Social Interaction

Video-Based Studies of Human Sociality

_____

When a Robot Comes to Life:
The Interactional Achievement of Agency as a Transient Phenomenon

_____

Hannah Pelikan, Mathias Broth & Leelo Keevallik

Linköping University

Abstract

Conceptualizing agency is a long-standing theoretical concern. Taking an ethnomethodological and conversation analytic perspective, we explore agency as the oriented to capacity to produce situationally and sequentially relevant action. Drawing on video recordings of families interacting with the Cozmo toy robot, we present a multimodal analysis of a single episode featuring a variety of rapidly interchanging forms of robotic (non-)agency. We demonstrate how agency is ongoingly constituted in situated interaction between humans and a robot. Describing different ways in which the robot's statuses as either an agent or an object are interactionally embodied into being, we distinguish "autonomous" agency, hybrid agency, ascribed agency, potential agency and non-agency.

Keywords: human-robot interaction, social robotics, agency, autonomy, sequence organization

1. Introduction

"We needed people to understand that that robot has agency, that it is a living thing."
Ben Gabaldon, Lead sound designer for the Cozmo robot

Agency, loosely understood as the capacity to act, has been widely discussed as a theoretical concept, with researchers exploring whether it is an individual (Davidson, 1980) or a distributed capacity (Latour, 2005; Enfield, 2013; Goffman, 1981). Much of this work is highly theoretical in nature and largely lacks empirical specification. In this paper, we take an ethnomethodological conversation analytic (EMCA) perspective to rethink agency as a members' phenomenon (Garfinkel, 1991). We are particularly interested in how agency manifests itself and how transitions from agency to non-agency are achieved by the interacting parties. We therefore study a "perspicuous setting" (Garfinkel, 2002: 181-182), where agency is not always taken for granted: humans interacting with a small toy robot—a plastic object that may only occasionally be treated as an agent. We propose concrete ways of analyzing agency in human-robot interaction, and in doing so we see various forms of agency, which are much more fine-grained than approaches outside EMCA have mapped so far. We demonstrate how the analytical procedure of sequential analysis contributes a practical operationalization of agency, also when concerning robots, that enables us to trace how agency emerges and changes during the course of an interaction.

Robots are frequently portrayed as autonomous agents in marketing discourse and the media: In movies such as *StarWars*, *Ex Machina*, and *Wall-E*, robots build complex relationships with humans and other robots. Companies pitch robots that autonomously drive passengers and goods to their destination, assist with household tasks and keep humans company. To manifest their status as social agents, many robots, including digital pets and toys, are physically designed with features such as a face and arms that suggest capacities to perceive and perform interactional moves (cf. Breazeal, 2002). When it comes to acting in the world, however, robots often lose their magic as they need to be continuously scaffolded by humans (Forlizzi & DiSalvo, 2006), not only in relation to programming and charging their batteries but also placing them in an appropriate environment and removing obstacles from their way, for example. While research has long since established that computers may indeed be treated as social agents (see, e.g., Reeves & Nass, 1996), Alač (2016) and Fischer (2021) remind us that robots remain things, and that their agent status may change dynamically. Human-robot interaction studies tend to take for granted that humans will treat the robot as an agent, and that this agent status is relatively stable. Thereby, the discussion often omits that machines may or may not be treated as agents in situated interaction.

Theorizing agency as a member's phenomenon, our work builds on a *dialogic* perspective (Linell, 2009). A growing body of work can be seen to take such a perspective, challenging a simplified and *monologic* (Linell, 2009) view of robots and other artificial agents as simple message broadcasters. Instead, dialogic theories see interaction with agents as a collaborative endeavor. Studies in human-agent interaction (HAI), human-robot interaction (HRI) and dialogue systems (SIGDIAL) criticize turn-taking models that resemble walkie-talkie interaction (see discussion in Skantze, 2021) and instead advocate dialogue systems that enable more dynamic, reflexive, and incremental interaction. Influenced by Lucy Suchman's (1987) ground-breaking work on situated interaction with copying machines, researchers are working towards more contextual awareness (see, e.g., Bohus & Horvitz, 2011; Hedayati, Szafir & Andrist, 2019; Otsuka et al., 2006). Current work is increasingly receptive to the idea that meaning is interactionally negotiated and accomplished (see, e.g., Traum, 1994; Axelsson et al., 2022; Jung, 2017).

From a dialogic perspective, "individual selves cannot be assumed to exist as agents and thinkers before they begin to interact with others and the world" (Linell, 2009:44). Agency never exists in a vacuum but always emerges in interaction with others. This perspective is particularly useful when studying robots, since it essentially takes a participatory stance (Asaro, 2000; Greenbaum & Kyng, 1991; Bødker et al., 2021) and focuses on how a robot is treated by its specific users rather than *a priori* assigning a label to its status. EMCA work that focuses on mundane interactions has started to shed light on how robotic agency is accomplished in everyday life (Alač, 2016; Krummheuer, 2015a) and demonstrates that it is interactionally constituted and that a robot may dynamically go from being treated as a thing to being treated as an agent (Alač, 2016). Our study aims to further this line of research and to show how robotic agency emerges in a variety of configurations.

We have studied a toy robot named Cozmo as it interacts with humans over the course of a 60 second episode, which was recorded in a family home. Our central concern is: How do humans attribute agency to Cozmo, and how do we best conceptualize this agency? By analyzing unfolding interaction in minute multimodal detail, we show how various forms of robotic agency are locally and interactionally embodied into being (cf. Heritage, 1984: 290 on how institutions are "talked into being") and oriented to by the participants in a manifest way. Our analysis identifies five different forms of robotic (non-)agency: "autonomous" agency, hybrid agency, ascribed agency, potential agency, and non-agency. Scrutinizing whether and how human participants treat a robot as having a capacity for action in real world interaction, we offer an empirical basis for theorizing agency in interactionally relevant ways. We also contribute to a rethinking of what designing "autonomous agents" means.

2. Perspectives on agency

Agency is a widely studied concept in different disciplines, including philosophy, cognitive science, psychology, sociology, and robotics. Generally, theories of agency can be divided into two groups: those which focus on delineating agency as an individual (typically human) capacity (e.g., Davidson, 1980; Dennett, 1987; see also discussion in Suchman, 2007), and those which highlight that agency is distributed across people and artifacts. Distributed agency has been theorized in a variety of ways, most notably in Actor-Network Theory (Latour, 2005) which suggests that agency is always mutually constituted and describes how "actants", that is anything that can influence action, emerge through the relationships between both humans and their environment. This view is also present in theories of distributed cognition (Hollan et al., 2000) and joint cognitive systems (Hollnagel & Woods, 2005). In terms of agency of speaking, Goffman has famously suggested that the roles of *animator, author*, and *principal* may be distributed across several individuals (Goffman, 1981; see Enfield, 2013, 2014 for a critical discussion). From this perspective, robot agency can be regarded as generally distributed between the designer(s) (the author and principal (responsible) of what a robot will say and do), and the robot (the animator of those actions in real time).

While not denying that agency can be distributed, ethnomethodological research focuses on how society, activities, and actions are accountably and reflexively produced by members, thus targeting practice rather than theory. It has, for instance, problematized the loose application of the term *actor* for larger systems or institutions, demonstrating instead how specific individual actions-in-interaction bring these about as a lived reality (cf. Button & Sharrock, 2010). Following this line of thinking, we consider that objects could only have *agency* if they are demonstrably treated as agents by participants in a particular event (see, e.g., Due, 2021b).

2.1 Agency: Contributing meaningful actions

In a widely used encyclopedia of philosophy, an *agent* is defined as "a being with the capacity to act, and 'agency' denotes the exercise or manifestation of this capacity" (Schlosser, 2019:1). In her seminal work on human-machine interaction, Lucy Suchman (2007:2) similarly treats agency as "capacity for action". Across and within academic disciplines, researchers may disagree on what *action* is and who or what can be an agent, even though many seem to agree that behavior must be in some way meaningful or relevant for achieving a goal. In an influential interactionist account (Enfield 2013, 2014), agency has been described as involving both the degree of flexibility in controlling the choice between different behaviors and the accountability for one's actions (i.e., being prepared and able to provide explanations for them). In philosophical discussion, the ability to contribute not random but

meaningful actions is also regarded as crucial, but motivated differently: Agency is typically treated as the (individual) capacity to perform actions that are intentional, that is, *directed to* or *about* something, motivated for example by a belief or will (cf. Davidson, 1980; Dennett, 1987; Bratman, 1987; see Schlosser, 2019 and Wilson & Shpall, 2016 for an overview). Social robotics research has taken up the idea of intentionality and strives to make robot actions (appear as) intentional (see, e.g., Breazeal, 2002).

