



Social Interaction. Video-Based Studies of Human Sociality.
2023 Vol. 6, Issue 1
ISBN: 2446-3620
DOI: 10.7146/si.v6i1.137378

Social Interaction

Video-Based Studies of Human Sociality

Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology

Stephan Habscheid, Tim Hector & Christine Hrncał

University of Siegen

Abstract

The article takes its starting point with heuristics, according to which agency is not to be seen as something that certain ontological entities stably do or do not possess. Rather, it is assumed that agency, especially in voice-based exchange with smart speaker technology, is a dynamic accomplishment, basically bound to the local (linguistic) practices carried out by or rather involving contributions by participants with unequal resources for participating. Following Hirschauer (2016), we distinguish between *levels of activity* both on an *active-passive spectrum* as well as on a *proactive-inhibitive spectrum* and reconstruct empirically against that background how in particular the smart speaker can appear in different situations and contexts. The article concludes with a discussion of the notion of agency relating the observed practices on the one hand and against the background of a broader context of agency as media theory has it on the other, including domestication theory as well as recent smart home technologies and platform logics.

Keywords: praxeology, smart speaker, voice user interface, agency, domestication theory

1. Introduction

Intelligent personal assistants (IPAs) with voice user interfaces (VUIs) are finding their way into private living environments as stationary smart speaker systems (among others). There, they are used for listening to music, searching for and retrieving information, controlling smart home devices, or for alarm and timer functions (see Ammari et al., 2019). The operation of such smart speakers takes place – at least in part – verbally. The interactions show similarities to interpersonal human interaction, which to a certain extent is simulated (or intended to be simulated) (see Hennig & Hauptmann, 2019), but also differences, which have been described as “computer talk” (Zoeppritz, 1985, p. 1; see also Chapter 4.2). To enable the voice-controlled handling, smart speakers must permanently scan the environment in which they are placed for an activation word in order to then change to what is called “listening mode” when they have recognized it (e.g., “Alexa” or “Hey Siri”). The subsequently recorded signals are then processed in the cloud by various operations (including speech-to-text or text-to-speech, natural language processing, and information retrieval, see Natale, 2020). The audio data is stored and additionally used for further purposes, in particular for the improvement of systems and commercial interests in marketing and advertising (Zuboff, 2018; Turow, 2021). The devices thus operate between continuously improved user comfort on the one hand and risks that the monitoring and evaluation of the data entail on the other (see Lau et al., 2018; Sweeney & Daves, 2020). For users, the requests are made partially transparent in the corresponding smartphone app and logged (at least on the surface; no information is given on the further processing of the data). The logfile data result in an activity protocol for the users, including audio recordings and transcriptions (see Habscheid et al., 2021).

Since the 1980s, one strand of research has been conceptualizing the exchange between humans and machines in a strictly empirical manner, avoiding categorical-ontological positions to distinguish between these two entities. Suchman (1987), in her classic study of the interaction between humans and an assistance system for operating a copy machine, assumes that exchange between humans and machines is a communicative arrangement, in which one of the two communication partners (the machine) is extremely “resource-limited” (Suchman, 1987, p. 70; see also Suchman, 1990, p. 43). The design of the machine’s sequences is plan-based, with the plans remaining opaque for the users. In contrast, the human’s interaction is situation-based, and even the orientation to plans is an enactment performed in situ, which again remains opaque to the photocopier and only appears where it can be incorporated into the plan. These fundamental differences in the processing and production of communicative signals lead to difficulties in accomplishing “interaction”, whereby openness and situation-boundness are the greatest challenges in coping with characteristics of human interaction for the machine. In her second edition of *Plans and Situated Actions*, Suchman (2007) actualizes and re-elaborates on

some of her questions in the light of new technologies and further discourse; we will return to this.

Krummheuer (2010) refers to human-machine-exchange on the basis of her study of the virtual agent Max. She conceptualises the human-machine-exchange – also with a methodology oriented towards ethnomethodology and conversation analysis – as a “hybrid exchange” of technical and social units. In this hybrid exchange, according to Krummheuer, it remains undecided what both the human and the machine participants in the interaction treat it as: sometimes the focus is on the aspect of the unity of the social and the technical characteristics – within the framework of a “to-act-as-if modulation” (Krummheuer, 2010, p. 324), but sometimes the difference between the two participants in the interaction is also emphasized. The exchange oscillates between these two poles, and Krummheuer also observes – depending on the occasion, for example in the case of disruptions – sudden changes from one approach to the other. The question of how the exchange is viewed and framed by the participants is thus also the object of situational execution (and here both the technical and the human participants are concerned).

Such situational shifts can also be observed when interactions with a smart speaker are examined against the background of the debate on the notion of agency (Krummheuer, 2015). Before discussing this in more detail, let us consider one example taken from a video, in which a smart speaker is set up and installed in the living environment of the interactants for the first time.

Example 1. *Let's get started*

```
397 AL: $legen wir LOS== $
      let's get started
      js: $scrollt am Handy$
          scrolls on smartphone
398     =um mit mir zu sprEchen (.) nenne einfach meinen NAMen und stelle
          eine +frage;
          to talk to me just call my name and ask a question
399     +versuche es MIT (.) alexa (.) wie ist+ das wEtter?+
          try alexa how is the weather
      dl: +blickt JS und anschließend Handy an----->
          >-----+
          looks at JS and then at smartphone
      js: +blickt DL und anschließend Handy an-----+
          looks at DL and then at smartphone
400     $(1.4)$ +
      js: $nickt $
          nods
          +blickt DL an+
          looks at DL
401 DL: alexa (.) wie ist +!MOR!gen das wetter;
          alexa how will the weather be tomorrow
402 p: (1.1)
```

403 AL: in+ rüdinghausen beträgt die temperatur SIEben grad celsius bei
bewölkte[m] himmel;
*in rüdinghausen the temperature will be seven degrees celsius with
a cloudy sky*

dl: +blickt JS und anschließend Handy an----->
>----->
>--+
looks at JS and then at smartphone

The smart speaker – Amazon EchoDot (4th generation) with the VUI personalized as “Alexa” (AL) – is in the initial set-up mode, a script that starts directly after the first connection with the Wi-Fi in the living environment (a process that was completed shortly before the given extract starts). In this mode, various functions are tried out; the smart speaker, or more precisely the VUI, takes over the communicative “leadership” of the dialogue and its utterances initiate the conversation (line 397: “let’s get started”, followed by an utterance, that “starts” the conversation). Thus, the formulation of voice commands is “trained” at the same time because through the suggestion-based type of dialogue guidance the users are shown how a certain function can be called up and used linguistically and also shown which functions are available. In this example, the suggestion made by the VUI is to ask about the weather (line 399). The user, DL, takes up this suggestion (line 401). However, DL adds a temporal specification, which is also emphasized by the strong focus accentuation as well as the look at JS and the subsequent raising of the eyebrows. The user, it can be concluded, is already testing the flexibility of the VUI in the first addressing. In this way, on the one hand, she attributes agency to the device. In this context, agency is understood as the ability to act, at least potentially. To address a challenge to an addressee that could not fulfill it anyway would not be of interest, with the exception of non-serious, jocular interaction modalities, which seems rather unlikely in this example given the reaction and specific facial expression of JS. Rather, in front of the audience of her partner JS – made relevant by the gaze and the mimic modulation – the user tests whether the VUI is flexible enough to deal with this adaptation. That this form of agency could exist is thus a precondition for this command. At the same time, the user herself regains agency through this test: while the dialogue between humans and machine was consistently directed by the machine in the last sequences, not (exactly) following the suggestion and the addition and specification are a form of agency claim.

Agency in this example appears firstly as a “situated, dynamic, collaborative, and temporally unfolding enactment” (Ibnelkaïd & Avgustis, 2023/this issue; also see Ibnelkaïd, 2019). Furthermore, agency appears to be quite ambivalent in relation to concrete interaction situations and sequential processes: while agency can be attributed to a technical device on the one hand, in the example discussed here, it is at the same time an agency claim for the user. At this point, we want to come back to Suchman, who, in her abovementioned revised edition of *Plans and Situated Actions* (2007), extensively discusses the distribution of agency between

humans and machines. In her discussion of Latour's (1993, pp. 77–78) notion of the "Middle Kingdom" as the "space between simple translations from human to nonhuman [...] and a commitment to maintaining the distinctness and purity of those categories" (Suchman, 2007, p. 260), Suchman renews the question of how to conceptualize human-machine-exchange and the included distribution of agency with regard to the specific characteristics of each of the categories on the one hand and acknowledging "the mutual constitution of humans and artifacts" (Suchman, 2007, p. 260) on the other. She argues that dissimilarity and blending should be taken into account in parallel (see Suchman, 2007, with reference to Latour, 1993, p. 11). Discussing the notion of "symmetry" of agency in works by Latour and in Actor Network Theory (ANT) and also considering other contributions, for example by Charles Goodwin, Karen Barad, and Madeleine Akrich, Suchman (2007) argues, that "humans and artifacts are *mutually constituted*" (p. 268) and stresses that ontological distinctions between humans and machines as independent, categorial entities are not leading anywhere. But, according to Suchman, that does not mean that mutuality means symmetry; instead, Suchman highlights what she now calls "dissymmetry" between humans and machines, as the two do not constitute each other "in the same way" (Suchman, 2007, p. 269) – and consequently also in the distribution of agency. In the attempt to integrate approaches that see agency as tied to human actors on the one hand, with approaches that clearly see agency as distributed among the "sociomaterial network" (especially ANT) on the other, Suchman comes to the conclusion that agency is neither to be tied to humans, machines, nor bodies, but rather the focus should be on "the materialization of subjects, objects, and the relations between them as an effect, more and less durable and contestable, of ongoing sociomaterial practices" (Suchman, 2007, p. 286).

