continuation of the display of sadness. In contrast to the display of happiness, participants do not produce a next action that progresses the activity or initiate a new action trajectory. Instead, the mother offers to Cozmo that it can do the fist bump with her, trying to re-do the action that previously failed.

Other families deal with the robot's display of sadness in similar ways. In Excerpt 5, mother, sister and brother are playing with

Cozmo. After a successful fist bump with the mother, the robot moves its arms up in front of the little boy.

Excerpt 5. FAM2_19-06-11_P2 [12:09-12:18]

01 MOM [försiktigt] %med robote+n%,
careful with the robot,
02 COZ &[wa] &[o:]
coz &arms down&
sis +.....+points at COZ+
mom %points at COZ---%
03 SIS [(a:)]
04 MOM [förs]iktigt med robo-
careful with the robo-
05 SIS han är les[sen]
he is sad
06 COZ & [wa]w&a[wawao:]
coz %turns2MOM&
07 MOM [%mj%a:::h] #är du leidsen?%
are you sad?
mom %..%pets COZs back---%,,,%
im #Image5a
08 (0.5)
09 MOM m(h)
10 (1.6)
11 MOM vill du %göra %fist# bump med me%j?%
would you like to do fist bump with me?
mom %....%fist towards COZ-%,,%
im #Image5b

The mother explained how to do a fist bump with the robot and just asked her son to try. She adds that he should be careful with the robot (l. 01). While she is saying this, the robot starts playing the sad animation (l. 02). The mother does not respond at first and proceeds to repeat her previous utterance (l. 04). This occurs in overlap with a sound that the sister produces, possibly realizing that the robot is sad (l. 03). The mother cuts off her talk as the girl offers her interpretation of the robot's action, saying "he is sad" (l. 05). The robot continues with the second part of the animation (l. 06), overlapping with the sister's speech. While the robot is still playing the animation, the mother starts to produce a prosodically marked "mjaa" (l. 07) similar to Excerpt 4, l. 08. She also pets Cozmo's back while doing so and then asks the robot "are you sad?" (l. 07, Image 5a), providing an account of how she interprets the robot's behavior. After some silence and a short laughter syllable (l. 09), the mother then asks the robot whether it wants to do a fist bump with her (l. 11, Image 5b).

There is a striking similarity in how the two mothers respond to the robot's display of sadness. They first produce a comforting sound, and then ask the robot whether it is sad. In both excerpts, an understanding of the animation as sad is proposed by participants already after the first part. In Excerpt 4, it is produced by the mother who also comforts the robot. In Excerpt 5, it is produced by the sister and then taken up by the mother. In both cases, the sequence does not end with the failed fist bump and a sad robot but both mothers offer to re-do the failed action.

However, similar to the display of happiness, the sad animation is not always interpreted as intended by the robot designers. In Excerpt 6, one of 3 such cases, the robot produces the sad animation after it has been offered a cube. In this context, the display of sadness is understood as negating a previous proposition by the human.

Excerpt 6. E19-01-05_1 [28:13-28:31]

01 COZ &(1.7) ädu? (1.6)
coz &lifts arms-->
02 HUS 'ch glau&be er will mit seinen kis&[tn spielen]
I think he wants to play with his boxes
03 COZ [kwa]ke]
coz -->&arms down, drives forward&lifts arms-->
04 + (0.2) + (2.4) +(0.2)&
hus +turns cube+puts cube on COZ's arms+
coz -->&
05 COZ &wa: &[o]
coz &arms down&
06 (0.4)
07 HUS NEI&N?=
no?
coz &turns away-->
08 COZ =wa]w&a[wawao:
coz -->&
09 HUS o:h nein
oh no
10 WIF hm m:
11 (0.3)
12 HUS das wars gar nich
that's not what it was ((not what he wanted))

Cozmo produces a sound with rising intonation and lifts its forklift arms (l. 01). The husband provides his interpretation of the action (l. 02). In overlap with the end of his turn, Cozmo produces another sound (l. 03) and the husband picks up a cube and places it on the robot's forklift arms (l.04). Once he has done so, the robot starts producing the sad animation and moves its arms down (l. 05), thereby also dropping the cube. The husband reacts by saying "no" (l. 07) with rising intonation, thereby offering it as a candidate understanding [1] of the robot's action that needs confirmation. Just a short moment later, the robot continues with the second part of the animation (l. 08). As in Excerpt 4, this is treated as an upgraded response in the current sequential position and the husband reacts with a prosodically marked "oh no" (l. 09), which sounds a bit sad in itself, similar to Excerpt 4, l. 10. His wife utters a "hm m" (l. 10), which can be heard as confirming the husband's interpretation. He concludes "that's not what it was" (l. 12), suggesting that the cube was not what the robot wanted.

