Stephan Habscheid, Tim Hector,
Dagmar Hoffmann, David Waldecker (eds.)

VOICE ASSISTANTS IN PRIVATE HOMES

Media, Data and Language
in Interaction and Discourse

[transcript] Media in Action

Stephan Habscheid, Tim Hector, Dagmar Hoffmann, David Waldecker (eds.)
Voice Assistants in Private Homes

Media in Action | 7

Editorial

The open access book series "Media in Action", conceived by the DFG Collaborative Research Centre "Media of Cooperation", examines the history and present of networked, data-intensive media and their social implications at the interdisciplinary interface of social and media studies. In the tradition of science and technology studies and actor-network theory, German and English monographs, edited volumes and dissertations of the series focus on practices, (co-)operations and procedures in the use, production and analysis of old and new media. A central challenge the series faces is the development of appropriate ethnographic, digital, sensor-based and design-oriented methods for a new conception of the description of distributed agency between people, computers, bodies and environments.
Funded by the German Research Foundation (DFG) – Project number 262513311 – SFB 1187.

Stephan Habscheid (Prof. Dr.) is a professor of German studies and applied linguistics at Universität Siegen. He is principal investigator of the interdisciplinary project "Un/desired Observation in Interaction: Smart Environments, Language, Body and Senses in Private Homes" at the Collaborative Research Center 1187 "Media of Cooperation", Universität Siegen (together with Dagmar Hoffmann). His research interests include media linguistics, linguistic praxeology, language in institutions and organizations as well as small talk and conversation.
Tim Hector (M.A.) works as a research assistant at the Collaborative Research Center 1187 "Media of Cooperation" in the project "Un/desired Observation in Interaction: Smart Environments, Language, Body and Senses in Private Homes" at Universität Siegen. He did a PhD in applied linguistics on the linguistic domestication of voice assistants in private homes. His research is focussed on media linguistics, conversation analysis and linguistic praxeology.
Dagmar Hoffmann (Prof. Dr.) is a professor of media sociology and gender media studies at Universität Siegen, Germany. She is principal investigator in the interdisciplinary project "Un/desired Observation in Interaction: Smart Environments, Language, Body and Senses in Private Homes" at the Collaborative Research Center 1187 "Media of Cooperation", Universität Siegen (together with Stephan Habscheid). Her research is focused on media and cultural sociology, digital literacy, and political participation.
David Waldecker (Dr.) is a sociologist and an academic librarian in training at Universitäts- und Landesbibliothek Darmstadt. He was a post-doc at the Collaborative Research Center 1187 "Media of Cooperation", Universität Siegen, and published his dissertation on Adorno in the recording studio in 2022.

Stephan Habscheid, Tim Hector, Dagmar Hoffmann, David Waldecker (eds.)

Voice Assistants in Private Homes

Media, Data and Language in Interaction and Discourse

[transcript]

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 262513311 – SFB 1187 "Media of Cooperation".

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://dnb.dnb.de

transcript Verlag | Hermannstraße 26 | D-33602 Bielefeld | live@transcript-verlag.de

Cover design: Maria Arndt
Cover Illustration: Bence Boros / Unsplash
Proofreading: Pip Hare
Printing: Elanders Waiblingen GmbH, Waiblingen
https://doi.org/10.14361/9783839472002
Print-ISBN: 978-3-8376-7200-8
PDF-ISBN: 978-3-8394-7200-2
ISSN of series: 2749-9960
eISSN of series: 2749-9979

Printed on permanent acid-free text paper.

Contents

II Linguistic Exchange with Voice Assistants as a Practical Problem

III Privacy and Data Protection as Practical Problems

IV Technical Infrastructures as a Practical Problem

Voice Assistants in Private Homes. Introduction to the Volume

Stephan Habscheid, Dagmar Hoffmann, Tim Hector, and David Waldecker

1. The Emergence of Voice Assistants

In 2011, together with the new iPhone 4S, Apple launched a voice assistant called "Siri"[1], which it claimed could understand questions and commands in spoken language (initially in English, German, and French) and respond to them as a human conversation partner would (Huq 2011). The announcement was met with great fascination: here was a talking technology, the first "intelligent personal assistance" system (IPA) to promise to make everyday life easier. With IPAs, it looked like a very popular, prototypical motif of science fiction was finally to become reality (Stresing 2011)[2]. Just a few years after "Siri", Amazon followed suit with "Alexa" (2015) and Google with its "Google Assistant" (2016) (cf. Dürscheid 2023), and by the end of the decade, the systems had become increasingly established in private households (Statista 2021). As well as in smartphone apps, voice assistants have been finding their way into various everyday devices, such as smart speakers, smart TVs, smart watches, or the media interfaces of digitally connected cars.

In recent years, however, the high-flying economic and technological hopes initially pinned to voice assistant technologies have been critically reappraised. In 2022, an article in *Business Insider* asserted that billions of invested dollars

1 Apple (under CEO Steve Jobs) had bought the company of the same name, which had been founded in 2007 and developed the product in 2010 (Wikipedia 2024; see also Dürscheid 2023).

2 This is not the place to retell the media history of talking machines (see, e.g., Volmar 2019). For a detailed media theoretical and linguistic description of the technologies relevant here with a specific focus on smart speakers, see Hector (in preparation).

had been lost, and hoped-for profits had not been achieved – a “colossal failure of imagination”, in the words of a former employee of Amazon (Kim 2022). According to the *Business Insider* article,

> most of those conversations were trivial, commands to play music or ask about the weather. That meant fewer opportunities to monetize. Amazon can’t make money from Alexa telling you the weather – and playing music through the Echo gives Amazon only a small piece of the proceeds. (Kim 2022)

In addition, since the introduction of new language processing technologies such as text-generating ChatGPT, which was launched in 2022, the earlier voice assistant technology risks appearing unwieldy in comparison (Dürscheid 2023). Against this backdrop, Amazon and other companies are apparently trying to integrate modern generative AI into older voice assistance systems. According to press reports, an assistance system presented by Amazon in 2023 spoke

> in a far more natural and conversational voice than the friendly-but-robotic one that hundreds of millions have become accustomed to communicating with for weather updates, reminders, timers, and music requests. (Goldman 2024)

According to the reports, this ‘new Alexa’ engaged more naturally in conversations, delivered more natural voice output, and had a more pronounced personality (ibid.). However, it seems that the version demonstrated has not yet been convincingly implemented into the real performance of the systems (ibid.). Thus, the American magazine *Fortune* has claimed that Amazon and Apple – once pioneers in the development of talking machines – are now “desperately behind [their] Big Tech rivals Google, Microsoft, and Meta in the race to launch AI chatbots and agents, and floundering (in [their] efforts to catch up)” (Goldman 2024). One reason given for this is that the characteristic technological architecture of older voice assistants is required to retain certain characteristics in order to maintain existing features, but therefore is no longer up to date enough for the integration of recent AI. In addition, these circumstances make it difficult to collect or synthetically produce suitable linguistic training data for the further development of the voice assistants. Citing former employees, the article reports that Amazon has therefore repeatedly deprioritized the further development of Alexa to focus on the development of generative AI for its cloud computing unit ("Amazon Web Services") – which could see the existing technology soon becoming a "digital relic" (ibid., see also Herbig 2024).

On the other hand, current usage studies show that the number of devices with voice user interfaces for different smart technologies is continually increasing. Amazon has not confirmed the reports about Alexa's economic failure, and it is evidently continuing to invest in such products (Amazon 2023). For example, further development of devices that combine voice user interfaces with camera, monitor, and touch interfaces seems to be ongoing. As Niklas Strüver (2023) points out, smart speakers are conceived as the central hub for the smart home – a field of consumer tech that is clearly continuing to gain ground. Thus, in the smart home, devices like smart speakers are what allow users to manage the entire orchestration of multiple interconnected smart home applications related to the kitchen, housekeeping, or security. While technology companies see internet-enabled devices in the home as a way to increase demand for many such products and associated services, critics point out that many of the devices are too expensive for most consumers and will take years to catch on.

Either way, there are ample reasons to examine language-processing machines and their future development from the perspective of interaction research and linguistics. Not only does human–machine dialogue offer a fruitful field for investigation, it also points to potential new approaches to research on human–human interactions, as Karola Pitsch (2015) has shown with the example of co-constructions: familiar conversational procedures are "broken open", making analytical access to more basic conversational phenomena possible. Furthermore, as Martin Porcheron, Joel Fischer, Stuart Reeves, and Sarah Sharples observed in 2018, social interaction among co-present participants changes when the use of machines is incorporated. The linguistic contributions in this volume address human–machine communication as well as human–human communication and can be read together as an overview of current research in this field.

However, many other academic disciplines also address the phenomenon of human–computer interaction (HCI), albeit from different research perspectives. In the social sciences, the focus tends not to be primarily on usability or usage modalities, the skills that people need to have in order to operate the devices, but above all on exploring how devices integrated into everyday life are changing the ways we live together, how new media and data practices are developing, and how privacy is being reinterpreted (e.g., Burgess et al. 2022; see Ochs, this volume). Within the social sciences, a field of research is emerging that builds on existing theoretical paradigms (including actor-network theory, diffusion research, science and technology studies, surveillance studies, and mediatization research), but which is also developing new innovative and complex methodological approaches.

2. Controversial Discourses, Household Publics, and Everyday Practices

Assessments of voice assistants in public discourse vary widely (see Habscheid, Hector, and Hrncal, this volume). On the one hand, they are advertised as an addition to a digitally-connected and thus smart lifestyle (Hennig and Hauptmann 2019). As assistance systems, they are said to have the potential to compensate for handicaps and facilitate a self-determined life for older people (Endter, Fischer, and Wörle 2023). On the other hand, they are also subject to critique, because the devices provide manufacturers with users' voice data from a particularly sensitive context, the private domestic sphere (Sadowski 2020; Turow 2021), largely as a result of "cooperation without consensus" (Waldecker, Hector, and Hoffmann 2023; for the concept see Star 1993). Although voice-controlled assistance systems are embedded in social interaction and everyday practice, those who want to make full use of their functional potential must adapt to technologized dialogue structures and platform logics. In doing so, they have to reveal a lot about themselves that is transmitted beyond the household as 'data' where it can be analyzed and exploited in ways and for purposes that are opaque to the user. Furthermore, the creation of social order in such contexts can be distorted by problematic biases (see, for example, Leblebici in this volume).

It is in smart home environments that assistance systems as central interfaces come into their own, while at the same time opening up the household to the outside world far more than ever before. Whereas classic smart speakers' capacity for surveillance was limited to the perceptual mode of hearing (on "eavesdroppers" in physical or electronically mediated presence, see Goffman 1981, 132), smart homes incorporate camera-, monitor- and sensor-based systems and networks including various stationary and mobile devices and infrastructures, which can massively expand the scope for data collection. Under certain circumstances, this is accompanied by a further dissolution of the boundaries of privacy, which on the one hand (e.g., in the case of surveillance of household members) may be perceived as abuse, but on the other hand (e.g., from a security perspective) may seem desirable.

As with all media, it is an open question as to how users will continue to adapt to new forms of media and how they make media adapt to the circumstances of their everyday lives. Individuals and households follow public discourse and interpret it in the light of their own household's public sphere, their concrete living conditions and interests. There is an ongoing debate within the humanities and social sciences, and especially within the domestication research paradigm (Hartmann 2023; Hector et al. 2023) on the adoption of digital media in household use settings. In principle, domestication research is based upon an analogy drawn between the process whereby media are appropriated and the process whereby cohabitation with farm animals or pets is established in the course of civilization. Domestication research, as summarized by Waldecker and Hector (2023, 5) "paints media as something that comes into the everyday life of users as foreign and wild, as something that has to be tamed and brought to relate to domestic routines". The metaphor of "taming" emphasizes the somewhat unpredictable and sometimes even threatening aspect of media technologies. This contrasts with the private household that often symbolizes a sense of security. With reference to Giddens (1984), Waldecker and Hector point out that this "ontological security" fundamentally establishes trust, supposedly guarantees the stability of one's own identity, the continuity of life and of the immediate environment. (Media) technologies that become entangled with this ontological security challenge it and can disrupt it: They become involved in everyday rituals, and even if ontological security is initially called into question by new media technologies (see Silverstone et al. 1992, 17), they (often) lose their threatening character as they are successively woven into everyday life, i.e., they become domesticated (see Bausinger 1984, 349–350).

In the process, everyday routines take new forms, and new practices emerge. First of all, new practices are required to get the novel devices and services to work at all. Further practices serve again and again to overcome the systems' technical unwieldiness and resistance. At the same time, the new usage practices become more or less deeply embedded in everyday life (see Waldecker and Hector 2023): They may (re-)shape, for example, the structuring of time between 'work' and 'leisure', ways of dealing with privacy, or the design of rooms and furnishings in the home. When users live together with other people in households, they must negotiate among themselves who uses which media, when, and how. In such contexts, economic decisions are also discussed in connection with political and ultimately moral issues, such as whether to subscribe to streaming services, and if so, from which provider(s). Deliberation of such questions involves not only members of the household with its own power dynamics, but also voices from beyond the home; advice may be asked of friends, or sought in online forums or among reviews in which "online warm experts" reflect in accessible language on the possible uses of consumer technology as well as their limitations (Neville 2021; see also Waldecker and Hoffmann 2023).

3. Media Appropriation as a Linguistically Mediated Practice

Changing everyday practices as a consequence of media use is also observable at the linguistic level of everyday practice, all the more so when the technology concerned has a linguistic surface. This is the case, for example, for television, which is one of the classic mass media that has attracted particular interest in domestication and appropriation research. Unlike smart speakers, television does not require verbal input from users, neither at the level of content nor at the level of operation. Television broadcasts unidirectional communication, yet users have been shown to participate nonetheless. For television, "parainteraction" is characteristic, as Ayaß (1993) – drawing on Horton and Strauss (1957) as well as on Horton and Wohl (1986) – has shown: In unidirectional communication, forms of direct address and staged connection to everyday practices are used by on-screen performers to create an impression of interaction with those watching. Such utterances counterfactually imply that bidirectional interaction 'through' the screen could be possible (see also Böckmann et al. 2019, 145), and under certain circumstances, viewers pick up on this with forms of "parasocial" pseudointeraction in front of the screen (Ayaß 1993, 36).

The fact that in many cases the use of media is anchored in linguistic and interactional practice has been emphasized especially strongly by linguistic studies. These have addressed, among other topics, speaking while watching television together (Holly, Püschel, and Bergmann 2001) and intermission talk in theater (Gerwinski, Habscheid, and Linz 2018). It has been shown that viewers use the semiotic material their TV brings into the home as a resource for mutual "orientation" with respect to public issues (Holly 2001, 11–13). The studies also revealed that the appropriation of media – technologies as well as content – is affected not least by the possibilities of linguistic interaction during and after reception. Examples include the format of "response cries" (Goffman 1981) and other forms of "terse speaking" (Baldauf 2002) in television-accompanied speech or reenactments and other reconstructive genres related to theater dialogues in intermission talk (Schlinkmann 2021). Accordingly, to study the appropriation of smart speaker technologies, it is necessary to ask how the linguistic conditions of their use enable and limit appropriation.

Unlike traditional television, internet technologies are two-way media: To put it bluntly, they not only bring the world into the household, but also the household into the world, with the latter in the form of specifically collected, aggregated, and processed data. The use of this data impacts on everyday life in ways that are noticed but cannot be traced, for example, in the form of personalized advertising or sensor-based environments in the smart home that adapt to usage habits. Thus, in the case of digital household technologies, not only are digital media domesticated in the home, households are also "externalized" (Brause and Blank 2020), or, in Hepp's words "deeply mediatized" (Hepp 2020).

The world that comes into the home with smart speakers is also linguistic on the surface – to a certain extent, it resembles the spoken language of interpersonal interaction. However, linguistic exchange with the machines differs not only in that dialogue involves non-human conversation partners, but that the technical language-processing systems upon which the latter depend have a limited ability to cooperate (Suchman 2007). The linguistic contributions to this volume discuss the range of forms such conversation can take: focusing on the human–machine dialogues, the social interaction they take place within, and the everyday practices that are realized – or not – as a result.

The sociotechnical relationships under discussion also raise fundamental questions for social theory. From a conversational linguistics perspective, the ANT approach, whereby all participating entities are conceptualized as equally significant actants (Latour 2005) seems unsatisfactory to us. For example, language-processing machines like Alexa are participants in practice, but not participants in social interaction as it is understood by conversation analysis (Habscheid 2023; Hector, in preparation; Habscheid, Hector, and Hrncal, this volume). From an ethnomethodological perspective, it can be shown that users orient towards machinic conversation partners with attitudes that, depending on the situation, sometimes reflect a more anthropomorphizing and at other times a more instrumentalized approach to the technology. Accordingly, Antonia Krummheuer (2010) characterizes the sociotechnical dialogue with an embodied conversational agent (ECA) as a "hybrid" or "ambiguous" exchange:

The exchange between human and machine shows similarities to interpersonal interaction, which is simulated to a certain extent (see also Hennig and Hauptmann, 2019), but also differences that require users to adapt to the limited communication capabilities of the machines (see also Lotze, this volume). Agency of the voice assistants is an object of negotiation both in everyday practice (Habscheid, Hector, and Hrncal 2023) and at the level of public discourse (Lind and Dickel 2024).

4. Smart Speaker Use and the Social Consequences for Everyday Reality

The use of digital technology is just as much a part of everyday life as the use of many other devices and communication with people who are physically present (Keppler 2018, 73). With the integration of a smart speaker into one's private household, this is extended by a technical artifact that is designed to function as a kind of interaction partner. Based on studies of social robotics, Michaela Pfadenhauer and Tobias Lehmann (2021) propose that a smart speaker can also be regarded as an "artificial companion" in everyday life. Smart speakers are expected to execute various commands as reliably as possible, search for and provide information, manage operation of networked devices, and offer services. Although their dialogue capabilities are still limited (Habscheid 2023) and communication is prone to disruption and often inconclusive or unpredictable (Pins et al. 2020; see also Lutz and Newlands, this volume), it can be assumed that this will improve significantly in the future, not least through the implementation of artificial intelligence. As an everyday companion, the smart speaker is certainly part of household communication: as an omnipresent third party. This participation at the locus of everyday life not only creates a social and emotional relationship with the device or with devices, but will also change how we communicate socially in everyday life. In the words of Hepp (2015), the communicative figuration of households, i.e., the communicative arrangement and role behavior of their members, is currently undergoing transformation. It is therefore of sociological interest to explore the extent to which the artificial companions can be regarded as "vehicles to cultural worlds of experience" (Pfadenhauer and Lehmann 2021) and prompt new fundamental questions of sociality (see also Hepp et al. 2022).

Furthermore, sociological investigation into sociotechnical practices and their consequences for the protection of privacy is called for. Through the appropriation and use of smart technology, users reveal personal data about themselves (such as their taste in music, their shopping behavior, their account data, their everyday routines, their address book) and allow their home environment to be (acoustically) recorded. Huge volumes of data are transmitted to tech companies, stored, and evidently used as training data or for other purposes. Users are not always aware of this and it is largely beyond their control, although within the EU at least the Digital Service Act is intended to ensure greater transparency (see the conversation with Nikolai Horn in this volume). On this matter, it is important to examine users' own attitudes and explanations for how they deal with data protection and privacy. The narrative 'nothing to hide, nothing to fear' is expressed by many users as a pragmatic approach to data protection settings and issues for a variety of reasons (see Waldecker, Martin, and Hoffmann, this volume). Existing studies of ways of dealing with and justifying decisions concerning the data protection settings of digital applications have tended to neglect to consider those indirectly affected, such as visitors to households in which such devices are installed and used as a matter of course (e.g., Hoffmann 2023). Discourses on media and critiques of corporate data practices not only shape public debate, but are also negotiated in the private sphere (see Vermeulen and Mols, this volume). It remains to be seen how these smart technologies and media practices will 'conventionalize' in the future and how social scientists will study the ongoing developments.

5. On the Contributions in this Volume

This volume presents a wide spectrum of recent research on voice-operated systems and services, including analyses focusing on their (linguistically mediated) use and appropriation, on users' appraisals of them, and on the question of the exploitative utilization of the data they transmit. Perspectives from conversation analysis and media linguistics, media sociology, media studies, surveillance studies, the critique of political economy and related aspects of consumer research, domestication research, pragmatist and praxeological sociology as well as critical theory are brought together to shed light on the practical entanglement of users, devices, algorithms, data, and corporate interests. By encompassing these diverse approaches, this volume sets out to analyze the phenomenon of IPAs at multiple levels: from that of interaction, to everyday practices in households, to the level of users' perceptions and evaluations, and not least in relation to global processes of data processing and exploitation. Our aim is to provide a comprehensive view of the transformation and persistence of everyday practices under platformized conditions and usage practices mediated by novel interfaces.

The majority of the contributions to this volume have evolved from presentations given at the conference "Voice Assistants in Private Homes. Media, Data, and Language in Interaction and Discourse", which took place on May 8 and 9, 2023, at the University of Siegen, Germany, organized by the research project "Un/desired observation in interaction: Intelligent Personal Assistants", which from 2020–2023 empirically investigated media practices with voice assistants as a key technology in the field of data-intensive digital media, taking a dual approach combining media sociology (Waldecker, Martin, and Hoffmann, this volume) and applied linguistics (Habscheid, Hector, and Hrncal, this volume). The project was part of the Collaborative Research Center "Media of Cooperation", which brings together numerous sub-projects investigating diverse phenomena but all taking as their point of departure a praxeological media theory paradigm that conceptualizes practice as the "mutual making of common goals, means and processes" and, in this context, media as "cooperatively created conditions of cooperation" or, in short, as "media of cooperation" (Schüttpelz 2017, 24). The "means" that can be cooperatively produced as "media" can – but do not have to – be of a linguistic nature (see Goodwin 2018; Habscheid, Hector, and Hrncal, this volume). In accordance with the interdisciplinary agenda of the Collaborative Research Center, as editors of this volume we seek to examine the complex phenomenon of data-intensive, AI-based assistance systems by addressing its multiple layers. The aim is to shed light on the intricate interrelationships between use and users, language, devices, algorithms, data, organizations, and economic exploitation.

The volume is structured in four parts. The first section – **Voice Assistants in Private Homes. Conceptual Considerations –** focuses on the theoretical foundations of key areas of IPA research and showcases various methodological approaches and findings of empirical studies. **Carsten Ochs** begins by examining the affective reactions of people who wonder why the users of smart speakers seem so unconcerned about their privacy. He traces the emergence of the modern practice of privacy protection, which was established in the 20th century, and now, since the advent of smart technologies in private homes, is being renegotiated. Ochs attempts to show what actually happens to the data collected and processed by smart speaker infrastructures that reach into private homes, and concludes that the term "surveillance capitalism" is an apt one under the circumstances. Taking a Marxist approach and drawing on feminist theory, **Markus Kienscherf**'s contribution investigates the role of voice assistants in the reproduction of labor and capital. The appropriation of user-generated voice data by smart speakers is positioned within a more general history of the role of surveillance in the (re)production of capitalist social relations. The author shows that surveillance is central to the appropriation of surplus value in the spheres of production, the social reproduction of labor power, and the management of circulation and consumption. He then looks at the business models of tech companies and argues that the appropriation of user-generated data transmitted via smart speakers represents an extension of capitalist surveillance into the sphere of social reproduction.

The chapter by **Caja Thimm**, **Phillip Engelhardt**, and **Julia Schmitz** deals with anthropomorphism and communication accommodation to voice assistants. The focus is on how assistance systems with VUIs (voice user interfaces) are used and affectively engaged with in multi-person households, based upon a case study with households including physically impaired people with special support needs. The authors observe that these users' assumptions, attitudes, and expectations were not stable but varied according to contextual factors. As a theoretical basis for the research, "Communication Accommodation Theory" (CAT) is developed and adapted for the study of HCI constellations, focusing on strategies of anthropomorphization, which are shown to partially – and perhaps increasingly – influence the ways people interact with machines as well as to shape the discourse, interface design, and self-image of users. Last but not least, the authors reflect on the different insights into usage gained by their methodological combination of interviews and media diaries. The last contribution in this first section by **Netaya Lotze** traces the development of a complex sociolinguistic model that can bring findings concerning the anthropomorphization of HCI technologies together with evidence of cognitive and linguistic adaptation to the (more or less) limited communicative capacity of machines. After a comprehensive research overview, Lotze presents the results of her own studies conducted since 2000, which she summarizes and interprets in the light of the model (and vice versa). The model integrates various approaches from the philosophy of language, computer science, cognitive science, and linguistics, and is structured to take into account 'external factors', 'system variables', and 'user variables', while incorporating a user typology as well as enabling diachronic analysis.

Section 2, **Linguistic Exchange with Voice Assistants as a Practical Problem** presents studies from the field of linguistics examining the practical use of and critical discourse about language assistants. The chapter by **Florence Oloff** provides an empirically underpinned perspective on the usability and learnability of voice assistants as everyday technologies. Oloff examines specific instances of older users' first encounters, during adult education courses, with hitherto unknown voice-operated applications. She shows how, in nonprofit, professionally guided practical training sessions, participants explore the potential benefits and problems of multimodal interfaces – the first stage of appropriation. Furthermore, a mismatch between the actual needs of the learners, the spatial, temporal, and medial limitations of the settings, and the teaching methods used by instructors to deal with these factors in an improvisational way becomes clearly evident. Oloff makes some suggestions on how to improve teaching and learning in these contexts. The contribution by **Didem Leblebici** provides insights into the experiences of Turkish-speaking users of non-Turkish-speaking voice assistants in Germany. The author expands upon a media linguistics interest in voice user interfaces by drawing on theoretical understandings of multilingualism from sociolinguistics and critical discourse analysis. The chapter, which is based on the linguistic analysis of ethnographic interview data, advances a critical discussion of the ways that language-processing technologies reinforce the standardization of language. Detailed examples are drawn upon to illustrate and analyze different phenomena of stylized language use in interaction with IPAs. A contribution by **Stephan Habscheid**, **Tim Hector,** and **Christine Hrncal** concludes this second section. The authors present an overview of the results to date from the linguistic strand of the project "Un/desired observation in interaction: Intelligent Personal Assistants". In the theoretical part of the chapter, the conceptual foundations of the "Media of Cooperation" Collaborative Research Center are further elaborated from a linguistic praxeological perspective, discussing approaches taken in interaction research on the one hand and linguistic media research on the other, as well as the domestication approach in media and communication research. This is followed by analyses of empirical findings from the research project, which underscore how instrumental linguistic practices are in embedding smart speakers into domestic routines, and illustrate how newly acquired technology reshapes social practices and communication within households.