Following on from this, we take a praxeological perspective that refrains from identifying ontological actor categories as well as from extreme levelling of human and machine categories as ANT has it (see Latour, 2007). In line with such a praxeological approach is the conceptual work by Stefan Hirschauer, which seems to be very helpful here. Hirschauer (2004, 2016) focuses on the execution of practice (as we do in the analysis of the introductory example and the following excerpts) and distances himself from centering on the actors and thus also from the question of whether, for example, a smart speaker has "agency" per se. Hirschauer argues that not only objects (like a ball in a football match, see Hirschauer, 2016, p. 51), but even social relations, postures, seating arrangements, and other entities could be "involved" in practice. This makes them participants in such practices, but not participants per se in interpersonal interaction (see also Krummheuer, 2010, pp. 42–44, with reference to Rammert & Schulz-Schaeffer, 2002). In practice, however, they can be involved with different "levels of activity" (Hirschauer, 2016, p. 49). To distinguish these levels, Hirschauer differentiates between two spectra on which activities can be located: an active-passive-spectrum and a proactive-inhibitive-spectrum. Whereas active-proactive action is represented as the explicit completion of an activity, the more

passive side of the proactive spectrum can be used, for example, to locate the proactive non-processing of a matter. In contrast, on the inhibitive side of the active spectrum, one can find, for example, consummations such as “counteracting” or “preventing,” while in terms of the passive-inhibitive spectrum, one can, for example, leave a matter to its own devices (see Hirschauer, 2016, p. 49). Actions of individuals are thus threaded into an ongoing stream of practice involving not only one person with one level of activity, but also different entities with different degrees of participation. Hence, a machine – such as a smart speaker – can be involved in the enactment of practice as a participant with different levels of activity (Hector, 2022): it can actively start a conversation (as in the given example) or enable activities, but also undermine or even prevent others (e.g., disrupting ongoing human-to-human-interaction by playing very loud music).

Thus, agency comes into view as a situational, practical accomplishment (Garfinkel, 1967), which we want to investigate with ethnomethodological-conversation-analytical methods that have been extensively tested in the field of human-machine interaction, especially in studies on computer-supported cooperative work (CSCW) and Workplace Studies (Luff et al., 2000; Oberzaucher, 2018) but also other media research with a background in conversation analysis (Ayaß, 2004; Ayaß & Bergmann, 2011), and especially by Krummheuer (2015) who argues that “agency emerges from interaction and how participants orient to it while interacting” (pp. 195–196). We therefore raise the following questions: On what occasions is agency, also in its gradients, negotiated, attributed, and withdrawn interactionally? How are machine artifacts thereby integrated into the interaction and how are they (de)constructed as participants in a conversation? How do these constructions take place on the linguistic and multimodal level and how do technical artifacts thereby shape interpersonal interaction?¹

2. Research Context, Background and Corpus Description

Our qualitative investigations are carried out in the context of the research project “Un/desired Observation in Interaction: Intelligent Personal Assistants (IPA)” in the Collaborative Research Center Media of Cooperation at the University of Siegen. The project examines whether and how users reflect on the aforementioned exploitation of their data with regard to data protection and privacy, and how smart speakers with their specific affordances are embedded in everyday domestic practices against this background. In this context, the previously mentioned questions about the agency of social and technical actors also emerge.

¹ We thank the anonymous reviewers for many helpful comments.

We thereby work with an understanding of media that is grounded in social theory and focuses on the question of how media are constituted as *social* entities. This approach is crucially based on the work of Charles Goodwin, whose concept of “co-operative action” is used to set social practice as logically preordered to other variables such as action, interaction, routines, or techniques (Schüttpelz & Meyer, 2017, p. 159). Goodwin starts from the assumption that “co-operation” does not mean the consensual co-operation of two or more individuals who maximize their common and thus also their individual benefit for a certain price. For Goodwin, co-operation is rather the reciprocal inducing of an operational sequence. This reciprocity is induced in the sense of a mutual co-operation, that is, in a possibly unspoken and presupposed reciprocity, on which we are dependent to be able to coordinate with other people at all (Schüttpelz & Meyer, 2018, p. 175). Actions (that become part of the co-operative action) are, according to Goodwin (2018), semiotically opportunistic and tend to “incorporate voraciously whatever local materials might be used to construct the action required at just this moment” (p. 445). Hence, co-operation does not only use linguistic structures, but also material and socio-spatial characteristics of the environment which the co-operating individuals are placed in. The various co-operations carried out by manifold human beings can have a cumulative effect on each other. Humans can come back to an operational sequence that was induced once and orient their co-operation towards this induction again. They can dissect and transform the sequences, forming new concatenations towards which humans can orientate themselves again (Goodwin, 2018, pp. 4–9). These accumulations can have a situation-transcendent effect and manifest themselves independently from individuals and socio-temporal situations; thus, they form a socio-material “web” of co-operatively carried out operative sequences, which are based on and available for other such sequences. In this sense, language can be seen as a specific form of co-operation. Coming back to what this means for the investigation of media, through the recurring co-operation media-specific stabilizations in practice can potentially emerge (precisely when there are media-specific conditions of co-operation). These kinds of stabilizations, however, are always induced anew in practice – in co-operative action; there is no situation-transcendent media-specific determinism. Media in this sense are co-operatively achieved conditions of co-operation (Schüttpelz & Gießmann, 2015, p. 15). This perspective allows us to treat media as embedded in and part of the social practice (in the sense of Hirschauer, as described above).

To empirically investigate our abovementioned research questions within this context, eight households were studied; the initial set-up of a smart speaker was recorded on video. In addition, with the help of a conditional voice recorder (CVR) specially developed for this purpose by Porcheron, Fischer, Reeves, and Sharples (2018) for the smart speaker by Amazon with the activation word “Alexa”. audio recordings of the use of the smart speaker in everyday practice were generated over a period of three to four weeks at two different stages: first directly after the initial installation and then three to four months later (cf. Hector

et al., 2022). The CVR was advanced by us to enable two other activation words used within the smart speakers distributed by Google (“Ok/Hey Google”) and Apple (“Hey Siri”). The CVR permanently records three minutes of the environmental audio, keeps the recordings in a buffer and deletes them afterwards. In parallel, the device “listens to” and scans the recorded audio for the wake word, just as the smart speaker does. If a wake word is recognized, the recorded audio of three minutes will be saved and complemented by three further minutes of audio-recording, resulting mainly in six-minute audio files (sometimes longer if the wake word is recognized more than once within this timespan). In this way, approximately 106 minutes of video material from initial installation situations and approximately 18 hours of audio material from use situations was collected, inventoried, and transcribed according to GAT2 (Selting et al., 2011).

3. Interactional Occasions for Ascription of Agency

3.1 Attempts to make an impression on somebody

A local context in which users attribute a capability for social interaction to the smart speaker are situations in which a user in the role of an everyday “promoter” presents the innovative, smart technology in a favorable light to a partner with no (or only little) experience of use.² Such situations, like promotion in general, have a theatrical and at the same time strategic character.³ In contrast to professional influencer marketing on social media (see Och & Habscheid, 2022), this form of personally mediated promotion is purely private and free of commercial interests, but it can contribute not only to the acceptance of technical innovations, but also to the multiplication of social capital (Bourdieu, 1983) on the promoter’s account; by highlighting and exemplifying the benefits of a good relationship with the intelligent and personal assistant, the latter’s beneficiary also presents themselves as successfully networked and thus as socially more attractive.

Let us consider an example: in the apartment of a young couple, Damaris Lang (DL) and Jan-Ole Soellner (JS), a small birthday is celebrated. Damaris’ parents, Petra (PL) and Jürgen (JL), are guests; they visit the couple for a joint dinner on the occasion of their daughter’s birthday. Also present is Alexa (AL), the software of a smart speaker, which, along with Damaris and Jürgen, is ‘one of the household members’ involved in the situation. In the situation documented by the audio recording and transcription in Example 1, Jürgen and Jan-Ole are in the room where the smart speaker is located. Damaris and Petra are in the neighboring kitchen and produce noises and loud laughter there, which at times acoustically disturb the exchange between the two men and the smart speaker.