Comparing Excerpts 4 and 5, in which the animation for sadness is explicitly verbalized by participants as "sad", to Excerpt 6, it is clear that the very same animation is understood in a different way. Participants did not recognize that the robot aimed to play fist bump, but interpreted its behavior as a request for the cubes. While Cozmo's subsequent emotion display is still interpreted as a negative response, it is not explicitly treated as sad, but rather as a refusal to participate in the proposed activity.

As we have seen in Excerpts 4-6, displays of sadness attract manifest reactions by participants. In contrast to displays of happiness, they are typically responded to verbally, treating the robot as just having produced a dispreferred responsive action [28, 39] that needs to be negotiated. While the happy animation is unproblematic for the interaction to proceed, displays of sadness need to be addressed before participants can move on. This becomes especially evident in the number of turns-at-talk that are needed to deal with the interactional trouble. While displays of happiness can be acknowledged in a single turn (see Excerpt 1, l. 10 and Excerpt 2, l. 07-08), participants respond to Cozmo's displays of sadness by first offering a potential (negative) understanding, which needs to get confirmed (l. 08 in Excerpt 4 and l. 07 in Excerpt 6). In subsequent

turns, participants may then attempt to address the interactional trouble by offering the missed action (Excerpts 4 and 5). The husband in Excerpt 6 also reconsiders his own action and identifies it as not what the robot wanted. However, as the robot provides no hint about what would be the desired action, the participants cannot offer the "correct" action and ultimately abandon the activity.

6 DISCUSSION

Using the examples of one happy and one sad animation, we have shown in detail how humans make sense of robot emotion displays in everyday interaction at home. When they occur as part of a recognizable action sequence, people typically attend to these emotion displays, even interrupting their speech when the robot produces its sounds. People tend to interpret the animation designed as happy in a positive way and the animation designed as sad as negatively charged, which suggests that designers can work with generic categories of emotion displays. However, we find that human interpretation of the robot's emotion displays is strongly dependent on their immediate context. People make sense of the emotion displays in relation to preceding actions and treat them as projecting specific ways to continue the interaction. As we have demonstrated in Excerpt 3, an animation designed to communicate "happy to see you" may get reinterpreted as displaying that the robot likes beer when it is preceded by the question "do you like beer?". Similarly, an animation designed as a display of sadness about a missing fist bump gets reinterpreted as refusing a toy cube when it is produced after an offer of a cube (Excerpt 6). The robot's emotion displays project particular next actions such as closing of the sequence [7] in the case of happiness and (re-)negotiation of participants' prior actions in the case of sadness. This provides further evidence that particular emotion displays are interpreted as social actions in their sequential context [12, 18, 51].

As we have shown, people may treat displays of happiness as "go-ahead-signals". Displays of sadness may work as an interactional "rewind button" that makes people reconsider their prior actions and help them to identify a missing action or to recognize that their previous action was not compatible with the robot's "expectations".

6.1 Implications for Design

The above observations have several implications for design of social robots, which we list below.

Observation 1: Emotion displays are interpreted in the context of immediately preceding actions.

Design recommendation: Design for sequential contexts. Do not use robot emotions as a random outburst but place them in a sequence of actions that humans can make sense of. During a face recognition activity: (1) scan user face, (2) indicate success through display of happiness, (3) say user name, (4) move on. Involve conversation analysts to design action sequences that humans recognize [58].

Observation 2: Participants move on to other matters before the designed robot sequence has finished (e.g. Excerpt 2), making the emotion display redundant.

Design recommendation: Temporal relationship is key. Emotion displays need to occur right after the action that they respond to. Do not use them to repeat what has already been communicated. Test the temporalities with people in real world contexts.

Observation 3: Generally, happiness implies that interaction can move on, while sadness seems to function as an interactional "rewind button".

Design recommendation: Design emotion displays as actions in a sequence. Like verbal information, emotion displays constitute a form of feedback that humans draw upon when asking "what next?" [54]. Having broader meaning potentials [29] and being less specific than lexical speech [21], emotion displays can make sense in a variety of contexts.