Linell's (2009) distinction between monologic and dialogic theories is useful in discussions about intentionality. From a monologic perspective, one needs to distinguish between a twitch of the eye versus closing the eye to wink at someone on the abstract basis of whether the move was intentional or not: Traditionally, only intentional, goal-driven moves are seen to involve agency. From the dialogical EMCA perspective, which is generally agnostic about mental states, a partner in interaction may treat some piece of behavior as intentional or not and in turn their interpretation can be challenged, for example, by claiming that the wink was in fact a twitch of the eye. An action is thus not brought about by an individual alone, but behavior is understood as some action only when a dialogical "other" recognizes it as relevant and accountable (Garfinkel, 1967), that is, meaningful. Note that EMCA does not oppose the concept of intentionality *per se* but rather does not see it as a constitutive element of action. When accounting for one's own or others' actions in everyday life, people may indeed refer to intentions and mental states (Coulter, 1979; Edwards, 1997; Sacks, 1992; Weatherall & Keevallik, 2016; Broth et al., 2019), but observing and describing these verbal practices are not the same thing as postulating "invisible" intentions that are causing observable behavior. Along the same lines, research on machines has repeatedly documented that people explain machine behavior through the *attribution* of beliefs and other intentional states to machines (Thellman & Ziemke, 2019; Parenti et al., 2021).

2.2 Autonomy: Who is acting in asymmetrical interaction?

Agency as a dialogical achievement is closely related to the question of who acts autonomously and what kinds of asymmetries are at play in different contexts. EMCA research has discussed agency in a range of settings where asymmetries between participants become particularly relevant, for instance settings involving representatives of various potentially disadvantaged groups, such as children (Sterponi, 2003, Demuth, 2021), students (Waring, 2011), crime victims (Weatherall, 2020), people with disabilities (Mikesell, 2010; Antaki & Crompton, 2015), and people in need of translations (Harjunpää, 2012; Warnicke & Broth, forthcoming). A stream of studies has targeted the grammatical formatting of turns that accomplish variable degrees of agency (Curl, 2006; Heritage & Raymond, 2012; Clayman, 2013;

Keevallik, 2017), while others have dissected various institutional practices that might curb the agency of the subjects concerned, most prominent among them being the medical establishment (Peräkylä, 2002) and the courtroom (Licoppe, 2021). Although agency may be very asymmetrically distributed between pets and their owners, even dogs manifest agency when they perform actions that fit the context of a specific moment (Laurier et al., 2006). In the current study, we will likewise look at a different and somewhat challenged party in interaction and their possible degrees of agency – a robot.

The idea that a speaker can autonomously choose what to do and then carry it out by themselves has been argued to be a Western ideal. Classic as well as recent work by Goodwin (2004) and Auer, Bauer and Hörmeyer (2020) has demonstrated that this ideal is so strong that it is upheld even in interaction with people whose speaking abilities are heavily restrained: Co-participants support them in such a way that they remain in control, at least as principals (in Goffman's terms) of their utterances (Goodwin, 2018). Auer and colleagues (2020) argue that if the interactional ideal is making speakers appear autonomous, it may be easy to overlook the work that co-participants engage in to uphold this impression. Similarly, in robotics *autonomy* is usually defined as the absence of human assistance, for example, that a robot can "perform intended tasks based on current state and sensing, without human intervention" (International Organization for Standardization [ISO], 2021:3.2). While such phrasing may suggest that no human is involved, roboticists highlight that this only needs to hold for a limited amount of time (cf. Bekey, 2005:1-2). Even at the highest level of autonomy, a human operator still initiates sequences of actions that are then performed without intervention (see Sheridan & Verplank, 1978:8–17ff). More recent work further problematizes the traditional view of autonomy, highlighting that no robotic system can ever be fully autonomous (i.e., self-sufficient and self-directed) (Bradshaw et al., 2013). A report by the Defense Science Board [DSB] suggests that "[i]nstead of viewing autonomy as an intrinsic property of an unmanned vehicle in isolation, the design and operation of autonomous systems needs to be considered in terms of *human-system collaboration*" (DSB, 2012:1-2). In other words, autonomy in robotics can be understood in the sense of 'no human intervention *in an ongoing action sequence*' but does not deny that at some level there is always a human involved. As long as robots need to be switched on, set to do certain tasks (see Alač et al., 2011), and assisted by moving objects around them (see Forlizzi & DiSalvo, 2006; Lipp, 2019), robots cannot be considered fully autonomous. This often seems to be overlooked in public discourses about autonomy. By scrutinizing the collaborative work by humans that is involved, our work elucidates the different forms of agency a machine can embody in interaction with humans.

2.3 Prior EMCA work on machine agency

By studying interaction in mundane settings, a small body of EMCA work has begun to shed light on how machine agency is accomplished on a moment-by-moment basis. Krummheuer (2015a) shows that agency is not a static feature but emerges over the course of interaction, depending on how humans treat a particular technical object: A conversational agent may be oriented to as possessing varying degrees of competence, and a walking aid may be talked about as being "alive" when it opposes ongoing action. Alač, Movellan and Tanaka (2011) studied how a robot's social agency is enacted in an extended laboratory when a robot meets a group of toddlers. Alač and her collaborators underline the role of the roboticists who are always co-present in achieving the social character of the robot. In a follow-up study, Alač (2016) argues that to fully understand robot agency it is important to also study moments in which the robot is treated as a thing. Alač (2016) demonstrates how the status of the robot dynamically changes and how both its agency and its materiality (as a thing with an inscribed culture) are achieved in interaction. While Fischer (2021) focuses on anthropomorphizing behavior rather than agency, she also demonstrates how humans may respond to robots as if they were humans in one moment and treat the robot like a machine in the next.

In this paper, we delve further into the variable dimensions of machine agency and its temporal development as humans move between treating a robot as an interaction partner or a physical object. We thereby show in detail how agency is dynamically achieved and negotiated through human interaction with the robot.

3. Data and method

This study is based on video data of people interacting with a palm-sized Cozmo robot in their homes in the context of a field experiment (see Pelikan et al., 2020 for details). In addition, the first author conducted a 45-minute interview with Ben Gabaldon, Cozmo's sound designer.

Cozmo was initially developed by Anki (now Digital Dream Labs) and marketed as a toy for children between the ages of 8–14. This robot does not have speech recognition and mainly interacts through beeps, movements, and animated eyes. It has forklift arms that can be raised and lowered, and wheels that enable mobility in space (see Figure 1, left). It is switched on and controlled through a smartphone app that offers different modes of interaction, such as letting the robot roam freely and playing games. The phone functions as a control interface, through which the robot can either be remote controlled, programmed, or started off in the execution of autonomous sequences. The activity studied in this paper is of the latter type: After the user selects the "Meet Cozmo" activity and enters a name to be learnt (Figure 1,

top right), the app switches to a loading screen with an instruction for how a user should position themselves to have their face scanned and learnt by the robot (Figure 1, bottom right, after "Name" has been entered). This screen remains the same for at least 20 seconds. Through its *face recognition* component Cozmo can match names typed into the app with scanned human faces and later greets the people with their names (see Extract 1).

Figure 1. *Left: Cozmo robot with its charging station, the app needed to control and switch it on and the robot's touch sensitive toy cubes. Right: screenshots of the app interface before and after entering the name*
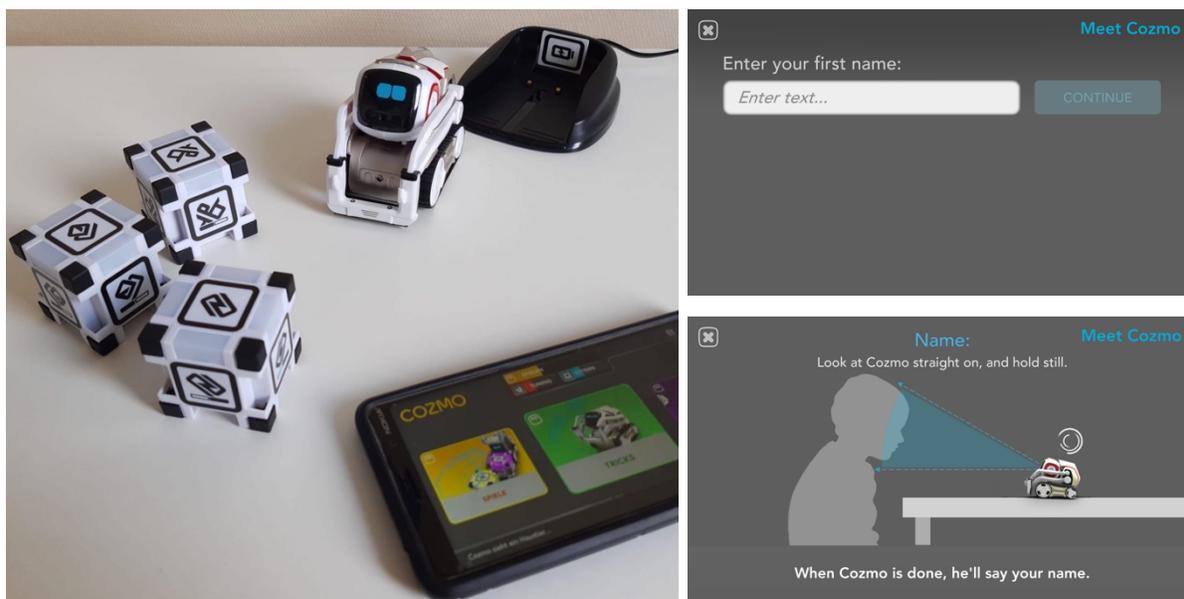
Data was collected in Sweden from six families who volunteered to host the robot. Participants were recruited via personal networks and local social media groups and kept the robot for 1–2 weeks. To participate in the study, at least one family member had to be aged 8–14, the group which the robot is marketed at, but all family members were encouraged to interact with the robot. In the episode used in this study, Oskar, who is in his early teenage years, was the main participant, but his mother, his older sibling Selma, and her partner Niklas also interacted with the robot. Oskar installed the app on his smartphone, and the family recorded themselves with a single camcorder on a tripod provided by the researcher. All participants agreed to be recorded and adults have given consent for their and their children's anonymized data to be used in this article. All names have been changed.