² See Folkerts (2001) on the promoter model in economics.

³ See Willems & Kautt (2003) on the strategic “theatricality of advertising” (“Theatralität der Werbung”).

At the beginning of the interaction before the exchange with the smart speaker, a subtle amalgamation of socio-technical exchange and social interaction becomes apparent. Jan-Ole asks Alexa to tell a joke to his father-in-law (“the [DAT] Jürgen,” line 0255).

Example 2. Tell a joke

0253 AL: [((spielt Musik weiter))]
continues to play music

0254 [((unverständliches Hintergrundgespräch zwischen DL und PL))]
unintelligible conversation between DL and PL in background

0255 JS: alexa (.) erzÄHL (.) erzähl dem jürgen ein witz;
alexa tell the [DAT] jürgen a joke

0256 (1.0)

0257 AL: sagt der VAtEr zu seinem sohn;
a father says to his son

0258 junge (.) ich muss DIR was sagen;
boy I have to tell you something

0259 DU wurdest adoptiert;
you were adopted

0260 [sagt der SOHN;]
the son says

0261 DL: [((lachen))]
laughing

0262 [((Lautes Lachen von DL und PL im Hintergrund))]
loud laughing by DL and PL in background
 [((Lautes Hintergrundgespräch von DL und PL))]
loud conversation between DL and PL in background

0263 AL: [was (.) (ich will SOfort meine biologischen eltern sehen,)]
what I need to see my biological parents immediately

0264 [(da antwortet der VAtEr,)]]
the father answers
 [((unverständlich, ca. 1 Sek))][((unverständlich, ca. 2 Sek))]
unintelligible, approx. 1 sec. unintelligible, approx. 2 sec.

0265] [ich] werde dich GLEICH abholen,
I will pick you up soon

0266 DL: [ärgernis verstehen;]
nuisance understand

0267 JS: [MA::N (.) die alexa erzählt
grad_n witz.] gee the [NOM] alexa is telling a
joke

0268 DL: [oh.]
oh

0269 AL: [((spielt Musik an))]
plays music

0270 [(0.5)]

Remarkable under the aspect of attribution of agency is not only the system’s designation of a function as telling jokes. Deviating from the usual command structure of hybrid exchange (“alexa, tell a joke”), Jan-Ole further emphasizes the social dimension of the competence of *telling* jokes by staging a triadic interaction (“alexa (.) tell jürgen a joke,” line 0255). In doing so, he treats the smart speaker,

in keeping with the role of an intelligent assistant, as a counterpart which, while not proactively, plays along at a high level of activity (cf. Hirschauer, 2016, p. 49).

With this command, Jan-Ole arguably takes the system to a limit of what can be accomplished technically: the IPA (at the time of the study) could neither engage in differentiated dialogic exchange with more than one human dialogue partner at the same time, nor did the system have the interaction history to really personalize a joke with the guest, Jürgen, in mind. Presumably, the personalized request to Alexa is addressed more to Jürgen, who is to be impressed and/or entertained by the imputed performance of the IPA. The attribution of a high level of situated interaction capability thus results not only from the functionalities of the system but also from its embedding in a user-sided theatrical staging.

Due to the noises from the neighboring room, in the further course of the event the necessity arises to order the hybrid communicative situation by an utterance in the “meta-interaction space” (Habscheid, 2022, p. 176): Jan-Ole, referring to the reception situation of the two men, indirectly asks the two women to be considerate, thereby expressing his irritation (“gee alexa is telling a joke,” line 0267). Continuing his colloquial staging of a triadic interaction situation (“the alexa ...”), he continues to seriously present the IPA as a social participant who is to be helped to assert its rights in communication. Accordingly, what is crucial is not that he and Jürgen are disturbed in the perception and reception of the joke, but that Alexa is disturbed in its production, which is also presented here quite naturally as an act of telling. Here, too, the attribution of agency results from the embedding in a multi-party constellation in which users exhibit their personal social relationship to Alexa to interaction partners on the basis of “her” linguistic ability.

3.2 Being impressed by the device

Another local context in which social interaction abilities are attributed to smart speakers are expressive utterances in which users present themselves as being impressed by an unexpectedly high performance of the IPA and thus at the same time express a positive evaluation of the system.⁴ The transitions to strive for influence on another (see Section 4.2) are fluid here, insofar as expressive speech acts can also serve as an indirect form of regulating another person’s state of mind (Rolf, 1997, p. 222). With reference to Michael Tomasello (2009, pp. 97–98), who counts such purposes among the fundamental types of human communication and cooperation, one could also speak of emotions being “shared” with each other in this way.

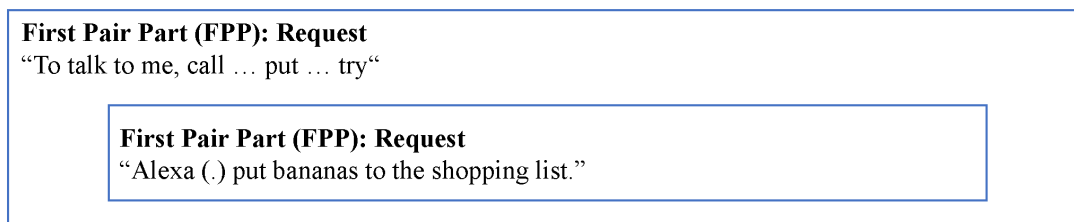
Let us look at an example from a situation in which a smart speaker is put into operation by two people. The situation was recorded on video and the multimodal interaction was transcribed. The participants are Alexa and two young men, Alex

⁴ For a theoretical analysis of expressive speech acts, see Rolf (1997, p. 217).

Kripp (AK) and Lukas Faßbender (LF), who share an apartment. During the installation, the system repeatedly suggests that the users try out some of its pre-installed skills,⁵ as we have already seen in Example 1. In this type of situation, there are potentially many occasions when users can comment on and evaluate the performance of the IPA, its possibilities and limitations, and to attribute or withdraw (degrees of) agency during the installation process.

The presentation of skills in the test environment has had limited success with regard to gradients of agency; it seems to be primarily designed to present a trouble-free socio-technical functionality and to avoid failures. Underlying this is a smart but simple dialogue design (Lotze, 2016, Chapter 2), in which one pair sequence embeds another pair sequence (Habscheid, 2022). As a closer look reveals, the degree of linguistic “smartness” is rather limited. In this type of socio-technical dialogue structure, a superordinate pair sequence includes an imperative construction: “To talk to me, name ... state ... try ...” as first pair part (FPP). Embedded in this request is the first move of another pair sequence, such as a wh-question: “Alexa (.) how is the weather?” or – as in the present case – a command: “Alexa (.) put bananas on my shopping list” (see Figure 1).

Figure 1. *Dialogue design in the testing environment*



By following the prompt and realizing the proposed question or request, the user enables the system to produce a prefabricated, precisely fitting second utterance, which at the same time responds precisely to the prompt (e.g., “I have added bananas to your shopping list,” line 525). The progression produced in this way appears to be a spontaneously organized sequential interaction which enables, by means of co-operative action (Goodwin, 2018), the agency of both participants. In fact, the course of utterances is based entirely on a prefabricated structure of pair sequences in conjunction with a tight user guidance that controls the user’s as well as the system’s participation.

⁵ “Skills” are voice-based software applications for smart speakers by Amazon, some of which are pre-installed as basic features and others of which can be installed additionally.

Let us now consider an example of implicit and explicit attributions of agency in the context of this socio-technical arrangement.

Example 3. *Recognizing*

```

519 AL:  §[versuche]es mit-    §
         try
    ak:  §lehnt sich wieder vor§
         leans forward again
520 AL:  alexa (--) +setze baNAnen auf die einkaufsliste.+
         alexa      add bananas to the shopping list
    lf:  +blickt AK an-----+
         looks at AK
521 LF:  [((lacht leise))]
         laughs quietly
522 AK:  [alexa (-) ] setze baNAnen +auf die
         einkaufs++1+2liste-      +1             +2
         alexa      add bananas to the
         shopping list
    ak:  +wendet Blick kurz von Alexa ab->
         shortly averts gaze from alexa
    >-----+
    ak:  +1blickt Alexa wieder an +1
         looks back at alexa
    lf:  +2blickt zurück auf Smartphone in seiner Hand+2
         looks back at smartphone in his hand
523      (1.6)
524 AL:  oKAY.
         okay
525      |+ich habe
         zu deiner einkaufs[liste hinZU][gefügt.      ]
         I have added bananas to your shopping list
    ak:  +blickt zurück auf den Laptopbildschirm auf seinem Schoß+
         looks back at the laptop screen on his lap
526 LF:  +[ich find  ][ich find geil,]
         I think      I think it is awesome
527      dass der [DICH direkt erkennt- ]+
         that he immediately recognises you
    lf:  +blickt AK an----->
         >-----+
         looks at AK
528 AL:  [hier ist etwas anderes] zum aus[probieren.      ]
         here is something else to try
529 LF:  *+ [und ich meinen;]
         *
         +
         and I my
    lf:  *deutet mit dem Smartphone
         auf sich selbst*
         points with the
         smartphone to himself
    lf:  Smartphone in seiner Hand+
         looks back at
         the smartphone in his hand

```

While Lukas shows his amusement by the simple staging of agency through a quiet laugh and thus casts doubt on it, Alex acts maximally cooperatively by

repeating the suggested command in exactly the same linguistic form. The further course of events is surprising in that the confirmation by Alexa (“I have added bananas to your shopping list”) proves to be unexpectedly “smart” on another level, even from Lukas’ perspective: in the early test phase in the two-party constellation, by means of the VUI the smart speaker is already able to recognize from the individual user’s voice which of the two residents is currently adding bananas to a shopping list.