The third thematic section brings together contributions that deal with the issues of **Privacy and Data Protection as Practical Problems**. Concerns relating to the extraction of personal data and its subsequent use have been raised for almost as long as digital media technologies and applications have been available, with the disclosure of data from the domestic sphere often attracting particularly critical attention. One instance when contextual privacy is called into question is when devices behave unexpectedly. Glitches occur, which can appear as technical anomalies and expose critical privacy vulnerabilities. Taking a glitch studies approach, **Christoph Lutz** and **Gemma Newslands** focus on users' experiences of malfunctions, which can also have wider societal implications and raise questions about surveillance, data security, and the ethical responsibilities of technology companies. Although glitch studies is an interdisciplinary field that tends to use qualitative methods, Lutz and Newslands draw on quantitative data to identify the four most common glitches experienced by Amazon Echo users and how they categorize the consequences of those glitches in relation to levels of trust and concerns about privacy. The findings highlight a critical aspect of smart speaker technology: the delicate balance between their perceived benefits and the fears of potential negative consequences of using them. Such considerations and fears also play a major role when people decide whether (or not) to purchase voice-operated devices in the first place. In their chapter, **Jasper Vermeulen** and **Anouk Mols** present a multi-methods study that investigated the privacy perceptions of users and non-users of smart speakers. Based upon data from in-depth interviews and focus groups, they elaborate on Dutch users' and non-users' assessments of risks and benefits. They found that users generally appreciated affordances such as controllability, support, conversation, linkability, and recordability, while some indicated they would prefer greater transparency regarding corporations' use of data. Non-users associated recordability and locatability with privacy risks that were seen as significant enough to not use such technology at all. In addition to rational considerations, the study also pointed to the role of emotions in shaping adoption considerations and decisions. **David Waldecker**, **Alexander Martin**, and **Dagmar Hoffmann** also look at users' attitudes towards data protection issues in connection with the use of smart speakers and, in particular, how they deal with them. In doing so, they draw on various studies that show the extent to which users of digital media technologies develop a kind of "online apathy", "data protection cynicism", or even "digital resignation". Based on qualitative interviews with smart speaker users in Germany, the authors report how users cultivate certain attitudes towards the devices and the discourse surrounding them and how they explain their usage routines and pragmatic considerations. In addition to the findings of the studies cited, the authors' analysis of their own interviews reveals an attitude that Andreas Pettenkofer has termed "pragmatic fatalism". Users who adopt such a stance more or less accept the data practices of companies and at the same time declare them to be irrelevant to their everyday lives. To conclude this section, **Nikolai Horn**, data protection expert and currently political advisor to iRights.Lab, discusses in a conversation with Dagmar Hoffmann and David Waldecker the legal and political aspects of protecting voice-based data. The new possibilities offered by AI and natural language processing are also addressed. Questions are raised about the extent to which voice recordings can be used to draw conclusions about identity characteristics of users and how voice recordings could be misused. The interview also explores the question of how users can be made more aware of data protection issues and how EU regulations such as the GDPR can ensure greater transparency in data use and give users more control of their own data.

The final (fourth) section – **Technical Infrastructures as a Practical Problem** – brings together a contribution from the field of social informatics and one from the sociology of technology to focus more explicitly on the IT processes and infrastructures that enable smart speaker technology but are not always transparent for users. Over a period of three years, **Dominik Pins**, **Fatemeh Alizahdeh**, **Alexander Boden**, **Sebastian Zilles,** and **Gunnar Stevens** used the living lab approach to investigate users' uncertainties with regard to the data collected as a consequence of their use of smart speakers in everyday life. Based on findings from interviews, field research, and participatory design workshops with 35 households, the authors developed a tool called "CheckMyVA" that supports users in accessing and visualizing their own VA data. The observations and findings presented in the chapter offer suggestions for tools and design strategies that could foster data literacy and enable users to reflect on their long-term interactions with VAs, ultimately "demystifying" the technology. The final chapter, by **Niklas Strüver,** takes a look behind the scenes to explore the practices involved in the ongoing development of automatic language processing. Amazon was once a pioneer in this field, but the launch of new large language models (LLMs) has posed major challenges for the company. Strüver conducted expert and narrative interviews with participants from university research teams who competed in the most recent Alexa Prize Competitions (APCs) to advance Alexa technology. These interviewees are able to offer fascinating insights into development practices, especially concerning the integration of LLMs into existing technology. Examining how the participants in these competitions deal with the conditions set by Amazon and the resources it makes available to competitors, Strüver outlines some of the path dependencies, risks, benefits, and structuring aspects that participants encountered in their attempts to innovate Alexa.

It can be summarized that research in the field of smart technologies will certainly continue to be necessary, and that lines of inquiry are always shaped by disciplinary conventions, hence interdisciplinary exchange should continue to be promoted in the future.

Acknowledgements

As stated above, most of the contributions to this volume were developed from presentations held at the conference "Voice Assistants in Private Homes. Media, Data and Language in Interaction and Discourse" organized by the research project "Un/desired observation in interaction: Intelligent Personal Assistants". We express our gratitude to all those who participated in the conference and to all the authors who have contributed to this book. We would also like to thank our student assistants Chris Dangelmaier, Sarah Diehl, Aileen Halbe, Johanna Klein, Alexander Martin, Franziska Niersberger-Guéye, and Leonie Tittel, who helped to make the conference such a success.

The research project was part of the Collaborative Research Center "Media of Cooperation" at the University of Siegen, Germany, in its second funding phase (2020–2023). In its first funding phase (2016–2019), the project was led by Wolfgang Ludwig-Mayerhofer and investigated the strategic practices by which young people using social media tried to attract 'requested' attention and avoid 'undesired' observation. After Wolfgang Ludwig-Mayerhofer's retirement, Stephan Habscheid and Dagmar Hoffmann took over as principal investigators in the second funding phase, continuing to follow the general idea of un/desired observation but applying it to smart speakers with VUIs. This volume is a key publication that resulted from the research conducted in those four years. In its third funding phase (2024–2027), the project is once again shifting its focus, this time to smart home environments, in order to explore further transformations accompanying the recent massive expansion of interfaces as well of registrable data in smart homes through sensor-based mechanisms.

The project team in the second funding phase of the Collaborative Research Center (2020–2023) was led by principal investigators Stephan Habscheid and Dagmar Hoffmann, with Tim Hector, David Waldecker, Christine Hrncal, and Kathrin Englert as (post-)doctoral researchers. As the editors of this volume, we would like to thank all of our colleagues for their long-term collaboration and their contribution to the success of this book. We want to thank Pip Hare for her extremely dedicated comprehensive copy editing as well as Sarah Diehl, Franziska Niersberger-Guéye, and Christopher Wegner for their extensive editorial work.

This research was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) – Project ID 262513311 – SFB 1187 Media of Cooperation.

References

Ayaß, Ruth.1993. "Auf der Suche nach dem verlorenen Zuschauer." In *Medienrezeption als Aneignung. Methoden und Perspektiven qualitativer Medienforschung*, edited by Werner Holly, and Ulrich Püschel, 27–41. Opladen: Westdeutscher Verlag.

Baldauf, Heike. 2002. *Knappes Sprechen*. Tübingen: Niemeyer.

Bausinger, Hermann. 1984. "Media, technology and daily life." *Media, Culture and Society* 6: 343–351. https://doi.org/10.1177/016344378400600403.

Böckmann, Barbara, Dorothee Meer, Michelle Mohn, Anastasia-Patricia Och, Ilaria Paltrinieri, Alina Renelt, Christine Ramdorf, Daniel Rettinghausen, Katharina Staubach, and Martin Tenz. 2019. "Multimodale Produktbewertungen in Influencer-Videos auf YouTube: Zur parainteraktionalen Konstruktion von Warenwelten." *Zeitschrift für Angewandte Linguistik* 70: 139–171. https://doi.org/10.1515/zfal-2019-2003.

Borbach, Christoph. 2018. "Sprachspiele|Stimmsynthesen: Zur Nachrichtentechnischen Genese Des Auditiven Pendants Von ELIZA." In *Hello, I'm Eliza: Fünfzig Jahre Gespräche Mit Computern*, edited by Marianna Baranovska and Stefan Höltgen, 177–98. Bochum: projektverlag.

Brause, Saba Rebecca, and Grant Blank: "Externalized domestication: smart speaker assistants, networks and domestication theory." *Information, Communication & Society* 23(5): 751–763. https://doi.org/10.1080/1369118X.2020.1713845.

Burgess, Jean, Kath Albury, Anthony McCosker, and Rowan Wilken. 2022. *Everyday data cultures*. Cambridge: Polity Press.

Dürscheid, Christa. 2023. "Mit Menschen und Chatbots. Digitale Kommunikation früher und heute. Conference Paper: LingColl 2022 in Tampere." *Re-*

search Gate. Accessed August 10, 2024. https://www.researchgate.net/publication/369506979_Mit_Menschen_und_ Chatbots.

Endter, Cordula, Florian Fischer, and Tobias Wörle. 2023. "Doing Home by Using Digital Assistive Technologies. On the Role of Meaning and Materiality in the use of Digital Assistive Technologies at Older People's Domesticity." *Digital Culture & Society* 9 (1): 179–200. https://doi.org/10.14361/dcs-2023-0109.

Englert, Kathrin, Dagmar Hoffmann, and David Waldecker. 2022. "Tut mir leid, ich verstehe nicht ganz – Smart Speaker als vermeintliche Gesprächspartner*innen." *merz | Zeitschrift für Medienpädagogik* (66) 2: 24–34.

Gerwinski, Jan, Stephan Habscheid, and Erika Linz. 2018. *Theater im Gespräch. Sprachliche Publikumspraktiken in der Theaterpause.* Berlin and Boston: de Gruyter.

Giddens, Anthony. 1984. *The Constitution of Society. Outline of a Theory of Structuration*. Cambridge: Polity Press.

Goffman, Erving. 1981. *Forms of Talk.* Philadelphia: University of Pennsylvania Press.

Goldman, Sharon. 2024. "How Amazon blew Alexa's shot to dominate AI, according to more than a dozen employees who worked on it." *Fortune*, June 13, 2024 https://fortune.com/2024/06/12/amazon-insiders-why-new-alexa-llm-generative-ai-conversational-chatbot-missing-in-action/.

Goodwin, Charles. 2018. *Co-Operative Action.* New York: Cambridge University Press.

Habscheid, Stephan. 2023. "Socio-technical dialogue and linguistic interaction. Intelligent Personal Assistants (IPA) in the Private Home." *Sprache und Literatur* 51 (2): 167–196. https://doi.org/10.30965/25890859-05002020.

Habscheid, Stephan, Tim Hector, and Christine Hrncal. 2023. "Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology". *Social Interaction. Video-Based Studies of Human Sociality* 6 (1). https://doi.org/10.7146/si.v6i1.137378.

Hartmann, Maren, ed. 2023. *The Routledge Handbook of Media and Technology Domestication.* London and New York: Routledge.

Hector, Tim. In preparation. *Smart Speaker als Gesprächsbeteiligte? Sprachliche Praktiken im Prozess der Domestizierung mit Voice-User-Interfaces.* Dissertation, University of Siegen, Faculty of Arts and Humanities – Faculty I. Defense: April 12, 2024.

Hector, Tim, David Waldecker, Niklas Strüver, and Tanja Aal, eds. 2023. *Thematic issue: "Taming Digital Practices – On the Domestication of Data-Driven Technologies". Digital Culture & Society* 9 (1).

Hennig, Martin and Kilian Hauptmann. 2019. "Alexa, optimier mich! KI-Fiktionen digitaler Assistenzsysteme in der Werbung." *Zeitschrift für Medienwissenschaft* 11 (2): 86–94. https://doi.org/10.25969/mediarep/12636.

Hepp, Andreas. 2020. *Deep mediatization*. Abingdon and New York: Routledge.

Hepp, Andreas. 2015. "Kommunikative Figurationen: Zur Beschreibung der Transformation mediatisierter Gesellschaften und Kulturen." In *Theorien des Medienwandels*, edited by Susanne Kinnebrock, Christian Schwarzenegger, and Thomas Birkner, 161–188. Köln: Herbert von Halem.

Hepp, Andreas, Wiebke Loosen, Stephan Dreyer, Juliane Jarke, Sigrid Kannengießer, Christian Katzenbach, Rainer Malaka, Michaela Pfadenhauer, Cornelius Puschmann, and Wolfgang Schulz. 2022. "Von der Mensch-Maschine-Interaktion zur kommunikativen KI. Automatisierung von Kommunikation als Gegenstand der Kommunikations- und Medienforschung." *Publizistik* 67 (4): 449–474. https://doi.org/10.1007/s11616-022-00758-4.

Herbig, Daniel. 2024. "Amazon: Verbesserte Alexa kommt laut Bericht nicht voran." *Heise online*, June 14, 2024. https://www.heise.de/news/Amazon-Verbesserte-Alexa-kommt-laut-Bericht-nicht-voran-9763110.html.

Hoffmann, Dagmar. 2023. Digitaler Alltag, Datenpraktiken und Subjektautonomie. In: *merz | Zeitschrift für Medienpädagogik*, 67 (1): 51–59.

Holly, Werner. 2001. "Der sprechende Zuschauer." In *Der sprechende Zuschauer. Wie wir uns Fernsehen kommunikativ aneignen*, edited by Werner Holly, Ulrich Püschel, and Jörg Bergmann, 12–24. Wiesbaden: Westdeutscher Verlag.

Holly, Werner, Ulrich Püschel, and Jörg Bergmann, eds. 2001. *Der sprechende Zuschauer. Wie wir uns Fernsehen kommunikativ aneignen*. Wiesbaden: Westdeutscher Verlag.

Horton, Donald, and Anselm Strauss. 1957. "Interaction in Audience Participation Shows." *American Journal of Sociology* 62 (6): 579–587.

Horton, Donald, and R. Richard Wohl. 1986. "Mass communication and parasocial interaction: Observation on intimacy at a distance." In *Inter/Media. Interpersonal communication in a media world* (3rd edition), edited by Gary Gumpert, and Robert Cathcart, 185–206. New York, Oxford: Oxford University Press.

Huq, Oliver. 2011. "iPhone 4S: Auf die inneren Werte kommt es an." *Heise online*, October 4, 2011. https://www.heise.de/news/iPhone-4S-Auf-die-inneren-Werte-kommt-es-an-1354456.html.

Keppler, Angela. 2018. "Medien, Lebenswelt und Alltagshandeln." In *Mediensoziologie. Handbuch für Wissenschaft und Studium*, edited by Dagmar Hoffmann and Rainer Winter, 71–85. Baden-Baden: Nomos. https://doi.org/10.5771/9783845264196-70.

Kim, Eugene. 2022. "Amazon is gutting its voice assistant, Alexa. Employees describe a division in crisis and huge losses on a wasted opportunity." *Business Insider*, November 19, 2022. https://www.businessinsider.com/amazon-alexa-job-layoffs-rise-and-fall-2022-11.

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

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Lind, Miriam, and Sascha Dickel. 2024. "Speaking, but having no voice. Negotiating agency in advertisements for intelligent personal assistants." *Convergence* 30 (3): 1008–1024. https://doi.org/10.1177/13548565231192100.

Neville, Stephen J. 2021. "The domestication of privacy-invasive technology on YouTube: Unboxing the Amazon Echo with the online warm expert." *Convergence* 27 (5): 1288–1307. https://doi.org/10.1177/1354856520970729.

Pfadenhauer, Michaela, and Tobias Lehmann. 2022. "Affects after AI: Sociological perspectives on artificial companionship." In *The Routledge Social Science Handbook of AI*, edited by Anthony Elliott, 91–106. London: Routledge.

Pins, Dominik, Alexander Boden, Gunnar Stevens and Britta Essing. 2020. "'Miss understandable' – A study on how users appropriate voice assistants and deal with misunderstandings." In *Proceedings of Mensch und Computer* 2020 (MUC20), 349–359. ACM, Magdeburg.

Pitsch, Karola. 2015. "Ko-Konstruktion in der Mensch-Roboter-Interaktion. Kontingenz, Erwartungen und Routinen in der Eröffnung." In *Ko-Konstruktionen in der Interaktion*, edited by Ulrich Krafft, Elisabeth Gülich, and Ulrich Dausendschön-Gay, 229–58. Bielefeld: transcript. https://doi.org/10.1515/9783839432952-013.

Porcheron, Martin, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. "Voice Interfaces in Everyday Life." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*, 1–12. https://doi.org/10.1145/3173574.3174214.

Sadowski, Jathan. 2020. *Too Smart. How Digital Capitalism is Extracting Data, Controlling Our Lives, and Taking Over the World*. Cambridge: MIT Press.

Schlinkmann, Eva (2021): *Kommunikative Rekonstruktion in der Theaterpause. Eine Gattungsanalyse von Pausengesprächen im Theater.* Stuttgart: Metzler.

Schüttpelz, Erhard. 2017. "Infrastructural Media and Public Media." *Media in Action* 1 (1): 13–61. http://dx.doi.org/10.25819/ubsi/7935.

Silverstone, Roger, Eric Hirsch, and David Morley. 1992. "Information and communication technologies and the moral economy of the household." In *Consuming Technologies. Media and Information in Domestic Spaces*, edited by Roger Silverstone and Eric Hirsch, 13–28. London: Routledge.

"Siri". Wikipedia article. Accessed August 10, 2024. https://en.wikipedia.org/wiki/Siri.

Star, Susan. 1993. "Cooperation Without Consensus in Scientific Problem Solving: Dynamics of closure in open Systems." In *CSCW. Cooperation or Conflict?*, edited by Dan Daiper, Closton Sanger, and Steve Easterbrock, 93–106. London: Springer. https://doi.org/10.1007/978-1-4471-1981-4_3.

Statista. 2021. "Statista Global Consumer Survey. ID 999790" Accessed August 21, 2024. https://de.statista.com/statistik/daten/studie/1271603/umfrage/anteil-der-haushalte-in-deutschland-mit-smart-speaker/.

Stresing, Laura. 2011. "Siri sagt: Warum die Apple-Sprachsoftware nicht nur Geeks fasziniert." *Tagesspiegel*, December 7, 2011. https://www.tagesspiegel.de/gesellschaft/medien-_-ki/warum-die-apple-sprachsoftware-nicht-nur-geeks-fasziniert-6708628.html.

Strüver, Niklas. 2023. "Frustration Free: How Alexa Orchestrates the Development of the Smart Home." *Digital Culture & Society* 9 (1): 99–123. https://doi.org/10.14361/dcs-2023-090106.

Suchman, Lucy. 2007. *Human-machine reconfigurations. Plans and situated actions* (2nd edition). Cambridge: Cambridge University Press.

Turow, Joseph. 2021. *The Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet.* New Haven and London: Yale University Press.

Volmar, Axel. 2019. "Productive Sounds." In *The Democratization of Artificial Intelligence*, edited by Andreas Sudmann, 55–76. Bielefeld: transcript. https://doi.org/10.1515/9783839447192-004.

Waldecker, David, and Tim Hector. 2023. "A Praxeological Approach Towards the Domestication of Connected Media Technologies. Introduction." *Digital Culture & Society* 9 (1): 5–22. https://doi.org/10.14361/dcs-2023-0102.

Waldecker, David, and Dagmar Hoffmann. 2023. "Inszenierung von kritischen Kompetenzen in Nischenöffentlichkeiten. Bewertungen von Smart Speakern auf YouTube." *kommunikation@gesellschaft* 23 (1). https://doi.org/10.15460/kommges.2022.23.1.1000.

Waldecker, David, Tim Hector, and Dagmar Hoffmann. 2024. "Intelligent Personal Assistants in practice. Situational agencies and the multiple forms of cooperation without consensus." *Convergence* 30 (3): 975–991.

I Voice Assistants in Private Homes. Conceptual Considerations

The DataEconomy@Home

The Private Sphere, Privacy, and the Embedding of Artificial Intelligence Systems into Everyday Life as an Expansion of Economic Data Grabbing

Carsten Ochs

Abstract *Although voice assistants have been adopted widely in private homes, they still cause bafflement among those who have a negative attitude towards smart speakers. But what is at stake in affective reactions such as these? And why does the issue of privacy frequently come to the fore in this context? This contribution sets out to somewhat unsettle the seeming naturalness of problematizing smart speakers as a "privacy issue", so as to offer a clearer understanding of the whys and wherefores of the issue in the first place. To this end, I first examine the astonishment that is frequently expressed in response to the dissemination of smart speakers (section 2). What is so astounding about installing smart speakers in the private sphere of the home? The next aspect to be investigated (section 3) concerns an essentially modern privacy practice: it is linked to the expectation that individuals have the right and the means to control which entities may receive which elements of their personal information. The idea that in order to constitute oneself as an individual one must have control over who can access one's personal information came to prevail as the dominant concept of data privacy in the 20th century. Having thus specified the notion of the private sphere on the one hand, and of privacy on the other, I proceed (section 4) by investigating why some of today's users willingly relegate these fundamental forms of privacy. And I analyze what actually happens to the data that is collected and processed by smart speaker infrastructures that reach into private homes. To conclude (section 5), I bring together the insights gained in order to support the argument that smart speakers in the private home form part of surveillance capitalism's expansion into as many social spheres as possible.*

1. Introduction

Over the last few years, I participated regularly in University of Kassel's winter semester lecture series "Der soziologische Blick" ("The sociological gaze"), a course that serves primarily to introduce new students to relevant research fields, topics, and debates addressed by contemporary sociology, but sometimes also attracts interested listeners from the general public. As I have spent many years investigating the digital transformation of information privacy (Ochs 2022), I was frequently assigned with the task of presenting to students a sociological perspective on the social role played by the distinction of private versus public in pre-modern, modern, and contemporary societies. At the end of my lecture in the 2022 series, I was approached by an elderly man; I assumed that he had either started to study sociology since retirement, or was simply interested in the topic. He expressed his appreciation of the lecture, before going on to raise some criticism regarding my bad habit of bridging pauses for thought by murmuring filler words like "exactly", "yes", "that's it", etc. After this assessment of the quality, he shifted to the lecture's content and pointed out to me that the major current threat to privacy was the implementation of smart speakers, "such as Alexa", in private homes. *That* was something that my research should focus upon, he advised, shaking his head with bafflement that anyone could be crazy enough to welcome such devices into their homes.

What the anecdote illustrates is a rather common reaction when it comes to voice assistants in private homes, common at least among people who have a negative attitude towards smart speakers and the infrastructures that enable their agency (for an impressive mapping of such an infrastructure, see Crawford and Joler's 2018 visual rendition and analysis of Amazon Echo's "anatomy"). It is perhaps unsurprising that the practice of using smart speakers seems particularly alarming to an elderly generation that has witnessed the state surveillance in East Germany and/or the resistance to the West German census in the 1980s and the Federal Constitutional Court's assertion of the right to informational self-determination. And yet, we should not presume that it is only the elderly who are concerned. But what is at stake in affective reactions such as these? What exactly was it that made the lecture attendee shake his head at the idea of allowing smart speakers into private homes? And why does the issue of privacy come to the fore in this context?

This contribution sets out to somewhat unsettle the seeming naturalness of problematizing smart speakers, such as Echo, and voice assistants, such as Alexa, as a "privacy issue". It is not my aim to applaud the proliferation of these devices and infrastructures, nor to absolve them of criticism, but rather to offer a clearer understanding of the whys and wherefores of the issue in the first place. To this end, I will distinguish three different aspects and consider them in succession before consolidating the insights gained to formulate the main argument of my contribution.

The first aspect to be examined in the next section (section 2) concerns the astonishment that is frequently expressed in response to the dissemination of smart speakers. What is so remarkable or astounding about installing smart speakers *in the home*? As I will explain, there is nothing "natural" about assumptions that the home as a private sphere should be shielded from techno-economic agencies such as the Amazon Echo. Yet, many people do perceive the idea of connecting their household to Amazon's global infrastructure as an invasion into the domestic private sphere that threatens the established norms of the private/public distinction in contemporary society.

While the notion that the sanctity of "local privacy" (Rössler 2001, 25; 255; cf. Roessler 2004) must be upheld already had genealogical precedents in premodernity even if it took on a more specific form in modern societies, the next aspect to be investigated (section 3) represents an essentially modern practice: it is linked to the expectation that individuals (the owners or residents of private homes, for example) have the right and the means to control which entities may receive which elements of their personal information. The idea that in order to constitute oneself as an individual one must have control over who can access one's personal information came to prevail as the dominant concept of data privacy in the 20th century. It is intimately tied to the idea that 'the individual' is not static or given but rather evolves over an individual trajectory of self-development, i.e., as an individual "career" (Luhmann 1989; 1997) with the self becoming a "project" (Giddens 1991). By this point, then, we should have gained a deeper understanding of the reasons that led to the lecture attendant's head-shaking after the 2022 lecture on the private/public distinction. The installation of smart speakers such as the Amazon Echo in people's private homes affects two basic types of privacy at the same time – and two that are guaranteed by basic rights: the security of the private spatial sphere, and that of personal information. For some, it is hard to imagine why anyone would willingly relegate these fundamental forms of privacy.

The next section (section 4) will present some explanations for this apparent "carelessness" on behalf of smart speaker users, and consider them alongside an analysis of what actually happens to the data that is collected and processed by smart speaker infrastructures that reach into private homes.

To conclude (section 5), I bring together the insights gained in order to support the argument that smart speakers in the private home form part of surveillance capitalism's expansion into as many social spheres as possible (Zuboff 2019). The difficulties that data protection bodies have in adapting to this expansion, I propose, are due to the historical context in which measures to protect privacy protection were originally developed – they were tailored to the sphere of labor, and to the practices of work. Whereas individuals' control of their own personal information is undermined by the requirements of digital, networked self-constitution, practices that take place in the private sphere of the home have only recently been dragged into the realm of social datafication.