The EMCA method used in this study involves paying close attention to minute details of behavior as they unfold in real time (Goodwin, 2018; Mondada 2019). By

observing how participants demonstrably orient to previous actions, our analyses aim to reflect the participants' own perspective (i.e., an emic perspective) on what is happening. We have scrutinized a single episode and followed the verbal and bodily behavior of the humans as well as Cozmo, and we explore the extent to which portions of this behavior play out regarding the accomplishment of various forms of robotic agency. Unpacking a single case enables us to provide all the relevant details and to trace the subtle changes in agency on a moment-by-moment basis.

4. Analysis: Shifting forms of agency in human–robot interaction

The focus episode involves Oskar, his older sister Selma with her boyfriend Niklas, and Cozmo, which Selma and Niklas meet for the first time. They are sitting on a sofa around a coffee table on which Cozmo is placed (as shown in Images 1.1 and 1.2 below, where the greater part of the image is visually modified for anonymizing purposes). The video snippet stems from a longer recording that starts when the robot is already switched on.

The analytical section presents four different extracts from the same episode, and each extract uses the episode's continuous line numbering (see Appendix for full transcription). The first two cases show moments when the participants orient to the robot as an autonomous agent (section 1) and as contributing to human-robot hybrid agency (section 2). Our third case demonstrates how humans ascribe agency to robotic behavior (section 3), whereas the fourth case (section 4) involves participants testing the robot's potential for agency and inspecting it as an object for which agency is not a relevant feature.

4.1 "Autonomous" agency: Greeting sequence

As our first point, we will show how Cozmo succeeds in accomplishing an action in a sequence. The human participants orient to the action as sequentially implicative, and thus retrospectively cast the robot as an accountable agent and a participant in the interaction. Before Extract 1 starts, Niklas has had his face scanned and his name fed into the Cozmo app on Oskar's phone. This means that Niklas' name is displayed in the phone interface when he appears in Cozmo's field of view. In the first transcript line, Niklas gazes at and points to the phone and formulates as a noticing that he has been "recognized" by Cozmo (l. 01). We are interested in what happens immediately following this.

Extract 1. **Niklas**

```
01 NIK    ni*klas, >de* kän%de ig^en# mej.<
         Niklas     it recognized me
   nik    »-- gaze app--> (l.03)
   nik                    %,,,,,,--,,,,--> ((points to app))
   osk     *,,......,* ((gazes from app to coz to app))
   sel    »-- gaze to coz        ^,,..app -->l.04
   img                           #img1.1
```


Image 1.1. Oskar, Niklas and Selma with Cozmo, its toy cubes as well as the smartphone that controls the robot through the Cozmo app. Niklas is pointing at the phone.

```
02 COZ    +/KLICK/*  + (.)%
   coz    +arm strike+
   nik             -->%
   osk          *   ,,,--> ((gazes to coz))


03 COZ    >ad*eo dae-eo< dA+o dE+o* A-Ao NI%Kl*[as].
04 NIK                                  [ja]:#, he*^j.
                                        yeah   hello
   coz                 +forw+ ((moves toward nik))
   osk    ,,,*......coz.........,,*gaze to app*,,,.coz.,,*.. ((gazes to app))
   nik                            %,,,.............--> ((gaze on coz))
   img                            #img1.2
```


Image 1.2. Niklas gazes at Cozmo.

The focus sequence begins by Cozmo producing a sound animation (l. 03) and rolling straight toward Niklas and facing him. This animation is immediately followed by the production of a name, *Niklas* (l. 03). When the robot has started to say his

name, Niklas turns his head, gazes at Cozmo and responds 'yeah, hello', (l. 04, Image 1.2). Niklas thereby treats Cozmo's contribution as a first greeting that makes relevant a response greeting with the robot as the appropriate recipient. Cozmo's first greeting is a multimodal and interactive accomplishment, as it is acknowledged (and interactionally accomplished) by Niklas as this specific type of social action. Furthermore, when the robot is uttering Niklas' name, Oskar also turns to Cozmo (l. 03-04). Oskar thereby orients to this ongoing action as relevant, as he is otherwise mostly engaged with the app on the phone. The success of Cozmo's move also relies on the fact that it is a reasonably fitted action at a moment when Niklas has just discovered that Cozmo can recognize faces. Through this greeting sequence, Cozmo emerges as an "autonomous" agent who, quite literally, comes to life (even though human activities such as programming and switching the robot on are of course scaffolding that "autonomy"). This is also witnessed as something of a spectacular event in Selma's affect-laden *NE::e:j* 'no way' (l. 05). Her gaze towards the phone during the greeting sequence and then briefly to her brother (l. 05) underlines that her reaction is not directed at Cozmo – it is thus not part of the greeting sequence but rather a comment on its success.

This relatively straightforward paired action sequence, where Cozmo greets Niklas and Niklas produces a return greeting, is preceded and followed by qualitatively different configurations. Just prior to the first greeting, Niklas refers to Cozmo as *de(n)* 'it' (l. 01), an inanimate object, while gazing at the app on the mobile phone. Even though Niklas has commented on him being "seen" in this turn designed for co-present humans, he does not respond to whatever he discovers in the app before Cozmo addresses him by uttering his name. Likewise, immediately after completing the greeting sequence, Niklas returns his gaze to the phone (l. 05), thus not displaying any expectation that Cozmo will continue the interaction with him. Instead, Niklas seems to be looking for further clues "behind the scenes" in the app while Cozmo remains idle. Talk continues about the robot, but it is not directed to it (line 05 and onwards). Summing up, we can thus observe that Niklas' verbal and embodied response treats Cozmo's action as a meaningful sequence initiation that requires a fitting next action, which he also produces. Admittedly, the timeframe during which Cozmo thereby emerges as an "autonomous" agent is brief, as it is restricted to only the two core parts of the greeting adjacency pair. While previous work has likewise argued that machine agency is a joint accomplishment (Alač et al., 2011, Alač, 2016, Krummheuer, 2015a), we show that it is thoroughly transient: It is embodied into being by finely coordinated orientations toward the robot and quickly dissolves again as participants shift their orientations elsewhere. Based on several similar sequences (see, e.g., Extract 2c, l. 45-46, Extract 3, l. 47, 49, 51), we suggest, in line with the conversation analytic mentality (Schenkein, 1978), that a robot emerges as an "autonomous" agent when it initiates and performs an action

that is treated as relevant and accountable by human participants, as evidenced in their next actions (cf. Sacks, Schegloff & Jefferson, 1974:45).

4.2 Hybrid agency: Orchestrating a prank via the robot

A few moments later, the robot is used as a partially teleoperated machine that accomplishes actions launched by a human. Rather than treating the robot's action as autonomous, it is now oriented to as the hybrid product of both human and robotic agency.

Between Extract 1 and Extracts 2a–c below, Oskar has offered to add his sister Selma, via the app on his phone, to the list of faces that Cozmo will recognize. He first needs to type her name[1] into the app interface, and as an indication of successful learning of her face, Cozmo will then read it out. As Extract 2a starts, Oskar has just finished typing (l. 20) and presses the "continue" button (Figure 1, top right), which initiates Cozmo's scanning procedure and switches to the loading screen (Figure 1, bottom right). After the activity has been launched via the phone, Cozmo will perform a number of actions during which the app screen remains the same.

---

[1] Most participants use their first name, although some use a family category ("mama") or a nickname ("the queen"), so in principle anything can be typed here.

Extract 2a. What did you write? Launching of the prank

```
20 OSK    *nh                    *nh ((smiles))
   osk    *presses button in app*

21        ¤ (0.3)¤  (0.2)   %¤*(0.1)^¤(0.2)%¤(0.1)
   osk    ¤......¤gaze to nik¤,,,,,,,¤coz-,,¤sel-->
   nik                      %smiles--------%
   osk                       *...-->
   sel                              ^gaze at osk-->

22 SEL    *vad skre¤v^  d+u?*
          what did you write?
   osk    *turns coz to sel-*
   osk         -->¤looks down-->
   sel          -->^,,,-->
   coz              +turns left-->

23        ¤(0.1)^¤  (0.1)    ¤(0.1)¤(0.2)+(0.4)*(0.2)
   osk    ¤......¤glance sel¤,,,,,¤coz-->
   sel    ,,,,,,^gazes at coz-->
   coz                          -->+
   osk                                 *turns coz to sel-->

24 OSK     n*(h) ((smiles))
   osk    -->*,,,-->

25        (0.3)¤(0.1)Δ¤  (0.3) ¤(0.1)*(0.2)¤(0.2)
   osk      -->¤,,,,,,¤gaze sel¤,,,,,,,,,,,¤gaze phone-->
   sel            Δ...raised eyebrows-->
   osk                        -->*

26        +^  (0.4)  +(0.1)
   coz    +drives fwd+
   sel     ^gazes at phone-->

27 COZ    OA:^:+oΔ^        +de-e ((face learning sequence starts))
   sel    -->^   ^gazes at COZ-->l.13
   coz         +lifts arms-+
   sel         -->Δ,,,-->

28 OSK    kolla raΔkt in i cozmos ögon
          look right into Cozmo's eyes
   sel    ,,,,,,,,Δ
```

When pressing the last button in the app, Oskar produces two laughter particles (loud nasal outbreaths) (l. 20). He briefly exchanges a glance with Niklas (l. 21), begins to smile and then repositions the robot to face Selma (l. 21-25). Selma seems to notice her brother's smile and, while turning to Cozmo, asks him *vad skrev du* 'what did you write' (l. 22). Oskar does not respond (l. 23), but after setting up Cozmo in a suitable position for scanning Selma's face, he produces another laughter particle (l. 24). Selma does not repeat her question but instead continues to look at Cozmo, raises her eyebrows (l. 25), and briefly glances at the phone in Oskar's hands (l. 26). At this very moment, Cozmo plays the first sound in the face learning sequence (l. 27). Oskar now instructs his sister with a firm voice *kolla rakt in i Cozmos ögon* 'look right into Cozmo's eyes' (l. 28). Successful face scanning is contingent on people facing the robot and sitting still, so Selma's cooperation is one of the aspects that make Cozmo's name-learning possible in the first place.