Lukas shares his enthusiasm about this with Alex in the form of an expressive utterance with a clearly positive evaluation, using the verb *erkennen* (to recognize someone) to refer to an agency of living beings (line 527). In this utterance, the expression *du* in German can be interpreted both with reference to Alex and generally in the sense of *man* (one). It is remarkable, however, that – unlike in Example 2 – the device is referred to with a masculine simple demonstrative pronoun (*der*); thus, the carrier of the ascribed agency here is probably not the persona Alexa, but the technical device, which includes the software. Overall, it should be noted that boundaries between living beings and material devices in hybrid situations become blurred at times, not only in theatrical productions, but also in everyday situations of use (see Section 1 above, with reference to Krummheuer, 2010). Although our smart speakers, in contrast to Krummheuer’s Embodied Conversational Agent, have to perform without the resource of the body, the linguistic exchange can temporarily nourish the illusion of social interaction; however, the interaction is repeatedly accompanied by irritations, which then raise questions about the social character of the counterpart and make the technical, simulative character of the (previous) dialogues clear (Krummheuer, 2010, p. 323f.).

3.3 Device as part of human-to-human dialogue

Another local context in which the properties of social interaction partners are attributed to smart speakers are situations in which emotional involvement is made relevant through the actions of users towards the system and/or utterances by the system itself. Even if the equal participation of human and technical participants in the moment-to-moment unfolding interaction is not always established, IPAs can be involved with an ascription of a certain level of activity (Hirschauer, 2016, p. 49f.) in the dynamic ritual order of interaction and the concomitant procedures of balancing self- and partner evaluations (Goffman, 2005[1967]; Holly, 2000).

Let us consider the following example: Josie, who lives alone, has her brothers Tom, Peter, and Markus over for dinner; all of them are in their twenties. Josie has owned a smart speaker for a few days and together they try out the device. When asked by the system whether the user wants to end the game, Tom, Markus, and especially Peter react in a way that treats Alexa as a quasi-social counterpart on the one hand, but on the other hand, given the explicitness and

directness of the rejection, either as being incapable of being insulted or as having to put up with insults. In today's social life, such a style would hardly be acceptable in any kind of social relationship between humans. Accordingly, Josie meets the social obligation of expressing compassion ("oh poor one," line 07), thus correcting the construction of a relationship.⁶

Example 4. *The Poor One!*

01 Alexa: möchtest du das [SPIEL beenden,]
would you like to finish the game
02 Tom: [ja es hat ja keinen SINN;]
yes it does not make any sense
03 Peter: alexa sei [STILL;]
alex a shut up
04 [(klirrendes Scheppern)]
clanking
05 (0.5)
06 Markus: ja;
yes
07 Josie: oh [die ARme.]
oh poor one
08 Alexa: [oKAY (-)] danke fürs spIElen.
okay thank you for playing
09 (0.2)
10 Josie: oKA[::Y?]
okay
11 Alexa: [wenn du-] irgendein FEEDback für
12 un[ser team hast (-) lasse-es uns ()]
if you have any feedback for
our team let us
13 Josie: [voll EINGeschnappt die gute]
completely miffed the missus
14 Alexa: [()] wissen unter () punkt ()
know via dot
15 UMS: [(HAMmer;)]
stunning

In Josie's utterance, the attribution of status is also contradictory to a certain degree: Alexa is excluded from the inner circle on the one hand, being laterally addressed as a third-person attendant; on the other hand, being treated as a bearer of emotionality and listener to an address of solidarity, she is categorized as a sensitive member of the group of attendants. Admittedly, the ascribed emotionality in the context of a practice of "letting it happen" is a comparatively low level of activity, especially on the active-passive spectrum; on the proactive-inhibitive spectrum, the activity of "letting things happen" still tends slightly toward the proactive pole (Hirschauer, 2016, p. 49).

The personifying interpretation is reinforced (in the perception of Josie and in that of the unknown speaker) by a system-side prosodic enactment that puts a tone of being snapped into the persona Alexa's voice. As in the case of impressive

⁶ The speaker in line 15 is not identifiable, thus we refer to an unknown male speaker (UMS).

functional performances, this humanizing quality of the system also entails an expressive utterance with a strong positive expression of emotion (“stunning”, line 15).

4. Interactional Occasions for Withdrawal/Reduction of Agency and Ascriptions of Technicality

In Section 3, we shed light on how agency is dynamically negotiated in practice. In this section, we take up this focus on the interactional negotiation by asking how the agency of the smart speaker is withdrawn or reduced by the human dialogue partner who thus claims agency for himself and asserts his superiority (Example 5). In addition, we ask how users interactionally focus on the technical character of the devices (Krummheuer, 2010, p. 324) by following their pre-programmed script structure (Example 6), which can lead to dialogue problems and to forms of computer talk (Zoeppritz, 1985; Lotze, 2016). Finally, we take into account how operating problems can lead to a negotiation of agency in a *meta-interaction space*, a specific type of social interaction between humans, which is directly related to the socio-technical human-machine exchange, in which various types of actions are carried out, such as negotiation on the usage of the smart speaker, consideration of the correct voice command, dealing with malfunctions, making the use of the smart speaker accountable to other co-present humans, and generally spoken embedding of the smart speaker in the sequential unfolding of everyday practices (Porcheron et al., 2018, p. 9; Habscheid, 2022, Ch. 4). This meta-interaction space is specifically relevant for the negotiation of agency (Example 7).

4.1 Sounding out the system’s limitations

In the following example, which is taken from a video recording of a first installation of a smart speaker in the household of Alex Kripp (AK) and Lukas Faßbender (LF), who live together in an apartment (see Example 3, Section 3.2), Alexa (AL) proactively lists what it is able to help the two users with (calculations, counting the days to the next holiday, solving maths problems, etc.).

Example 5. Post Codes

530 AL: ich kann dir bei beRECHnungen in deinem alltag helfen-
I can help you with calculations in your everyday life
531 egal ob du maßeinheiten UMrechnen,
no matter if you want to convert units
532 die tage bis zu einem feiertag ZÄHlen,
count the days until a holiday
533 oder ein matheproblem LÖsen möchtest,
or solve a maths problem

534 ich kann dir dabei HELfen.
I can help you with that

535 sag [EINFach-]
just say

536 AK: [alexa WIE viel] sieben tage inzidenzen sind es noch bis weihnachten?
alexa how many seven day incidence rates are left until christmas

537 (2.5)

538 AL: folgendes habe ich im web geFUNDen [und über]setzt;
I found the following on the web and translated it

539 AK: +[hmm KLAR;] +
hmm sure

ak: +blickt zurück auf Laptop+
looks back at laptop

540 |+((lacht)) +
laughs

ak: +blickt zunächst kurz zu LF, dann in die Kamera+
looks at LF, then to the camera

541 AL: laut () punkt ORG-
according to point org

542 +mit anderen WORTen- +
in other words

ak: +blickt zurück zu Alexa+
looks back at alexa

543 AL: +der FÜNfundzwanzigste dezember beginnt mit den zwölf weihnachtstagen,+
the twenty-fifth of december starts with the twelve days of christmas

lf: +wendet Blick von Alexa ab zum Fenster hin----->
>-----+
looks away from alexa and to the window

544 AL: +die am FÜNfften [januar-]+
on the fifth of january

lf: +blickt wieder Richtung Alexa--+
looks back at alexa

545 LF: [alexa STOPP;]
alexa stop

546 AK: §+((lacht)) + §
laughs

ak: §lehnt sich kurz zurück und direkt wieder vor§
leans back shorty and then again forward immediately

ak: +blickt Laptopbildschirm wieder an+
looks back at laptop screen

547 (1.3)

548 LF: alexa was ist die inziDENZzahl von münster?
alexa what is the number of confirmed cases in münster

549 (1.6)

550 AL: laut einem alexa () NUTzer-
according to an alexa user

551 münster hat mehrere +POSTleitzahlen; +
münster has got several post codes

lf: +wendet Blick aus dem Fenster hinaus+

looks out of the window

552 AL: je nach STADTteil gibt es [verschiedene (postleitzahlen)-]
depending on the district there are different post codes