2. Private Spheres: Genealogical Remarks on the Private Home

In a 2018 essay accompanying their impressive analytical mapping of the sociotechnical planetary infrastructure that constitutes Amazon's Machine Learning (ML)-based Artificial Intelligence (AI) agent Alexa, Kate Crawford and Vladan Joler sketch out the underlying user scenario propagated by Amazon:

> A cylinder sits in a room. … It is silently attending. A woman walks into the room, carrying a sleeping child in her arms, and she addresses the cylinder. 'Alexa, turn on the hall lights?' The cylinder springs into life. 'OK.' The room lights up. … A brief interrogative conversation – a short question and a response – is the most common form of engagement with this consumer voice-enabled AI device. But in this fleeting moment of interaction, a vast matrix of capacities is invoked: interlaced chains of resource extraction, human labor and algorithmic processing across networks of mining, logistics, distribution, processing, prediction and optimization. The scale of this system is almost beyond human imagining. (Crawford and Joler 2018, 1)

What is so astonishing about the idea of implementing a technical agent that is "silently attending" in one's private home? Why do some people shake their heads when Echo/Alexa users connect their private homes to a sociotechnical global system that "is almost beyond human imagining"? The first and almost automatic response to this question is that many people find it disturbing to envisage inviting a silent listener that is connected to some infrastructure 'out there' into their private homes. Do we not usually expect external listeners to remain firmly outside our private sphere, the spatial privacy of our homes, where we engage with family and friends, i.e., with those who do not play a functional role, but with whom we choose to share our lives with? Do we not expect these domestic interactions, which constitute our lifeworld, to be none of the economy's business?

Indeed, upholding the spatial privacy of the home is a long-standing social practice that can be traced back to the ancient world of Greco-Roman antiquity and is still performed today, with the sanctity of the home in Germany guaranteed by article 13 of German constitutional law[1]. While it therefore might seem somewhat natural to us to expect the private sphere to form a separate realm within society, there is nothing natural about this separation whatsoever. In fact, the status of the spatial private sphere as an experiential realm in its own right, shielded from authorities' access, and clearly separated from the world of work, is a product of the social history of European societies from antiquity to the present day.

As Hannah Arendt has explained, in ancient Greek society, the "oikos" was the homestead of the extended families of Greek patriarchy. It served both as a discrete spatial realm in which families went about their daily business, and as the site of economic reproduction that guaranteed the social position and standing of the family head in the public agora, and thus in Greek society (Arendt 2002, 76–77). In this way, "the distinction between private and public correspond[ed] to a division between two institutional domains – the private domain of the household and the public domain of the body politic" (Gobetti 1997, 104).

Notwithstanding that European medieval societies differed, of course, in many respects from those of Greco-Roman antiquity, the family and its homestead in the Middle Ages continued to play the role of a base from which to operate. Even if the head of this medieval type of family did not act in any realm

1 As this remark indicates, the issues dealt with in this chapter are approached from a European perspective by an author based in Germany. The ideas and explanations presented thus relate to the social history of the 'province' of Europe, which is not to say that similar developments might not have occurred elsewhere too. For example, it seems that the US approach to privacy is based upon a similar idea of the home, at least this is suggested by Warren and Brandeis' considerations in their classic "The Right to Privacy" in which they discuss "the sacred precincts of private and domestic life" (Warren and Brandeis 1890, 195).

that may be reasonably called "public" or "private" as these terms are used in industrialized times, it was nonetheless the function of the family "to strengthen the authority of the head of the household, without threatening the stability of his relationship with the community" (Ariès 1977, 228). What is more, just as in ancient times, the homestead featured a certain openness compared to the private sphere we have become accustomed to now: "The medieval household mixed up young and old, men and women, servants and masters, friends and family, intimates and strangers. It was open, almost like a café or pub, to the comings and goings of a multitude of diverse types of people, intent upon a bewildering variety of tasks concerned with business or pleasure" (Kumar 1997, 209).

Leaving aside structural differences between ancient and medieval "oikos" (see Ochs 2022, 116), it is important to note that medieval family life was practiced within the stratified social order of feudalism. Significantly, for nobles, the family was not positioned in dichotomous opposition to the polis (as in Greek antiquity) or the state (as it is to a certain degree in modernity), but was part of a competitive landscape with all the other families that ruled a particular territorial dominion, always striving to expand their territory (Elias 1997, 95). As territories constantly changed hands, for a long time, medieval forms of rule remained decentralized – there was no overarching central power that could establish itself as a kind of quasi-public counterpart to some quasi-private familial sphere (Elias 1997, 28; Habermas 1990, 58; Ariès 1991, 7)[2].

Although sociological (e.g., Habermas 1990) and social history analyses of medieval privacy (e.g., Brandt 1997) disagree as to whether a specifically medieval type of privacy can be distinguished, the current state of research invites the conclusion that the development of the familial private sphere occurred as part of the processes of social differentiation that were observable in all areas of early modern society. The compartmentalization of social life (Shibtuani 1955, 567) had a lasting effect on the private sphere:

> Gradually, starting sometime in the early seventeenth century, this promiscuous world was ordered and tidied up. Houses – upper-class houses to

2 This is not to say that medieval societies did not recognize any form of privacy at all. Shaw (1996), for example, identifies practices relating to property and to the body in medieval London that reference privacy both in semiotic (use of the word) and practical terms (distinct practices). Nonetheless, there are marked differences between ancient and modern ways of enacting privacy practices (Ochs 2022, 150).

> start with – began to reflect a marked degree of segregation of the status and functions of husband and wife, parents and children, masters and servants, friends and family. Boundaries were more strictly drawn – in paths and hedges, bricks and mortar, as well as in social customs – between the private and intimate world of the home and family, and the public world of acquaintances, business associates, and strangers. Work and nonwork ('living') were rigidly separated. (Kumar 1997, 209)

In the 18th century, the private sphere of the family once again comprised a closed realm, separate from public space and life (Sennett 2008, 18–19; 89–91). A range of oppositional counterparts distinguished themselves from the private sphere of the home and the family. First, the state evolved from an absolutist regime of surveillance (Elias 1997, 282) – "loath to accept the fact that there were certain areas of life beyond its sphere of control and influence" (Ariès 1977, 228) – into the public monopoly on violence and taxation that we are familiar with today (Ochs 2022, 108). Second, and quite relevant for my argument here, the private sphere of the family became gradually separated from the realm of labor. The structural force driving this separation, as many scholars assert, was the sociotechnical drive towards industrialization. In the pre-industrial economies of the Middle Ages, the whole "oikos" of the extended family's homestead had been the site of economic reproduction (hence the term "economy" as derived from "oikos"), where economic and other social activities consolidated as a family's spatial-economic unit (Meier-Gräwe 2008, 116; Lundt 2008, 60–61). When the means and processes of production increasingly shifted to factories and sweatshops, this unit fell apart: the result was a "split between home and factory, a split between economic and other aspects of the parent-child relationship" in workers' families (Smelser 1967, 31), while in bourgeois society in general, work was separated from the private realm and families' homes were conceived as a private sphere, shielded from labor (Burkart 2001, 403).

As the spatial private sphere thus evolved in structural opposition to the state (representing public authority); to the private economy and working world; and also to "public life" in general, a gendering of the separated sphere occurred. The male *homo eoconomicus* was deemed to belong "naturally" to public life in all its varieties, while females were considered *domina privata* (Meier-

Gräwe 2008, 117)[3]. At the same time, there was a shrinking of the family, which in the 19th century increasingly came to play the role of a "bulwark against the buffets of a rapidly changing world" (Kumar 1997, 222). With the transition to capitalism inducing massive transformations that unsettled established expectations and practices, actors retreated into the idealized private sphere of the familial homestead, which came to be seen as a refuge from the vagaries of public social and economic life (Sennett 2008, 20)[4].

Over the course of the 20th century, the shrinking of the "staff" operating in the spatial private sphere continued:

> The twentieth century has seen the decline and disintegration of the family as a community, as a collectivity expressing the common purposes of its members. Individualism's progress, interrupted and held in check in various ways, has continued apace. It has now invaded the family as well as other sectors of society. In the end it's individualism, not the family that has triumphed. (Kumar 1997, 222)

Whether or not one agrees with the idea that the family is in a process of dissolution (the patchwork character of many families rather suggests a de-naturalization of the form called 'family'), most will accept that the private sphere nowadays can be occupied by different constellations such as single persons, familial groupings, or flatmates. But whoever the actors are that claim the privacy of their homes, the closed-shop character of the private sphere as a realm distinct from the working world, from the attention of public authorities, and from uninvited listeners representing the economy or the general public, remains a widespread normative expectation[5].

---

3 The picture drawn here is an accurate, yet simplified one, as empirical reality is always more messy than historical analysis suggests. For detailed and at the same time controversial accounts of the gendering of public and private spheres in industrial society see Hausen (1976); Pleck (1976); and Lundt (2008). Please note that despite the ways in which these researchers' views differ, they largely agree on what counts for the argument of this chapter: the spatial private sphere (of the family) began to separate from that of work in the 17th century and gradually became a distinct realm.

4 At the same time, the private sphere of the family became the site of gendered violence, especially against women and children (Müller 2008); the 20th-century "women's movement" therefore re-politicized the private in order to render patriarchal violence accessible to public intervention (Lundt 2008, 51).

5 The phenomenon of the 'home office' in the course of the COVID-19 pandemic temporarily blurred the boundaries between the private home and the working world.

So, here we have our first explanation for the head-shaking of people who feel disturbed by the introduction into the home of listening devices that are deemed at least potentially capable of transmitting recorded audio to an unknown audience: such persons are uneasy about the unsettling of the closed shop that they still expect the private sphere of their homes to encapsulate.

3. Information Control: Privacy in the 20th Century

AI-equipped smart speakers and the infrastructures they form part of disturb people's entrenched expectations concerning the exclusivity of the private home; its separation from the economy, from the realm of work, and from external observation in general. A further aspect that normative attitudes towards smart speakers relate to are issues of privacy and data protection. What is called "information privacy" in social theory (e.g., Rössler 2001, 45) usually goes under the name of "data protection" in regulation. Smart speakers seem to affect this idea of privacy/data protection, because as

> human agents we are visible in almost every interaction with technological platforms. We are always being tracked, quantified, analyzed and commodified. But in contrast to user visibility, the precise details about the phases of birth, life and death of networked devices are obscured. With emerging devices like the Echo relying on a centralized AI infrastructure far from view, even more of the detail falls into the shadows. (Crawford and Joler 2018, 12)

This may well be true, but why is it at all noteworthy that we "are always being tracked, quantified, analyzed and commodified"? Couched in social theory terms: why should information privacy (to be distinguished from the private sphere) be an issue at all? What is the meaning of "information privacy" in the first place? And, how did information privacy become an entrenched practice in contemporary digital society's genealogical forerunner – 20th century European modernity? To answer these questions, I will begin by offering a general sociological characterization of 20th century high modernity, before focusing on the issue of self-constitution and privacy.

However, I will not discuss here whether these developments have had structural consequences for people's normative expectations concerning the privacy of their homes.

According to Andreas Reckwitz (2006, 275), the early decades of the 20th century marked the end of bourgeois cultural rule. The period witnessed a massive expansion of space–time relations, enabled by innovations in technologies of transport, communication, media, and production (Berger, Berger, and Kellner 1975; Beniger 1986). At the same time, social life came to be increasingly structured by large organizations, such as unions, associations, people's parties, huge corporations etc. – an observation that has led sociological analysis to characterize, roughly speaking, the first half of the 20th century as "Organized Modernity" (Wagner 1998). Nazi barbarism, totalitarianism, and the two industrialized world wars of the "short 20th century" (Hobsbawm 1994) could not have taken place without Organized Modernity's capacity to assemble people by sociotechnical means at a huge scale; and to construct for them *collective identities* based on the sometimes violent and lethal exclusion of "othered" (i.e., purposefully generated) "outsiders" (Bauman 1989; Wagner 1998, 68–69; Arendt 1975). After World War II, European post-war societies passed into what has been called "Reflexive" or "Second Modernity" (Beck 1986; Beck, Giddens, and Lash 1994), within which self-constitution became an ever more *individualized* process that was to be realized by neo-liberalism's structurally "released" – and also isolated – actors themselves.

The shifting logic of self-constitution mirrors the transition from Organized to Second Modernity. The beginning of the short 20th century witnessed the appearance of "organization man", a social figure who tended to follow a career largely predetermined by organizational environments (Reckwitz 2006). A typical trajectory of "organization man" would lead him through organizations that aim to provide their members with a "corporate identity" (Whyte 2002). In such settings, organizations strive to *fix* their members' identities (Mönkeberg 2014), because stable – or rather *stabilized* – identities can be easily integrated into large organizations and formalized sequences of operation (e.g., production under Taylorism). However, while organizations demanded stable identities, the mass media (radio, TV) and urbanization began to make it plain for all to see that "[m]ost people live more or less compartmentalized lives, shifting from one social world to another as they participate in a succession of transactions" (Shibutani 1955, 567). For 20th century subjects, it came to be taken for granted that "[d]ifferent sectors of their everyday life relate them to vastly and often severely discrepant worlds of meaning and experience" (Berger, Berger, and Kellner 1975, 63). Whereas in the 19th century, everybody had implicitly known that they lived "compartmentalized lives", radio and television rendered visible this compartmentalization of life by putting the pluralism of social worlds on display simultaneously (Berger, Berger, and Kellner 1975, 64 ff.; Goffman 1959). Now, everybody knew that *everybody knows* that everybody lives compartmentalized lives.

As a result, the idea of the self as an undivided coherent whole, which defined early modernity's notion of the individual, begins to seem increasingly unsustainable. Sociologists monitor closely how actors moved in everyday life and over the life course through different social worlds and organizational contexts that offer contradicting rules and roles. Pierre Bourdieu (1987) elaborates in a virtuoso manner how people in European post-war societies came to terms with the different social worlds and areas they passed through, how they continually adapted themselves and developed further instead of self-constituting as a static self with some singular once-and-for-all core identity. In 20th century high modernity, processes of self-constitution were obliged to incorporate frequent changes of subjectification schemes as well as organizations' identity fixations. The mechanism that allows people to reconcile continuous change with the constancy of corporate identity is the *career mode* (Luhmann 1997, 742). Facilitating the organizational channeling (fixation) of developmental trajectories (movement) through society, it became subjectification's key mechanism. Giddens (1991) accounts for 20th century self-constitution with the concept of the "reflexive project of the self", while Goffman sheds light on the informational aspects of practicing such a self. The project-self is habitually bound to play contradictory roles, for "[i]n each [social] world there are special norms of conduct, a set of values, a special prestige ladder, characteristic career lines, and a common outlook toward life – a Weltanschauung" (Shibutani 1955, 567). Given the potential contradictions between contexts, it becomes imperative for individual project-selves to separate the audiences associated with different roles from one another, and to hide internal inconsistencies. Individuals are obliged to establish "audience segregation", and to do this, the project-self takes measures to control which audiences have access to which elements of their personal information (Goffman 1959). Hence, over the course of the 20th century in Euro-American society, boundaries came to be drawn between different types of information. As long as these boundaries were not crossed, "contextual integrity" (Nissenbaum 2010) remained intact.

In the 1980s, a conflict arose in Germany that led to the practice of individual information control becoming a case of legal dispute: the right to informational self-determination. At the time, "new social movements" were evolving, addressing issues such as women's rights, environmental protection, discrimination, etc. (Beck 1986). Extending the objectives of German social movements beyond labor issues, these movements contributed to a generally politicized atmosphere, marked by the Cold War and accompanying controversies.

It was in this tense political atmosphere that the German government announced its intention to conduct a census (Berlinghoff 2013). Fueled by the politicized *Zeitgeist*, a large-scale controversy erupted. Before long, advocates of data protection who were worried about government surveillance had filed a suit to the German Federal Constitutional Court. Crucially, the conflict unfolded against the backdrop of the computerization of administration and heated debate about data protection (Frohman 2013). The Constitutional Court's response was sensitive to this and explicitly pointed out the potential dangers of the *networking of data across informational contexts*. It argued that, as citizens, people might feel pressurized to hide their political commitments if they knew they were being monitored from a central point of observation. For this reason, the court ruled, information about persons' political activities must remain private (BVerfG 1983).

The verdict of this *Volkszählungsurteil* asserted that any German citizen has the general right to control who knows what about them, at what point in time and for what purpose – because if they did not, they might not be able to engage freely in self-development, and in the processes of self-constitution. This is ultimately a legalistic articulation of the view that any individual actor, in order to self-constitute as a Giddensian "project-self" (Giddens 1991), or to follow a Luhmannian "career" (Luhmann 1989), must be able to regulate what information concerning their person is accessible to actors from the various social contexts and worlds that that individual passes through. Arguing along similar lines of reasoning, the court translated the everyday practice of information control into the right to information self-determination (Rössler 2010, 45).

Individual information control became the dominant privacy practice of the 20th century because it allowed the project-self to deal with the contradiction between corporate identity fixation and ongoing personal development. The court mobilized this practice and turned it into a legally guaranteed right when the practice appeared to be coming under threat from a novel type of emergent public enabled by digital networking – which was already discussed in the data protection discourse of the 1980s, although the internet at that time was but a far cry from being part of digital everyday practices (Steinmüller 1988). Even so, a technological innovation that facilitated the flow of information across borders was already on the horizon, threatening to disrupt "contextual integrity" (Nissenbaum 2010). Nevertheless, the right to control who has access to one's own personal information still forms the basis of current data protection law, and Amazon's Echo and Alexa operate in a techno-legal environment that is still largely informed by the idea of individual information control. This raises the question of whether these technologies contribute to the border-crossing of information flows, and, if so, what the consequences are in terms of social structuration. Perhaps those who shake their heads at the thought of Alexa implicitly assume that there will indeed be consequences? Let's render this assumption explicit.

4. Digital Self-Constitution and Machine Learning@Home

Having gained some clarity regarding the different conceptualizations of the private that seem to be somehow affected by the integration of smart speakers and AI assistants into private homes, we can now move on to consider the functionality of these technical apparatuses, i.e., the purposes they serve and operations they perform once they have been installed in people's homes. From the perspective of Echo/Alexa users, smart speakers are there to increase automation and convenience. At least, that is Amazon's great promise. Describing a 2017 promotional video advertising the Echo, Kate Crawford and Vladen Joler observe:

> The video … explains that the Echo will connect to Alexa (the artificial intelligence agent) in order to 'play music, call friends and family, control smart home devices, and more.' … The shiny design options maintain a kind of blankness: nothing will alert the owner to the vast network that subtends and drives its interactive capacities. The promotional video simply states that the range of things you can ask Alexa to do is always expanding. (Crawford and Joler 2018, 3)

As the authors go on to point out, firstly, the smart speaker itself appears to be just an "'ear' in the home" but is actually far more than that: "a disembodied listening agent that never shows its deep connections to remote systems" (Crawford and Joler 2018, 5), by means of which the private home of the Alexa user is connected to an extensive infrastructure that is inaccessible to the user. Second, the device seems to have been designed to remain unnoticed, and is notably unrevealing of its connection to the external infrastructure. And third, the number of tasks Alexa can fulfill is promised to increase over time. How can that be possible?

All three aspects refer to the nature of the Echo's/Alexa's functionality and thus shed light on the question of what the system actually *does* in people's private homes. Starting with the third aspect, the system's increasing capabilities, we should note that the quite brief interactions between user and device (the user issues a command, the system executes it, or, if it fails to do as required, the user attempts to articulate their command more clearly) not only serve to deliver an immediate required response (e.g., switching on a light, playing a particular song, warming up the living room); crucially, the interaction sequences serve as a training material to expand the system's capabilities: "For each response that Alexa gives, its effectiveness is inferred by what happens next: Is the same question uttered again? … Was the question reworded? … Was there an action following the question?" (Crawford and Joler 2018, 3). In this sense, the service that *users* provide is to "supply … Amazon with the valuable training data of verbal questions and responses that they can use to further refine their voice-enabled AI systems" (Crawford and Joler 2018, 5).[6]

In providing data to train the device, users and their homes are integrated into the infrastructure and process of value-creation that is organized, managed, and exploited by Amazon. This is made possible by the first aspect highlighted above in Crawford and Joler's characterization of Alexa: the connection of users' private homes to Amazon's extensive sociotechnical and techno-economic infrastructure. The second aspect mentioned above, the attempt to render this infrastructural connection unnoticeable, points to Amazon's strategy to make the Echo a sociotechnical actor that forms part of users' everyday practices in a seemingly 'natural' way. By shaping practices, the Echo becomes an entity that operates on the level of what Giddens (1984, 7) has called "practical consciousness" as distinct from "discursive consciousness". That is, the device is generally perceived as merely part of the background. It may occasionally become the focus of attention if it does not function as expected, for example, but by and large its presence is simply taken for granted within everyday life and practices.

6 In retrospect, we can say that Amazon's strategy has not yet paid off, as users' simple utterances turned out to be of limited value for training AI systems (Lindner 2023). The recent announcement by Facebook, however, that it would use personal data for machine learning purposes (Spiegel 2024) indicates that the data economy's drive to collect (personal) data is indeed to a large extent motivated by the desire to improve their machine learning systems – even if this desire is not always satisfied, as in the case of Amazon's Alexa. Regardless of its degree of success, what counts for my argument here is what *causes* the drive towards increased data collection.

There is anecdotal evidence that Amazon's strategy has been successful, at least in some instances. As part of a research project exploring the social negotiation of artificial intelligence, privacy, and democracy[7], we interviewed two Echo/Alexa users. When asked whether they switched their Alexa off when they had visitors, such as friends or family, the first interviewee, who worked as a software engineer and presented a more business-oriented mindset in the interview, said "no", adding that "such devices have simply become too normal to do so."[8] The second respondent, who expressed a more critical attitude to the data economy, also answered "no", but went on to reflect:

> Actually we should have warned our guests … as one would in the case of CCTV … actually we should do that. But we just don't – not out of maliciousness. Who would do something like that? But because it's so natural to us. And perhaps that's the crux of the matter, that it's become so natural that you don't even mention the device anymore. Like having an oven in your kitchen. You wouldn't tell anyone: 'Beware, there's an oven', or 'there's a toaster, you might burn yourself'; these are devices that are simply natural to us, but, of course, for those who visit us, they might seem not natural at all.[9]

As the second quote indicates, when prompted by the interviewer to reflect on the Echo's presence in social situations that include visitors, the interviewee focused their attention on the device, thus shifting it from the realm of practical consciousness to that of discursive consciousness. The problematization

---

7 The project "Democracy, AI, and Privacy" forms part of the long-running research association "Forum Privatheit." I would like to thank the Federal Ministry of Education and Research (BMBF) for funding the project (16KIS1379) and thus enabling me to write this article.

8 In German: "dafür sind solche Geräte zu normal geworden."

9 The original quote: "[Z]umindest müsste man mal darauf hinweisen … so, wie man es bei Videoüberwachung auch macht. Das müsste man eigentlich tun. Wir tun es explizit nicht; gar nicht mal aus böser Absicht heraus. Wer würde sowas schon machen? Sondern eher, weil es für uns so selbstverständlich ist. Und das ist vielleicht auch die Krux, dass es so selbstverständlich ist, dass man schon gar nicht mehr darauf hinweist. Also so quasi wie man in der Küche einen Backofen hat. Da würde man auch nicht sagen: 'Achtung, da ist ein Backofen', oder 'hier steht ein Toaster, du kannst dich verbrennen', sondern das sind Geräte, die mittlerweile schon für uns so selbstverständlich sind, aber natürlich für die, die uns besuchen, möglicherweise mitnichten selbstverständlich sind."

that followed is precisely what usually remains in the shadows of practical consciousness – just as Amazon's strategy strives to achieve.

Amazon's obscuring of the Echo's/Alexa's infrastructural connection has the convenient side effect – or perhaps it is even the main objective – that users rarely reflect on the sociotechnical relations they and their operations form part of. The seemingly isolated magic of Alexa's AI is in fact the product of the real-life actions of a whole variety of embodied beings (Engemann 2018; Crawford and Joler 2018, 14) who provide the material, physical, intellectual, etc., resources that make the system run in the first place. From the data economy's point of view, devices such as the Echo can be understood as agents of "datafication": the expansion of socio-digital agencies into all areas of social life and society (Houben and Prietl 2018; van Dijck 2014). Insofar as "datafication" is driven by the data economy's interest in profit (Zuboff 2019), it results, as Till Heilmann (2015) has aptly stated, in the systematic expansion of the realm of economic utilization (*Ausweitung der Verwertungszone*).

How does this expansion structurally affect the institutionally protected privacy of the private sphere, as well as the degree to which individuals are able to control who can access their data? These are the questions to be addressed in my conclusion.

5. Conclusion: How Surveillance Capitalism Taps into Just Another Realm of Experience

According to Crawford and Joler (2018, 14) the goal that motivates corporations to persuade consumers to install their devices, such as the Amazon Echo, in private homes, is the expansion of the infrastructure by means of which they can engage in "data extractivism". Succeeding industrial society, contemporary digital society is populated by new players that aggressively aim to maximize data-based profits:

> The new infinite horizon is data extraction, machine learning, and reorganizing information through artificial intelligence systems of combined human and machinic processing. The territories are dominated by a few global mega-companies, which are creating new infrastructures and mechanisms for the accumulation of capital and exploitation of human and planetary resources. (Crawford and Joler 2018, 14)

To the extent that profit-oriented companies engaged in advancing datafication expand their infrastructures of value generation into private homes, activities performed at home should be classified as work, surmise Crawford and Joler (2018, 7) and others including Heilmann (2015) who talks about "data work". So, if activities undertaken at home are drawn into economic schemes of value creation, i.e., those activities that in modern society were recognized as part of one's "lifeworld" and were (ideally) to remain undisturbed by the imperatives of private economic agencies and public authorities (Habermas 1995, 473), what are the implications for contemporary privacy and the private sphere? There are at least two possible interpretations:

- First, we might interpret this process as an *expansion of work*, insofar as human activities are utilized to generate a product – data – that is appropriated and translated into exchange value (Heilmann 2015, 43). From this perspective, then, the proliferation of voice assistants helps to expand the realm of work, thereby breaking down the historically evolved demarcation of the private sphere of the home as a zone separate from institutionalized labor, productivity, and economic imperatives.
- Second, an alternative interpretation would not so much portray the infrastructural expansion into private homes as the transformation of whatever activities are done there into work, but as the *appropriation of the realm of non-work by the agencies of surveillance capitalism's data economies*. The author whose work supports this perspective is, of course, Shoshana Zuboff (2019), who argues that surveillance capitalism has expanded its exploitation of human labor to capitalize on human experience itself.