Skipping a few lines, we resume the analysis at the end of the scanning process, when Cozmo utters what Oskar typed into the interface.

Extract 2b. Se-ma. The performance of the prank

```
((7 seconds omitted))

38        (0.4)

39 COZ    ¤see*#-maaa?
   osk    ¤..gazes sel-->
   osk         *smiles-->
   img            #img2.1
```



```
Image 2.1. Oskar monitors Selma's face while Cozmo utters "Se-ma"
```

```
40        ¤(0.2)%(0.1)^¤
   osk    ¤............¤gazes to phone-->
   nik         %...........-->
   sel             -->^gazes to phone-->

41 OSK    S%E¤^MA[ΔA    ^ah^e][he#he%hehehe]
42 COZ          [oaaaow    ]
43 SEL              [ m(h)%   ehe] he% he
   nik    .%gazes at sel-----------%..........%coz-->
   osk    -->¤.gazes at sel-->>
   sel     -->^..gz osk^--^((blinks))gazes at cozmo-->>
   sel         -->Δsmiles-->>
   img                        #img2.2
```



```
Image 2.2. Selma joins in with her brother's laughter.
```

Instead of *Selma*, the robot utters *Se-ma* (l. 39) – a distorted version of her name – with rising intonation. Oskar, monitoring his sister's face (see Image 2.1), now

repeats what the robot said ("Se-ma") and starts laughing (l. 41). Oskar and Selma look at each other and Selma joins in with her brother's laughter (l. 43, Image 2.2). While still laughing, Selma gazes back toward Cozmo, which has just uttered a sound resembling "wow" (l. 42) and now continues with its face learning script by repeating the name a second time, this time with falling intonation (l. 45). The continuation of the sequence is shown in Extract 2c:

Extract 2c. Continuing the prank

```
44 SEL    [heha]
45 COZ    [se -]maaa.

46 SEL    he he he+he [°hej°=]+
          ha ha ha ha   hi
47 COZ                [>adeo]+[dae-eo<=      ]+ /KLICK/  +
48 OSK                        [=det låter som]+#
                               it sounds like
   coz              +fork up---+"waves"--------+fork down +
   img                                     #img3.1


49 COZ    ^=da[o       d[eo      ]a-a][o  ]
50 OSK        [som han sä[ger (.) se]nap]
              like he says mustard
51 SEL                  [he-   hej]    [nej] vänta
                         he-  hello     no wait
```

While Cozmo is continuing with its pre-scripted happy animation (l. 47), Oskar uses the repetition as an opportunity to further highlight the robot's performance as a laughable, suggesting that it sounds as if Cozmo said *senap*, the Swedish word for mustard (l. 48–50).

In Extracts 2a–2c, a human participant sets up and scaffolds an activity in which Cozmo is partly teleoperated and carries out a specific sequence of actions. Oskar, who is already familiar with the robot and in control of the Cozmo app on his phone, is orchestrating a prank through the robot – calling his sister by a distorted name (either by changing the spelling or just relying on the relatively exotic name coming out wrong in the English speech generator). By launching a prank via the phone, Oskar manifests himself as the one who has ultimate control over what the robot will do, while simultaneously relying on the robot's independent capability to enact specific recognizable behaviors within a certain timeframe. Although Cozmo is carrying out actions that are pre-planned and launched by a human *in situ* (see also Suchman, 1987), the robot is more flexible (Enfield, 2013) than a wind-up doll, a dishwasher or remote-controlled toy (which tend to be treated as artifacts that extend human actions rather than as agents on their own). Oskar treats the robot as a collaborator rather than a mere animator of his own action as he repeats Cozmo's pronunciation (l. 41) and further comments on the utterance (l. 48-50), highlighting the fact that part of the performance is beyond his (i.e., the human operator's) control.

The success of the face learning activity launched by the robot and Oskar – together forming a hybrid agent – is crucially dependent on Selma's collaboration because she needs to turn to Cozmo and keep still while Cozmo is scanning her face. Cozmo's action is ultimately made possible by human participants pressing buttons on the phone and keeping still, respectively. By gazing at the phone, at Oskar and subsequently at Cozmo (l. 40 and 43), Selma displays an orientation to the distributed nature of this action as initiated by her brother via the phone, and then produced by the robot. Hybrid agency involves thus not only the launching of actions by a human operator but emerges when others treat the robot–human assemblage as a hybrid agent that jointly achieves the action.

4.3 Ascribed agency: Accounting for lack of action

Another way in which Cozmo is treated as an agent in our episode is through a reference to its current mental state. Mental state accounts of behavior take for granted the capacity to act and place the cause for the behavior in the agent (see, e.g., Dennett, 1987). In Extract 3, which starts in the aftermath of the prank sequence, two of the human participants orient to Cozmo's current inaction as meaning "not wanting" to participate in a projected activity, that is, they treat Cozmo as being endowed with a will of its own that prevents it from acting.

In Extract 3, (where l. 46–51 are repeated from Extract 2c for convenience), we can see Selma verbally completing two consecutive greeting sequences with Cozmo through return greetings (l. 46 and l. 51). Just prior to Selma's second return greeting, Cozmo has raised its forklift arms (l. 47, Image 3.1) to which Selma may be responding when she then gently jabs her fist toward Cozmo (l. 51). As we will see, Selma's jabbing gesture is accountable as an invitation to do a fist bump.

Extract 3. He doesn't want to.

```
46 SEL    he he he+he [°hej°=]+
         ha ha ha ha   hi
47 COZ              [>adeo]+[dae-eo<=      ]+ /KLICK/  +
48 OSK                  [=det låter som]+#
                          it sounds like
   coz              +fork up---+"waves"---------+fork down +
   img                                        #img3.1
```


Image 3.1. Cozmo is raising its forklift arms during the happy animation

```
49 COZ    ^=da[o        d[eo       ]#a-a^][o  ]
50 OSK        [som han sä[ger (.) se]#nap^]
              like he says mustard
51 SEL               [he-   h^ej]#      [nej] vänta^
                      he-    hello       no wait
   sel    ^moves fist toward coz-^jabs---^fist down-up^
   img                        #img3.2
```


Image 3.2. Selma jabs fist.   Image 3.3. Selma moves fist in circles.

```
52       (0.2)

53 SEL    ^hur gör han. (0.2)^(0.2)#
            how does he do it
   sel    ^jabs twice--------^fist in circles x 5-->
   img                        #img3.3
```

Selma may have learned about or observed how Cozmo can participate in fist bumping (not on recording). Her jabbing gestures embody the possibility that Cozmo could produce the appropriate embodied response in return. However, as this is not forthcoming – instead Cozmo remains basically idle – she begins redoing her gesture with small variations, while uttering *nä vänta, hur gör han, hur gör du* 'no

wait, how does he do it, how do you do it' (l. 51–53 and 55). This turn-at-talk, as well as the fist movement variations in the consecutive re-trials (l. 53 to 57), ascribe the responsibility for the robot's failed response not to Cozmo, but to her own way of performing the gesture.

At this point, Oskar takes the turn, stating with authority that Cozmo does not want to engage in a fist bump (l. 54), thereby accounting for the missing sequentially relevant next action "on behalf of" the robot. This does not necessarily mean that Oskar really thinks that Cozmo can have its own intentions. In practice, though, Oskar's account demonstrably orients to, and construes, the robot's current behavior as embodying a refusal of an invitation, that is, it is treated as a dispreferred response (see, e.g., Heritage 1984: 265–269). The upshot of Oskar's account omits details like whether the robot is able to participate in a fist bump sequence or understands what Selma is after, and simply highlights that it does not want to do this right now. Continuing her engagement with the robot by moving her fist (l. 53-57), Selma however insists *men ja vill* 'but I want' (line 56), an utterance that implicitly accepts the premise that the robot does not want to go along with her invitation. She thereby treats Cozmo as the relevant prospective partner in the activity (rather than challenging her brother, who might in fact launch the activity in the app).

In the first two extracts we described how actual robotic behavior was treated as embodying either autonomous or hybrid agency. In this third extract, we have seen how humans, by referring to the robot's current mental state, may likewise cast the robot's behavior as displaying agency even in the complete absence of any relevant responsive behavior.