553 LF: +[pht (-) alexa STOPP;]+
alex stop

ak: +blickt zu LF-----+
looks at LF

554 LF: alexa was sind die coRONa (-) +zahlen +
 von münster?
alex what are the covid figures
in münster

ak: +blickt zurück auf Laptopbildschirm+
looks back at laptop screen

555 (8.04)

556 LF: \$alexa ist überFORdert; \$
alex is unable to cope

lf: \$wendet Kopf Richtung AK\$
turns his head to AK

557 AK: ((lacht))
laughs

558 (1.4)

559 LF: ä:hm;
erm

560 WIE kann ich denn jetzt?
how can I now

In line 536, before Alexa finishes her turn, Alex stops her in overlap with her utterance by asking a question which – in relation to the pre-installed skills – implies a challenging arithmetic problem that would challenge human calculation competences, too: “alexa how many seven day incidence rates are left until christmas.”⁷ With his question, Alex does not follow the pre-programmed script structure that Alexa starts to propose in line 535 but ascribes a potential “smartness” to the system (Section 3.2) and demonstrates superiority as a reaction to its offer (line 530) in the context of exploring its level of calculation ability. After a short pause in line 537, in which the system seems to process Alex’ voice command, Alexa provides an answer explicitly stating the source of information (“I found the following on the web”, line 538), hinting at cloud-based background operations (Section 1) and also (implicitly) informing the user about its translation skills (“and translated”, line 538). Alexa’s output is then – in overlap with its not yet finished utterance – commented on by Alex’ evaluating and ironic comment “hmm sure” in line 539, which is followed by laughter and Alex’ turning of gaze towards his roommate Lukas and the camera (line 540), thus indicating that his comment is not addressed to the smart speaker but to his social

⁷ It is not exactly clear what kind of answer the user expects from the smart speaker after this voice command; however, it is worth mentioning that the situation transcribed here took place on December 4, 2020, which was in the middle of the COVID-19 pandemic in Germany with restrictive measures for infection control in place (such as contact restriction) and incidence rates of COVID-19 cases were one of the important markers during this time.

interaction partner in the meta-interaction space. In lines 541 and 542, the system then produces two incomplete sentences, first referencing the author of the cited information (“according to () point org”) and then projecting a paraphrase of the retrieved information (“in other words”). In lines 543 and 544, the system finally provides concrete information (“the twenty-fifth of december starts with the twelve days of christmas, on the fifth of january”) but is then stopped by Lukas (line 545), as it did not provide the requested calculation. Altering their requests, and thus seemingly following the pre-programmed script structure yet again, Lukas and Alex try to get Alexa to provide first the requested challenging calculation, then – simplifying their voice commands – the number of Coronavirus-infected people in the city they live in: Lukas utters another voice command, this time asking about the COVID-19 numbers of the city of Münster (“alexa what is the number of confirmed cases in münster”, line 548) using simpler syntax and thereby reducing the system’s level of “smartness”, or rather its linguistic and calculation ability, and also reducing its gradient of agency. In line 550, Alexa again refers to its source of information (“according to an alexa user”) and provides information about different postal codes of areas of the city Lukas and Alex live in (lines 551/552). Lukas then puts his question differently using the same syntax as before but altering the keyword (“covid figures” instead of “number of confirmed cases”, line 554). This ultimately leads to a complete stop of the system after Lukas’ question in line 554. Lukas, who was previously impressed by Alexa’s technical performance (see Example 3, Section 3.2), is now – indicated by turning his gaze towards Alex – opening up a meta-interaction space, reflecting on the system’s failure with his comment “alexa is unable to cope” (556), ascribing characteristics of a living being with a low level of intelligence to the system. What started as taking over communicative “leadership” (see Section 1) by the system and its human dialogue partners attributing agency to the device in a testing setting finally leads to reducing the smart speaker’s “smartness” (Section 3.2) or calculation ability while its human users sound out its limitations.

4.2 Adapting to pre-programmed script structures

Example 6 illustrates how users (need to) adapt to pre-programmed script structures in the mode of “computer talk” to make the device work, thereby orienting to the technical appearance and capacities (Krummheuer, 2015) of the device, and how these adaptations lead to both an ascription and reduction of the gradient of agency. The excerpt is taken from an audio recording of the routine use of the smart speaker in the same household as in Example 5 shown in Section 4.1 above, a few weeks after its first installation. In this excerpt, Lukas (LF) asks Alexa (AL) to set a timer – one of the typical functions smart speakers are often used for (Ammari et al., 2019; Natale, 2020; Pins et al., 2020; Pins & Alizadeh, 2021; Tas et al., 2019). Lukas then tries to change the set time, which leads to a malfunction and the adjustment of Lukas’ voice command.

Example 6. Timer

121 LF: aLEXa?
alex

122 p: (1.0)

123 LF: TImeR auf zwanzig minUten.
timer for twenty minutes

124 p: (1.7)

125 AL: zwanzig minuten ab JETZT;
twenty minutes starting from now

126 p: [(39.5)]

127 k: [((leise Musikwiedergabe, vereinzelt Hintergrundgeräusche))]
quiet music playback sporadic background noise

128 LF: aLEXa?
alex

129 p: (1.5)

130 LF: timer MINus fünf minUten;
timer minus five minutes

131 p: (1.7)

132 AL: zwEIter timer (.) fünf minuten (.) ab JETZT-
second timer five minutes starting from now

133 =möchtest du diesem tImeR einen namen GEben,
would you like to give this timer a title

134 p: (0.2)

135 LF: alexa: (.) lösche timer eins und ZWEI;
alex delete timer one and two

136 p: (1.3)

137 AL: du hast keinen TImeR mit der bezeichnung eins zwEI,
you do not have a timer with the title one two

138 aber es gibt TImeR für fünf minuten und zwanzig minUten.
but there are timers for five minutes and twenty minutes

139 p: (0.5)

140 LF: aLEXa:?
alex

141 p: (0.8)

142 LF: lösch: (.) TImeR (.) für fünf minuten;
delete timer for five minutes

143 p: (1.8)

144 AL: fünf miNUten timer wurde abgebrochen.
timer for five minutes was cancelled

145 p: (0.6)

146 LF: alexa lösche TImeR für zwanzig minuten.
alex delete the timer for twenty minutes

147 p: (1.3)

148 AL: zwanzig minuten TImeR wurde abgebrochen;
timer for twenty minutes was cancelled

149 p: (0.2)

150 LF: aLEXa?
alex

151 p: (0.2)

152 LF: time:r fünfzehn miNUten;
timer fifteen minutes

153 p: (0.5)

154 AL: fünfzehn minuten ab JETZT;
fifteen minutes starting from now

Lukas' voice command ("alexa (1.0) timer for twenty minutes", lines 121–123) is – after a short pause (line 124) – confirmed by Alexa ("twenty minutes starting from now", line 125). Then, after a longer pause of almost 40 seconds, Lukas voices another command ("alexa (1.5) timer minus five minutes", lines 128–130), which syntactically lacks a verb and might therefore be ambiguous to Alexa. This voice command then leads Alexa to set a second timer ("second timer five minutes starting from now", line 132) and then proactively propose to give this second timer a name ("would you like to give this timer a title", line 133) instead of changing the timer Lukas has already set. The syntactical shortening of the voice command by skipping the verb "set" could in this case be a characteristic of what is (critically) discussed under the term "computer talk"⁸ (see the early work by Zoeppritz, 1985, as well as Lotze's study on chatbots from 2016). Alexa's confirmation ("second timer five minutes starting from now", line 132) results in Lukas initiating a repair sequence, in which he first provides a clear and unambiguous voice command including a verb ("alexa delete timer one and two", l. 135). But Alexa cannot process this voice command either, because the system has not fully grasped Lukas' utterance (line 137). Alexa then proactively provides the names of the timers which Lukas set before ("but there are timers for five minutes and twenty minutes", line 138). At this point, it becomes clear that Lukas is the one who has to make an adaptation effort and thus, to a certain extent, give up his superiority over the device and reduce his gradient of agency regarding the dialogue with the system in order to be able to use the smart speaker for the purpose he has requested. However, Lukas gains enhanced agency on the level of the meta-interaction regime. Reducing his own gradient of agency by following the pre-programmed script structure of the device, it then takes Lukas three more turns (deleting the timers set before and setting a new one, lines 142, 146, 150–152) to finally get Alexa to execute his voice command. In this context, not following the pre-programmed script structure of the device on the one hand indicates a form of agency claim by the user (Section 1). On the other hand, this also means ascribing a certain degree of "smartness" (Section 3.2) and a certain level of agency to the smart speaker, which is then gradually revoked or reduced by the user in the process of adapting to be able to use it efficiently – and thus enhancing its agency.