Happy animations can be part of sequences of the general form: (1) Robot prompts human action, if (2a) correct action is detected, (3) the robot indicates success by a display of happiness. If instead (2b) a subsequent human action is missing or inappropriate, (3) the robot can indicate trouble by a display of sadness (4) followed by repairing actions (e.g. lifting arms for fist bump again, verbal clarification of what needs to be fixed, etc.). After a sad animation participants should not be left struggling without possibilities to fix the problem (see e.g. Excerpt 6). Instead, remedial actions should follow. Humans are obviously willing to deal with the trouble but they first need to receive a cue of what needs to be redone.

Observation 4: Participants may interpret emotion animations in ways that are different from the intended design.

Design recommendation: Exploit their vagueness. Treat the inherent vagueness as a perk rather than a problem. Emotion displays leave room for participants' own interpretation and creativity [14, 26], which may be particularly desirable in contexts of play or when it is difficult to generate a semantically and pragmatically correct verbal response for a robot. We propose that emotion displays can communicate "quickly and vaguely", i.e. indicating quickly in a semantically underspecified way how the interaction may evolve in general terms. In contexts of long-term human-machine collaboration such generic signals may suffice, while in less habitual contexts where correct understanding is crucial, subsequent ("slower and more specific") actions can serve to elaborate.

7 CONCLUSION

Taking a multimodal conversation analytic perspective, we have investigated how humans interpret robot emotion displays in everyday interaction in their homes. Focusing on two particular animations, one designed to display happiness and the other to display sadness, we find that participants make sense of robot emotions only in their specific context of occurrence, in a way that may well differ from what designers had in mind. Further, we find that while displays of happiness are generally unproblematic and move the interaction forward, displays of sadness make participants reconsider their previous actions, treating them as problematic in their interaction with the robot. We discuss how these findings can inform design of robot emotion displays as a resource for communicating "quickly and vaguely" whether everything is going according to plan or whether humans need to reconsider their actions.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments and our participants for welcoming us into their homes. Thank you, Malte Jung, for inspiring discussions on robot affect. This work was funded by the Swedish Research Council, project no. 2016-00827.