4.4 Potential agency and non-agency: Capacity testing and inspecting the robot as an object

In contrast to what we have seen so far, a few moments later Cozmo is no longer approached as a participant with agency. In Extract 4, Cozmo is rather dealt with as an object that can be exposed to testing and talked about without any social consideration. In what follows, Niklas and Selma engage in a detailed inspection of Cozmo's constitution.

Extract 4. Here is the little camera

```
65 OSK    =sister (ko^l)la*hah*a
            sister, look here
   osk                   *...*gazes sel-->
   sel             ^gazes mobile, leans fw-->

66         (0.3)+*(0.1)* (0.2)  +(0.1)
   osk       -->*,,,,,*
   coz        +small turn left+

67 OSK    *se%^: ma:%::::#
   osk    *gazes mobile-->
   sel     -->^gazes mobile-->
   nik       %......%points coz-->
   img                 #img4.1
```



```
Image 4.1. Niklas' finger in front of Cozmo's screen|Image 4.2. Cozmo drives
backwards
```

```
68 OSK    a(h)hn^△[hn]
69 SEL           [SE]+-  M[  +A+△:: ]
70 COZ                   [  +d+udi ]#dui
   sel     -->^coz-->
   sel        -->△leans back-----△
   coz             +fl up,bw+-+ ((head up))
   img                         #img4.2

71         (0.2)%(0.4)+(0.2)    +
   nik     -->%turns finger, nail covers screen-->
   coz             +moves fw+

72 COZ    se::-

73         (2.1)

74 SEL    △#mpt a:: [här är        lilla#+△ ka+m%△eran]
            mpt yea: here is the little camera
75 COZ               [>adeo dae-eo< dAo   +dEoA+A%△o  s]e:: ma::::%
   sel    △...............................△points△,.screen-->l.90
   nik                                  -->%,,,,,,,,,,,,,,,,,%
   coz                                  +fwd-+
   img     #img4.3               #img4.4
```

At the beginning of Extract 4 (l. 65–67), Oskar invites Selma to look at something in the mobile interface. While Selma leans over to look at it, Niklas begins to engage in a kind of testing activity by putting his index finger in front of Cozmo's screen face (l. 67). Niklas turns his hand so that the upper side of his finger ends up only a centimeter or two from Cozmo's "eyes" on the screen. As Niklas switches between

covering Cozmo's display face and removing his finger, he seems to be exploring the robot's field of "vision" and how it may respond to the manipulation, thus to some extent Niklas is still dealing with the robot as having reactive powers, that is, it is still being treated as a potential agent.

Soon, however, Cozmo becomes a mere object. In lines 74–75 Selma, who has by now returned to closely monitor her boyfriend's testing activity, points at Cozmo's camera lens at the lower part of the display. In this way she shifts the focus to the very premise of Niklas' testing – that the robot is equipped with a technology for detecting visual information. By designing her turn in terms of having made a discovery about a detail in Cozmo's physical constitution (*a:: här är lilla kameran* 'oh here is the little camera', l. 74), she clearly deals with the robot as an object that can be scrutinized and explored rather than as a participant relevant to interact with.

Through this brief analysis we show that Cozmo may quickly cease to be an agent and turn into a mere object instead. Niklas' attempts to explore the robot's abilities to react still imply that the robot is being tested for potential agency. Exploring whether the robot will react to a finger that is covering its eyes is a way to investigate its perceptual abilities and thereby its potential to act (see Braitenberg's (1986) classic account of simple agents). However, by indiscreetly scrutinizing the robot's body, the human participants demonstrate that they do not treat Cozmo as having the integrity germane to living beings. By pointing out that there is a camera, Selma categorizes the robot as a technical object that registers its environment through a tiny camera rather than the somewhat human-like eyes modeled on its screen. In short, in this extract the robot is far from treated like an accountable agent with a moral right to react, even when it is being poked in the eyes and talked about.

5. Discussion

We inspected robotic agency as it intermittently emerges in various forms in family interaction (see Table 1 for a summary). By analyzing a short episode in detail, we demonstrated that Cozmo's behavior may at times be treated as a relevant contribution to an emergent course of action by human participants. The analytical procedure of sequential analysis deployed above contributes a practical operationalization of agency and a means to pinpoint the precise moments in which the robot emerges as an "autonomous" agent, or whether there is anything else happening in terms of agency.

Table 1. *Summary of identified forms of robotic agency*

Form of agency	Behavior	Response	Example
"Autonomous" agency	Robot performs a situated action	Human treats robot behavior as relevant	Human responds to a greeting by the robot (Extract 1)
Hybrid agency	Human operator launches action that is embodied by robot, partly autonomously	Human treats robot behavior as partly instructed by a human operator	Human holds another human accountable for robot behavior (Extract 2)
Ascribed agency	Any robot behavior (including robot staying idle)	Human accounts for robot behavior	Human explains robot behavior as intentional (Extract 3)
Potential agency	Human performs a move that could make the robot react	Robot reacts or not	Human performs actions directed at the robot's sensors (Extract 4)
Non-agency	Any robot behavior (including robot staying idle)	Human does not treat robot behavior as relevant	Human inspects robot's physical-material body (Extract 4)

For brief moments, the robot gained the status of an "autonomous" (albeit programmed and switched-on) agent, contributing relevant actions to an ongoing interaction without being instructed *in situ* to do so by a human. While this is indeed the form of agency promised by entrepreneurs, prevalent in media and folk understanding, we demonstrated that even with the current cutting-edge technology that Cozmo represents, autonomous agency was only briefly achieved in our data. The concept of *autonomy* has been problematized in interaction research, insofar as not even humans individually control all aspects of action (Goodwin, 2004; Auer et al., 2020). Likewise, roboticists have argued that autonomous systems always need to be scaffolded by humans (see Bradshaw et al., 2013; DSB, 2012). Our analysis highlights that even in moments when the machine is technically acting on its own initiative, the robot's status as an agent is an interactional accomplishment. Taking a dialogic stance, agency can never be assigned *a priori* or in a vacuum but only emerges as a social reality through interaction with others. Applying this to robots, the ability to move and play sounds does not in itself imply meaningful action, but the robot's moves become actions only when they are recognized by other participants as relevant contributions. It requires human engagement for the robot's

behavior to emerge as relevant, meaningful, and accountable actions. For us, then, autonomous agency is equivalent to the recognized ability to *participate* (i.e., contribute relevant actions) in an evolving sequence (Schegloff, 2007).

We also showed an instance of *hybrid agency*, in which a human and a robot form a collaborative unit that consists of the human launching relevant actions and the robot performing them through its own body (i.e., the robot is a distinct entity). While Alač and colleagues (2011) have discussed the role of roboticists in bringing a robot to "life", we demonstrate that even a child (Oskar) may form a hybrid agent with the robot, in which the robot performs the sequential actions while the human designs them so that they fit the situated context and may be oriented to as relevant by other participants. As robots leave constrained laboratory settings, this operator role may increasingly be taken on by lay persons, including children. While robotics is rhetorically concerned with autonomous action, it often practically targets this form of hybrid agency, in which a human operator selects and initiates a pre-programmed sequence of robotic actions as she sees fit in the given context. The robot becomes a Goffmanian animator (Goffman, 1981:226) of sorts that performs the action a programmer has authored and is directed by a human operator who takes the role of the principal in selecting the type of action in context. However, we would like to highlight that Goffman's model only loosely captures the asymmetric relationship between humans and robots (see also Enfield, 2013). Pitsch (2016) suggests that human and robot could be regarded as forming one interactional system, recognizing that situatedness and sequentiality are centrally human competences. Unlike technology that transmits human actions (such as buttons and levers or microphones/speakers in a phone), robots can increasingly adjust their performance of an action sequence to the specific context, shifting some of the control away from the operator towards the robot. In our empirically informed conceptualization of hybrid agency, the robot is not merely mediating human actions but instead gains a degree of autonomy through its distinct body and sensors that allow for situated adjustments of the human-initiated sequence. Future work could explore the agentive affordances of robots that are presented as mediating human interaction, such as telepresence robots (see, e.g., Due, 2021a; Boudouraki et al., 2019; Jakonen & Jauni, 2021) and robots that are presented as autonomous agents while being teleoperated by an invisible operator (such as in Wizard-of-Oz settings in which the human operator is deliberately hidden, see, e.g., Porcheron et al., 2020). In human-human interaction, this may be applied to cases in which a human animates the actions of another, such as in theatre. Similarly, hybrid agency may occur in assistive settings, in which a caretaker may animate actions that a client cannot perform alone (Krummheuer, 2015b; Auer et al., 2020).

Interestingly, even when the robot fails to produce a relevant next action, it may be cast as an agent that deliberately refuses to do so. Cozmo's sound designer mentioned that the robot was designed to make mistakes to blur the boundaries between failure to act and intentional refusal to act: "if you intentionally build in all these mistakes and have this charming character, when the robot actually makes mistakes it just kinda blends in. […] imperfect felt more human." In fact, Krummheuer argues (2015a) that the ability to oppose actions may be a crucial part of technical agency. In our analysis, participants deal with the absence of action through verbal *attribution of intentions.* While others have explored the attribution of mental states and intentional agency to robots in experiments (see Thellmann & Ziemke, 2021; Parenti et al., 2021), our work shows how mental states may be attributed in mundane settings when dealing with some practical problem. In our case, the *attribution of a mental state* is done to account for the robot's missing action, which at that precise moment safeguards the robot's status as an agent in interaction with humans. Importantly, even though the robot's agent status varies from moment to moment, participants may aim to maintain a consistent sense of agency and try to keep the robot 'alive'. This is how humans are treated by default, even when they are not currently acting.