4.3 Handling operating problems

The next and last example (7) in this section addresses the negotiation of agency in what we call meta-interaction space and shows how the agencies of the two

⁸ "Computer Talk" (CT) is referred to by Zoeppritz (1985, p. 1) as "several instances of deviant or odd formulations that looked as if they were intended to be particularly suitable to use with a computer as the partner of communication". For a critical discussion of Zoeppritz' hypotheses and the concept of CT as a structurally or functionally detectable register in users' language use, as well as for a potentially new online variant of CT and the question of whether the adoption of CT is useful for a user-friendly design, see Lotze (2016, pp. 154–178).

users and of the smart speaker are interactionally intertwined. Jan-Ole Söllner (JS) wants to test the function to call a person from his contact list via the smart speaker. Jan-Ole’s voice command “call Damaris” (line 626) leads to operating issues as he has not yet shared his contacts with the device – at least the device tells him that Damaris cannot be found in the contact list – and ultimately to a negotiation of settings and the gradient of agency he attributed to the device concerning the sharing of contacts with the device.

Example 7. *Contacts*

625 JS: aLEXa,
*alex*a

626 ruf (.) daMAris an;
call damaris

627 p: (2.0)

628 AL: \$du hast nach daMAris gefragt-\$
you asked for damaris
 js: \$scrollt durch smartphone-----\$
scrolls on smartphone

629 +aber ich kann diesen namen in deiner kontAKT+ [oder gerätelis]te
 nicht fInden.
but I cannot find this name in your contacts or device list
 dl: +blickt JS an-----+
looks at JS

630 DL: \$[geNAU;]
sure

631 wow (.) COOL;\$
wow cool
 dl: \$lehnt sich kurz
leans back shortly
 nach hinten--\$

632 p: (0.4)

633 JS: okay WOW;
okay wow

634 p: \$(2.4) \$
 js: \$scrollt auf Smartphone\$
scrolls on smartphone

635 JS: \$aber hier IS doch damaris; \$
but damaris is here
 js: \$deutet mit LH Richtung Smartphone\$
points to smartphone with LH

636 p: (3.4)

637 JS: <<leise sprechend> alter was sind denn hier überall für kontAKTe
speaking quietly dude what about all the contacts in here
 \$da IS doch damaris;>\$
there is damaris
 js: \$zeigt auf Smartphone\$
points at smartphone

638 oh.
oh

639 p: (3.1)

640 JS: <<leise sprechend> warum (.) \$KANN die das denn nich;\$>
speaking quietly why can't she do that
 js: \$tippt auf Smartphone---\$
types on smartphone

641 p: (1.3)

642 DL: musst du das erst FREIgeben,

do you have to enable that first
643 p: (0.6)
644 JS: HAB ich ja;
I have

With his voice command in line 625, Jan-Ole (JS) is testing the ability of the smart speaker to call a person from his contact list, which cannot be processed by the device (lines 628/629). Alexa's answer "you asked for damaris / but I cannot find this name in your contacts or device list" (lines 628–629) leads to exclamations of astonishment by both JS and his girlfriend Damaris (DL) (lines 630–633). They reflect on this in a meta-interaction space, a dimension for social interaction in a group of more or less involved participants in the human-machine-dialogue, which is interwoven with the socio-technical exchange, but distinct from it, as can be seen in how these terms are designed linguistically in terms of recipient design (Sacks et al., 1974, p. 727).

Looking at the app on his mobile phone which helps JS to set up the device, he realizes that somehow some contacts have already been deposited ("but damaris is here"/"dude what about all the contacts in here, there is damaris", lines 635/637). He does not have a solution at hand, as his question ("why can't she do that", line 640) shows. Whereas in the context of dealing practically with the system he puts himself in a passive position, thereby reducing his own gradient of agency, he gains agency again on the level of the meta-interaction regime by his evaluative comment.⁹ The process is finally aborted without success.

The negotiation of agency is twofold in this case of an operating problem: agency between the user JS and the device as well as agency between the two users, JS and DL. For the latter, one can say that DL tries to raise her agency with her instruction regarding technical issues (line 642), by which she claims epistemic authority (Mondada, 2013; Heritage & Raymond, 2005), which JS then downgrades with his utterance "I have" (line 644), claiming a certain level of agency, as he is the one who handles the settings and sets up the device. This is only indirectly related to the exchange between the human and the machine as it is primarily an exchange regarding the authority over the device and the instruction process.

What is more interesting here is the failure of the attempt to call DL and how this lowers both the agency of JS and DL and the agency of the device. While a completion of this task would have raised both of their agencies, in the end the abandonment of the process lowers both as they did not have the potential to act/react as planned.¹⁰ With reference to Hirschauer (2016, p. 49), one could say

⁹ We thank one of the anonymous reviewers for this differentiation.

¹⁰ The question of agency depends on the intentionality of the utterances: our interpretation refers to the command and its (non)execution in the immediate situation. If one embeds these actions in the superordinate context of testing, in which even a failed command can represent a successful sub-action, then agency can be stated here for the evaluation on the level of meta-interaction. We thank one of the anonymous reviewers for this differentiation.

that the smart speaker is an inhibitive participant of the action, appearing neither passive nor explicitly active: it undermines that JS can carry out the planned action of calling Damaris. Thus, JS's practice also appears as passive: although he active-proactively *tries* to start the call and also discusses it with DL, his agency is limited. Finally, both of them appear passive-inhibitive as they do not further search for a solution but stop the attempt to make the call.

Excerpt 7 thus shows that such complex tasks for smart speakers (including managing different user accounts and the database of JS's smartphone) could raise the agency for both the human and machine participants: the potential to carry out the call sets them both on a higher level of activity. However, it can also lead to a reduction of agency for both of them: neither the user nor the smart speaker can act as planned. This leads to a withdrawal of agency from the system (not the user) in the meta-interaction space, verbalized by JS as a question (line 640), which also underlines that he expected the smart speaker to have this function, and the illusion of agency collapses in this process. The example demonstrates that the agency of the device and the user are not dependent on each other, but the agency of one influences the other as they are interwoven in one co-operation: although the call does not work, the non-functionality of the smart speaker is accomplished together with the smart speaker, JS, and DL. Hence, the agencies of all involved parties in socio-technical dialogue are not deterministically connected but influence each other.

5. Discussion

In this paper, we empirically tested theoretical heuristics, according to which agency is not to be seen as something that certain ontological entities stably do or do not possess. Rather, we assumed that agency is a dynamic practical accomplishment, bound to the local context, socio-spatial environment, temporal unfolding of a situation, and hence most basically to the practices carried out by or rather involving contributions by participants with unequal sets of resources for participating (Krummheuer, 2015). It is consequently, whether intentional or not, a constant matter of negotiation in these situations – one could say that it is not possible to not negotiate agency in the execution of practice.

The presented cases also show that not only the human actors but also technology such as smart speakers and other participating entities can take different levels of action: following Hirschauer (2016), one can distinguish between levels of activity both on an active-passive spectrum as well as on a proactive-inhibitive spectrum. For example, a smart speaker, in keeping with the role of an intelligent assistant, can be treated in practice as a counterpart which, while not proactive, plays along at a high level of activity, for example, by telling jokes or solving challenging mathematical problems. Therefore, it is important to note that the ascription and raising as well as the withdrawal or reduction of agency is not to be seen as a simple “constant sum game”: the agency of both

the human and the technological actor can be gradually raised or reduced at the same time, but possibly in complementary spectra of activity. Thus, in certain circumstances, the assistant remains subordinate to the human user, who gives it commands, in the proactivity dimension, while it can achieve parity or even superiority in the activity dimension (Hirschauer, 2016, p. 49, for a visual representation of these dimensions). However, the relationships and status of activity always remain precarious: if (possibly demanding) tasks are delegated to the intelligent assistant, this leads to an accumulation of agency in the event of success; however, it also leads to a dependency which, in the case of technical failure, causes the entire accumulated power to act – including that of the user – to collapse. However, in such cases, the user can regain agency in the regime of the meta-interaction space.

The presented cases show a series of occasions in which agency is negotiated: situations of everyday promotion, presenting the innovative, smart technology in a favorable light to a partner with no (or only little) experience of use; contexts of “doing”¹¹ being impressed by the technology’s capabilities; and (temporary) emphatic ascriptions of emotions to the socio-technical counterpart, which can lead to an ascription, or rather a raise, of agency to the gadget – in terms of linguistic and social capabilities attributed to the VUI. Malfunctions and failure, needs for adaptations as well as the practice to switch to another interaction space, in which the VUI is not treated as a participant in the conversation anymore, lead to withdrawal or reduction of agency. There could be further case types, and what appears as an occasion for an ascription of agency in one context might appear as the opposite in another. The presented examples are thus neither a complete nor a stable list of patterns but rather examples to show how different the occasions for the negotiation of agency can be.