While I have little difficulty accepting the diagnosis that, in the last two decades or so, we have witnessed the digital expansion of the realm of economic utilization (*Ausweitung der Verwertungszone* in the words of Heilmann 2015), I believe there is also substantial indication that it is the second interpretation that accounts for what is novel about this expansion. As many commentators have observed, techno-economic expansion into people's everyday social lives is often not experienced as an extension of work at all (Heilmann 2015, 41), but rather as the incorporation of social life into the digital realm (Ochs 2021). Moreover, while users and their social lives are indeed exploited, insofar as they provide the resources for the profitable activities of the data economy, they do not participate in crafting *the product* itself that is then sold. Users whose data is utilized do not themselves generate advertisement space, ads, or attention; neither do they produce predictions, devise strategies, or impose manipulations. As Dolata and Schrape (forthcoming) clarify, the platforms of the data economy use data as raw material, but the value of the data is only realized when it has undergone further processing by those platforms' commodification processes – processes that users are not at all involved in.

Perhaps it would be even easier to come to terms with the constitution of digital society if the digital expansion of the realm of economic utilization was indeed transforming all social activities into work. For one, that would simplify the measures needed to regulate the data economy. But it is not so simple. As shown above, at least when it comes to voice assistants and smart speakers in the private home, the datafication of social life affects *both* the privacy of the home and individual information privacy at once. In the working world, there are well-established regulatory bodies and legal protection that can be mobilized to address data protection. But once datafication expands its scope to access the social realms of human experience, established concepts and boundaries become hard to enforce. As Werner Steinmüller, a German pioneer of data protection, already warned in the 1980s:

> As yet, there is no legal term to describe the spread of IT beyond the sphere of labor into the grey zone of illicit work, into the lifeworld that is not about wage-earning … and even into children's worlds of play; nor does any work-like legal protection exist, and even less so when it comes to the newly emerging interrelationships between the world of work and that of 'life'. It is not easy to legally and politically support those affected. (Steinmüller 1988, 157; my translation[10])

At first glance, these considerations might seem to suggest that we should simply adapt and expand the regulatory and political measures imposed in response to digital capitalism, which themselves were based upon those created to address certain consequences of industrial capitalism. However, the digitally-enabled expansion of the realm of economic utilization traverses established forms of structuration. This is exemplified by the way it simultaneously

10 The German original reads: "Für die Ausbreitung der IT in die Grauzonen der Schattenarbeit und in die Lebenswelt außerhalb des Erwerbslebens (…) bis hinein in die Spielewelt der Kinder gibt es noch keinen recht(lich)en Namen und keinen Arbeitsrecht-ähnlichen Schutz – erst recht nicht für die neuartigen Verbindungen zwischen Arbeits- und ‚Lebens'welt. Die Lage der Wohnweltbetroffenen ist rechtspolitisch nicht einfach zu würdigen" (Steinmüller 1988, 157).

affects the spatial-institutional private sphere and individual information privacy. We will therefore need regulatory innovation that builds upon, but also goes beyond established regulatory schemes that have co-evolved with industrial society. Hence, the socio-digital restructuration of society and the digital expansion of the realm of economic utilization as it materializes in the deployment of smart speakers in private homes urges us to use our heads in more creative ways than just in shaking them.

Acknowledgments

I want to thank the editors and reviewers for thoroughly reviewing this chapter, thus having supported this contribution in terms of content and language. Still, the author is responsible for remaining flaws, of course.

References

Arendt, Hannah. 1975. *Elemente und Ursprünge totaler Herrschaft, Teil 3: Totale Herrschaft.* Frankfurt am Main: Ullstein.

Arendt, Hannah. 2002. *Vita Activa oder Vom tätigen Leben.* München: Piper.

Ariès, Philippe. 1977. "The Family and the City." *Daedalus* 106 (2): 227–35.

Ariès, Philippe. 1991. "Einleitung: Zur Geschichte des privaten Lebens." In *Geschichte des privaten Lebens, Band 3. Von der Renaissance zur Aufklärung*, edited by Philippe Ariès and Roger Chartier, 7–19. Frankfurt am Main: S. Fischer.

Bauman, Zygmunt. 1989. *Modernity and the Holocaust.* Cambridge (UK): Polity.

Beck, Ulrich. 1986. *Risikogesellschaft. Auf dem Weg in eine andere Moderne.* Frankfurt am Main: Suhrkamp.

Beck, Ulrich, Anthony Giddens, and Scott Lash. 1994. *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order.* Stanford (Cal.): Stanford University Press.

Beniger, James. 1986. *The Control Revolution. Technological and Economic Origins of the Information Society.* Cambridge, MA, London: Harvard University Press.

Berger, Peter L., Brigitte Berger, and Hansfried Kellner. 1975. *Das Unbehagen in der Modernität.* Frankfurt am Main: Campus.

Berlinghoff, Marcel. 2013. "Totalerfassung im Computerstaat. Computer und Privatheit in den 1970er und 1980er Jahren." In *Im Sog des Internets. Öf-*

fentlichkeit und Privatheit im digitalen Zeitalter, edited by Ulrike Ackermann, 93–110. Frankfurt am Main: Humanities Online.

Bourdieu, Pierre. 1987. *Die feinen Unterschiede: Kritik der gesellschaftlichen Urteilskraft.* Frankfurt am Main: Suhrkamp.

Brandt, Rüdiger. 1997. "… his stupris incumbere non pertimescit publice: Heimlichkeit zum Schutz sozialer Konformität im Mittelalter." In *Geheimnis und Öffentlichkeit*, edited by Aleida Assmann, Jan Assmann, Alois Hahn, and Hans-Jürgen Lüsebrink, 71–88. München: Fink.

Burkart, Günter. 2002. "Stufen der Privatheit und die diskursive Ordnung der Familie." *Soziale Welt* 53 (4): 397–413.

Crawford, Jane, and Vladan Joler. 2018. "Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Human Labor, Data and Planetary Resources." https://anatomyof.ai/img/ai-anatomy-publication.pdf.

Dolata, Ulrich, and Jan-Felix Schape. Forthcoming. "Bewertung, Inwertsetzung, Regulierung: Eine kurze politische Ökonomie des plattformbasierten Internets." In *Varianten digitaler Bewertung: Rahmungen, Aushandlungen und Widersprüche digitaler Bewertungspraktiken*, edited by Karoline Krenn, Jonathan Kropf, Stefan Laser, and Carsten Ochs. Wiesbaden: VS.

Elias, Norbert. 1997. *Über den Prozeß der Zivilisation. Soziogenetische und psychogenetische Untersuchungen: Band 1.* Frankfurt am Main: Suhrkamp.

Frohman, Larry. 2013. "Rethinking Privacy in the Age of the Mainframe: Integrated Information Systems, the Changing Logic of Privacy, and the Problem of Democratic Politics in Surveillance Societies." In *Im Sog des Internets. Öffentlichkeit und Privatheit im digitalen Zeitalter*, edited by Ulrike Ackermann, 71–92. Frankfurt am Main: Humanities Online.

Giddens, Anthony. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley, Los Angeles: University of California Press.

Giddens, Anthony. 1991. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Cambridge, UK, Oxford, UK: Polity.

Gobetti, Daniela. 1997. "Humankind as a System: Private and Public Agency at the Origins of Modern Liberalism." In *Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy*, edited by Jeff Weintraub and Krishan Kumar, 103–32. Chicago, London: University of Chicago Press.

Goffman, Erving. 1959. *The Presentation of Self in Everday Life.* New York: Doubleday.

Habermas, Jürgen. 1990. *Strukturwandel der Öffentlichkeit: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft.* Frankfurt am Main: Suhrkamp.

Habermas, Jürgen. 1995. *Theorie des Kommunikativen Handelns, Band 2: Zur Kritik der funktionalistischen Vernunft.* Frankfurt am Main: Suhrkamp.

Hausen, Karin. 1976. "Die Polarisierung der ‚Geschlechtscharaktere': Eine Spiegelung der Dissoziation von Erwerbs- und Familienleben." In *Arbeiterexistenz im 19. Jahrhundert: Lebensstandard und Lebensgestaltung deutscher Arbeiter und Handwerker*, edited by Werner Conze and Ulrich Engelhardt, 363–93. Stuttgart: Klett-Cotta.

Heilmann, Till. 2015. "Datenarbeit im „Capture"-Kapitalismus: Zur Ausweitung der Verwertungszone im Zeitalter informatischer Überwachung." *Zeitschrift für Medienwissenschaft* 7 (13): 35–48.

Hobsbawm, Eric J. 1994. *Age of Extremes. The short twentieth century 1914–1991.* London: Michael Joseph.

Kumar, Krishan. 1997 "Home: The Promise and Predicament of Private Life at the End of the Twentieth Century." In *Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy*, edited by Jeff Weintraub and Krishan Kumar, 204–36. Chicago, London: University of Chicago Press.

Lindner, Roland. 2023. "Amazon baut weitere Stellen ab." In: *Frankfurter Allgemeine Online*. Accessed June 13 2024. https://www.faz.net/aktuell/wirtschaft/unternehmen/amazon-baut-weitere-stellen-rund-um-alexa-ab-19323202.html.

Luhmann, Niklas. 1989. "Individuum, Individualität, Individualismus." In *Gesellschaftsstruktur und Semantik. Studien zur Wissenssoziologie der modernen Gesellschaft, Band 3,* edited by Niklas Luhmann, 149–258. Frankfurt am Main: Suhrkamp.

Luhmann, Niklas. 1997. *Die Gesellschaft der Gesellschaft. Zweiter Teilband.* Frankfurt am Main: Suhrkamp.

Lundt, Bea. 2008. "Öffentlichkeit" und "Privatheit" in der historischen Geschlechterforschung." In *Das Private neu denken. Erosionen, Ambivalenzen, Leistungen*, edited by Karin Jurczyk and Mechtild Oechsle, 48–69. Münster: Westfälisches Dampfboot.

Meier-Gräwe, Uta. 2008. "Familie, Ökonomie und Gesellschaft – über die Wirkungsmächtigkeit des vermeintlich Privaten." In *Das Private neu denken. Erosionen, Ambivalenzen, Leistungen,* edited by Karin Jurczyk and Mechtild Oechsle, 113–32. Münster: Westfälisches Dampfboot.

Mönkeberg, Sarah. 2014. "Feststellungen der Identität? Über Nutzen und Laster digitaler Sichtbarkeit." *Der Bürger im Staat* (4): 268–75.

Müller, Ursula. 2008. "Privatheit als Ort geschlechtsbezogener Gewalt." In *Das Private neu denken. Erosionen, Ambivalenzen, Leistungen*, edited by Karin Jurczyk and Mechtild Oechsle, 224–45. Münster: Westfälisches Dampfboot.

Nissenbaum, Helen. 2010. *Privacy in Context. Technology, Policy, and the Integrity of Social Life.* Stanford, CA: Stanford University Press.

Ochs, Carsten. 2021. "Digital Infrastructures Suck: Zur digitalen Absorption des Sozialen." In: "Welt. Wissen. Gestalten. 42. Kongress Der Deutschen Gesellschaft Für Volkskunde (dgv) 2019." Special issue, *Hamburger Journal für Kulturanthropologie* (13): 127–38. https://journals.sub.uni-hamburg.de/hjk/issue/view/92.

Ochs, Carsten. 2022. *Soziologie der Privatheit: Informationelle Teilhabebeschränkung vom Reputation Management bis zum Recht auf Unberechenbarkeit.* Weilerswist: Velbrück. https://doi.org/10.5771/9783748914877.

Pleck, Elizabeth H. 1976. "Two Worlds in One: Work and Family." *Journal of Social History* 10 (2): 178–95.

Reckwitz, Andreas. 2006. *Das Hybride Subjekt: Eine Theorie der Subjektkulturen von der bürgerlichen Moderne zur Postmoderne.* Weilerswist: Velbrück Wissenschaft.

Roessler, Beate. 2004. *The Value of Privacy.* Stanford, CA: Stanford University Press.

Rössler, Beate. 2001. *Der Wert des Privaten.* Frankfurt am Main: Suhrkamp.

Rössler, Beate. 2010. "Privatheit und Autonomie: Zum individuellen und gesellschaftlichen Wert des Privaten." In *Die Grenzen des Privaten*, edited by Sandra Seubert and Peter Niesen, 41–57. Baden-Baden: Nomos.

Sennett, Richard. 2002. *The Fall of Public Man.* London: Penguin.

Shaw, Diane. 1996. "The Construction of the Private in Medieval London." *Journal of Medieval and Early Modern Studies* 26 (3): 447–66.

Shibutani, Tamotsu. 1955. "Reference Groups as Perspectives." *American Journal of Sociology* 60 (6): 562–69.

Smelser, Neil J. 1967. "Sociological History: The Industrial Revolution and the British Working-Class Family." *Journal of Social History* 1 (1): 17–35.

Spiegel. 2024. "Datenschützer legen Beschwerde gegen Meta wegen KI-Trainings mit Nutzerdaten ein." In: *Der Spiegel Online.* Accessed June 13 2024. https://www.spiegel.de/netzwelt/netzpolitik/noyb-datenschuetzer-legen-beschwerde-gegen-meta-wegen-ki-training-mit-nutzerdaten-ein-a-31d98e77-621f-4cfb-b9ab-4aee207b0c7f.

Steinmüller, Werner. 1988. "Rechtspolitik im Kontext einer informatisierten Gesellschaft." In *Verdatet und Vernetzt. Sozialökologische Handlungsspielräu-*

me in der Informationsgesellschaft, edited by Werner Steinmüller, 142–64: Fischer.

Volkszählungsurteil BVerfG 22. Bundesverfassungsgericht, December 15. Accessed October 13, 2022. https://www.bundesverfassungsgericht.de/SharedDocs/Downloads/DE/1983/12/rs19831215_1bvr020983.pdf?__blob=publicationFile&v=1.

Wagner, Peter. 1998. *A Sociology of Modernity: Liberty and Discipline.* London: Routledge.

Warren, Samuel D., and Louis D. Brandeis. 1890. "The Right to Privacy." *Harvard Law Review* 4 (5): 193–220.

Whyte, William Hollingsworth. 2002. *The Organization Man*. Philadelphia: University of Pennsylvania Press.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism. The Fight for the Future at the New Frontier of Power.* London: Profile.

Voice Assistants, Capitalism, and the Surveillance of Social Reproduction

Markus Kienscherf

Abstract *Drawing on the Marxian critique of political economy and feminist social reproduction theory, this contribution examines the role of voice assistants (VAs) or intelligent personal assistants (IPAs) in the reproduction of labor and capital. I argue that the appropriation of user-generated voice data serves the purpose of streamlining and accelerating the circulation and consumption of commodities, and, thus, ought to be understood as a continuation, or even radicalization, of classical capitalist accumulation. I reach this conclusion in two steps. Firstly, I locate the appropriation of user-generated voice data captured by smart speakers within a more general history of the role of surveillance in (re)producing capitalist social relations. Indeed, surveillance has been, and continues to be, central to (a) the appropriation of surplus value in the sphere of production; (b) the social reproduction of labor power; and (c) the management of circulation and consumption. In short, surveillance has been key in trying to fix some of capitalism's most important contradictions. Secondly, I analyze the business models of the corporations behind the three most prominent brands of smart speakers – Apple, Amazon, and Google – to show how the appropriation of user-generated data via smart speakers marks an extension of capitalist surveillance into the sphere of social reproduction.*

1. Introduction

In recent decades, social theory has seen a proliferation of diagnoses of novel forms of capitalism or even proclamations that we have reached the end of capitalism (as we know it). Labels such as cognitive capitalism (Couldry and Mejias 2019; Fumagalli 2010; Vercellone 2010), data capitalism (Sadowski 2019), digital capitalism (Fuchs 2018; Fuchs and Mosco 2017; Sadowski 2020; Schiller 1999), platform capitalism (Langley and Leyshon 2017; Srnicek 2017) and surveillance capitalism (Foster and McChesney 2014; Zuboff 2019) imply that networked digital technologies have facilitated the emergence of new forms of capitalism or have even led to a fundamental break with the logic of capitalist accumulation (most recently, Varoufakis 2023).

According to Shoshana Zuboff (2019), the appropriation of user-generated voice data by means of smart speakers exemplifies the logic of 'surveillance capitalism', which, for her, marks a clear break with classical capitalist accumulation. Focusing on the case of voice assistants (VAs) or intelligent personal assistants (IPAs), which have entered many private households in the form of so-called smart speakers, I propose that we ought rather to understand the socio-economic role of networked digital technologies as well as their surveillance function in more traditional Marxian terms. I will show that the appropriation of user-generated voice data serves the purpose of streamlining and accelerating the circulation and consumption of commodities, and must therefore be understood as a continuation, or even radicalization, of classical capitalist accumulation. Firstly, surveillance capitalism in general, and the appropriation of user-generated voice data captured by smart speakers in particular, ought to be located within a more general history of the role of surveillance in (re)producing capitalist social relations. Building on the work of Andrejevic (2007), Fuchs (2013), Ferguson (2020), and Fortunati (1995), I will show in the following that surveillance has been, and continues to be, central to (a) the appropriation of surplus value in the sphere of production; (b) the social reproduction of labor power; and (c) the management of circulation and consumption. In short, surveillance has been key in *trying* to fix some of the central contradictions of capitalism. Secondly, I will analyze the business models of the three most well-known providers of digital voice assistants – Apple, Amazon, and Google – to demonstrate that the appropriation of user-generated data attained by smart speakers is part of a wider extension of capitalist surveillance into the sphere of social reproduction in order to sell more commodities more quickly.

2. Capitalist Accumulation and Social Reproduction

To map out the role of surveillance in and for both capitalist accumulation and social reproduction, I will take a brief detour through the Marxian critique of political economy. In the first volume of *Capital*, Marx (Marx 1976) argues that commodities with different qualitative use values can only enter purely quantitative exchange relations because they are all products of human labor. "Socially necessary labor time" (129) determines a commodity's value, which is, in turn, represented by its exchange value in relation to other commodities, and ultimately expressed in terms of a price. The peculiarity of the commodity of labor power is that it is the only commodity that can produce more value than it itself has. Labor power also has a value, namely the socially necessary labor time for producing the commodities needed to sustain a laborer at a historically and geographically specific standard of living (275). The value of labor power is reproduced after a certain time (necessary labor time), but if laborers are made to work longer and/or more productively (surplus labor time) than required to reproduce the value of their labor power, capital has obtained surplus value (325). For Marx, exploitation is expressed in the contractual obligation of laborers to work longer (absolute surplus value) and/or more productively (relative surplus value) than necessary to produce the value of the commodities they need to sustain themselves at a historically and geographically specific standard of living (643–654). In this sense, exploitation is the sole source of surplus value, and the continuous productive reinvestment of at least some portion of surplus value – what Marx calls capital accumulation or valorization – is what ultimately defines the capitalist mode of production (725–734).

This is a powerful critique of capitalism, but, as many feminist theorists and activists have argued, it falls somewhat short, because it fails to address the additional work necessary for reproducing both individual workers and the working class (Bhattacharya 2017; Bakker 2007; Dalla Costa and James 1972; Ferguson 2020; Fortunati 1995; Fraser 2014; Glenn 1992; Katz 2001; Kienscherf and Thumm 2024; Mezzadri 2021; Naidu 2022; Mies 2014; Picchio 1992; Vogel 2013). Workers receive a money wage that is supposed to cover all the expenses required to sustain a specific standard of living. But this money wage needs to be converted into readily consumable use values. The adage that you cannot eat money holds particularly true here. For example, buying groceries and preparing a meal after a day of work requires additional labor. Hence, all sorts of additional labor processes and labor times are necessary for (re)producing both workers and the working class, on top of the labor time spent earning the wage:

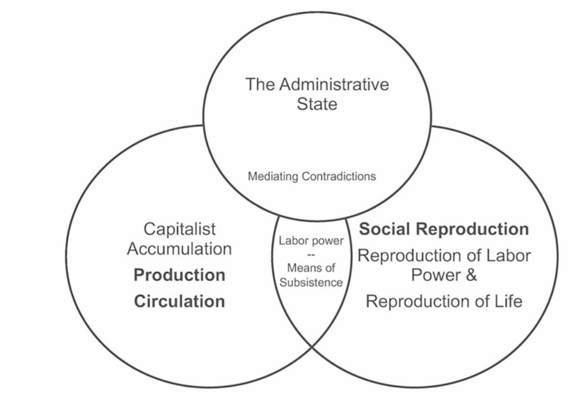
> [Marx] does not realize that the individual male worker's consumption is not a direct consumption of the wage, that the wage does not have an immediate use-value for the male worker and that the consumption of the

> wage's use-value presupposes that some other work has taken place […] (Fortunati 1995, 49).

The work that goes into producing and reproducing both workers and labor power is termed social reproductive work, while the overall process is called social reproduction. While much social reproductive work is performed within the household, a large amount is also performed by public and private sector organizations. Social reproductive work, moreover, may be commodified to a greater or lesser extent and may be waged or unwaged. Moreover, most social reproductive work has been and continues to be disproportionately performed by women. And this holds particularly true for household-based unwaged social reproductive work (see also Kienscherf and Thumm, 2024).

Under capitalism, employers do not simply want as much of their employees' labor time as possible but also labor of a particular quality, intensity, and productivity. Employers pursue absolute surplus value by having their workers work longer than needed to reproduce the value of their labor power, and relative surplus value by having them work as intensively and/or productively as possible. Capitalist accumulation, therefore, pivots on controlling labor in terms of both duration and intensity. This has serious repercussions for social reproduction. The more time workers spend performing waged labor to generate capital, the less time they have for engaging in reproductive work, either for themselves or for their families and communities. The more intensive their work hours, the less energy they have for performing reproductive work. When subject to the capitalist logic of value, then, increasing the duration and intensity of labor time severely undermines workers' capacity for social reproduction. On the other hand, having too little or no access to waged labor may also undermine workers' capacities for social reproduction, because under capitalism they must buy their means of subsistence with the money wage they receive in exchange for their labor power. Hence, there is not only a contradiction *between* capitalist accumulation and social reproduction but also a contradiction *within* social reproduction between the (re)production of human life and the (re)production of labor power. Capitalist accumulation depends on the availability of labor power, but its exploitation of labor undermines the conditions not only for the reproduction of labor power but also for the reproduction of life itself, so that the state has to step in to secure the condition of capitalist accumulation. This simple Venn diagram (Figure 1) serves to illustrate the contradictory relations between capitalist accumulation and social reproduction:

Figure 1: The relations between capitalist accumulation, social reproduction, and the state.



Some feminist theorists tend to reduce social reproduction to the production of labor-power-as-commodity (e.g., Dalla Costa and James 1972; Fortunati 1995). Yet, social reproduction also produces human life itself – in biological, social, and cultural terms. It is only under capitalism that human beings become the bearers of the commodity of labor power. In fact, there are many aspects of social reproductive work that point beyond the capitalist imperative of value (see Ferguson 2020). For one, even in its waged forms, social reproductive work is not nearly as susceptible to the treadmill effect as commodity-producing types of labor are – although that is not for lack for trying. Indeed, productivity metrics often fall short when applied to labor processes that deal with human beings. It is precisely because it does not directly produce value for capital that so much social reproductive labor is either relatively badly paid or completely unwaged

3. Surveillance of Production, Circulation, and Social Reproduction

Every mode of production that seeks to extract surplus from producers requires some form of surveillance – at least in the sense of basic supervision – to ensure that workers perform the required work. This held just as true for slave production in ancient Greece and feudalism in the medieval period as it does for capitalism. What distinguishes the capitalist mode from other modes of production is that under capitalism surplus production is no longer a means to an end but becomes an end in itself. As Ellen Meiksins Wood puts it, "the production of goods and services is subordinate to the production of capital and capitalist profit. The basic objective of the capitalist system, in other words, is the production and self-expansion of capital" (2002, 9). This is why capitalism aims to constantly increase labor productivity and thereby extract more relative surplus value. This leads to a particular type of labor extraction problem, as the extraction of relative surplus value requires the extraction from workers of not just a particular kind of labor for a specific amount of time, but of labor of a particular intensity and productivity. Increasing the extraction of relative surplus value thus requires not only the supervision of workers to ensure they work, but also the collection and analysis of data about the production process in order to evaluate it and, based upon the assessment, take measures to boost productivity. This evaluation process is what ultimately gives rise to the infamous treadmill effect whereby each productivity gain becomes the new baseline against which productivity is subsequently measured. Increases in productivity raise "the amount of value produced per unit of time – until this productivity becomes generalized; at that point the magnitude of value yielded in that time period, because of its abstract and general temporal determination, falls back to its previous level" (Postone 1993, 289). Taylor's *Principles of Scientific Management* provide perhaps the best-known analysis of the use of surveillance for the purpose of extracting relative surplus value from labor (Taylor 1911; see also Braverman 1974). Over time, surveillance of and control over workers has been progressively inscribed into the very technological design of the labor process (Braverman 1974). In the early days of capitalism, capitalists took control of traditional labor processes and appropriated them for the purpose of accumulation. This is what Marx (1976, 645) calls the *formal subsumption* of labor by capital. But, as the capitalist mode of production expanded, capitalists began to (re)design labor processes in order to meet their objectives to extract ever more relative surplus value. This is what Marx (1976, 645) calls the *real subsumption* of labor by capital. Surveillance, initially in the form of close direct supervision and later in the sense of data collection and analysis, has played and continues to play a central part in facilitating capital's *real* subsumption of labor. Capitalist surveillance in the sphere of production thus helped consolidate the capitalist mode of production. We could call the period of the consolidation of capitalism in the late 19th/early 20th century Taylorism – characterized by intensive accumulation without mass consumption (see Jessop and Sum 2006).

Once capital had established tight control over the production process, some capitalists also tried to extend factory-floor-like surveillance to their workers' reproductive sphere, i.e., to their private households. Henry Ford's (in)famous sociology department is a case in point. Capitalists' surveillance of their workers' social reproduction served the general purpose of ensuring that workers' lifestyles would not interfere with the imperative of producing surplus value. Employers, therefore, surveilled working-class consumption habits in order to promote conventions around sobriety, cleanliness, good housekeeping, and the like (Meyer III 1981; quoted in Roediger and Esch 2012). Early capitalist surveillance of working-class households was also driven by the distinct paternalism of particular capitalists who sought to shape their workers' behaviors according to their own religious and political beliefs. Capitalist surveillance of workers' social reproduction persists, for example in dormitory production systems (Schling 2017). Yet, in the Global North, capitalist surveillance of working-class social reproduction has for the most part been replaced by state surveillance, which emerged in response to the dislocations brought about by unfettered capitalist accumulation. In fact, unfettered capitalist accumulation ends up undermining the very conditions for accumulation. As Marx (1976, 375–6) writes in *Capital, Vol. I*:

> But in its blind and measureless drive, its insatiable appetite for surplus labour, capital oversteps not only the moral but even the physical limits of the working day. […] By extending the working day, therefore, capitalist production, which is essentially the production of surplus value, the absorption of surplus labour, not only produces a deterioration of human labour-power by robbing it of its normal moral and physical conditions of development and activity but also produces the premature exhaustion and death of this labour-power itself.