REFERENCES

- [1] Charles Antaki. 2012. Affiliative and disaffiliative candidate understandings. *Discourse Studies* 14, 5 (October 2012), 531–547. <https://doi.org/10.1177/1461445612454074>
- [2] Christian Becker-Asano and Hiroshi Ishiguro. 2011. Evaluating facial displays of emotion for the android robot Geminoid F. In *2011 IEEE Workshop on Affective Computational Intelligence (WACI)*. IEEE, 1–8. <https://doi.org/10.1109/WACI.2011.5953147>
- [3] Emma Betz and Andrea Golato. 2008. Remembering relevant information and withholding relevant next actions: The German token achja. *Research on Language and Social Interaction* 41, 1 (2008), 58–98. <https://doi.org/10.1080/08351810701691164> arXiv:<https://doi.org/10.1080/08351810701691164>
- [4] Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers. 2007. How emotion is made and measured. *International Journal of Human-Computer Studies* 65, 4 (2007), 275 – 291. <https://doi.org/10.1016/j.ijhcs.2006.11.016> Evaluating affective interactions.
- [5] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59, 1 (2003), 119 – 155. [https://doi.org/10.1016/S1071-5819\(03\)00018-1](https://doi.org/10.1016/S1071-5819(03)00018-1)
- [6] Mathias Broth. 2002. *Agents Secrets: Le Public Dans la Construction Interactive de la Représentation Théâtrale*. Ph.D. Dissertation. Uppsala University, Uppsala.
- [7] Mathias Broth and Lorenza Mondada. 2013. Walking away: The embodied achievement of activity closings in mobile interaction. *Journal of Pragmatics* 47, 1 (Feb 2013), 41–58. <https://doi.org/10.1016/j.pragma.2012.11.016>
- [8] Jessica Rebecca Cauchard, Kevin Y. Zhai, Marco Spadafora, and James A. Landay. 2016. Emotion encoding in human-drone interaction. In *The Eleventh ACM/IEEE International Conference on Human-Robot Interaction (HRI '16)*. IEEE Press, Piscataway, NJ, USA, 263–270. <http://dl.acm.org/citation.cfm?id=2906831.2906878>
- [9] Elizabeth Couper-Kuhlen. 2009. A sequential approach to affect: The case of disappointment. In *Talk in Interaction: Comparative Dimensions*, Minna Laakso Markku Haakana and Jan Lindström (Eds.). Finnish Literature Society (SKS), Helsinki, 94–123.
- [10] Ezequiel A. Di Paolo. 2003. Organismically-inspired robotics: Homeostatic adaptation and teleology beyond the closed sensorimotor loop. In *Dynamic Systems Approach to Embodiment and Sociality*. Advanced Knowledge International, Adelaide, Australia, 19–42.
- [11] Paul Ekman. 1999. *Basic emotions*. Wiley, New York, USA, 45–60.
- [12] Kerstin Fischer, Malte Jung, Lars Christian Jensen, and Maria Vanessa aus der Wieschen. 2019. Emotion Expression in HRI - When and Why. In *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, United States, 29–38. <https://doi.org/10.1109/HRI.2019.8673078>
- [13] Harold Garfinkel. 1967. *Studies in Ethnomethodology*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.
- [14] William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 233–240. <https://doi.org/10.1145/642611.642653>
- [15] Phillip Glenn. 2003. *Laughter in Interaction*. Cambridge University Press, Cambridge, UK.
- [16] Andrea Golato and Emma Betz. 2008. German ach and achso in repair uptake: Resources to sustain or remove epistemic asymmetry. *Zeitschrift für Sprachwissenschaft* 27, 1 (2008), 7–37. <https://doi.org/10.1515/ZFSW.2008.002>
- [17] Charles Goodwin. 2017. *Co-Operative Action*. Cambridge University Press, Cambridge, UK. <https://doi.org/10.1017/9781139016735>
- [18] Marjorie H. Goodwin and Charles Goodwin. 2001. Emotion within situated activity. In *Linguistic Anthropology: A Reader*. Blackwell, Malden, MA, USA, 33–54.
- [19] Elliott M. Hoey and Robin H. Kendrick. 2017. Conversation analysis. In *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide*, A. M. B. De Groot and Peter Hagoort (Eds.). 151–173.
- [20] Guy Hoffman, Gurit E. Birnbaum, Keinan Vanunu, Omri Sass, and Harry T. Reis. 2014. Robot responsiveness to human disclosure affects social impression and appeal. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI '14)*. ACM, New York, NY, USA, 1–8. <https://doi.org/10.1145/2559636.2559660>
- [21] Emily Hofstetter. 2020. Non-lexical 'moans': Response cries in board game interactions. *Research on Language and Social Interaction* 51, 3 (2020).
- [22] Ruud Hortensius, Felix Hekele, and Emily S. Cross. 2018. The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (Dec 2018), 852–864. <https://doi.org/10.1109/TCDS.2018.2826921>
- [23] Gail Jefferson. 2004. Glossary of transcript symbols with an introduction. In *Conversation Analysis: Studies from the first generation*, Gene H. Lerner (Ed.). John Benjamins, Amsterdam, 13–31. <https://doi.org/10.1075/pbns.125.02jef>
- [24] Gail Jefferson, Harvey Sacks, and Emanuel A. Schegloff. 1987. Notes on laughter in the pursuit of intimacy. In *Talk and Social Organisation*, Graham Button and John R. E. Lee (Eds.). Multilingual Matters, Clevedon, 152–205.
- [25] Malte F. Jung. 2017. Affective Grounding in Human-Robot Interaction. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, USA, 263–273. <https://doi.org/10.1145/2909824.3020224>
- [26] Leelo Keevallik and Richard Ogden. 2020. Sounds on the margins of language, at the heart of interaction. *Research on Language and Social Interaction* 51, 3 (2020).
- [27] Iolanda Leite, Rui Henriques, Carlos Martinho, and Ana Paiva. 2013. Sensors in the Wild: Exploring Electrodermal Activity in Child-Robot Interaction. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI '13)*. IEEE Press, Piscataway, NJ, USA, 41–48. <http://dl.acm.org/citation.cfm?id=2447556.2447564>
- [28] Stephen C. Levinson. 1983. *Pragmatics*. Cambridge University Press, Cambridge, UK.
- [29] Per Linell. 2009. *Rethinking Language, Mind, and World Dialogically*. Information Age Publishing, Charlotte, NC, USA.
- [30] Diana Löffler, Nina Schmidt, and Robert Tscharn. 2018. Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 334–343. <https://doi.org/10.1145/3171221.3171261>
- [31] Lorenza Mondada. 2006. Video recording as the reflexive preservation-configuration of phenomenal features for analysis. In *Video Analysis: Methodology and Methods Qualitative Audiovisual Data Analysis in Sociology*, H. Knoblauch, J. Raab, H.-G. Soeffner, and B. Schnettler (Eds.). Lang, Bern, Switzerland. http://www.ispla.su.se/iis/Dokument/Mondada_Video5diff.pdf
- [32] Lorenza Mondada. 2016. Conventions for multimodal transcription. https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf
- [33] Lorenza Mondada. 2019. Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics* 145 (2019), 47–62. <https://doi.org/10.1016/j.pragma.2019.01.016>
- [34] Hannah R.M. Pelikan and Mathias Broth. 2016. Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4921–4932. <https://doi.org/10.1145/2858036.2858478>
- [35] Anssi Peräkylä, Pentti Henttonen, Liisa Voutilainen, Mikko Kahri, Melisa Stevanovic, Mikko Sams, and Niklas Ravaja. 2015. Sharing the Emotional Load: Recipient Affiliation Calms Down the Storyteller. *Social Psychology Quarterly* 78, 4 (2015), 301–323. <http://spq.sagepub.com/content/78/4/301.short>
- [36] Anssi Peräkylä and Johanna Ruusuvuori. 2012. Facial expression and interactional regulation of emotion. In *Emotion in interaction*, Anssi Peräkylä and Marja-Leena Sorjonen (Eds.). Oxford University Press, Oxford, 64–91.
- [37] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.
- [38] Danielle Pillet-Shore. 2018. Arriving: Expanding the Personal State Sequence. *Research on Language and Social Interaction* 51, 3 (2018), 232–247. <https://doi.org/10.1080/08351813.2018.1485225>
- [39] Anita Pomerantz. 1984. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In *Structures of Social Action: Studies in Conversation Analysis*, John Heritage J. Maxwell Atkinson (Ed.). Cambridge University Press, Cambridge, U.K., Chapter 4, 57–101.
- [40] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages. <https://doi.org/10.1145/3173574.3174214>
- [41] Robin Read and Tony Belpaeme. 2014. Situational Context Directs How People Affectively Interpret Robotic Non-linguistic Utterances. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI '14)*. ACM, New York, NY, USA, 41–48. <https://doi.org/10.1145/2559636.2559680>
- [42] Robin Read and Tony Belpaeme. 2016. People interpret robotic non-linguistic utterances categorically. *International Journal of Social Robotics* 8, 1 (2016), 31–50. <http://dx.doi.org/10.1007/s12369-015-0304-0>
- [43] Elisabeth Reber. 2012. *Affectivity in Interaction: Sound Objects in English*. John Benjamins, Amsterdam / Philadelphia. <https://www.benjamins.com/#catalog/books/pbns.215/main>
- [44] Johanna Ruusuvuori. 2012. Emotion, affect and conversation. In *The Handbook of Conversation Analysis*, Jack Sidnell and Tanya Stivers (Eds.). Wiley-Blackwell, Oxford, U.K., Chapter 16, 330–49.
- [45] Martin Saerbeck and Christoph Bartneck. 2010. Perception of Affect Elicited by Robot Motion. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI '10)*. IEEE Press, Piscataway, NJ, USA, 53–60. <http://dl.acm.org/citation.cfm?id=1734454.1734473>
- [46] Jelle Saldien, Kristof Goris, Bram Vanderborght, Johan Vanderfaeillie, and Dirk Lefeber. 2010. Expressing emotions with the social robot Probo. *International Journal of Social Robotics* 2, 4 (2010), 377–389.
- [47] Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W. McOwan, and Ana Paiva. 2011. Automatic Analysis of Affective Postures and Body Motion to Detect Engagement with a Game Companion. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI '11)*. ACM, New