Last, we showed how the robot's capabilities for perception and action can be put to the test. Previously, Alač (2016) has argued that the status of a robot dynamically alternates between being a thing and an agent. This duality becomes evident when the capacities for action are explored and the robot is treated as merely a *potential agent*. In the moment of inspection, however, the robot is treated as an object that can be scrutinized without regard to integrity, which constitutes the robot as a *non-agent*. As we argued above, there are clear differences and temporal shifts between moments of treating the robot as more of an agent or more of an object. In a sense, anthropomorphizing behavior could be described as allowing for potential agency, which may be grounded in the inherently dialogic nature of humans (see Linell, 2009). Even humans may temporarily be treated as non-agents, for instance, in specific institutional settings such as surgery, in which practices such as draping seem to facilitate the focus on the physical body that can be inspected and cut into (Hirschauer, 1991).

Overall, our findings show that the status of an interactional agent may be highly transient, meaning that its variable forms of agency are accomplished on a moment-by-moment basis. While EMCA work has already highlighted that agency is accountably constituted in interaction (Alač et al., 2011; Krummheuer, 2015a), we show how rapidly robotic agency may change during a mere 60 seconds of multi-party interaction. The transitions that we describe may occur in every new interactional move: A robot that was in one moment a potential agent may turn into a non-agent, an autonomous agent or an ascribed agent in a matter of

(milli)seconds. The brief moments when the robot is indeed treated as an autonomous agent, however, seem to be the focus of most academic interest, effectively mystifying robots' agentive capacities. We hope to have demonstrated both the interactional work that goes into bringing a robot to life and the playfulness and willingness on the human side to do so.

While Cozmo is a children's toy, it contains many features of state-of-the-art robots (in embryotic form): The robot uses its computer vision system for detecting faces and objects, can maneuver through its surroundings and move objects it recognizes. Cozmo does not speak but can communicate through sounds (many of them designed as emotion displays), and utter names. Cozmo can alternate between full teleoperation and activities without human intervention. We therefore strongly believe that our findings will apply to robots that consist of similar sensors, actuators, and control units. We hope to have covered a continuum of most common forms of agency, which does not exclude the possibility of further, more specific forms of agency, especially, perhaps, in the various levels of control by the parties in hybrid agency.

6. Conclusion

Much research in human-computer interaction and social robotics builds on the Media Equation, that is, the finding that people may respond to computers in social ways, applying human interactional rules (Reeves & Nass, 1996; Nass & Moon, 2000). This claim has been criticized as too simplistic, but others have offered further nuance to it (see, e.g., Fischer, 2011, 2021; Klowait & Erofeeva, 2021). We would like to underline, in line with Reeves and Nass' (1996) original argument, that our analysis of agency does not necessarily suggest that humans really believe that the robot is something like a living being. Instead, robotic agency emerges as an inherently interactional and collaborative achievement that is contingent on participants' "good will" to treat the robot's contributions as relevant, perhaps parallel to the suspension of disbelief in theatre (Coleridge, 1817/2004) or to treating the robot as a depiction of a social agent (Clark & Fischer, 2022).

As we have demonstrated in this study, social responses to robots are an expression of human orientation to the fact that the robot can perform actions that move the ongoing activity forward, intermittently rendering the robot as an agent. We therefore advocate sequential analysis of action as the basis for studying agency, testing autonomy, and designing interactional agents. This acknowledges that agency is not a static property of individual actors or agents but can only be dialogically constituted with co-present others. By studying how agency manifests itself in a brief video recording of humans interacting with a robot, we hope to have furthered an

alternative and more empirically informed conceptualization of agency as a members' phenomenon.

Acknowledgements

References

Alač, M. (2016). Social robots: Things or agents? *AI & SOCIETY*, *31*(4), 519–535. https://doi.org/10.1007/s00146-015-0631-6

Alač, M., Movellan, J., & Tanaka, F. (2011). When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics. *Social Studies of Science*, *41*(6), 893–926. https://doi.org/10.1177/0306312711420565

Antaki, C., & Crompton, R. J. (2015). Conversational practices promoting a discourse of agency for adults with intellectual disabilities. *Discourse & Society*, *26*(6), 645–661. https://doi.org/10.1177/0957926515592774

Asaro, P. M. (2000). Transforming society by transforming technology: the science and politics of participatory design. *Accounting, Management and Information Technologies*, *10*(4), 257–290. https://doi.org/https://doi.org/10.1016/S0959-8022(00)00004-7

Auer, P., Bauer, A., & Hörmeyer, I. (2020). How Can the 'Autonomous Speaker' Survive in Atypical Interaction? The Case of Anarthria and Aphasia. In R. Wilkinson, J. P. Rae, & G. Rasmussen (Eds.), *Atypical Interaction: The Impact of Communicative Impairments within Everyday Talk* (pp. 373–408). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-28799-3_13

Axelsson, A., Buschmeier, H., & Skantze, G. (2022). Modeling Feedback in Interaction With Conversational Agents—A Review. *Frontiers in Computer Science, 4*. https://doi.org/10.3389/fcomp.2022.744574

Bekey, G. A. (2005). *Autonomous Robots: From Biological Inspiration to Implementation and Control*. MIT Press.

Bohus, D., & Horvitz, E. (2011). Multiparty Turn Taking in Situated Dialog: Study, Lessons, and Directions. In *Proceedings of the SIGDIAL 2011 Conference* (pp. 98–109). USA: Association for Computational Linguistics.

Boudouraki, A., Fischer, J. E., Reeves, S., & Rintel, S. (2021). "I can't get round" Recruiting assistance in mobile robotic telepresence. *Proceedings of the ACM on Human-Computer Interaction, 4*(CSCW3), 1–21. https://doi.org/10.1145/3432947

Bradshaw, J. M., Hoffman, R. R., Johnson, M., & Woods, D. D. (2013). The Seven Deadly Myths of "Autonomous Systems." *IEEE Intelligent Systems, 28*(3), 54–61. https://doi.org/10.1109/MIS.2013.70

Braitenberg, V. (1986). *Vehicles. Experiments in Synthetic Psychology*. MIT Press.

Bratman, M.E., 1987, *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.

Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Broth, M., Cromdal, J., & Levin, L. (2019). Telling the other's side: Formulating others' mental states in driver training. *Language & Communication, 65*, 7–21. https://doi.org/10.1016/j.langcom.2018.04.007

Button, G., & Sharrock, W. W. (2010). The structure problem in the context of structure and agency controversies. In A. Denis & P. J. Martin (Eds.), *Human Agents and Social Structures* (pp. 17–33). Manchester: Manchester University Press.

Bødker, S., Dindler, C., Iversen, O. S., & Smith, R. C. (2021). Participatory Design. Synthesis Lectures on Human-Centered Informatics, 14(5), i–143. https://doi.org/10.2200/S01136ED1V01Y202110HCI052

Clark, H. H., & Fischer, K. (2022). Social robots as depictions of social agents. *Behavioral and Brain Sciences*, 1–33. https://doi.org/10.1017/S0140525X22000668

Clayman, S. E. (2013). Agency in response: The role of prefatory address terms. *Journal of Pragmatics*, *57*, 290–302. https://doi.org/10.1016/j.pragma.2012.12.001

Coleridge, S. T. (1817/2004). *Biographia Literaria*. https://www.gutenberg.org/files/6081/6081-h/6081-h.htm

Coulter, J. (1979). *The social construction of mind: Studies in ethnomethodology and linguistic philosophy*. New Jersey: Bowman & Littlefield.

Curl, T. S. (2006). Offers of assistance: Constraints on syntactic design. *Journal of Pragmatics*, *38*(8), 1257–1280. https://doi.org/10.1016/j.pragma.2005.09.004

Davidson, D., 1980, *Essays on Actions and Events*, Oxford: Clarendon Press.

Defense Science Board. (2012). *The role of autonomy in DoD Systems* (Task Force Report No. ADA566864). Department of Defense United States of America. https://dsb.cto.mil/reports/2010s/AutonomyReport.pdf

Demuth, C. (2021). Managing accountability of children's bodily conduct: Embodied discursive practices in preschool. In S. Wiggins & K. O. Cromdal (Eds.), *Discursive Psychology and Embodiment: Beyond Subject-Object Binaries* (pp. 81–111). Cham: Palgrave Macmillan. https://doi.org/http://link.springer.com/10.1007/978-3-030-53709-8_4

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Due, B. L. (2021a). RoboDoc: Semiotic resources for achieving face-to-screenface formation with a telepresence robot. *Semiotica*, *2021*(238), 253–278. https://doi.org/10.1515/sem-2018-0148

Due, B. L. (2021b). Distributed Perception: Co-Operation between Sense-Able, Actionable, and Accountable Semiotic Agents. *Symbolic Interaction*, *44*(1), 134–162. https://doi.org/10.1002/symb.538

Edwards, D. (1997). *Discourse and Cognition*. London: Sage Publications. Retrieved from http://www.uk.sagepub.com/books/9780803976979

Enfield, N. J. (2013). *Relationship Thinking: Agency, Enchrony, and Human Sociality*. USA: OUP.