The various social dislocations caused by capitalist accumulation – in terms of not only working-class health and well-being but also of overall societal health and well-being – gave rise to what Karl Polanyi (1957, 151–157) famously called the double movement: the enactment of protective legislation to secure not just the reproduction of labor power but also the reproduction of life itself. This occurred partly in response to the class-based demands for shorter working days, occupational health and safety measures, and various forms of welfare (see Mohandesi and Teitelman 2017; Piven Fox and Cloward 1993). Yet, many social protective measures and regulations also arose out of concerns that were not class-specific. The bourgeoisie was also concerned about pollution and the quality of industrially-produced foodstuffs. Just consider the reception of Upton Sinclair's *The Jungle* (1973 [1905]), a muckraking novel about labor conditions and capitalist exploitation in the Chicago meatpacking industry at the beginning of the 20th century. Most bourgeois readers, including President Theodore Roosevelt, were far more disturbed by the stomach-turning description of industrial food production than by that of the labor conditions (see, for example, Pickavance 2010). They were, after all, much more likely to eat industrially-produced meat products than to work in a meatpacking plant. *The Jungle* thus played a key role in raising concerns that led to the passing of the Pure Food and Drug Act in 1906. Faced with the central contradiction between capitalist accumulation and social reproduction, the modern administrative state arose as the *formally* neutral protector of the conditions for capitalist accumulation. The administrative state thus came to mediate between the imperative of accumulation and the need for stable social reproduction. This historical process unfolded with considerable local variation across the Global North between the second half of the 19th century and the end of World War II. In the case of the US, the development began with the rise of the progressive movement at the end of the 19th century and culminated with the New Deal in the early 1930s. The following – far from complete – list shows that the modern administrative state has developed enormous domestic surveillance capabilities: the administrative state surveils the sphere of *production* to enforce environmental standards, health and safety standards, food and drug purity standards, labor practices, etc.; the sphere of *circulation* to make and enforce market rules, to guarantee consumer safety, etc.; and the sphere of *social reproduction* to assess citizens' eligibility for welfare programs, to guarantee the safety and well-being of children, to police working-class lifestyles, etc. (see Kienscherf 2019, 2021). In brief, by way of surveillance, the modern administrative state seeks to mediate the contradiction between capitalist accumulation and social production, as well as the contradiction between the reproduction of life itself and the reproduction of labor power within social reproduction (see Figure 1).

Over the course of the 19th century, capital came to deploy increasingly sophisticated forms of surveillance to gain almost full control over labor in the process of production. But for capital to accumulate, it must also successfully pass through the sphere of circulation. Rising productive throughputs thus prompted the need to exert more control over the sphere of circulation (see Beniger 1986). This brings us to the period of Fordism, which was characterized by intensive accumulation with mass consumption. While under Taylorism techniques of surveillance had been developed and deployed to manage the production process, Fordism saw these techniques of surveillance extended into the sphere of circulation, as well as the development of new ones, as can be seen in the rise of the mass communication, market research, and advertising industries (see Andrejevic 2007).

The political, economic, and social crisis of the 1970s, however, precipitated the contemporary Post-Fordist capitalist period (see Jessop 2002), which is characterized by flexible accumulation alongside customized production and consumption. This includes the extension of precarious employment situations to hitherto relatively privileged populations, alongside changes in production, which have been both facilitated by and have given rise to new transportation, information, communication, and surveillance technologies. The shift from just-in-case to just-in-time production and the advance of mass customization – "high volume and high mix" production (Eastwood 1996) – hinges on the collection, sharing, and analysis of data within and across corporations in order to manage increasingly complex production processes and supply chains. On the one hand, real-time surveillance of intricate supply chains has become essential to manage the geographically dispersed production and circulation processes that characterize Post-Fordism. On the other, the production of ever more customizable commodities at ever higher volumes also required managing consumer demand by deploying increasingly precision-targeted techniques of marketing and advertising. Sabine Pfeiffer (2022) calls this the development of the *distributive* forces of capitalism that, unlike productive forces, are not geared towards *producing* value but towards *realizing* value as efficiently as possible.

This is the context in which capitalist surveillance in the sphere of circulation has been extended into the sphere of social reproduction. The algorithmic selection, combination, and analysis of data produced by people's interactions on and with digital platforms has facilitated the analysis of who is exposed to which advertisements and how that exposure affects their consumptive behavior (Andrejevic 2007; Dyer-Witheford 2015; Srnicek 2017). Micro-targeted advertising, if it is to be based on reliable information about the preferences, wishes, and desires of ever more finely-grained consumer demographics, requires access to data not only about people's patterns of consumption but also about their more general patterns of social reproduction. This is how the digital platform-based surveillance of consumers differs from the 'mere' surveillance of consumption at sites of consumption, such as supermarkets. Supermarket loyalty cards, for instance, monitor only one particular aspect of people's social reproduction: their interaction with capitalist markets (Trurow 2017). Digital platforms like Google and Facebook, by contrast, collect data about all the interactions that the platforms facilitate. While capitalist surveillance of circulation under Fordism may indeed be construed as an early foray into the monitoring of social reproduction in order to accelerate the circulation of commodities, in its early days, Fordist marketing and advertising surveillance was focused more on markets than on market actors, more on consumption than on consumers. Significantly, the shift from tracking consumption to monitoring consumers marks the extension of capitalist surveillance of circulation into the sphere of social reproduction.

Many social reproductive activities now take place online, and digital advertising platforms, like Google and Facebook, facilitate the extraction of data that users generate while interacting with one another via these platforms and/or with the platforms themselves:

> Platforms allow surveillance capital to channel activities that happen outside the logic of capitalist accumulation (but are still a condition for its reproduction) into processes of valorization. By engaging in these activities on platforms, users produce data that surveillance capital then expropriates through almost ubiquitous surveillance. (Kienscherf 2022, 23)

This is what Shoshana Zuboff (2019) calls "surveillance capitalism": which is not, I argue, a new form of capitalism, but rather the extension of capitalist surveillance into the sphere of social reproduction with the aim of shaping and controlling consumer demand (see Kienscherf 2022).

4. Personal Digital Assistants in Capitalist Accumulation and Social Reproduction

Despite attempts to channel ever more human behavior through digital platforms, many processes of social reproduction still take place offline and, thus, have eluded the reach of platform surveillance – until recently. Now platforms have acquired 'eyes and ears' that extend into offline spaces. This is where the internet-of-things and 'smart' everyday objects, such as smartphones, smart watches, smart fridges, smart thermostats, and smart speakers, enter the equation. What all these 'smart' everyday objects have in common is that they are connected to online platforms and they are equipped with sensors that allow for the appropriation of offline data (Sadowski 2020; Turow 2021). Indeed, "the 'personal digital assistant' is revealed as a market avatar, another Trojan horse in which the determination to render and monetize your life is secreted under the veil of 'assistance' and embellished with the poetry of 'personalization'" (Zuboff 2019, 260). Voice assistants thus play a key role in endeavors to subject offline social reproduction to capitalist surveillance. Waldecker and Volmer (2022) point out that voice data, due to its embodied quality, may contain information on age, gender, mood, health, or personality. This is why the prospect of appropriating vast amounts of voice data is so appealing to the advertising industry.

At the same time, as Waldecker and Volmer (2022) show, in practice, voice assistants are often perceived as somewhat obtuse maids. Indeed, it is no coincidence that voice assistants tend to have female names (Alexa and Siri) and feminine voices: this situates them squarely in the feminized domestic sphere within a gendered division of labor (see Strengers & Kennedy 2020). Moreover, in everyday interaction voice assistants may seem somewhat obtuse, because voice recognition software does not always work as advertised and users often need to repeat their commands several times in order to get the required response. Yet, voice assistants ultimately elude the control of their users not because of their 'obtuseness' but because of their 'smartness': voice assistants are embedded in distributed digital platforms and, as such, serve the extraction, analysis, and ultimately the monetization of everyday household communication. On the one hand, the voice data generated by the interactions between users and voice assistants provide training data used to help optimize a given system's acoustic intelligence (rendering them less obtuse). On the other hand, the same data can also be used for producing fine-grained consumer profiles that are a prerequisite for targeted advertising.

The situation is further complicated by important differences between the business models of the providers of voice assistants. Apple's voice assistant, Siri, is part of its range of upscale and high-margin gadgets. Apple claims to only use user-generated data as training data to improve its own systems. In its legal guidelines, Apple explicitly states that "Siri data is not used to build a marketing profile, and is never sold to anyone" (Apple 2023). However, Apple does not specify what happens to data shared through third-party integration with Siri, because "When Siri interacts with a third-party app on your behalf, you are subject to that app's terms and conditions and privacy policy" (Apple 2023). In Marxian parlance, it seems as if Apple operates as *industrial capital* that includes voice assistants within its range of strongly branded, high-margin commodities proclaiming high standards of data security, while third parties that gain access to voice data by integrating their apps with Siri may still use that voice data for the purpose of targeted advertising. External app providers are able to receive, store, and exploit relevant voice data if their app is integrated with Siri and if the user grants the app the necessary access permissions – which is obligatory in order to use the Siri feature with the app (Apple 2021).

Amazon's voice assistant, Alexa, serves first and foremost as a direct interface to its online retail platform. This is why the "Alexa Terms of Use" go to great lengths to legally specify the practices of "voice purchasing" that it facilitates (Amazon 2023a). However, Amazon's general Privacy Notice, to which its Alexa products are also subject, clearly states:

> We provide ad companies with information that allows them to serve you with more useful and relevant Amazon ads and to measure their effectiveness. We never share your name or other information that directly identifies you when we do this. Instead, we use an advertising identifier like a cookie, a device identifier, or a code derived from applying irreversible cryptography to other information like an email address. … While we do not share your specific shopping actions like purchases, product views, or searches with ad companies, we may share an advertising identifier and an estimate of the value of the ads they show you on our behalf so they can serve you with more effective Amazon ads. Some ad companies also use this information to serve you relevant ads from other advertisers. (Amazon 2023b)

No information is offered indicating to what extent parameters associated with voice data in particular feed into the construction of "an advertising identifier and estimate of the value of the ads" that are shown to users, but it can be assumed that they do. For the most part, then, Amazon operates as *commercial capital* (Marx 1981, 379–393) that sells its voice assistant systems at cost in the hope that Alexa may ultimately help increase and accelerate the turnover of commodities. Thus far, however, Amazon has been losing money on its Alexa venture, with users showing reluctance to make voice purchases (Olson 2022; Kim 2022).

Google's voice assistant is an integral part of its overall digital architecture for the extraction of user data. In its privacy policy, Google stresses that it collects data – including "voice and audio data" – primarily for the purpose of "building better service" which, notably, also includes personalized ads (Google 2023). The "Google Privacy Policy" mentions personalized ads as something of an afterthought to its mission to constantly improve people's digital lives, while failing to mention that advertising revenue makes up more than 80 percent of Google's total revenue (Statista 2023). Google's voice assistant, then, is best viewed against the backdrop of Google's operation as *surveillance capital* – a subset of commercial capital that produces commodities of a very specific kind: finely-grained consumer profiles based on data extracted from people's everyday lives (see Kienscherf 2022).

The characterization of Apple as industrial capital, Amazon as commercial capital and Google as surveillance capital is a distinction of ideal types. Nonetheless, it highlights important differences in how voice assistants figure within these corporations' specific business models. Yet this differentiation ought not be read as a moral judgment in the vein of Zuboff's (2019, 28–31) distinction between Apple's benign form of capitalist disruption and Google's 'bad' surveillance capitalism. Indeed, these different business models are not indicative of different types of capitalism but of different – albeit closely entangled – processes *within* capitalist accumulation. In short, all three major providers of voice assistants harness voice data to optimize their own systems and they all – albeit with some variations –allow for the sharing of voice data with third parties, but they appear to differ in how voice data figure within their respective business models. Ultimately, the use to which these corporations put user-generated data in general and voice data in particular depends on how they position themselves within the overall circuit of capitalist accumulation.

5. Conclusion

Surveillance has long played and continues to play a key role in smoothing out the overall cycle of capitalist accumulation. In the sphere of production, surveillance facilitates capital's extraction of relative surplus value from labor, while in the sphere of circulation, it speeds up the exchange of commodities. In raising labor productivity and in cutting both production and circulation time, surveillance thus accelerates the overall turnover of capital and hence helps boost capitalist accumulation (see Marx 1978, 316–33). State surveillance of production, circulation, and social reproduction, on the other hand, aims to mediate the more general contradiction between capitalist accumulation and social production, as well as the contradiction between the reproduction of life itself and reproduction of labor power within social reproduction.

In the contemporary Post-Fordist era, the production of ever more customizable commodities at ever higher volumes makes it necessary to manage consumer demand by means of increasingly precision-targeted techniques of marketing and advertising. To this end, corporations have sought to extend commercial surveillance into the sphere of social reproduction, enabling them to tap their users' data for the purpose of micro-targeted advertising. The first step in this process was to channel increasing numbers of social reproductive activities to flow via digital platforms so that the data produced could be easily appropriated. The second step was to roll out smart technologies with platform-linked sensors that allow them to capture data generated in hitherto offline spaces of social reproduction such as private homes (Sadowski 2020; Turow 2021; Zuboff 2019). Digital voice assistants have thus become a tool to capture voice data from within private households.

The three major providers of digital voice assistants, Apple, Amazon, and Google, ultimately harness their users' voice data as part of a more general effort to accelerate the turnover of their specific commodities: high-margin electronic gadgets in the case of Apple, all sorts of different commodities in the case of Amazon, and fine-grained behavioral profiles alongside digital advertising space in the case of Google. The respective business models of Apple, Amazon, and Google thus operate within the overall imperative of capitalist accumulation and by no means herald a radically new form of capitalism. In fact, the appropriation of voice data 'merely' marks a further extension of capitalist surveillance, which was previously limited to market-based social reproduction (buying commodities) and is now deeply embedded within the sphere of reproduction, an ideal vantage point from which to surveille as many aspects of social reproduction as possible.

References

Amazon. 2023a. "Alexa Terms of Use." Available at: https://www.amazon.com/gp/help/customer/display.html?ref_=hp_left_v4_sib&nodeId=G201809740, (accessed November 15, 2023).

Amazon. 2023b. "Amazon.com Privacy Notice." Available at: https://www.amazon.com/gp/help/customer/display.html?nodeId=GX7NJQ4ZB8MHFRNJ#GUID-8966E75F-9B92-4A2B-BFD5-967D57513A40__SECTION_3F77537F901B4157B0CBD26834827652, (accessed November 15, 2023).

Andrejevic, Mark. 2007. *iSpy: Surveillance and Power in the Interactive Era*. Lawrence, KS: University of Kansas Press.

Apple. 2021. “SiriKit security for iOS, iPadOS and WatchOS.” https://support.apple.com/en-euro/guide/security/sece2da44807/web (accessed May 20, 2024).

Apple. 2023. “Legal: Ask Siri, Dictation & Privacy.” Available at: https://www.apple.com/legal/privacy/data/en/ask-siri-dictation/ (accessed November 15, 2023).

Bakker, Isabella. 2007. “Social Reproduction and the Constitution of a Gendered Political Economy.” New Political Economy 12 (4): 541–556.

Beniger, James R. 1986. *The Control Revolution. Technological and Economic Origins of the Information Society*. Cambridge, Mass.: Harvard University Press.

Bhattacharya, Tithi. 2017. “How Not to Skip Class: Social Reproduction of Labor and the Global Working Class.” In *Social Reproduction Theory: Remapping Class, Recentering Oppression*, edited by Tithi Bhattacharya, 68–93. London, UK: Pluto.

Braverman, Harry. 1974. *Labor and monopoly capital: the degradation of work in the twentieth century*. New York, NY: Monthly Review Press.

Couldry, Nick, and Ulises A. Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.

Dalla Costa, Mariarosa, and Selma James. 1972. *The power of women and the subversion of the community*. Edited by Selma James. 3. ed. ed. Bristol: Falling Wall Press.

Dyer-Witheford, Nick. 2015. *Cyber-Proletariat*. London: Pluto Press.

Eastwood, Margaret A. 1996. “Implementing mass customization.” *Computers in Industry* 30 (3): 171–174.

Ferguson, Susan. 2020. *Women and Work: Feminism, Labour, and Social Reproduction*. London: Pluto Press.

Fortunati, Leopoldina. 1995. *The Arcane of Reproduction: Housework, Prostitution, Labor and Capital*. Translated by Hilary Creek. Brooklyn, NY: Autonomedia.

Foster, John Bellamy, and Robert W. McChesney. 2014. Surveillance Capitalism: Monopoly-Finance Capital, the Military-Industrial Complex, and the Digital Age. Monthly Review 66 (3). Available at: https://monthlyreview.org/2014/07/01/surveillance-capitalism/ (accessed 9 March 2021).

Fraser, Nancy. 2014. “Behind Marx’s Hidden Abode: For an Expanded Conception of Capitalism.” New Left Review 86 (March-April): 55–72.

Fuchs, Christian. 2013. "Political Economy and Surveillance Theory." Critical Sociology 39 (5):671-687.

Fuchs, Christian. 2018. "Capitalism, Patriarchy, Slavery, and Racism in the Age of Digital Capitalism and Digital Labour." Critical Sociology 44 (4–5): 677–702.

Fuchs, Christian, and Vincent Mosco, eds. 2017. Marx in the Age of Digital Capitalism. Chicago, IL: Haymarket Books.

Fumagalli, Andrea. 2010. "The Global Economic Crisis and Socioleconomic Governance." In The Crisis in the Global Economy: Financial Markets, Social Struggles, and New Political Scenarios, edited by Andrea Fumagalli and Sandro Mezzadra, 61–83. Los Angeles, CA: Semiotext(e).

Glenn, Evelyn Nakano. 1992. "From Servitude to Service Work: Historical Continuities in the Racial Division of Paid Reproductive Labor." Signs 18 (1): 1–43.

Google. 2023. "Google Privacy Policy, effective November 2023." Available at https://policies.google.com/privacy?hl=en-US, (accessed November 16, 2023).

Jessop, Bob. 2002. The Future of the Capitalist State. Cambridge, UK: Polity Press.

Jessop, Bob, and Ngai-Ling Sum. 2006. Beyond the Regulation *Approach: Putting Capitalist Economies in Their Place*. Cheltenham, UK: Edward Elgar Publishing.

Katz, Cindi. 2001. "Vagabond Capitalism and the Necessity of Social Reproduction." *Antipode* 33 (4):709–728.

Kienscherf, Markus. 2019. "Race, class and persistent coloniality: US policing as liberal pacification." *Capital & Class* 43 (3): 417–436.

Kienscherf, Markus. 2021. "Classifying and dividing labor: the political economy of racializing surveillance." In *Trust and Transparency in an Age of Surveillance*, edited by Lora Anne Viola and Pawel Laidler, 85–103. London: Routledge.

Kienscherf, Markus. 2022. "Surveillance Capital and Post-Fordist Accumulation: Towards a Critical Political Economy of Surveillance-for-Profit." *Surveillance & Society* 20 (1): 18–29.

Kienscherf, Markus, and Clara Thumm. 2024. "Social Reproduction Under Lockdown: Capital, Labor, and Gender." *Feminist Economics* [online before print] https://doi.org/10.1080/13545701.2024.2334231

Kim, Eugene. 2022. "Amazon Is Gutting Its Voice Assistant, Alexa. Employees Describe a Division in Crisis and Huge Losses on 'a Wasted Opportunity.'"

Business Insider. Available at : https://www.businessinsider.com/amazon-alexa-job-layoffs-rise-and-fall-2022-11 (accessed November 22, 2022).

Langley, Paul, and Andrew Leyshon. 2017. "Platform Capitalism. The Intermediation and Capitalisation of Digital Economic Circulation." *Finance and Society* 3 (1): 11–31.

Marx, Karl. 1976. *Capital: A Critique of Political Economy Vol. I*. Translated by Ben Fowkes. London: Penguin Books. Original edition, 1867.

Marx, Karl. 1978. *Capital: A Critique of Political Economy Vol. II*. Translated by David Fernbach. London: Penguin Books.

Marx, Karl. 1981. *Capital: A Critique of Political Economy Vol. III*. Translated by David Fernbach. London: Penguin Books. Original edition, 1894.

Meyer III, Stephen. 1981. *The Five Dollar Day: Labor Management and Social Control in the Ford Motor Company, 1908–1921*. Albany, NY: SUNY Press.

Mezzadri, Alessandra. 2021. "A Value Theory of Inclusion: Informal Labour, the Homeworker, and the Social Reproduction of Value." *Antipode* 53 (4): 1186–1205.

Mies, Maria. 2014. *Patriarchy and Accumulation on a World Scale: Women in the International Division of Labour*. London: Zed Books.

Mohandesi, Salar, and Emma Teitelman. 2017. "Without Reserves." In *Social Reproduction Theory: Remapping Class, Recentering Oppression*, edited by Tithi Bhattacharya, 37–67. London, UK: Pluto.

Naidu, Sirisha C. 2022. "Circuits of Social Reproduction: Nature, Labor, and Capitalism." *Review of Radical Political Economics* 55 (1): 93–111.

Olson, Parmy. 2022. Alexa, Will You Ever Make Money. *The Washington Post* (November 22). Available at: https://www.washingtonpost.com/business/alexa-will-youever-make-money/2022/11/22/53caa54c-6a82-11ed-8619-0b92f0565592_story.html, (accessed November 15, 2023).

Picchio, Antonella. 1992. *Social reproduction: the political economy of the labour market.* Cambridge: Cambridge University Press.

Pickavance, Jason. 2010. "Gastronomic Realism: Upton Sinclair's *The Jungle*, The Fight for Pure Food and the Magic of Mastication." *Food and Foodways* 11 (2–3): 87–112.

Piven Fox, F., and Richard A. Cloward. 1993. *Regulating the Poor: The Function of Public Welfare*. New York, NY: Vintage.

Pfeiffer, Sabine. 2022. *Digital Capitalism and Distributive Forces*. Bielefeld: transcript.

Polanyi, K. 1957. *The Great Transformation: The Political and Economic Origins of Our Time*. Boston, MA: Beacon Press.

Postone, Moishe. 1993. *Time, Labor, and Social Domination: A reinterpretation of Marx's critical theory*. Cambridge: Cambridge University Press.

Roediger, David R., and Elizabeth D. Esch. 2012. *The Production of Difference: Race and the Management of Labor in U.S. History*. Oxford: Oxford University Press.

Sadowski, Jathan. 2019. "When data is capital: Datafication, accumulation, and extraction." *Big Data & Society* 6 (1): 1–12.

Sadowski, Jathan. 2020. *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives, and Taking Over the World*. Cambridge, MA: The MIT Press.

Schiller, Dan. 1999. *Digital Capitalism: Networking the Global Market System*. Cambridge, MA: The MIT Press.

Schling, Hannah. 2017. (Re)Production: Everyday Life In The Workers' Dormitory. *Society + Space* November 7. Accessed November 21, 2023.

Sinclair, Upton. 1973 [1905]. *The Jungle*. 18. print ed. New York, NY: New American Library.

Srnicek, Nick. 2017. *Platform Capitalism*. Cambridge: Polity Press.

Statista. 2023. "Distribution of Google segment revenues from 2027 to 2022." Available at: https://www.statista.com/statistics/1093781/distribution-of-googles-revenues-by-segment/ (accessed November 14, 2023).

Strengers, Yolande, and Jenny Kennedy. 2020. *The Smart Wife. Why Siri, Alexa, and Other Smart Home Devices Need a Feminist Reboot*. Cambridge, MA: MIT Press.

Taylor, Frederick W. 1911. *The Fundamentals of Scientific Management*. New York, NY: Harper & Brothers.

Trurow, Joseph. 2017. *The Ailes have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power*. New Haven, CT: Yale University Press.

Turow, Joseph. 2021. *The Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet*. New Haven, CT: Yale University Press.

Varoufakis, Yanis. 2023. *Techno Feudalism: What Killed Capitalism*. London: Bodley Head.

Vercellone, Carlo. 2010. "The Crisis of the Law of Value and the Becoming-Rent of Profit." In *The Crisis in the Global Economy: Financial Markets, Social Struggles, and New Political Scenarios*, edited by Andrea Fumagalli and Sandro Mezzadra, 85–118. Los Angeles, CA: Semiotext(e).

Vogel, Lise. 2013. *Marxism and the Oppression of Women: Toward a Unitary Theory*. Chicago, Il.: Haymarket Books.

Waldecker, David, and Axel Volmar. 2022. „Die zweifache akustische Intelligenz virtueller Sprachassistenten zwischen verteilter Kooperation und Datafizierung." In *Acoustic Intelligence*, edited by Schürmer Anna, Haberer

Maximilian and Brautschek Tomy, 161–182. Berlin, Boston: Düsseldorf University Press.
Wood, Ellen Meiksins. 2002. *The Origin Of Capitalism: A Longer View*. London: Verso.
Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London: Profile Books.

Machines as Partners

Anthropomorphism and Communication Accommodation to Voice Assistants in Disability Contexts

Caja Thimm, Phillip Engelhardt, and Julia Schmitz

Abstract *This chapter introduces a theoretical approach for the analysis of verbal interaction between humans and machines, and demonstrates its application in a specific social situation. Based on the well-established sociolinguistic model, Communication Accommodation Theory (CAT), we introduce the "CAT Technology Equivalence Model", which helps to identify specific convergence and divergence strategies in verbal communication with machines. Conceptualizing VAs as social actors, a qualitative study was carried out with four households with VAs used by people with care needs. The participants documented their activities with the VAs in media journals and commented on their communication strategies in semi-structured interviews. The aim of the study was to demonstrate implicit and explicit ways of communicative accommodation towards voice assistants in order to better understand how verbal AI systems are anthropomorphized in everyday interactions. Results demonstrate that participants consciously and/or unconsciously adjust their linguistic behavior to accommodate their anthropomorphic framing of Alexa and accommodate it to the perceived logics of the technology. The chapter concludes that, as technology adopts ever more human-like qualities including physical form and voice, the question of 'human-likeness' in shaping speech behaviors will become an even more significant area of study.*

1. Introduction

In millions of homes, voice assistants (VAs) have become the technology of choice for orchestrating an impressive variety of everyday tasks. Operating in response to voice commands, the devices can manage smart home appliances, provide traffic and weather updates, and perform many other duties according to individualized personal preferences. As such, smart speakers represent a form of machine technology that has facilitated widespread access to personalized technological functionalities in the home, with some researchers even calling them "game changers" (Vlahos 2019, 3). Authors argue that the pervasive integration of these assistants has fundamentally transformed our interactions within the home environment and has opened up AI-controlled technology for mass usage. At the same time, smart speakers are seen as a security risk and a threat to privacy. Various scandals involving Amazon employees listening via Alexa to household conversations as they take place in real time have stoked such fears and led to a lack of trust, particularly in data-conscious countries like Germany. In dealing with the tension between the desire for convenience and the unease of mistrust, anthropomorphizing tendencies have been observed, especially among younger individuals. Alexa herself is absolved of responsibility for the alleged privacy breaches ("It is not her fault", Fetterolf and Hertog 2023, 7). This is just one example of how smart speakers attract academic interest not only for their pragmatic utility but also for their capacity to critically reshape the dynamics of communication between humans and machines.