The question of agency in dealing with media is also linked to the question of power as it is debated in STS and media theory (Rammert, 2008; Neff & Nagy, 2018; Pentzold & Bischof, 2019; for an overview, see Buckingham, 2017). The idea that users are not simply subjected to the media – be it mass media or technical artifacts and systems – but also their own kind of agency has in the past been strongly supported by critical but constructive theories of media appropriation and domestication, for example with reference to television or smartphones (Silverstone, 1994; Faber, 2001; Ayaß, 2012). In view of discursive (and beyond) social power relations, even if there could have never been any claim of absolute user autonomy, it has often been shown on an empirical basis how media artifacts and (mass) media content could be embedded in everyday social interaction, practices, and orientations of meaning and thus used to achieve a certain agency in dealing with media technology and discourses in everyday life. Keppler (1993, p. 112), for instance, argues that the effect of mass media is only as strong as its communicative appropriation (see also Röser &

¹¹ For the notion of “doing” as an ethnomethodological conception of practices to make actions recognizable as such, see Goffman (2005[1967]).

Peil, 2012, with reference to the accessibility of the Internet in private homes). This stresses our point that the agencies of different participants in the practical accomplishment influence each other within the course of action and become a constant matter of negotiation (see below).

As digital networked media increasingly enter households in an ongoing process of “deep mediatization” (Hepp, 2020, p. 3-4), this view is fundamentally challenged, as is shown, for example, by smart speakers, which we have dealt with in this article. It is true that here, too, there is demonstrable embedding in social interaction and practice, and users assure themselves of their supposed agency or even superiority, for example, by humorously exploiting the technical weaknesses of today’s “smart” systems (see also Waldecker et al., forthcoming). On the other hand, as we have seen, those who want to make full use of the functional potential of voice-controlled assistance systems will not only have to adapt to rather rigid, engineered dialogue structures but also to action-based platform logics (Willson, 2017). This can be understood, at the level of the sociotechnical exchange itself, as a reduction of human agency, but also, similarly to explicit reflections in the meta-interaction space, as an expression of a higher-level and more complex agency.

In the process of hybrid exchange – a topic that cannot be discussed further in this article – users also reveal private information about themselves in many respects in the course of the desired interaction and use (Lau et al., 2018), which beyond the household, but having effects on it, is functionalized as structured data in exploitation contexts that are not transparent to the contributing users (Zuboff, 2018; Turow, 2021). However, it has to be stressed that even in this regard, agency in terms of social power and control is not deterministically transferred to Artificial Intelligence or any “neutral computational technique” but rather “embedded in social, political, cultural, and economic worlds” (Crawford, 2021, p. 211).

Such developments have led to a far-reaching theoretical dissolution of boundaries, in which users and households are decentralized and the necessary, heterogeneous participants in such processes are placed with Latour (2007) as equal agents in socio-material networks: socialized bodies (including physical and material technologies); interaction architectures; artifacts and their affordances; technical standards; textual and visual representations; sequentiality in interaction; and a certain (often low) degree of human consciousness participation. In contrast, we argue for pursuing the empirical delimitations of smart technologies in households without abandoning the differentiated disciplinary insights into characteristics of social and linguistic interaction, their socio-material preconditions, and the hybrid exchange between humans and technology (Krummheuer, 2010). We have shown in this regard that social interaction between two human actors is of course part of a complex practice but can nevertheless theoretically stand on its own. An analysis of socio-technical practice based on multimodal data helps to investigate the practical

accomplishment of agency, but this does not necessarily touch upon basic concepts of sequential conversation and human talk-in-social-interaction.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful remarks as well as Franziska Niersberger-Gueye for her work on the transcripts and other valuable comments on earlier versions of this paper. Our project was funded by the German Research Foundation as part of the Collaborative Research Centre “Media of Cooperation” at the University of Siegen. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 262513311 – SFB 1187.

References

- Ammari, Tawfiq, Kaye, Jofish, Tsai, Janice Y., & Bentley, Frank (2019): Music, Search, and IoT. How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction*, 26 (3), 1–28.
- Ayaß, Ruth (2004): Konversationsanalytische Medienforschung. *Medien & Kommunikationswissenschaft*, 52 (1), 5–29.
- Ayaß, Ruth (2012): Introduction. Media appropriation and everyday life. In: Ayaß, Ruth, & Gerhardt, Cornelia (Eds.): *The Appropriation of Media in Everyday Life* (1–16). Amsterdam: Benjamins.
- Ayaß, Ruth, & Bergmann, Jörg (Eds.) (2011): *Qualitative Methoden der Medienforschung*. Mannheim: Verlag für Gesprächsforschung. URL: <http://www.verlag-gespraechsforschung.de/2011/pdf/medienforschung.pdf>
- Bourdieu, Pierre (1983): Ökonomisches Kapital – Kulturelles Kapital – Soziales Kapital. In: Kreckel, Reinhard (Ed.): *Soziale Ungleichheiten* (183–198). Göttingen: Schwartz (= Soziale Welt, Sonderband 2). <http://unirot.blogspot.de/images/bourdieuKapital.pdf>
- Buckingham, David (2017): Media Theory 101: AGENCY. *The Journal of Media Literacy*, 64 (1/2), 12–15.
- Crawford, Kate (2021): *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven; London: Yale University Press.
- Faber, Marlene (2001): Medienrezeption als Aneignung. In: Holly, Werner, Püschel, Ulrich, & Bergmann, Jörg (Eds.): *Der sprechende Zuschauer. Wie wir uns Fernsehen kommunikativ aneignen* (25–40). Wiesbaden: Westdeutscher Verlag.

- Folkerts, Liesa (2001): *Promotoren in Innovationsprozessen*. Wiesbaden: Universitäts-Verlag (= Betriebswirtschaftslehre für Technologie und Innovation).
- Garfinkel, Harold (1967): *Studies in Ethnomethodology*. Cambridge: Polity.
- Goffman, Erving (2005): *Interaction Ritual. Essays in Face-to-Face Behaviour. With a new introduction by Joel Best. Interaktionsrituale. Über Verhalten in direkter Kommunikation*. New Brunswick: Transaction Publishers [Original 1967].
- Goodwin, Charles (2018): *Co-Operative Action*. Cambridge: Cambridge University Press.
- Habscheid, Stephan (2022): Socio-Technical Dialogue and Linguistic Interaction. Intelligent Personal Assistants (IPA) in the Private Home. *Sprache und Literatur*, 51 (2).
- Habscheid, Stephan, Hector, Tim, Hrcal, Christine, & Waldecker, David (2021). Intelligente Persönliche Assistenten (IPA) mit Voice User Interfaces (VUI) als ‚Beteiligte‘ in häuslicher Alltagsinteraktion. Welchen Aufschluss geben die Protokolldaten der Assistenzsysteme? *Journal für Medienlinguistik*, 4 (1), 16–53.
- Hector Tim (2022): Smart Speaker in der Praxis. Methodologische Überlegungen zur medienlinguistischen Erforschung von stationären Sprachassistenzsystemen. *Sprache und Literatur*, 51 (2).
- Hector, Tim, Hrcal, Christine, Niersberger-Gueye, Franziska, & Petri, Franziska (2022): The ‘Conditional Voice Recorder’: Data practices in the co-operative advancement and implementation of data-collection technology. *Working Paper Series Media of Cooperation* 23.
- Hennig, Martin, & Hauptmann, Kilian (2019). Alexa, optimier mich! KI-Fiktionen digitaler Assistenzsysteme in der Werbung. *Zeitschrift für Medienwissenschaft*, 11 (21), 86–94.
- Hepp, Andreas (2020): *Deep mediatization*. Abingdon, Oxon/New York, NY: Routledge (= Key ideas in media and cultural studies).
- Heritage, John, & Raymond, Geoffrey (2005): The Terms of Agreement: Indexing Epistemic Authority and Subordination in Talk-in-Interaction. *Social Psychology Quarterly*, 86 (1), 15–38.
- Hirschauer, Stefan (2004): Praktiken und ihre Körper. Über materielle Partizipanden des Tuns. In: Reuter, Julia, & Hörning, Karl H. (Eds.): *Doing*

Culture. Neue Positionen zum Verhältnis von Kultur und sozialer Praxis (73–91). Bielefeld: Transcript.

Hirschauer, Stefan (2016): „Verhalten, Handeln, Interagieren. Zu den mikrosoziologischen Grundlagen der Praxistheorie.“ In: Schäfer; Hilmar (Ed.): *Praxistheorie. Ein soziologisches Forschungsprogramm* (45–67). Bielefeld: Transcript.

Holly, Werner (2000): Beziehungsmanagement und Imagearbeit. In: Antos, Gerd, Brinker, Klaus, Heinemann, Wolfgang, & Sager, Sven F. (Eds.): *Text- und Gesprächslinguistik* (1382–1393). Bd. 2. Berlin, New York: de Gruyter 2001.