- York, NY, USA, 305–312. <https://doi.org/10.1145/1957656.1957781>
- [48] Bob R. Schadenberg, Dirk K. J. Heylen, and Vanessa Evers. 2018. Affect bursts to constrain the meaning of the facial expressions of the humanoid robot Zeno. In *Joint Proceedings of the Workshop on Social Interaction and Multimodal Expression for Socially Intelligent Robots and the Workshop on the Barriers of Social Robotics take-up by Society*, Christiana Tsiourti, Sten Hanke, and Luis Santos (Eds.). CEUR, 30–39.
- [49] Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: A primer in Conversation Analysis, Volume 1*. Cambridge University Press, Cambridge.
- [50] Sichao Song and Seiji Yamada. 2017. Expressing Emotions Through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, USA, 2–11. <https://doi.org/10.1145/2909824.3020239>
- [51] Marja-Leena Sorjonen and Anssi Peräkylä. 2012. Introduction. In *Emotion in Interaction*, Anssi Peräkylä and Marja-Leena Sorjonen (Eds.). Oxford University Press, Oxford, 3–15. <https://doi.org/10.1093/acprof:oso/9780199730735.003.0001>
- [52] Stefan Sosnowski, Ansgar Bittermann, Kolja Kuhnlenz, and Martin Buss. 2006. Design and Evaluation of Emotion-Display EDDIE. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3113–3118. <https://doi.org/10.1109/IROS.2006.282330>
- [53] Lucy A. Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, Cambridge, UK.
- [54] Lucy A. Suchman. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, Cambridge, UK.
- [55] Beatrice Szczepek-Reed. 2014. Prosodic, lexical and sequential cues for assessments with German süß: Assemblages for action and public commitment. In *Prosodie und Phonetik in der Interaktion / Prosody and Phonetics in Interaction*, Dagmar Barth-Weingarten and Beatrice Szczepek Reed (Eds.). Verlag für Gesprächsforschung, Mannheim, 162–186. <http://verlag-gespraechsforschung.de/2014/pdf/szczepek-audio.pdf>
- [56] Sue Wilkinson and Celia Kitzinger. 2006. Surprise As an Interactional Achievement: Reaction Tokens in Conversation. *Social Psychology Quarterly* 69, 2 (06 2006), 150–182. <https://doi.org/10.1177/019027250606900203>
- [57] Wendy Wilutzky. 2015. Emotions as pragmatic and epistemic actions. *Frontiers in Psychology* 6 (2015), 1593. <https://doi.org/10.3389/fpsyg.2015.01593>
- [58] Allison Woodruff, Margaret H. Szymanski, Rebecca E. Grinter, and Paul M. Aoki. 2002. Practical Strategies for Integrating a Conversation Analyst in an Iterative Design Process. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '02)*. Association for Computing Machinery, New York, NY, USA, 255–264. <https://doi.org/10.1145/778712.778748>
- [59] Keiichi Yamazaki, Akiko Yamazaki, Mai Okada, Yoshinori Kuno, Yoshinori Kobayashi, Yosuke Hoshi, Karola Pitsch, Paul Luff, Dirk vom Lehn, and Christian Heath. 2009. Revealing Gauguin: Engaging Visitors in Robot Guide's Explanation in an Art Museum. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1437–1446. <https://doi.org/10.1145/1518701.1518919>