With this broader perspective in mind, our chapter aims to explore a specific aspect of interaction between humans and voice assistants: *types of communicative social interaction* in which individual interlocutors regard the machine as a *social actor* (Lombard and Xu 2021, 29). Building upon the premise that technologies are becoming increasingly 'intelligent' in the sense that they are perceived to be gaining increasingly human-like capacities across various domains including general agency, verbal interaction, and emotion recognition, as well as offering an expanding array of services, we examine selected communication strategies in order to systematically analyze human–machine relationships. For this purpose, we develop a model based on *communication accommodation theory (CAT)*. The model provides a framework for integrating social cues and the social situations within which interactions take place into the analysis of human–VA (or other machines) communication. Furthermore, we propose that the notion of 'anthropomorphism' is a key element that can aid our analysis of human–machine communication. As we elaborate below, we understand anthropomorphism as a *bridging principle* that elucidates the various strategies employed by humans to adapt to the distinctive attributes and uncertainties inherent to communication with machines.

In order to apply this theoretical work to a specific social situation, we chose a setting in which a user's relationship with their VA is not simply supplementary or playful, but characterized by a certain degree of dependency. This social scenario pertains to individuals with disabilities, who face challenges associated with limited control over certain functionalities and diminished personal autonomy. Among the various technological solutions deployed to support individuals with disabilities or cognitive impairments, off-the-shelf conversational agents or voice assistant systems like Amazon's Alexa play an important role in increasing personal autonomy by supporting the management of everyday domestic life (Purington et al. 2017, 2858; Kramer et al. 2013, 1105; Albert et al. 2013, 19). To understand how people with disabilities incorporate VAs into their homes and routines and how they regard their communicative relations with those machines, we carried out a qualitative study in four households that were home to four participants with special needs ('test persons') who used VAs in their homes. Two caregivers from two of the households also took part in the study (B2 and D2).

Table 1: Participants and households

<table>
<tr><th>Household</th><th>Age</th><th>Gender</th><th>Care needs</th><th>Alexa (quantity)</th></tr>
<tr><td>A</td><td>25</td><td>Female</td><td>Yes</td><td>3</td></tr>
<tr><td>B1</td><td>58</td><td>Female</td><td>Yes</td><td rowspan="2">1</td></tr>
<tr><td>B2</td><td>57</td><td>Male</td><td>No</td></tr>
<tr><td>C</td><td>51</td><td>Male</td><td>Yes</td><td>1</td></tr>
<tr><td>D1</td><td>23</td><td>Male</td><td>Yes</td><td rowspan="2">1</td></tr>
<tr><td>D2</td><td>54</td><td>Male</td><td>No</td></tr>
</table>

The participants documented their user experience in media journals and also reflected upon their perceptions and attitudes towards VAs in semi-structured interviews before and after the journaling period. Our aim with this chapter is to illustrate that the ways these participants communicate with the machines demonstrate typical accommodation strategies on a technical and a personal level. Before doing so, we present the theoretical basis of our investigation by briefly outlining the role of anthropomorphic ascriptions and attributions and introducing 'communication accommodation theory' (CAT).

2. Machines as Partners: Computers as Social Actors

As the development of artificial intelligence advances unabated, ever more diverse possibilities for transforming relationships between humans and technology are being highlighted (Thimm 2019, 17). Social robots for care contexts (Henschel et al. 2021, 14), generative language programs (Large Language Models or LLMs) like ChatGPT, and interactive voice assistant systems (such as Alexa or Siri) simulate 'authentic' interpersonal interactions, mimic cognitive processes of emotion recognition, and some even present themselves in humanoid physical forms. With the continuously expanding functional spectrum of artificial intelligence, new scenarios are being addressed and AI systems are operating in ever more social contexts in diverse roles, from a simple executive tool to a more complex 'social companion'. The idea of the so-called 'social robot' in particular has attracted wide attention in recent years (Mahdi et al. 2022, 1; Thimm and Thimm-Braun 2024).

Since 1996, the idea of the social machine has been discussed under the *Computers are Social Actors* (CASA) paradigm (Nass et al. 1994, 72). Machines are no longer perceived within communication processes as media for merely storing, visualizing, and/or distributing information, but are designed, utilized, and studied as communication partners (Guzman and Lewis 2020, 71). Over the years of their development, their features have been categorized as increasingly interactive and responsive, to the extent that they have even been viewed as family friends who deserve legal protection (Darling 2016, 22). In many of these instances, such machines are objects of a technologically-induced anthropomorphization process (Epley et al. 2007, 864; Złotowski et al. 2015, 347).

The drive to develop and interact with technologies that appear to reflect the human condition in practice or physical appearance has increased considerably in recent years. Robots in particular have been designed to display varying degrees of human-like features such as stylized facial expressions or human-like voices, supposedly in order to facilitate anthropomorphization: the process by which human characteristics like motivation, behaviors, and social roles are attributed to nonhuman entities (Ezenkwu and Starkey 2019, 340; Coeckelbergh 2023, 2). As shown by Caporael (1986, 218) or Darling (2016, 22), framing technological artifacts through anthropomorphic language and design can influence human perception and behavior and oftentimes ameliorates human–machine relationships.

Closely connected to the role of anthropomorphism is the notion of trust and trustworthy systems. Humans desire a trustworthy (Kok and Soh 2020, 297), friendly (Fröding and Peterson 2021, 207), transparent (Larsson and Heintz 2020, 1), and emotionally intelligent (McStay 2020, 10) machine that not only meets instrumental criteria such as effectiveness and user-friendliness, but also supplements its functional spectrum with a (para-)social dimension in a human-like manner. Many chatbots are not only regarded as trustworthy, but also present themselves as personal (Cai et al. 2023, 24). ChatGPT, for example, excuses itself for mistakes, acts politely, and addresses users in different ways. When Olasik (2023, 269) titled her paper "Good morning, ChatGPT, Can We Become Friends?", she provided a vivid example of the expectations regarding relationships with a technological interface.

Many other researchers confirm that users exhibit behaviors that can be interpreted as showing empathy with the technical counterpart (Malinowska 2021, 361). Anthropomorphization is not seen as an active projection a priori, but as a passive inference in the moment of sociotechnical interaction experience. This (psychological) process of anthropomorphism is described by Damiano and Dumouchel (2018, 2): "The underlying idea is to actively involve users in the social performances and presence of the robots, by designing robotic agents that stimulate users to attribute human feelings and mental states to robots, which should enhance familiarity and promote social interactions".

We regard anthropomorphization as one of the central modes for bridging the gap between machines and humans. By anthropomorphizing machines, individuals engage in a form of accommodation whereby they adapt their linguistic and physical behavioral cues to better align with the supposed social qualities and performances of the technology. In most human-to-human communication, sociolinguists argue, people adapt their language and behavior according to a desire to establish rapport, reduce social uncertainty, and facilitate smoother interactions. This accommodation process involves both conscious and unconscious adjustments, and is exhibited in interactions with machines as well. Studies have shown that the level of anthropomorphism applied to machines can vary from moment to moment and is influenced by factors such as the machine's design, voice, behavior, and the interaction context. Systems with human-like features, such as humanoid robots or natural-sounding voices, tend to elicit higher levels of anthropomorphism from users (Darling 2016, 22; Wagner and Schramm-Klein 2019, 1). Furthermore, users often employ anthropomorphic language and behavior when interacting with such systems, treating them as social actors rather than as mere tools.

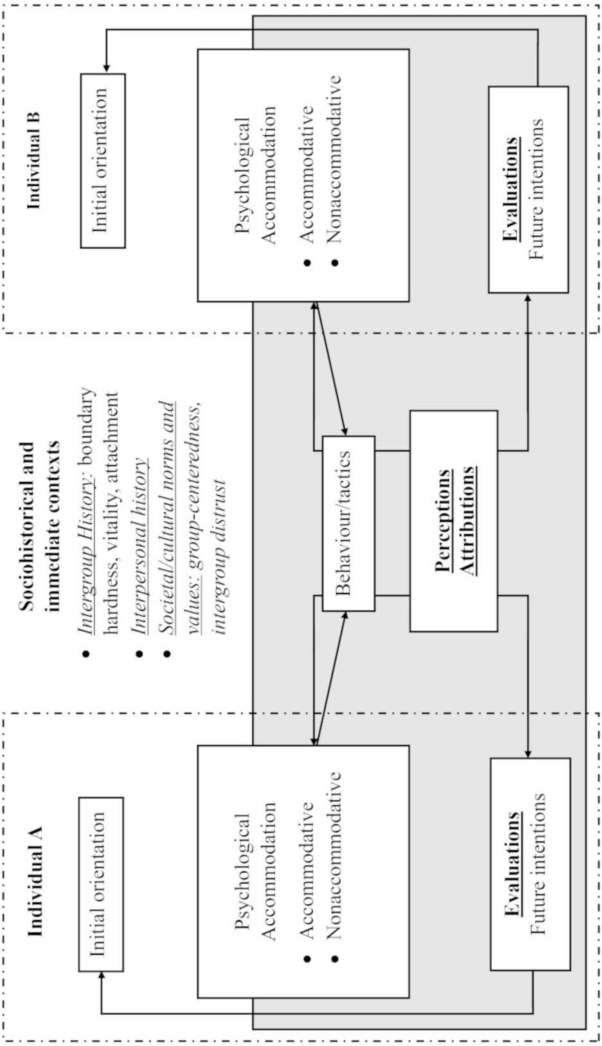
We therefore assume that anthropomorphization not only shapes individual interactions with technical devices but also influences societal perceptions and norms concerning technology on a more general level. Hence, we argue, it is crucial to understand the mechanisms and implications of anthropomorphization in order to design effective human–machine interfaces and to create conditions that promote positive user experiences. This might also include self-reflection on behalf of humans: Guzman and Lewis (2020, 78), for example, suggest that digital interaction partners – such as Alexa – can be instrumentalized as a stimulus to "reimagine the self". Overall, seeing human qualities in machines is as a fundamental aspect of human–machine communication. It can facilitate smoother interactions, but may also potentially provoke feelings of anxiety; thereby shaping the way individuals perceive and interact with technology across different contexts. There remain, however, many open questions concerning emotional and communicative relations between diverse technologies and the humans that interact with them.

In order to systematically study communicative relations between humans and machines, we adapt the idea of *Communication Accommodation Theory* and expand it for machine technologies.

3. Talking with Machines – the "CAT Technology Equivalence Model"

Initially developed as Speech Accommodation Theory (SAT), Communication Accommodation Theory (CAT) describes how a person adapts their communicative acts towards those of their (human) counterpart. This occurs not only at the linguistic level but includes social relations as well (Schreuter et al. 2021, 535). As Edwards et al. (2023, 2) summarize, "CAT proposes that individuals adjust their communication behaviors in response to the actions of others, on the assumption that communication fosters and maintains interpersonal and group relationships", and Giles et al. (2023, 4) explain that "accommodation regulates social interaction by decreasing or increasing social distance between communicators, thereby often reflecting relative social status and power differentials". The functionalities of communication in interpersonal exchange are complex and are not limited to the verbal. Rather, interpersonal negotiation and attribution of social roles play an important role. CAT asserts that this negotiation process implicitly manifests itself on diverse levels: "Communication is not only a matter of merely and only exchanging information about facts, ideas, and emotions (often called referential communications), but salient social category memberships are often negotiated during an interaction through the process of accommodation" (Giles and Ogay 2007, 294).

Figure 1: Revised model of Communication Accommodation Theory (Gallois et al. 2005, 135)



The ways in which individuals accommodate to their human communication partners have been characterized in relation to social status, language variety, and individual speakers' characteristics, among others (cf. Gallois and Giles 1998), with the following convergence strategies identified:

- upwardly or downwardly converging towards the degree of prestige, where relevant, of the language variety used by the communication partner;
- fully or partially accommodating a specific speaker characteristic or a particular constellation of characteristics;
- symmetrically or asymmetrically accommodating such that both or only one partner converges;
- converging at different paces and/or to a varying degree within a single conversation or over a longer time period.

Important factors that can influence the effectiveness of communication according to CAT are the sociohistorical and immediate contexts as well as perceptions and attributions. Gallois et al. (2005, 135) map out the different levels in the following model:

The capacity of CAT to further our understanding and observation of the effects of accommodation has been empirically tested in numerous ways (Gallois et al. 2016, 192). In addition to linguistically-focused studies of accents and dialects, the interaction patterns of convergence and divergence have been the subject of much socio-psychological research. Convergence, as an interpersonal goal, describes the alignment of one's own communicative behavior with the patterns and communication habits (conscious or unconscious) of the other person. Convergent linguistic styles contribute to the formation of sympathy and familiarity, reduce feelings of insecurity and social anxiety, and increase the chances of correctly predicting the behavior of the counterpart and thus aligning the social interaction with one's own need for compliance (Soliz and Giles 2014, 4). Divergent interaction patterns emphasize differences in language and expressive behavior, highlighting the differences between one's own and another's personal or group identity. Convergence and divergence strategies share a common normative starting point with the psycho-affective need for coherence, as divergence strategies often reflect an attempt to uphold the authenticity and integrity of one's own personality against environmental influences.

Whereas these categories have been well researched and the substantial body of research addressing human-to-human communication continues to expand, it is only recently that CAT has been applied to communication with technology (Giles et al. 2023). The starting point of such CAT-based research is the observation of how linguistic styles or linguistic behaviors are adapted in communication scenarios involving a (technological) interaction partner. In interpersonal human-to-human communication research, accommodative behavior is seen as an attempt to incite attitudes of recognition or acceptance; to increase the efficiency of communicative exchange; to create, maintain, or reduce social distance; and to enable the negotiation and maintenance of shared personal and collective identities (Gallois et al. 2005, 127). Research on lexical alignment in particular in human–machine communication (HMC) suggests that here too, users adapt their lexical choices to accommodate their partner's perceived limitations as interlocutors, with greater adaptation to partners perceived as less capable or eloquent (Branigan et al. 2011, 41). Branigan et al. (2010, 2360) suggest that people see agents with human-like qualities as more intelligent and competent than non-anthropomorphic agents. The tendency to align therefore appears to be mediated by evaluations concerning an interlocutor's perceived communicative capacities and deficits, with most humans implicitly assuming that humans' communicative capacities are superior to those of machines.

In recent years, a number of studies have investigated human interactions with social robots (Ahmad et al. 2017, 21; van Pinxteren et al. 2023, 537), productively employing the *computer as social partner* approach (Fortunati and Edwards 2022, 17). The launch of commercialized voice-operated agents like the Google Assistant (2012), Microsoft's Cortana (2013), and Amazon's Alexa (2014) for use in homes and domestic living spaces has added a fruitful context for this perspective as well as for CAT by introducing new communication partners, new modes and norms of communication, and new challenges (Etzrodt and Engesser 2021, 57; Gallois et al. 2016, 206). Studies on communication accommodation to VA systems shed light on how human speakers adapt their communication styles towards those of the devices, particularly in terms of speaking speed and vocal imitation. Linguistic analyses such as Cohn et al. (2019, 1816; 2021, 10) or Cohn et al. (2023, 14) demonstrate particularly clearly that linguistic performance levels are highly dependent on the perception of the sociotechnical interaction as a social situation. And Schreuter et al. (2021, 535) have shown that a VA's voice quality influences the degree to which humans adapt to or even obey it. This supports the conclusion of other studies that it is very much a question of communication attitudes toward machines that guides actual behavior in human–machine interaction (Etzrodt et al. 2022, 439). This extends beyond linguistic convergence: Etzrodt and Engässer (2021, 73) observed how participants modify and hybridize their ontological differentiation between object and subject to facilitate interaction with voice assistant systems.

If social actors such as VAs (Nass et al. 1994, 72) are to engage meaningfully in a social way to enable and support autonomous agency and decision-making, and if successful communication with them is a precondition for achieving just that, then convergent and divergent acts of accommodation should be regarded as an important factor in human–machine relations. In order to examine our approach in practice, this chapter attempts to apply the principles of CAT to interactions between humans with special needs and their VAs. Our core interest is to explore how participants themselves perceive, describe, and critically assess their own convergence towards the communicative styles and capacities of their speech assistants. Our approach is conceptionalized as the "CAT-Technology Equivalence Hypothesis": we assume that users apply similar social expectations and behaviors to technology as they do to humans. If this is the case, we can assume that individuals need to apply certain anthropomorphization strategies to the technical object. Epley et al. (2007, 866) identify three psychological triggers for anthropomorphic thinking:

a) *elicited agent knowledge*: the accessibility and applicability of anthropocentric knowledge
b) *effectance motivation*: the motivation to explain and understand the behavior of other agents
c) *sociality motivation*: the desire for social contact and affiliation

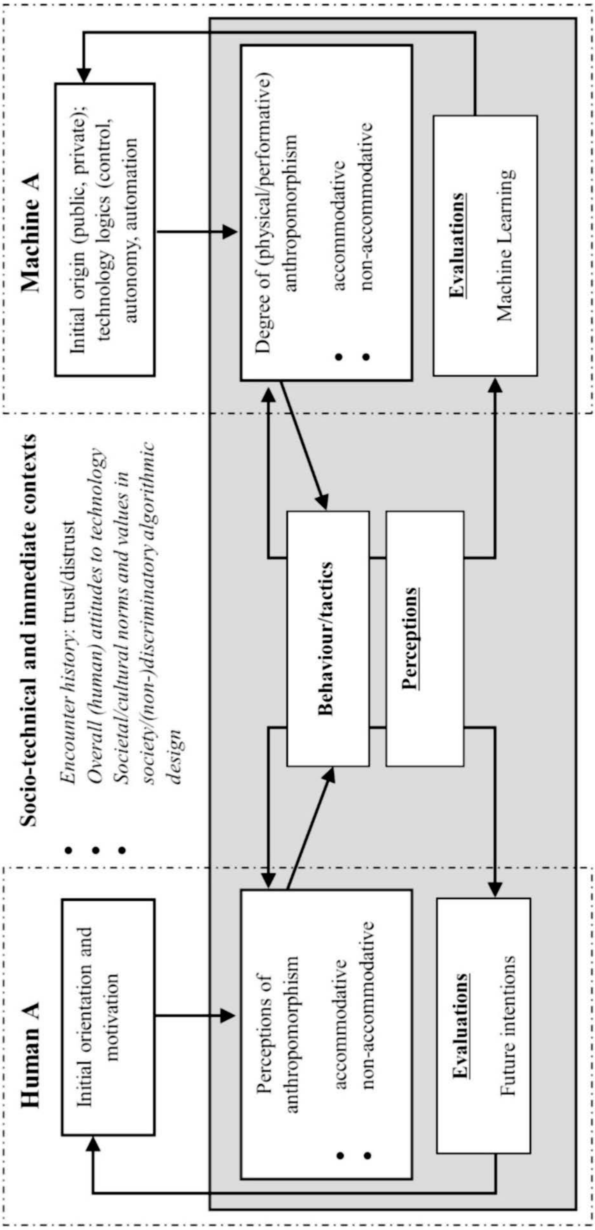
They claim that "people are more likely to anthropomorphize when anthropocentric knowledge is accessible and applicable, when motivated to be effective social agents, and when lacking a sense of social connection to other humans" (Epley et al. 2007, 864).

In our study, we employed an adapted model of CAT, based on the basic premises of accommodation and non-accommodation. As the interaction partner in HMC is technology, it is essential to reflect upon the qualities and limitations of the logics of the technology. As explained elsewhere (Thimm 2018, 116), the concept of *technology (or media) logic* refers to the affordances and limitations of a specific technology on various levels. To investigate accommodation practices with technological artifacts, it is necessary to recognize the distributed agency of humans and nonhumans that is at play in sociotechnical situations. Rather than thinking of the affordances of technology as a one-way relationship whereby either the technology affords something to users, or users afford things to technology, the important role played by algorithms renders notions of unidirectionality obsolete. Interactions with AI-driven chatbots, such as ChatGPT, present a dynamic landscape that defies simple linear explanations. These interactions are influenced by a variety of factors, including the sophistication of the AI, user expectations and experiences, contextual nuances, and cultural influences. As AI technology continues to advance, the intricacies of these interactions evolve, making it ever more inappropriate to try to reduce them to a linear framework. Successfully navigating this terrain requires a comprehensive understanding of the multifaceted elements at play. Hence, we propose the "CAT Technology Equivalence Model":

The accommodation-related activities clearly exhibit greater complexity on the part of the human interlocutor, at least at present. Humans not only possess *culturally and norm-based values and expectations towards technology*, they also harbor personal histories, experiences, and needs concerning the relevant machines within *sociotechnical and immediate contexts*. Moreover, immediate contextual factors, such as special needs on the part of humans as in our sample, influence human behavior and strategies. *Encounter history* denotes the trend toward personalized technologies tailored to the specific needs and preferences of human users. Occasionally, users implement adaptations of the original technology in order to facilitate communication. An illustrative example of such an adaptative measure was reported in our case study, in which a person's specific handicap rendered verbal interaction with the VA impossible, necessitating the use of an amplifying device to enable functionality. As our model is primarily rooted in CAT principles, it is inclined to attribute less agency to the machine. As machines are developed to incorporate ever more human-like characteristics, with social robots gaining enhanced competencies and finding broader application contexts, constraints on the side of machines may diminish over time.

For our own study, however, the current sociotechnical restrictions of Alexa reflect the state of the art of the VAs in use at the time of our study in 2023.

Figure 2: CAT Technology Equivalence Model



4. Communication with Machines in Contexts of Dependency

4.1. The Study: VAs in Households with Individuals with Special Needs

In order to investigate the hypotheses proposed above, we conducted a qualitative case study with four households, which each had at least one smart speaker and a person with care needs due to physical disabilities. The participants at the core of the study, referred to here as test persons, all had a diagnosed disability that impaired their mobility and physical action. Two further interlocutors had no care needs but lived together with two of the test persons in a supportive role. Care, support, and assistance were provided by these carers, relatives, or assisted living facilities. All households owned at least one Amazon Alexa VA.

Since the use of VAs in closed environments such as private households is strongly influenced by subjective impressions, adaptations, and adjustments, we employed a qualitative-ethnographic design for our study. The aim was to record exemplary individual attitudes, impressions, and interaction patterns, and in this way to explore sociotechnical practices and practices of accommodation in daily usage patterns.

Methodologically, the study combined two qualitative, semi-standardized procedures:

a) Individual interviews with all participants including the two caretakers (n = six interviewees)
b) Media journals, filled out by the participants with care needs themselves or by their assistants

In the first semi-structured interview, participants were asked about their attitudes towards the VA itself and about their general usage habits. They then kept a structured media diary for one week to document their usage patterns. Through this process, participants noted their individual media consumption in daily life, which showed implicit routines and interaction dynamics that they might not have consciously thought about before. To ensure thorough documentation and data integrity, no specific time intervals were set for when to note activities in the diary; interaction with the VA served as the sole criterion for when to do so. In a final interview conducted after the survey period, respondents revisited discussions on their usage behaviors, perceptions, privacy concerns, and future outlooks. The interviews were transcribed, structured, summarized, and subjected to qualitative descriptive content analysis to ensure a comprehensive examination.

The six participants' ages ranged from 23 to 58 years. Two were female and four were male. The households are referred to as households A to D for the sake of anonymity. Household A has three VAs, households B, C, and D have one device each. VAs had been purchased following recommendations from family or friends (A, C), based on personal research (B), or in order to address a specific problem in domestic living (D). The technical installation was carried out independently (A), in cooperation with involved friends (C), spouse (B), or the family environment (D); with the participation of the test persons in all cases. All households had owned their VA for several years at the time of the study (A=3, B=8, C=1.5, D=5).

4.2. Types and Frequency of Interactions

The duration of the study was one week. During this period, the participants kept a media diary and categorized their interactions with the VAs according to a set of criteria such as time of day, duration of interaction, communication objectives, (dis)satisfaction, or verbalization strategies.

In total, 759 interactions with the VA were logged, with the highest interaction rates noted in the morning and late afternoon to evening. The diaries showed that VAs were integrated into daily routines as an inherent part of everyday life. 332 interactions were classified as entertainment, 211 as planning and organizing tasks, 156 interactions operated smart home devices, and 34 were requests for information (see Figure 3). Respondents reported that they would not have been able to perform 667 of the total 759 actions without the VA's help. They deemed the remaining 92 actions would have been possible without technical assistance. However, it is worth noting that in some such cases, like the example of respondent A, the activity would have otherwise been performed by a caregiver.

These results affirm the remarkable relationship between users with care needs and their VAs: users rely heavily on the smooth functionality and effectiveness of their VAs to facilitate the organization and structuring of their daily lives. This creates a communicative situation in which adapting to the machine is crucial: any lack of 'understanding' between user and VA, whether due to machine malfunctions or disability-related communication barriers, would be more than just an inconvenience and could even, as remarked by a participant, significantly reduce participants' quality of life. Figure 4 emphasizes visually the importance attributed to successful interactions by participants with care needs.