Enfield, N. J. (2014). Human agency and the infrastructure for requests. In P. Drew & E. Couper-Kuhlen (Eds.), *Requesting in Social Interaction* (pp. 35–54). Amsterdam: John Benjamins Publishing Company.

Fischer, K. (2011). Interpersonal variation in understanding robots as social actors. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 53–60). https://doi.org/10.1145/1957656.1957672

Fischer, K. (2021). Tracking Anthropomorphizing Behavior in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction, 11*(1). https://doi.org/10.1145/3442677

Forlizzi, J., & DiSalvo, C. (2006). Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (pp. 258–265). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1121241.1121286

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs NJ: Prentice-Hall.

Garfinkel, H. (1991). Respecification: evidence for locally produced, naturally accountable phenomena of order*, logic, reason, meaning, method, etc. in and as of the essential haecceity of immortal ordinary society (I): an announcement of studies. In G. Button (Ed.), *Ethnomethodology and the Human Sciences* (pp. 10–19). Cambridge, U.K.: Cambridge University Press.

Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Goodwin, C. (2004). A Competent Speaker Who Can't Speak: The Social Life of Aphasia. *Journal of Linguistic Anthropology*, *14*(2), 151-170. https://doi.org/https://doi.org/10.1525/jlin.2004.14.2.151

Goodwin, C. (2018). *Co-Operative Action*. Cambridge: Cambridge University Press.

Greenbaum, J., & Kyng, M. (Eds.). (1991). *Design at Work: Cooperative Design of Computer Systems*. CRC Press.

Harjunpää, K. (2021). Brokering co-participants' volition in request and offer sequences. In A. P. Jan Lindström Ritva Laury & M.-L. Sorjonen (Eds.), *Intersubjectivity in Action: Studies in language and social interaction* (pp. 135–159). Amsterdam: John Benjamins. Retrieved from https://www.jbe-platform.com/content/books/9789027259035-pbns.326.07har

Hedayati, H., Szafir, D., & Andrist, S. (2019). Recognizing F-Formations in the Open World. In *2019 14th ACM/IEEE International Conference on Human-*

Robot Interaction (HRI) (pp. 558–559).
https://doi.org/10.1109/HRI.2019.8673233

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge
[Cambridgeshire]; New York, N.Y: Polity Press.

Heritage, J., & Raymond, G. (2012). Navigating epistemic landscapes:
acquiesence, agency and resistance in responses to polar questions. In J. P.
de Ruiter (Ed.), *Questions: Formal, Functional and Interactional Perspectives*
(pp. 179–192). Cambridge, U.K.: Cambridge University Press.
https://doi.org/10.1017/CBO9781139045414.013

Hirschauer, S. (1991). The manufacture of bodies in surgery. *Social Studies of
Science*, 21(2), 279–319.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition. *ACM
Transactions on Computer-Human Interaction*, *7*(2), 174–196.
https://doi.org/10.1145/353485.353487

Hollnagel, E., & Woods, D. (2005). *Joint Cognitive Systems: Foundations of
Cognitive Systems Engineering*. Boca Raton, FL: CRC Press.

International Organization for Standardization. (2021). *Robotics — Vocabulary*
(ISO Standard No. 8373). Retrieved from
https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-3:v1:en

Jakonen, T., & Jauni, H. (2021). Mediated learning materials: visibility checks in
telepresence robot mediated classroom interaction. *Classroom Discourse*,
12(1–2), 121–145. https://doi.org/10.1080/19463014.2020.1808496

Jung, M. F. (2017). Affective grounding in human-robot interaction. In *Proceedings
of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*
(pp. 263–273). New York, NY, USA: ACM.
https://doi.org/10.1145/2909824.3020224

Keevallik, L. (2017). Negotiating deontic rights in second position. In L. R. M.-L.
Sorjonen & E. Couper-Kuhlen (Eds.), *Imperative Turns at Talk: The Design of
Directives in Action* (pp. 271–295). Amsterdam: John Benjamins.
https://doi.org/10.1075/slsi.30.09kee

Klowait, N., & Erofeeva, M. A. (2021). The Rise of Interactional Multimodality in
Human-Computer Interaction. *The Monitoring of Public Opinion: Economic &*

Social Changes, 1(1), 46–70.
https://doi.org/10.14515/monitoring.2021.1.1793

Krummheuer, A. (2015a). Technical agency in practice: The enactment of artefacts as conversation partners, actants and opponents. *PsychNology Journal*, *13*(2–3), 179–202.

Krummheuer, A. (2015b). Performing an action one cannot do: Participation, scaffolding and embodied interaction. *Journal of Interactional Research in Communication Disorders*, 6(2), 187–210.
https://doi.org/http://dx.doi.org/10.1558/jircd.v6i2.26986

Laurier, E., Maze, R., & Lundin, J. (2006). Putting the Dog Back in the Park: Animal and Human Mind-in-Action. *Mind, Culture, and Activity*, *13*(1), 2–24.
https://doi.org/10.1207/s15327884mca1301_2

Latour, B., 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford UP.

Licoppe, C. (2021). The politics of visuality and talk in French courtroom proceedings with video links and remote participants. *Journal of Pragmatics*, *178*, 363–377. https://doi.org/10.1016/j.pragma.2021.03.023

Linell, P. (2009). *Rethinking Language, Mind, and World Dialogically*. Charlotte, NC: Information Age Publishing.

Lipp, B. (2019). *Interfacing RobotCare – On the Techno-Politics of Innovation*. Technische Universität München, Munich, Germany. Retrieved from https://mediatum.ub.tum.de/doc/1472757/1472757.pdf

Miksell, L. (2010). Repetitional responses in frontotemporal dementia discourse: Asserting agency or demonstrating confusion? *Discourse Studies*, *12*(4), 465–500. https://doi.org/10.1177/1461445610370127

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics, 145*, 47–62.
https://doi.org/10.1016/j.pragma.2019.01.016

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103.
https://doi.org/10.1111/0022-4537.00153

Otsuka, K., Yamato, J., Takemae, Y., & Murase, H. (2006). Conversation Scene Analysis with Dynamic Bayesian Network Basedon Visual Head Tracking. In *2006 IEEE International Conference on Multimedia and Expo* (pp. 949–952). IEEE. https://doi.org/10.1109/ICME.2006.262677

Parenti, L., Marchesi, S., Belkaid, M., & Wykowska, A. (2021). Exposure to Robotic Virtual Agent Affects Adoption of Intentional Stance. In *Proceedings of the 9th International Conference on Human-Agent Interaction* (pp. 348–353). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3472307.3484667

Pelikan, H. R. M., Broth, M., & Keevallik, L. (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* (pp. 461–470). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3319502.3374814

Peräkylä, A. (2002). Agency and authority: extended responses to diagnostic statements in primary care encounters. *Research on Language and Social Interaction*, *35*(2), 219–247. https://doi.org/10.1207/S15327973RLSI3502_5

Pitsch, K. (2016). Limits and Opportunities for Mathematizing Communicational Conduct for Social Robotics in the Real World? Toward Enabling a Robot to Make Use of the Human's Competences. *AI & Society*, *31*(4), 587–593. https://doi.org/10.1007/s00146-015-0629-0

Porcheron, M., Fischer, J. E., & Reeves, S. (2020). Pulling back the curtain on the Wizards of Oz. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3). https://doi.org/10.1145/3432942

Reeves, B., & Nass, C. (1996). *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge: Cambridge University Press.

Sacks, H. (1992). *Lectures on conversation*. (G. Jefferson, Ed.). Oxford: Blackwell.

Schegloff, E. A. (2007). *Sequence Organization in Interaction. Sequence Organization in Interaction: A Primer in Conversation Analysis I*. Cambridge: Cambridge University Press.

Schlosser, M. (2019). Agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition). Metaphysics Research Lab, Stanford

University. Retrieved from
https://plato.stanford.edu/archives/win2019/entries/agency/

Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. Cambridge, MA. Retrieved from https://apps.dtic.mil/sti/citations/ADA057655

Skantze, G. (2021). Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language*, 67, 101178. https://doi.org/https://doi.org/10.1016/j.csl.2020.101178

Sterponi, L. (2003). Account episodes in family discourse: the making of morality in everyday interaction. *Discourse Studies, 5*(1), 79–100. https://doi.org/10.1177/14614456030050010401

Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press.

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions, 2nd ed.* New York, NY, US: Cambridge University Press.

Thellman, S., & Ziemke, T. (2021). The Perceptual Belief Problem: Why Explainability Is a Tough Challenge in Social Robotics. *ACM Transactions on Human-Robot Interaction,10*(3). https://doi.org/10.1145/3461781

Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Rochester University, Rochester, NY, USA. Retrieved from https://apps.dtic.mil/sti/citations/ADA289894

Waring, H. Z. (2011). Learner initiatives and learning opportunities. *Classroom Discourse, 2*(2), 201–218. https://doi.org/10.1080/19463014.2011.614053

Warnicke, C., & Broth, M. (forthcoming). Making Two Actions at Once**:** How interpreters address different parties simultaneously in the Swedish Video Relay Service. Translation and Interpreting Studies.