A TRANSCRIPTION SYMBOLS

A.1 Transcription symbols for verbal interaction, adapted from [23]

COZ	Participant who is performing the turn at talk
((comment))	Transcriber's descriptions
(0.2)	Timed pause in tenths of seconds
[a]	Overlapping talk
=	Latching of utterances, no interval or overlap
a:	Lengthening of sound
A	Utterance louder than surrounding talk
°a°	Utterance softer than surrounding talk
<u>a</u>	Stress through pitch and/or amplitude
>a<	Utterance speeded up
?	Rising intonation
,	Continuing intonation
-	Cut off
↑	Rise in intonation of next syllable
↓	Drop in intonation of next syllable
(h)	Hearable aspiration, often associated with laughter
.h	Hearable inbreath
h	Hearable outbreath
.pt	Lip smack
ha/he/hi	Laughter
(a)	Transcription in doubt

A.2 Transcription symbols for embodied conduct, adapted from [32]

coz	Participant who is performing the action
im	Indicates exact time point of screen shot (image)
#	Indicates position of screen shot within turn at talk
& &	Delimit description of Cozmo's embodied actions
+ +	Delimit description of participants' gestures and actions
* *	Delimit description of participants' gestures and actions
% %	Delimit description of participants' gestures and actions
+-->	Action continues across subsequent lines
-->+	Action continues until symbol is reached
>>	Action begins before start of the transcript
-->>	Action continues until after end of the transcript
...	Preparation of gesture
, , ,	Retraction of gesture