Figure 3: (left): Alexa: Categories of use

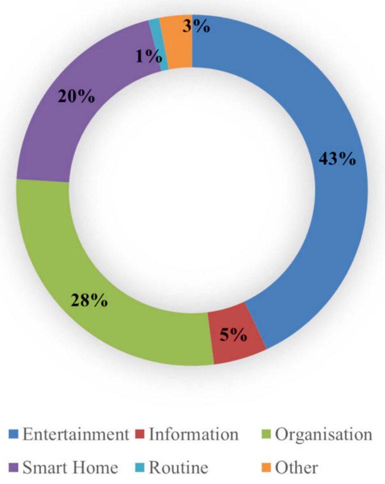
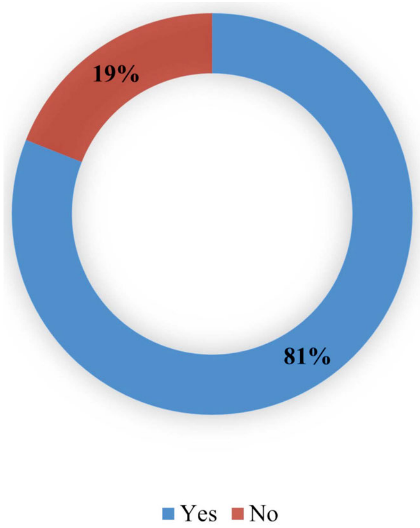


Figure 4: (right): Activity only possible with Alexa

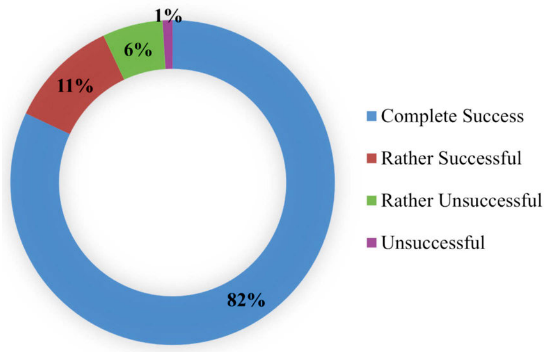


4.3. Verbal Communication

Verbal communication with the VAs was rated as completely successful in 621 of the 759 interactions. In 79 instances, communication yielded reasonably favorable outcomes, while in 48 instances, it proved less than satisfactory. Moreover, in 11 instances, communication endeavors were so unsatisfactory that they were discontinued without the VA having performed the desired task. All 135 interactions involving person D necessitated the involvement of a technical intermediary: the OSC Talker.

Users with disabilities often encounter challenges with speech recognition when interacting with voice assistants. Questions and commands spoken softly or in areas with poor internet connectivity are frequently not processed or answered accurately. Moreover, unclear pronunciation, regional dialects, background noise, or speech impediments related to disabilities further complicate interaction. In situations in which they experience frustration with their VAs, users are less inclined to view the VA as a partner. Instead, they perceive the VA as a mere machine or service provider, and adjust their behavior accordingly, often accompanied by negative emotions.

Figure 5: Success of Communication



A special case is D. He was initially not understood by his VA due to his unclear manner of pronunciation. To facilitate communication, D made use of an additional technical assistance tool known as the OSC Talker (OnScreen Communicator). The OSC Talker serves to enhance communication capabilities for individuals with disabilities, offering operability through eye movement, button input, or touch interaction. D utilizes the OSC Talker via his computer, leveraging its features, which include email functions, an on-screen keyboard, and various communication interfaces. Of particular value for D, the OSC Talker offers voice output, enabling him to utilize its synthesized speech to engage with Alexa. Tailored communication interfaces have been configured specifically for D, facilitating interaction with the VA and facilitating the activation and management of smart home devices and other functions. Furthermore, D utilizes a joystick on his wheelchair to regulate the power socket of his computer, enabling him to switch it on or off.

4.4. Accommodation to 'Technical Alexa' or 'Anthropomorphic Alexa'

As outlined in section 1 above, we regard anthropomorphization as a bridging concept that can help to explain some of the specifics of human–machine relations. Hence, we used the concept in our categorization of VA usage practices, identifying patterns that suggest when the VA was perceived as a machine/technological device and when it was approached more as a person/anthropomorphized entity.

In the course of the interviews, we found interesting reflections by our participants which corresponded to elements of our 'CAT Technology Equivalence Model'. In general, the participants exhibited very diverse relationships with their VAs – not just between participants but also for one and the same person in different situations. Participants' characterizations of the VA ranged from the purely technical – "merely a machine" – to the intimately personal – "a trained family member". Each participant demonstrated intrapersonal fluidity in their attributions, sometimes viewing their VAs as solely technical implements, while at other times regarding them with near affection as social companions and aides. This ambivalence is exemplified in the following dialogue between participant B and the interviewer (I):

B: So, you can also whisper to her, and she whispers back.
I: Really?
B: Yes! She can even get offended… didn't you know that?
I: No, I didn't know that.
B: [to Alexa] Alexa, you're a stupid cow. [Alexa doesn't understand B]
B: [to Alexa] Alexa, you're dumb.
Alexa: I don't know everything, but I'm always getting better.
B: But sometimes it also says: That wasn't very nice of you.
I: Okay, so she can also get offended.
B: But she can also be nice. When you thank her, then [speaks to Alexa]: Alexa, that was very kind of you.
Alexa: It was my pleasure. I wish you a lovely Monday.
I: Oh okay, so very polite.
B: Exactly, she also always mentions the day of the week.

Explicit acknowledgment of this ambivalence between human-like performances and the inherent technological nature of the instrument is evident in several other comments, such as:

A: She sometimes acts like a human, but it's just a robot.

Further comments corroborated this inclination of participants to engage with their VAs in a parasocial manner. Person C, for instance, noted that she had begun to use anthropomorphic sign-offs when concluding interactions at the end of the day:

C: Lately, I've said more often: 'Good night, Alexa', and then she says, 'Likewise, thank you, and have nice dreams.

Others mentioned conversing with their VAs simply for entertainment, i.e., using the machine as a substitute for human companionship:

C: Well… sometimes I chat with Alexa just for fun. When I feel like it, when I want to have a chat, I get a slightly metallic voice, but it's okay.
I: Okay. But for you, she's already a bit… well, someone to talk to… to chat with.
C: Yeah, exactly. Like a trained family member, you could almost say.
I: Okay. A trained family member… so almost… would you say not just any technical device, but already approaching becoming a real family member.

In many instances, we observed flexible interchangeability between the two kinds of personae attributed to the VA. A single participant did not consistently address the technological persona, nor an anthropomorphized one; rather, there often appeared to be a fluid switching between the two. Some authors argue that more stable routines of communication and status ascription need to be developed over time (Krummheuer 2010, 105).

When reflecting upon verbal accommodation, participants raised numerous concerns; above all, difficulties in mutual comprehension. Users frequently encountered the need to rephrase commands multiple times in order to achieve a successful interaction. For instance, Person C consistently experienced difficulties when inquiring about the weather report for his location. Likewise, Person B reported similar issues with Alexa. B suspected that these problems might be due to her unreliable internet connection, or that she didn't always speak loud enough for the VA to pick it up correctly. Person A reflected upon the need to accommodate when engaging with the VA in order to achieve successful results:

A: Maybe not differently, but more consciously. And what I also find interesting is that she made more mistakes than I was aware of. So, I feel like I had to repeat things more often without realizing it…

Typical technology based behavior when interacting with VAs mentioned by our participants centered on voice and pronunciation accommodation; comments pointed to accommodations of pitch, speaking volume, repetition, and dialect:

C: Yes, or only after pointing it out clearly three or four times about [location]… then she understands it.

B: No, you have to speak more clearly, otherwise she won't understand. So mumbling or speaking in a strong regional accent, like Colognian, she doesn't understand that at all!

Clearly, the participants learned to converge their verbalization styles towards the capabilities – or rather, incapabilities – of their technological interaction partners. Moreover, the context and purpose of interactions were reflected upon explicitly:

A: No, I mean, I do give her commands. I would never talk to people like that.
I: Okay. So, you order Alexa around, too. Would you say that?
A: Yeah, well, to me, she's not human. And then I don't see the point in having to talk to her like that.

It is worth remembering that for individuals with special needs, the relationship with technology, which serves to support, enhance, and in some instances, facilitate personal autonomy, is distinctly different from that experienced by non-disabled individuals, as explained by B:

B: I don't use them for fun like many others do, but because I need them.

The participants made it clear that without their VAs they would need significantly more help from other people, and they all asserted that their VAs played a very important role in organizing their daily lives. One interlocutor went as far as to say he "*could not imagine everyday life without Alexa*"; another even described his VA as a "*trained family member*". However, when discussing the usefulness of VAs, despite expressing their appreciation for the reduced need for human assistance that the devices facilitate, all participants insisted that they would never want to become dependent on VAs. Indeed, all the participants emphasized that their VA was not a substitute for the social contact they have with their human caretakers. Nevertheless, A, C, and D did assert that voice assistance systems make a significant contribution to equality within society and to improving accessibility.

5. Conclusion and Outlook

The experiences that participants reported in their everyday use of Alexa show a variety of convergence activities undertaken to adapt their communication to the requirements of the VA's system. Describing their own social practices in interaction with Alexa, the participants portrayed their VAs as a helpful friend, indispensable organizational helper, means of contact with the outside world, and as a safety net. In interactions with participants, Alexa emerged as a companion of shared agency, effectively blurring any distinction between external/instrumental and internal/integral use of technological objects. The study indicates that for some users with disabilities, systems involving AI such as VAs can enhance their personal autonomy and help them to maintain a level of control over their daily activities.

The results of the interviews and the one-week media diaries also highlight a degree of ambiguity characterizing the relationships between users with disabilities and their VAs. On the one hand, participants stressed that Alexa had become an irreplaceable part of their everyday lives, that they could not and would not want to live without her support, and that the voice assistant increased their sense of freedom and independence. Hence, lack of functionality or loss of Alexa was perceived as an enormous limitation. The interviewees, all of whom had been interacting with Alexa for several years, described a high level of familiarity with Alexa and emphasized that she was an integral part of everyday domestic life. The comparison of our methods (guided interviews and media diaries) showed how, in practice, VAs are so deeply embedded into routines that their involvement in actions is often not consciously reflected upon – except when something goes wrong. Additionally, non-communicative adaptations, such as the purchase of additional smart home devices or the acquisition of technical skills to set up and use them, illustrate the practical value of VA systems for people with disabilities.

At the same time, it became clear that users had often been obliged to take drastic measures to adapt their usage and communication behavior to the functional and operational logic of VAs. The spectrum of adaptations ranged from simple to complex accommodations of a convergent and divergent nature. All respondents described the interaction as limited and observed that when communicating with Alexa, idiosyncratic linguistic habits such as regional dialects should be avoided. The limited technical capacities of VAs' voice recognition software often necessitate multiple repetitions, which in turn influence users' attitudes towards VAs and caused frustration for some of our participants. Communication behavior was also adapted at a more general level, with participants describing how they adopted a more demanding, direct, and authoritative tone of voice than they would when interacting with a human counterpart.

Commands and requests in particular have to be articulated clearly, distinctly, and slowly – a communication hurdle that is sometimes difficult for people with physical disabilities to overcome. In the case of D, the necessary convergent accommodation was achieved by means of a technical solution. Because of his own limited speech capacity, he had to install the OSC Talker as technical intermediary that enabled him to communicate with Alexa verbally. The investment of financial resources, the installation of an additional technical device, and the corresponding double adaptation of usage behavior in order to communicate in the mode foreseen by Alexa's designers all illustrate the one-sidedness of this accommodation process: in this example, the human was obliged to adapt to the inflexible technology.

The data presented in this chapter support some of the concepts laid out in the "CAT-Technology Equivalence Model". Most notable are the diverse ways that an attitude of anthropomorphism in dealing with devices manifests itself as part of implicit performative accommodation of communication behavior. Not only does this affirm our contention that anthropomorphization is an unconscious tendency, it also emphasizes the influence of users' emotions upon the status they ascribe to nonhuman communication partners. The attribution of human-like qualities is not only significant in interaction with technological agents such as VAs but also characterizes the reciprocal performances of communication patterns with many other machines (Malinowska 2021). In order to increase the efficacy of communication, both technological and human practices draw upon established patterns and customs of interpersonal communication, but it is the reiterative bilateral exchange of (para-)social cues that evokes anthropomorphic perceptions and, at the same time, ambivalent feelings about the status of smart devices like VAs. In response to Giles et al. (2023, 11) we conclude that users' communicative strategies when interacting with Alexa are initially primarily adaptive towards the technological logics of operation in order to facilitate functionality, but with repetition and long-term exposure they increasingly encompass anthropomorphic experiences and attributions, which in turn shape future interaction practices. Conceptualized through CAT, anthropomorphism bridges the gap between technological usability and status-relevant attributions.

However, the potential of CAT to theoretically map this relational performativity has so far rarely been explored within a communication science framework in studies of human–machine interaction. When Fortunati and Edwards (2022, 8) defined HMC as a form of communication between humans and digital interlocuters, or machines, they proposed that these machines act as human surrogates, simulating humans' "biological and psychological abilities to formulate, issue, and receive a message and on the basis of this message, to elaborate another message." Recognizing technology as a social partner, our proposed "CAT-Technology Equivalence Model" replaces the second human interlocutor with a communicative machine. Especially in contexts of social robotics, disability, or elderly care, this reintroduction of CAT can build upon prior research that has identified a positive correlation between the human willingness to socialize and the projection of human-like qualities onto robots (Christoforakos et al. 2022, 1059). Focusing on such contexts also has the potential to increase the visibility of marginalized user groups when it comes to developing, integrating, and adjusting AI technologies to individual needs and to furthering our understanding of these groups as early adapters to the functional spectrum of future innovations (Bigham and Carrington 2018, 1). Finally, as technology adopts ever more human-like qualities, including physical form, human voice quality, and ever more human-like verbal fluency, such as in LLMs (e.g., ChatGPT), the question of 'human-likeness' in shaping speech behaviors, and specifically accommodative behaviors, will become an even more significant area of study.

"Is it really accommodation after all, when we tailor our speech and language to what a virtual agent can understand? Or is it simply a matter of verbally learning what buttons to press?" asked Giles et al. (2023, 10), the founder of communication accommodation theory. Perhaps this question will soon be answered by our intelligent partners.

Acknowledgements

Parts of the translation of this chapter were completed with the support of ChatGPT.

References

Ahmad, Muneeb Imtiaz, Omar Mubin, and Joanne Orlando. 2017. "A Systematic Review of Adaptivity in Human-Robot Interaction." *Multimodal Technologies and Interaction* 1 (3): 1–25. https://doi.org/10.3390/mti1030014.

Albert, Saul, Magnus Hamann, and Elizabeth Stokoe. 2023. "Conversational User Interfaces in Smart Homecare Interaction: A Conversation Analytic

Case Study" In *Proceedings of 5th International Conference on Conversational User Interfaces*, edited by Minha Lee et al., 1–12. New York: ACM. https://doi.org/10.1145/3571884.3597140.

Bigham, Jeffery P., and Patrick Carrington. 2018. "Learning from the front: People with disabilities as early adaptors of AI." *Proceedings of the 2018 HCIC Human-Computer Interaction Consortium*: 1–5.

Branigan, Holly P., Martin Pickering, Jamie Pearson, and Janet McLean. 2010. "Linguistic alignment between people and computers." *J. Pragmat* 42 (9): 2355–68.

Branigan, Holly P., Martin Pickering, Jamie Pearson, Janet McLean, and Ash Brown. 2011. "The role of beliefs in lexical alignment: evidence from dialogs with humans and computers." *Cognition*, 121 (1): 41–57.

Cai, Zhenguang G., David Haslett, Yuefeng Duan, Shuqi Wang, and Martin Pickering. 2023. "Does ChatGPT resemble humans in language use?" *arXiv*, 1–33. https://doi.org/10.48550/arXiv.2303.08014.

Caporael, Linnda. 1986. "Anthropomorphism and mechanomorphism: Two faces of the human machine." *Computers in Human Behaviour*, 2 (3): 215–34. https://doi.org/10.1016/0747-5632(86)90004-x.

Christoforakos, Lara, and Sarah Diefenbach. 2023. "Technology as a Social Companion? An Exploration of Individual and Product-Related Factors of Anthropomorphism." *Social Science Computer Review*, 41 (3): 1039–62. https://doi.org/10.1177/08944393211065867.

Coeckelbergh, Mark. 2023. "How to do robots with words: a performative view of the moral status of humans and nonhumans." *Ethics Inf Technol*, 25 (44): 1–9. https://doi.org/10.1007/s10676-023-09719-5.

Cohn, Michelle, Brubno Ferenc Segedin, and Georgia Zellou. 2019. "Imitating Siri: socially-mediated vocal alignment to device and human voices [Conference Paper]." *Proceedings of the 19th Int.Congress of Phonetic Science, Melbourne*: 1813–17. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_1862.pdf.

Cohn, Michelle, and Georgia Zellou. 2021. "Prosodic differences in human- and Alexa-directed speech, but similar local intelligibility adjustments." *Frontiers in Communication* 6: 1–13. https://doi.org/10.3389/fcomm.2021.675704.

Cohn, Michelle, Ashley Keaton, Jonas Beskow, and Georgia Zellou. 2023. "Vocal accommodation to technology: the role of physical form." *Language Sciences* 99: 1–17. https://doi.org/10.1016/j.langsci.2023.101567.

Damiano, Luisa, and Paul Dumouchel. 2018. "Anthropomorphism in Human–Robot Co-evolution." *Front. Psychol* 9 (468): 1–9. https://doi.org/10.3389/fpsyg.2018.00468.

Darling, Kate. 2016. "Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects." In *Robot Law*, edited by Ryan Calo, Michael Froomkin, and Ian Kerr, 213–34. Cheltenham and Northampton: Edward Elgar. https://doi.org/10.2139/ssrn.2044797.

Edwards, Chad, Autumn Edwards, and Varun Rijhwani. 2023. "When in doubt, lay it out: Over vs. under-accommodation in human-robot interaction." *Language Sciences* 99: 1–10. https://doi.org/10.1016/j.langsci.2023.101561.

Epley, Nicolas, Adam Waytz, and John Cacioppo. 2007. "On Seeing Human: a Three-factor theory of Anthropomorphism." *Psychological Review* 114 (4): 864–86. https://doi.org/10.1037/0033-295x.114.4.864.

Etzrodt, Kathrin, and Sven Engesser (2021). "Voice-based agents as personified things: Assimilation and accommodation as equilibration of doubt." *Human-Machine Communication* 2: 57–79. https://doi.org/10.30658/hmc.2.3.

Etzrodt, Kathrin, Peter Gentzel, Sonja Utz, and Sven Engesser. 2022. "Human-machine-communication: introduction to the special issue." *Publizistik* 67: 439–48. https://doi.org/10.1007/s11616-022-00754-8.

Ezenkwu, Chinedu, and Andrew Starkey. 2019. "Machine Autonomy: Definition, Approaches, Challenges and Research Gaps." In *Intelligent computing*, edited by Kohei Arai, Rahul Bhatia, and Supriya Kapoor, 335–58. Wiesbaden: Springer International Publishing. https://doi.org/10.1007/978-3-030-22871-2_24.

Fetterolf, Elizabeth, and Ektarina Hertog. 2023. "It's not her fault: Trust through anthropomorphism among young adult Amazon Alexa users." *Convergence* 0 (0): 1–18. https://doi.org/10.1177/13548565231200337.

Fortunati, Leopoldina, and Autumn Edwards. 2022. "Framing the psycho-social and cultural aspects of human-machine communication." *Human-Machine Communication* 4: 7–26. https://doi.org/10.30658/hmc.4.1.

Fröding, Barbro, and Martin Peterson. 2021. "Friendly AI." *Ethics Inf Technol* 23: 207–14. https://doi.org/10.1007/s10676-020-09556-w.

Gallois, Cindy, and Howard Giles. 1998. "Accommodating mutual influence in intergroup encounters." In *Mutual influence in interpersonal communication: Theory and research in cognition, affect and behavior, Vol. Progress in Communi-*

cation Sciences, edited by Mark T. Palmer, and George A. Barnett, 135–62. Stamford: Ablex Publishing Corporation.

Gallois, Cindy, Tania Ogay, and Howard Giles. 2005. "Communication Accommodation Theory: A Look Back and a Look Ahead." In *Theorizing about intercultural communication*, edited by William Gudykunst, 121–48. Thousand Oaks: Sage.

Gallois, Cindy, Jessica Gasiorek, Howard Giles, and Jordan Soliz. 2016. "Communication Accommodation Theory: Integrations and New Framework Developments." In *Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities across Contexts,* edited by Howard Giles, 192–210. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316226537.010.

Giles, Howard, and Tania Ogay. 2007. "Communication Accommodation Theory." In *Explaining communication: Contemporary theories and exemplars,* edited by Bryan B. Whaley, and Wendy Samter, 293–310. Lawrence Erlbaum.

Giles, Howard, America Edwards, and Joseph B. Walther. 2023. "Communication accommodation theory: Past accomplishments, current trends, and future prospects." *Language Sciences* 99: 1–16. https://doi.org/10.1016/j.langsci.2023.101571.

Guzman, Andrea L., and Seth C. Lewis. 2020. "Artificial intelligence and communication: A human– machine communication research agenda." *New Media & Society* 22 (1): 70–86. https://doi.org/10.1177/1461444819858691.

Henschel, Anna, Guy Laban, and Emiliy Cross. 2021. "What makes a robot social? A review of social robots from science fiction to a home or hospital near you." In *Current Robotics Reports* 2: 9–19. https://doi.org/10.1007/s43154-020-00035-0.

Kok, Bing Cai, and Harold Soh. 2002. "Trust in Robots: Challenges and Opportunities." *Curr Robot Rep* 1: 297–309. https://doi.org/10.1007/s43154-020-00029-y.

Kramer, Marcel, Ramin Yaghoubzadeh, Stefan Kopp, and Karola Pitsch. 2013. "A conversational virtual human as autonomous assistant for elderly and cognitively impaired users? Social acceptability and design considerations." *Informatik:* 1105–19.

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

Larsson, Stefan, and Fredrik Heintz. 2020. "Transparency in artificial intelligence." *Internet Policy Review* 9: 1–16. https://doi.org/10.14763/2020.2.1469.

Lombard, Matthew, and Kun Xu. 2021. "Social responses to media technologies in the 21st century: The media are social actors paradigm." *Human-Machine Communication* 2: 29–55. https://doi.org/10.30658/hmc.2.2.

Malinowska, Joanna K. 2021. "What Does It Mean to Empathise with a Robot?" *Minds & Machines* 31: 361–76. https://doi.org/10.1007/s11023-021-09558-7.

Mahdi, Hamza, Sami Akgun, Shahad Saleh, and Kerstin Dautenhahn. 2022. "A survey on the design and evolution of social robots – Past, present and future." *Robotics and Autonomous Systems* 104193: 1–48. https://doi.org/10.1016/j.robot.2022.104193.

McStay, Andrew. 2018. *Emotional AI: The Rise of Empathic Media*. London: Sage.

Nass, Clifford, Jonathan Steuer, and Ellen Tauber. 1994. "Computers are social actors." *Proceedings of CHI Conference on Human Factors in Computing Systems*: 72–78. https://doi.org/10.1145/191666.191703.

Olasik, Marta. 2023. "'Good morning, ChatGPT, Can We Become Friends?' An Interdisciplinary Scholar's Experience of 'Getting Acquainted' with the OpenAI's Chat GPT: An Auto Ethnographical Report." *European Research Studies Journal*, XXVI (2): 269–84.

Purington, Amana, Jessie Taft, Shruti Sannon, Natalya Bazarova, and Samuel Hardman Taylor. 2017. "'Alexa is my new BFF': Social Roles, User Satisfaction, and Personification of the Amazon Echo." *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*: 2853–59. https://doi.org/10.1145/3027063.3053246.

Schreuter, Donna, Peter van der Putten, and Maarten H. Lamers. 2021. "Trust me on this one: conforming to conversational assistants." *Minds and Machines* 31: 535–62.

Soliz, Jordan, and Howard Giles. 2014. "Relational and Identity Processes in Communication: A Contextual and Meta-Analytical Review of Communication Accommodation Theory." *Annals of the International Communication Association*, 38 (1): 107–44. https://doi.org/10.1080/23808985.2014.11679160.

van Pinxteren, Michelle M.E., Mark Pluymakers, Jos Lemmink, and Anna Krispin. 2023. "Effects of communication style on relational outcomes in interactions between customers and embodied conversational agents." *Psychol. Market* 40 (5): 537–59. https://doi.org/10.1002/mar.21792.

Thimm, Caja. 2018. "Media Technology and Media Logic(s): The Media Grammar Approach." In *Media Logic(s) Revisited. Modelling the Interplay between Media Institutions, Media Technology and Societal Change*, edited by Caja Thimm, Mario Anastasiadis, and Jessica Einspänner-Pflock, 111–32. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-65756-1_6.

Thimm, Caja. 2018. 2019. "Die Maschine – Materialität, Metapher, Mythos. Ethische Perspektiven auf das Verhältnis zwischen Mensch und Maschine." In *Maschine: Freund oder Feind? Mensch und Technologie im digitalen Zeitalter*, edited by Caja Thimm, and Thomas Bächle, 1–40. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-22954-2_2.

Thimm, Caja, and Laura Thimm-Braun. 2024. "Policies, regulation and legal perspectives on social robots." In *DeGruyter Handbook of Robots in Society and Culture*, edited by Autumn Edwards, and Leopoldina Fortunati, 161–76. Berlin: De Gruyter.

Vlahos, James. 2019. *Talk to Me: How Voice Computing Will Transform the Way We Live, Work, and Think*. Houghton Mifflin Harcourt.

Wagner, Katja, and Hanna Schramm-Klein. 2019. "Alexa, Are You Human? Investigating Anthropomorphism of Digital Voice Assistants-A Qualitative Approach." *ICIS Fortieth International Conference on Information Systems:* 1–17.

Złotowski, Jakub, Diana Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. "Anthropomorphism: opportunities and challenges in human-robot interaction." *Int. J. Soc. Robot* 7: 347–60. https://doi.org/10.1007/s12369-014-0267-6.