Weatherall, A. (2020). Constituting Agency in the Delivery of Telephone-Mediated Victim Support. *Qualitative Research in Psychology, 17*(3), 396–412. https://doi.org/10.1080/14780887.2020.1725951

Weatherall, A., & Keevallik, L. (2016). When claims of understanding are less than affiliative. *Research on Language and Social Interaction, 49*(3), 167–182. https://doi.org/10.1080/08351813.2016.1196544

Wilson, G., & Shpall, S. (2012). Action. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Retrieved from https://plato.stanford.edu/archives/win2016/entries/action

Appendix

FAM6_day4_P1 [11:26-12:32]

```
((music from Cozmo app in background))
01 NIK   ni*klas, >de* kän%de ig^en# mej.<
         Niklas    it recognized me
   nik   ≫-- gaze app -->(l.03)
   nik                      %,,,,,,--,,,,-> ((point to app))
   osk     *,,…....,* ((gaze from app to coz to app))
   sel   ≫-- gaze to coz        ^,,..app -->l.04
   img                          #img1.1

02 COZ   +/KLICK/* + (.)%
   coz   +arm strike+
   nik              -->%
   osk           *   ,,,-> ((gaze to coz))

03 COZ   >ad*eo dae-eo< dA+o dE+o* A-Ao NI%Kl*[as].
04 NIK                                    [ja]:#, he*^j.
                                           yeah   hello
   coz                 +forw+ ((moves toward nik))
   osk   ,,,*......coz..........,,*gaze to app*,,,.coz.,,*.. ((gaze to app))
   nik                              %,,,.............--> ((gaze on coz))
   img                                       #img1.2

05 SEL   NE%‰^::e^:%j.
         no way
   nik     %blink%
   nik   ....%,,,,,‰ ((gaze to app))
   sel        ^,.,^ ((gaze from app to osk to app))

06       (0.3)
```

```
07 OSK?  hn
08       (0.8)
09 SEL   mn ↑jag då?
         but what about me then?
10       (0.4)
11 OSK   skavi lägga till dej?
         should we add you?
12 SEL   a:::
         yes
13       (2.5) ((BRO types in app))
14 NIK   °han trodde inte du var human%△°
         he didn't think you were human
   nik                            %pats SEL on her back w left hand-->
   sel                            △pouting her lips-->
15 COZ   nik[la]s
16  ?       [nh]
17       (0.8)
18 NIK     %hej hej%
             hello
   nik   -->%waves-%
19       (2.9)
```

```
20 OSK   *nh                     *nh ((smiles))
   osk   *presses button in app*
```

```
21        ¤ (0.3)¤  (0.2)   %¤*(0.1)^¤(0.2)%¤(0.1)
   osk    ¤…...¤gaze to nik¤,,,,,,,¤coz-,,¤sel-->
   nik                     %smiles-------%
   osk                        *...-->
   sel                                ^gaze at osk-->

22 SEL  *vad skre¤v^  d+u?*
        what did you write?
   osk  *turns coz to sel*
   osk        -->¤looks down-->
   sel          -->^,,,-->
   coz               +turns left-->

23        ¤(0.1)^¤  (0.1)   ¤(0.1)¤(0.2)+(0.4)*(0.2)
   osk    ¤......¤glance sel¤,,,,,¤coz-->
   sel    ,,,,,,^gazes at coz-->
   coz                              -->+
   osk                                    *turns coz to sel-->

24 OSK    n*(h) ((smiles))
   osk    -->*,,,-->

25        (0.3)¤(0.1) Δ¤  (0.3) ¤(0.1)*(0.2)¤(0.2)
   osk     --> ¤,,,,,,,¤gaze sel¤,,,,,,,,,,¤gaze phone-->
   sel              Δ...raised eyebrows-->
   osk                           -->*

26        +^  (0.4)  +(0.1)
   coz    +drives fwd+
   sel     ^gazes at phone-->

27 COZ  OA:^:+oΔ^        +de-e ((face learning sequence starts))
   sel  -->^    ^gazes at COZ-->13
   coz        +lifts arms-+
   sel        -->Δ,,,-->

28 OSK  kolla raΔkt in i cozmos ögon
        look right into Cozmo's eyes
   sel  ,,,,,,,,Δ

29 COZ                    [ rrrrrrrrr]rrr aaAaaaaaaoooo rrrrrrr
30        ++ (0.1)Δ         ++(0.1) +
   coz    +head down, disengage gaze+
   coz     +drives backwards++lifts arms-->
   sel            Δsmile-->35
31 COZ   di+ng:¤::::::
   coz  -->+
   osk          ¤...gaze@COZ-->
32        (0.2)^|(0.1)^ (0.3)
   mus          |music change: more metallic
   sel      -->^     ^gaze@phone-->
33 COZ  +e¤e^                 + o       + ((motor sound))
   coz  +turns away, head down+
   coz                        +arm strike+
   osk  -->¤,,gaze to sel-->
   sel   -->^gaze to coz-->40
34        (0.1)¤ (0.1)
   osk      -->¤,,,,, -->
35 COZ  u¤uuuΔhh
   osk  ,¤gazes at coz-->
   sel     -->Δ...-->
36        (0.2)¤Δ (0.3)  ¤  (0.2)
   osk    .....¤gazes sel¤,,coz-->
   sel    ......Δraises eyebrows very high-->54
37 COZ  +°rrr°      + ((motor sound))
   coz  +lifts head+

38        (0.4)
```

```
39 COZ    ¤see*#-maaa?
   osk    ¤..gazes sel-->
   osk       *smiles-->
   img         #img2.1

40        ¤(0.2)%(0.1)^¤
   osk    ¤............¤gazes to phone-->
   nik          %..........-->
   sel              -->^gazes to phone-->

41 OSK    S%E¤^MA[ΔA    ^ah^e][he#he%hehehe]
42 COZ          [oaaaow    ]
43 SEL                [ m(h)%   ehe] he% he
   nik    .%gazes at sel-----------%.........%coz-->
   osk    -->¤.gazes at sel-->
   sel    -->^..gz osk^--^((blinks))gazes at cozmo-->>
   sel        -->Δsmiles-->>
   img                     #img2.2

44 SEL    [heha]
45 COZ    [se -]maaa.

46 SEL    he he he+he [°hej°=]+
          ha ha ha ha hi
47 COZ             [>adeo]+[dae-eo<=    ]+ /KLICK/  +
48 OSK             [=det låter som]+#
                    it sounds like
   coz         +fork up---+"waves"---------+fork down +
   img                           #img3.1

49 COZ    ^=da[o        d[eo      ]#a-a^][o  ]
50 OSK       [som han sä[ger (.) se]#nap^]
              like he says mustard
51 SEL             [he-  h^ej]]#     [nej] vänta^
                    he-    hello        no wait
   sel    ^moves fist toward coz-^jabs---^fist down-up^
   img                           #img3.2

52        (0.2)

53 SEL    ^hur gör han. (0.2)^(0.2)#
           how does he do it
   sel    ^jabs twice--------^fist in circles x 5-->
   img                        #img3.3

54 OSK    NEJ- h[an--   han ] vill inte göra::^#
          no   he      he   doesn't want to do
55 SEL          [hur gör du,]
                 how do you do it
   sel                               -->^
   img                           #img3.4

56 SEL    men j^a vill.^
          but i want
   sel       ^sml jab^

57        ^(1.2)
   sel    ^fist in ready position-->>

58 NIK    hehehe[he       ]
59 OSK          [(>kolla<)] ((points at phone))

60 OSK    se: [::  ma::]::: ((looks at NIK))
61 NIK        [(xxxxxx)]
62 OSK    a(h)heh[ehehe]
63 COZ          [se:: ][:   ma:]:
64 NIK                 [se-ma]= ((singing voice))
```

```
65 OSK    =sister (ko^l)la*hah*a
             sister, look here
   osk                   *...*gazes sel-->
   sel               ^gazes mobile, leans fw-->

66         (0.3)+*(0.1)* (0.2)  +(0.1)
   osk         -->*,,,,,*
   coz           +small turn left+

67 OSK     *se%^: ma:%::::#
   osk     *gazes mobile-->
   sel      -->^gazes mobile-->
   nik        %......%points coz-->
   img                 #img4.1

68 OSK     a(h)hn^Δ[hn]
69 SEL             [SE]+-  M[   +A+Δ:: ]
70 COZ                  [   +d+udi ]#dui
   sel     -->^coz-->
   sel       -->Δleans back-----Δ
   coz             +fl up,bw+head up+
   img                           #img4.2

71         (0.2)%(0.4)+(0.2)   +
   nik      -->%turns finger, nail covers screen-->
   coz              +moves fw+

72 COZ     se::-

73         (2.1)

74 SEL     Δ#mpt a:: [här är        lilla#+Δ ka+m%Δeran]
             mpt yea: here is the little camera
75 COZ                [>adeo dae-eo< dAo    +dEoA+A%Δo  s]e:: ma::::%
   sel     Δ.............................Δpoints Δ,..screen-->90
   nik                                  -->%,,,,,,,,,,,,,,,,%
   coz                                   +fwd-+
   img     #img4.3                      #img4.4

76         (0.3)%(1.2)
   nik         %....-->

77 NIK     mhm%:.
   nik     -->%nail covers screen-->>

78         Δ(0.7) Δ(2.3)
   sel     Δ,,,,,,Δ

79 OSK     selma, så ((types on phone))
             selma, like that

80         (0.2)

81 OSK     continue ((puts phone down))
```