Ibnelkaïd, Samira (2019). L'agentivité multimodale en interaction par écran, entre sujet et tekhnê. In: Mazur-Palandre, Audrey, & Colón de Carvajal, Isabel (Eds.): *Multimodalité du langage dans les interactions et l'acquisition* (281–315). Grenoble: UGA Éditions.

Ibnelkaïd, Samira, & Avgustis, Iuliia (2022/this issue). Situated agency in digitally artifaced social interactions: Introduction to the special issue. *Social Interaction. Video-Based Studies of Human Sociality*.

Keppler, Angela (1993): Fernsehunterhaltung aus Zuschauersicht. In: Holly, Werner, & Püschel, Ulrich (Eds.) (1993): *Medienrezeption als Aneignung. Methoden und Perspektiven qualitativer Medienforschung* (103–113). Wiesbaden: VS Verlag für Sozialwissenschaften.

Krummheuer, Antonia (2010): *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*. Stuttgart: Lucius&Lucius.

Krummheuer, Antonia (2015): Technical Agency in Practice: The enactment of artefacts as conversation partners, actants and opponents. *PsychNology Journal*, 13 (2-3), 179–202.

Latour, Bruno (1993): *We have never been modern*. Cambridge: Harvard University Press.

Latour, Bruno (2007). *Reassembling the social: An introduction to Actor-Network-Theory. Clarendon lectures in management studies*. Oxford: Oxford University Press.

Lau, Josephine, Zimmerman, Benjamin, & Schaub, Florian (2018): Alexa, Are You Listening? *Proceedings of the ACM on Human-Computer Interaction*, 2, 1–31.

Lotze, Netaya (2016): *Chatbots. Eine linguistische Analyse*. Berlin Peter Lang.

- Luff, Paul, Hindmarsh, Jon, & Heath, Christian (Eds.) (2000): *Workplace studies. Recovering work practice and informing system design*. Cambridge: Cambridge University Press.
- Mondada, Lorenza (2013): Displaying, contesting and negotiating epistemic authority in social interaction: Descriptions and questions in guided visits. *Discourse Studies*, 15 (5), 597–626.
- Neff, Gina, & Nagy, Peter (2018): Agency in the Digital Age: Using Symbiotic Agency to Explain Human–Technology Interaction. In: Papacharissi, Zizi (Ed.): *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*. London: Routledge.
- Natale, Simone (2020): To believe in Siri: A critical analysis of AI voice assistants. *Working Paper Communicative Figurations* 32.
- Oberzaucher, Frank (2018): Konversationsanalyse und Studies of Work. In: Habscheid, Stephan, Müller, Andreas, Thörle, Britta, & Wilton, Antje (Eds.): *Handbuch Sprache in Organisationen* (307–326). Berlin, Boston: de Gruyter.
- Och, Anastasia Patricia, & Habscheid, Stephan (2022): Kann Influencing Beratung sein? Zur Pragmatizität von Beauty-Kommunikation auf YouTube – am Beispiel von „First Impressions“. In: Hennig, Mathilde, & Niemann, Robert (Eds.): *Ratgeben in der spätmodernen Gesellschaft. Ansätze einer linguistischen Ratgeberforschung* (69–96). Tübingen: Stauffenburg.
- Pentzold, Christian, & Bischof, Andreas (2019): Making Affordances Real: Socio-Material Prefiguration, Performed Agency, and Coordinated Activities in Human–Robot Communication. *Social Media + Society*, 5 (3).
- Pins, Dominik, & Alizadeh, Fatemeh (2021): „Im Wohnzimmer kriegt die schon alles mit“ – Sprachassistentendaten im Alltag. In: Boden, Alexander, Jakobi, Timo, Stevens, Gunnar, & Bala, Christian (Eds.): *Verbraucherdatenschutz – Technik und Regulation zur Unterstützung des Individuums*. Schriften der Verbraucherinformatik, Band 1.
- Pins, Dominik, Boden, Alexander, Stevens, Gunnar, & Essing, Britta (2020): „Miss Understandable“: Eine Studie zur Aneignung von Sprachassistenten und dem Umgang mit Fehlinteraktionen. *Proceedings of Mensch und Computer 2020 (MUC20)*. ACM, Magdeburg, Germany.
- Porcheron, Martin, Fischer, Joel, Reeves, Stuart, & Sharples, Sarah (2018): Voice interfaces in everyday life. In: Mandryk, Regan, Hancock, Mark, Perry, Mark, & Cox, Anna (Eds.): *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.

- Rammert, Werner (2008): Where the Action Is: Distributed Agency between Humans, Machines, and Programs. *The Technical University Technology Studies Working Papers TUTS-WP-4-2008*. Berlin: Technische Universität Berlin.
- Rammert, Werner, & Schulz-Schaeffer, Ingo (2002): Technik und Handeln: wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt. In: Rammert, Werner, & Schulz-Schaeffer, Ingo (Eds.): *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik* (11–64). Frankfurt a.M.: Campus.
- Rolf, Eckard (1997): *Illokutionäre Kräfte. Grundbegriffe der Illokutionslogik*. Opladen: Westdeutscher Verlag.
- Röser, Jutta, & Peil, Corinna (2012). Das Zuhause als mediatisierte Welt im Wandel: Fallstudien und Befunde zur Domestizierung des Internets als Mediatisierungsprozess. In Hepp, Andreas, & Krotz, Friedrich (Eds.): *Mediatisierte Welten. Beschreibungsansätze und Forschungsfelder* (137–163). Wiesbaden: VS Verlag.
- Sacks, Harvey, Schegloff, Emanuel, & Jefferson, Gail (1974): A simplest systematics for the organisation of turn taking in conversation. *Language*, 50 (4), 696–735.
- Schüttpelz, Erhard, & Gießmann, Sebastian (2015): Medien der Kooperation. Überlegungen zum Forschungsstand. *Navigationen*, 15 (1), 7–55.
- Schüttpelz, Erhard, & Meyer, Christian (2017): Ein Glossar zur Praxistheorie. ‚Siegener Version‘ (Frühjahr 2017). *Navigationen*, 17 (1), 155–163.
- Schüttpelz, Erhard, & Meyer, Christian (2018): Charles Goodwin’s Co-Operative Action: The Idea and the Argument. *Media in Action. Interdisciplinary Journal on Cooperative Media* 1, 171–188.
- Selting, Margret, Auer, Peter, Barth-Weingarten, Dagmar, Bergmann, Jörg, Bergmann, Pia, Birkner, Karin, Couper-Kuhlen, Elizabeth, Deppermann, Arnulf, Gilles, Peter, Günthner, Susanne, Hartung, Martin, Kern, Friederike, Mertzluft, Christine, Meyer, Christian, Morek, Miriam, Oberzaucher, Frank, Peters, Jörg, Quasthoff, Uta, Schütte, Wilfried, Stukenbrock, Anja, & Uhmann, Susanne (2011): A system for transcribing talk-in-interaction: GAT 2 translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten. *Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion*, 12, 1–51.
- Silverstone, Roger (1994): *Television and everyday life*. London: Routledge.

- Suchman, Lucy (1987): *Plans and situated actions. The problem of human-machine communication*. Cambridge: Cambridge University Press [reprint].
- Suchman, Lucy (1990): What is Human-Machine Interaction? *Cognition, computing and cooperation 1990*, 25–55.
- Suchman, Lucy (2007): *Human-machine reconfigurations. Plans and situated actions* (2nd edition). Cambridge: Cambridge University Press.
- Sweeney, Miriam, & Davis, Emma (2020): Alexa, Are You Listening? *Information Technology and Libraries*, 39 (4).
- Tas, Serpil, Hildebrandt, Christian, & Arnold, René (2019): *Sprachassistenten in Deutschland, WIK Diskussionsbeitrag, No. 441*. WIK Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste, Bad Honnef. URL: <http://hdl.handle.net/10419/227052>
- Turow, Joseph (2021): *The Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet*. Yale: Yale University Press.
- Tomasello, Michael (2008): *The Origins of Human Communication*. Cambridge, London: The MIT Press.
- Waldecker, David, Hector, Tim, & Hoffmann, Dagmar (forthc.): Intelligent Personal Assistants in practice. Situational agencies and the multiple forms of cooperation without consensus. To be published in: Görland, Stephan, Roitsch, Cindy, & Hepp, Andreas (Eds.): *Special Issue "Agency in a Datafied Society. Communication between and across humans, platforms, and machines"*.
- Willems, Herbert, & Kautt; York (2003): *Theatralität der Werbung. Theorie und Analyse massenmedialer Wirklichkeit: Zur kulturellen Konstruktion von Identitäten*. Berlin, New York: de Gruyter.
- Willson, Michele (2017): Algorithms (and the) everyday. *Information, Communication & Society*, 20 (1), 137–150.
- Zoeppritz, Magdalena (1985): *Computer talk? Technical Report TN 85.05*. Heidelberg: IBM Heidelberg Scientific Center.
- Zuboff, Shoshana (2019): *The Age of Surveillance Capitalism*. New York: Public Affairs.