Human-Machine Interaction as a Complex Socio-Linguistic Practice

Netaya Lotze

Abstract *This paper presents a socio-linguistic model for Human-Machine Interaction (HMI), examining the interplay of technological affordances, user cognitive awareness, and language strategies. The model features three continua: technological affordances, users' cognitive awareness, and language strategies. The first dimension evaluates the anthropomorphism degree of the system, including linguistic anthropomorphism and therefore tries to integrate Ruijten's et al. (2014/2019) Rasch-scale of human perception of anthropomorphic designs. The second dimension explores users' cognitive awareness, ranging from pre-conscious alignment to conscious strategies. The third dimension depicts a continuum of user language, from pre-conscious alignment (Gandolfi et al. 2023) and linguistic routines and behaviors, transferred from HHC (CASA: Reeves and Nass 1996; MASA: Lombard and Xu 2021) to various simplification strategies as robot-directed speech (RDS), simplified registers (SR) (Fischer 2011) and computer talk (CT) (Zoeppritz 1985). The paper argues from a diachronic perspective that HMI language evolution is influenced not only by anthropomorphic technology and user awareness but also by language variation and change, and societal factors. Therefore, the results of numerous studies of my own research group conducted between 2000 and the present (with a particular focus on Lotze 2016) will be summarized and interpreted in light of the model, and vice versa.*

1. Introduction: The AAS-Model of HMI as a Complex Socio-Linguistic Practice

Users seem to interact with AI either as they would with a human conversation partner or in a simplified form specifically designed for operating a machine. The fact that empirical studies on HMI (Human-Machine Interaction) continue to yield contradictory results regarding alignment, politeness markers, and routines similar to those found in HHC (Human-Human Communication) on the one hand, and simplifications, imperatives, and isolated keywords as bot-directed language on the other (Fischer 2006), has led to the emergence of two competing research paradigms. Each of these paradigms only captures and explains partial aspects of the phenomenon: a) the CASA/MASA approach ('All media are social actors'), which assumes that users always attribute social characteristics to the system (see Reeves and Nass 1996; Lombart and Xu 2021), and the 'Simplified Registers' approach, which focuses primarily on simplifications, repair and related phenomena (Fischer 2006; see also 'Computer Talk' (CT) Zoeppritz 1985). Both approaches can only analyze minor portions of the whole complexity of the empirical phenomena.

Therefore, in Chapter 3 of this article, I propose my AAS-model of HMI as a complex socio-linguistic practice that can claim broader validity. In my opinion, 'user awareness' seems to be the most relevant cognitive key concept for this purpose, which is often missed by other approaches. Accordingly, my model is structured into three dimensions: the degree of anthropomorphism of the technology, the cognitive awareness of the users, and the user language. This creates a decision space in which users can position themselves on three continua, always keeping the current interaction situation with the AI and its cultural and pragmatic implications in mind.

The model also aims to capture the dynamic nature of HMI language development, emphasizing metaphorical language use after Krause and Hitzenberger (1992) as it undergoes diachronic transformations with technological shifts – always with a little delay (c.f. Schmitz 2015 "stilistisches Trägheitsgesetz" ("stylistic inertia")). External influences on user language, such as dialog design, data foundations, and linguistic models, are supplemented by socio-linguistic factors like language change, cultural dynamics, and societal shifts. The model provides at this point in time a first glimpse into the intricate web of language evolution in HMI, offering a balanced perspective on both technological and socio-linguistic dimensions in human-machine interaction.

The most significant contribution of the model lies in its ability to a) not only integrate but also partially explain traditionally contradictory tendencies in user behavior (e.g. anthropomorphization vs. simplification), and b) maintain connections to the historical academic discourses on CASA/MASA and "computer talk" (CT) as a "simplified register" (SR), while going far beyond those approaches by conceptualizing HMI as inherently dynamic, negotiable, and subject to socio-cultural change. Consequently, the model does not rely on static categories but instead opens a decision space composed of continua in which the user can position themselves. The approach breaks with structuralist or positivist theories, yet it does not seek to entirely discard those aspects that were useful for our better understanding of HMI. Rather, it aims to demonstrate that HMI is more complex and diverse than traditional theories have recognized.

In the following two chapters, the research landscape of traditional approaches in HMI research (CASA/MASA vs. SR) will be presented, followed by an overview of the empirical studies conducted by my research group over the past twenty years. The focus will be on aspects that extend beyond CASA/MASA and SR to illustrate why my new model is necessary.

2. The CASA/MASA Approach as One of the Earliest Reference Points for Interpreting Linguistic User Behavior

Despite from a philosophical and sociological perspective contemporary artificial intelligences, even in dialogue systems employing GPTs, still lacking the characteristics of a social actor, as they neither possess self-reflective consciousness, emotions, empathy, nor exercise a free will capable of autonomously setting and being held accountable for their own goals, linguistic analyses of user language since the late 1990s suggest that individuals exhibit a tendency to transfer linguistic concepts from human-human communication (HHC) onto human-machine interaction (HMI) (cf. Reeves and Nass 1996; Nass and Moon 2000; Nass and Brave 2005; Reeves et al. 2020). Thus, cognitively, they engage in a certain form of anthropomorphization of systems on a conceptual level, triggered by the natural language dialogue serving as depiction of a human interlocutor (Clark and Fischer 2023). Accordingly, the anthropomorphic design of the system on a technological level, triggers an anthropomorphization on the level of user cognition and a conceptualization as a social actor. Reeves and Nass (1996) analyze this user behavior in their early studies as “mindless behavior”, interpreting it as a preconscious transfer of concepts, schemas, and action routines from HHC (Reeves and Nass 1996; Nass and Moon 2000; Nass and Brave 2005; Reeves et al. 2020). Reeves and Nass (1996) argue that individuals apply a social model when confronted with a complex entity whose mechanisms they do not immediately comprehend. Linguistically, this phenomenon manifests, for instance, in the transfer of certain ritualized protocols (Sacks et al. 1992), frames and scripts (Fillmore 1976), or levels of politeness (Brown and Levinson 1987) from HHC to HCI.

They consolidate this research stance into their "Computers are Social Actors (CASA) Paradigm", elucidated and systematized in "The Media Equation" Reeves and Nass 1996). Subsequently, they and other research groups discovered numerous cross-cultural pieces of evidence for social effects of dialogue systems that can be interpreted through the CASA framework (Nass and Moon 2000; Nass and Brave 2005; Reeves et al. 2020). However, the precise role of preconscious attribution of social attributes to the system in the context of HMI remains a subject of ongoing controversial discourse (see Lotze 2016 for a deeper exploration; Dippold 2023).

In more recent times, the CASA approach has been expanded and refined into MASA ("Media are Social Actors", Lombard and Xu 2021), which can be applied to various contemporary media and it incorporates the degree of anthropomorphism associated with these media. Ruijten et al. (2014) proposed a Rasch-like anthropomorphism scale for their psychology of AI perception, systemizing objects in general (and specifically robots, agents, and assistance systems), with varying effects on user reception. Since 2019, they have tested and confirmed this idea in different scenarios with diverse participants, yielding replicable results. Ruijten's et al. approach could show, that the technological level of anthropomorphic design and the user's perception of it as more or less social are closely intertwined in a systematic way. And only because of that, we can combine the otherwise separate levels in the Rasch-like anthropomorphism scale, they suggest. Lombard and Xu (2021) adopt this scale of degrees of anthropomorphism from psychology and integrate it into CASA. Unlike Nass and Reeves' early approach, MASA considers the degree of anthropomorphism in AI design, which significantly influences individuals, especially when the representation is more humanoid.

In my opinion, CASA is an exceptionally fruitful idea and model that can explain a significant number of user utterances across various contexts. Nevertheless, having scrutinized hundreds of real interactions from anonymous users with customer support bots in the field, I contend that, on the flip side, there still exists a stable corpus of user expressions over different applications and decades that unfortunately eludes explanation through the CASA/MASA paradigm. Consequently, I have come to the conclusion that CASA (and to a lesser extent, MASA) constitutes a position within the research community that adeptly captures only one singular driving force behind user behavior towards AIs – specifically, the transfer of behaviors from Human-Human

Communication (HHC), with all its implications for dialogue (as mentioned earlier: routines and protocols, frames and scripts, linguistic politeness). However, it exhibits a blind spot for all aspects of user behavior that deviate from HHC and are currently evolving: simplifications in the form of audience-specific "Simplified Registers" (Fischer 2006; 2011), such as syntactically simplified commands or questions observed in RequestandResponse systems like Amazon Alexa (Greilich, in preparation, see below), or isolated keywords, popular among Digital Natives, for instance in Social Bots used in customer service (Lotze and Ohrndorf, in preparation, see below), leading quickly to the dialogue goal, especially in written media.

In my opinion, it is particularly crucial to emphasize that aspects of linguistic economy (including Ronneberger-Sibold 1980; Köhler 2005) play a crucial role here, as they have become relevant in the context of digitization in real-time written communication among people (regarding IR chats and SMS: Siever 2011; concerning messenger apps: König 2019). Some of the simplifications observed in the field can be explained depending on the technological affordances of the respective language system, while others appear to represent emerging socio-linguistic practices for interacting with AI, evolving as variants currently just in the process of formation.

With this article, my intention is to strongly advocate for the notion that Human-Machine Interaction (HMI), at least at present (future systems might become even more human-like), constitutes a heterogeneous form of interaction. It incorporates preconscious to ritualized transfers from Human-Human Communication (HHC) but also exhibits numerous new stylistic parameters addressing the utilitarian nature of the application. These include simplifications, the absence of politeness, increased use of vulgarisms, and considerations of its representational character (cf. Lotze 2016), or its performative nature (the staging of interaction with AI as a philosophical game, discussing the AI with others during an ongoing dialogue, etc.). Clark and Fischer (2023) similarly underscore that dialogue systems and robots are always "depictions" of humans, and modern users (in contrast to those of Weizenbaum's ELIZA in the 1960s) are indeed conscious of this representational character. They provide numerous example dialogues that effectively illustrate how individuals intermittently engage more or less in this role-play. Fischer's (2006) user types – a "Player" who embraces the portrayal of an anthropomorphic conversational partner and a "Non-Player" who conceptualizes the application more as a tool –, in my opinion, are valuable key concepts for systematizing the heterogeneity of user strategies.

In this article, I aim to present HMI as a multi-dimensional socio-linguistic practice, considering not only varying degrees of anthropomorphism in system design but also increasing levels of user awareness. Our studies, particularly those focused on interactive alignment, have demonstrated that users exhibit more or less preconscious or routinized behavior in different dialogue sequences and phases. Importantly, users maintain such behavior only as long as the sequence proceeds without disruptions (cf. Lotze 2016; Krummheuer 2010). Thus, preconscious transfers from HHC depend on the user type according to Fischer (2006), the degree of linguistic anthropomorphism in the interface (Ruijten et al. 2014; 2019), the dialogue phase, and disruptions in the dialogue (Lotze 2016). Only a model that additionally incorporates the user's levels of awareness can adequately address the heterogeneity of HMI, as opposed to models that consider individual aspects in isolation. HMI, therefore, must be conceptualized in three dimensions: a) as user language (in variation and evolution), b) as user awareness (on a continuum from preconscious to conscious/strategic), and c) in relation to the degree of anthropomorphism in system design (both visually and linguistically). This approach creates a three-dimensional decision space, wherein users position themselves with each contribution to the conversation. Simultaneously, this framework serves as a model for the linguistic AI research community to better interpret linguistic user behavior in HMI.

In the article, my model of HMI will be introduced as a complex and heterogeneous socio-linguistic practice, grounded in theory (see Chapter 3) and motivated by the results of my research group (empirical evidence, see Chapter 2). It is imperative to firmly connect our research in both empirical evidence and theory to the existing and current research landscape.

In Chapter 2, I will present relevant studies conducted by my research group on linguistic user behavior, discussing those aspects (Alignment, Acceptance, Simplification (AAS)) that have been incorporated into the model:

a) Lotze (2016): Corpus study on rule- and plan-based chatbots.
b) Lotze and Ohrndorf (in preparation): Corpus study on Socialbots in customer service.
c) Greilich (in preparation): Psycho-linguistic experiment on Amazon Alexa.
d) Lotze and Aydin (in preparation): Qualitative-explorative study on ChatGPT following an ethnomethodology.

Subsequently, in Chapter 1, the research horizon will be outlined. In Subsection 1.2, the scientific-historical foundations of the discourse on "Computer-Talk" (CT, Zoeppritz 1985) and "Simplified Registers" (Fischer 2006; 2011) will be presented to better understand the relevance of simplifications for HMI. Considering the historical background of both terms, new and more nuanced conceptualizations will be explored. Section 1.3 will then focus on the heterogeneity of HMI (following Lotze 2016) in detail. Subsets of HMI will be delineated, described, and categorized.

Chapter 2, as mentioned earlier, follows as the empirical section, with Subsection 2.2 focusing on our current studies on ChatGPT concerning simplifications, addressing and discussing relevant aspects.

In Chapter 3, the model of HMI as a complex socio-linguistic practice is presented. It will be discussed within the context of a diachronic perspective on communication in the age of digitization. This chapter aims to provide a comprehensive understanding of HMI, drawing on the theoretical foundations and empirical findings outlined in the preceding chapters.

2.1 The Academic Discourse on "Simplified Registers" as a Counterpoint to CASA/MASA?

Fischer's (2011) framework of "Simplified Registers" emerges as a crucial starting point for analyzing strategic simplifications by users. When faced with a robot or agent, individuals engage in strategic actions, consciously simplifying their language. Fischer's benchmarks for HMI include other highly simplified registers such as child-directed or animal-directed language, along with intercultural communication. In these scenarios, speakers intentionally simplify their communication, tailoring it appropriately to the respective audience. While there is a certain level of intuition involved when interacting with AI, given users' prior experiences with other 'Simplified Registers,' the process, in my opinion, primarily constitutes a strategic and conscious decision rather than a preconscious behavioral mechanism. Therefore, it is crucial to conceptually distinguish between preconscious behavior and conscious action in the ensuing discussion. These states of consciousness should not be perceived as a dichotomy but rather as poles within a continuum of degrees of awareness (see Chapter 3).

2.2 Historical foundations of the academic discourse on "Computer-Talk" (Zoeppritz 1985)

Fischer's conceptualization of "Simplified Registers" emerged within the context of the much older academic discourse on "Computer Talk", instigated by Magdalena Zoeppritz in 1985 based on initial experiments with users of early rule-based systems. Zoeppritz observed "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" (Zoeppritz 1985, 1). She explained these linguistic acts by proposing that users had a concept of the system's functioning in mind, tailoring their utterances accordingly, with a focus on the system's tool-like nature. To describe this phenomenon, she introduced the term "Computer Talk" (CT), drawing parallels to "Baby-Talk" or "Foreigner-Talk."[1]

Krause and Hitzenberger (1992) found numerous instances in their early German-language DICOS experiments with users of an early system for grade recording with a speech interface that supported Zoeppritz's (1985) concept of "Computer Talk." They observed simplifications of syntactic constructions, an increasing number of overspecifications, a growing amount of formal coding, a decreasing number of frame elements in the dialogue, a diminishing number of politeness phrases, a declining number of partner-oriented dialogue signals, and a reduced use of particles as markers for the speaker's personal disposition toward the spoken content. Krause (in Krause and Hitzenberger 1992) interpreted these as "metaphorical language use," wherein the actual metaphor lies in users tailoring their language use to the concept they have of the internal processes of the early language processing system.

Example 1: Krause and Hitzenberger (1992, 159–60)

User: Welche Deutschnote in Quarta hat wie viele Schüler?
[What is the German grade distribution in the fourth grade?]

User: Wieviele Schüler repetieren 1 Klasse?
[How many students repeat 1 grade?]

1 However, both terms are now problematic, as "Talk" inherently carries a derogatory, paternalistic connotation. In L1 and L2 acquisition research, these terms have been discarded in favor of "child-directed language" and "intercultural communication" (as mentioned above).

System: nicht verstanden
[not understood]
User: Wieviele Schüler repetieren 2 Klassen?
[How many students repeat 2 grades?]
System: 25
User: Wieviele Schüler repetieren 1 Klassen?
[How many students repeat 1 grades?]
System: 99

In this early phase of AI history, this cognitive concept is directed towards formal expressions (in programming language). However, an actual understanding of the system's architecture and programming is only partially present and varies significantly among users. Nevertheless, in Krause's early experiments, users tend to align more with a tool metaphor (AI as a tool) rather than an assistant metaphor (AI as an anthropomorphic conversational partner).

Krause and Hitzenberger (1992) characterize "Computer-Talk" based on their DICOS experiments as a structural register. Fischer, drawing on research data from the Verbmobil project, expands upon this assumption and conceptualizes "Computer Talk" more broadly as a "functional variety" (Fischer 2006) and later as a "Simplified Register" and "Robot-Directed Speech" (Fischer 2011, 261). Similar to Womser-Hacker's earlier observations on a structural level in Krause and Hitzenberger (1992) Fischer (2006) notes that HMI, in comparison to HHC, is distinguished by either an increase or decrease in lexical variety, syntactic complexity, and politeness markers. Her significant contribution lies in shifting the interpretative perspective from empirically structural features of CT to functional parameters and concepts of user cognition. "By looking at the peculiarities observable as strategies, we stop thinking of CT as a particular product and turn instead to the process in which it is created – a negotiation process" (Fischer 2006, 78). Before Fischer's 2006 analysis, early studies in HMI had a far too broad focus and a structuralist bias.

I am fully aware, that the methodological implications of Krause and Hitzenberger (1992) and Fischer (2006) are not neatly compatible with my praxeological attempt, but in order to create a model, that can include a broader range of empirical linguistic parameters I choose a more open approach.

HMI was then and remains a highly asymmetric interaction situation in which humans and machines process dialogicity quite differently, utilizing rather distinct resources. When we compare HMI and HHC, the asymmetry is immediately apparent and manifests empirically across all linguistic levels (refer to lexicon, syntax, semantics – particularly disruptions in dialogue coherence and grounding attempts – and linguistic politeness: Lotze 2016; Lotze and Ohrndorf, in preparation, concerning phonetics/phonology in Amazon Alexa: Greilich, in preparation). In all of our rather diverse studies we can observe, that users seem to transfer only the basic principles of communication from HHC (preconscious alignment, adjacency principle, frame sequences, concepts of registers, concepts of repairs, grounding, and framing), as long as the assistant metaphor is successful. When the dialogue design is geared towards it (e.g., in the case of social bots through textuality and multimodality in the form of clickable areas or in the case of Alexa through Voice User Interface (VUI) and RequestandResponse architecture) or when disruptions occur during the ongoing dialogue, users across all system types increasingly resort to simplifications or other markers of Computer-Talk (CT), such as vulgarisms, abrupt terminations of the conversation, etc.

2.2.1 User types according to Fischer (2006)

Another notable contribution by Fischer is the introduction of two user types: Players and Non-Players. What is particularly valuable about this distinction is that it involves open categories based on functional criteria. Characteristic of the Player type is treating the system as if it were a human interlocutor. The Player engages in the metaphorical game, addressing the system with personal pronouns like "du" [you without social distance] or "Sie" [you with social distance], offering greetings, and/or providing information about their own well-being when prompted by the system. On the other hand, the Non-Player type views the bot as a tool and utilizes it accordingly. They do not greet the bot, nor use personal pronouns to address it, and avoid politeness indicators. While the Non-Player demonstrates fewer transfers from HHC that can be interpreted through CASA/MASA, there are more instances of a "Simplified Register." Therefore, it is crucial to consider both approaches together.

Both types are defined by the conversational strategies they employ based on their assumptions about the AI. Consequently, their utterances become somewhat predictable. Fischer suggests that one can infer the user's category based on their behavior in the opening sequence. If the user responds to the system's greeting, they are a Player; if they ignore it, they are a Non-Player. Lotze (2016) was able to replicate this fundamental distinction between Players and Non-Players, but states that user types are more complicated and not always dichotomous.

2.2.2 The heterogeneity of HMI (Lotze 2016)

How should we define HMI then? As the attribution of social characteristics to the system, accompanied by the transfer of linguistic behavior from HHC (CASA/MASA)? Or as, in any case, partially a strategic user decision for a simplified register in the sense of bot-directed speech according to Fischer (2011)? Does the interpretation of conflicting linguistic evidence in different studies lead to a dilemma?

In my dissertation (Lotze 2016, 346–47), I argue that this perceived dilemma can be easily resolved. HMI is, after all, a genuinely heterogeneous form of interaction that varies depending on system architecture, application context, user type, and awareness level. Therefore, we need a model that accommodates the entire variability of HMI by considering all relevant parameters and not focusing solely on individual aspects. The following variables must be taken into account when interpreting HMI data, as they all have a significant impact on HMI and contribute to the variation in user language. Accordingly, the asymmetry of HMI is not a monolithic feature but manifests in very different factors and variables that are all interconnected and have an important influence on the users' language behavior and stratetgies.

Levels of asymmetry (Lotze 2016, 346):

a) External Factors

- The scenario of the application domain determines the interaction situation.

b) System Variables

- Persona
- Robot, avatar, or interface design
- Hardware
- Input channel
- Dialog design
- System architecture

c) User Variables

- Technical expertise
- User type
- Assumptions about the system
- Dialog goals
- Pre-conscious priming
- Conscious action strategies

The heterogeneity of the HMI can only be addressed by a multi-dimensional conceptualization, taking into account that the HMI is influenced by numerous factors, and these factors are highly asymmetrical on the part of both the system and the users.

Table 1: Dimensions of the dialogically inherent heterogeneity of HMI (Lotze 2016, 348)

System Architecture	**User Guidance: guided – free – hybrid**
Dialog Design	· Textuality – Orality · Social Distance – Proximity · Different Phases of Dialogue (Introduction – Middle – Farewell) · Handling of Disruptions: Incoherences, Quasi-coherences, Default Responses, or Follow-up Questions
User	· User Type · Conscious Strategic – Preconscious or Routinized · CT („Computer Talk") – HHC (Human-Human Communication)

The factors I listed in 2016 remain relevant for current systems with Natural Language Processing/Understanding (NLP/U) and Machine Learning (ML), as well as GPTs. In the HMI, individuals who are self-aware and wish to freely choose and negotiate their dialogue goals still encounter machines that still exhibit significant challenges in these aspects. As mentioned above, this asymmetry manifests across all linguistic levels. Design decisions regarding system architecture and dialogue design are, of course, impactful for interaction. However, the type of user and whether they consciously act or preconsciously react are equally consequential for dialogue. Therefore, we must consider the following levels of asymmetry in the HMI with the following effects on dialogue:

Levels of asymmetry of HMI and their effects

User – System: Humans and machines fundamentally differ in terms of "world knowledge" (Habermas 1993), emotions, (first-) language acquisition, and self-reflective consciousness. This has significant implications for dialogue semantics and coherence.

System$_A$ – System$_B$: Systems differ significantly from one another. Different technical approaches are currently used for various applications, and their functionalities should not be generalized. System architectures, dialogue designs, and the mediality of interfaces (oral, literal, embodiment) vary. This affects the chosen simplification strategies of users. The technological affordances of the system, in general, are just as relevant as its degree of anthropomorphism.

SystemA$_{error\text{-}free}$ – SystemA$_{error\text{-}prone}$: Errors are a particularly relevant factor contributing to the heterogeneity of the HMI because users must reconsider their dialogue strategy in such situations. One possible consequence is that users transfer repair strategies from the HHC to the HMI (for grounding, see Fischer 2006). Since these often fail, an extreme outcome may involve a user type switch, where a cooperative, flexible, polite player transforms into a non-player who vulgarly insults the system and abruptly ends the dialogue without a farewell. The reverse principle we currently observe in users of ChatGPT who initially attempt to operate the system with isolated keywords but then switch to more elaborate prompts when they realize that the system is capable of generating longer sequences of disruption-free dialogues (see below).

User$_A$ – User$_B$: User types (Player / Non-Player) according to Fischer (2006; 2011) have implications for the level of conscious cognitive reflection and, consequently, the chosen linguistic register (see above).

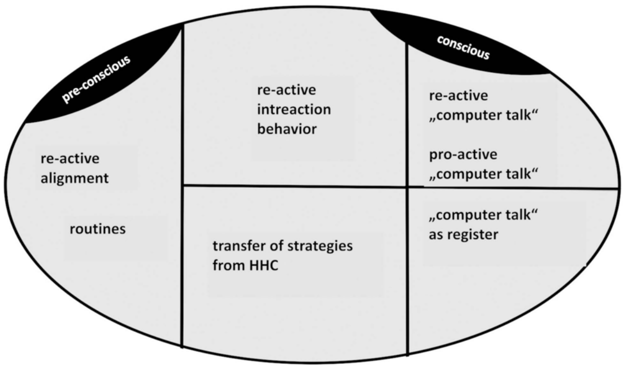
User$_{t1}$ – User$_{t2}$: Users fundamentally change their strategy when it fails (see errors and disruptions). This can occur in specific sequences without an immediate full user-type switch (extreme case). This also has implications for the level of conscious cognitive reflection, consequently affecting the chosen register, and makes the contributions of the same user in different sequences heterogeneous.

Context$_A$ – Context$_B$: Application contexts can vary extremely, impacting the attribution of social proximity or distance to the system, which linguistically reflects in politeness levels, etc.

Time$_{t1}$ – Time$_{t2}$: Mediatization manifests in diachronic variation and change affecting both systems (technological history from rule(+plan)-based systems to Big-Data approaches with NLP/U and ML, as well as GPTs) and users (Digital Non-Natives, Digital Natives, GenZ), who develop different strategies/styles/registers to interact with respective system types. Thus, in our diachronic corpora for the past 20 years, we can observe, in my opinion, how Krause's metaphor in user reception has shifted from "code" to "natural language 'enter'-key" (confirming a predefined dialog script) and "isolated keywords" (as a concept from Google search).

Within this theoretical framework that considers all relevant variables of HMI, the heterogeneity of HMI manifests empirically as user language as follows:

Figure 1: Composition of HMI (Lotze 2016, 359)



Preconscious behavior: In the preconscious realm, we find numerous instances of *re*-active alignment as a lower-level priming effect; i.e., humans adapt to the system – phonetically, syntactically, and lexically. We cannot speak of *inter*-active alignment (Pickering and Garrod 2004) here because it is not built up interactively or collaboratively. With Lotze (2016), I am still not referring to interactive alignment in the strict sense (Latin: *inter-agere*) but rather to the user's reactive alignment to the system. Also, routines transferred from HHC, such as turn construction and allocation, politeness levels, greeting sequences, etc., can be substantiated through our studies. These two aspects can only be interpreted within the CASA/MASA paradigm. However, the interaction has both preconscious and conscious components, and the better the illusion of a natural dialogue is maintained, the more "mindless behavior" (cf. Reeves and Nass 1996; Nass and Moon 2000; Nass and Brave 2005) is exhibited by the users. In contrast, during disruptions, the artificial dialogue situation must be reflected upon, and conscious strategic behavior is the logical response (cf. Fischer 2006).

Transitional behavior / strategies: Not only preconscious aspects of the HMI can be interpreted within the CASA/MASA framework, but also some of the conscious proactive action strategies involve transfers from HHC. On the functional level, for instance, all attempts by users to establish common ground or create dialog coherence (grounding, repair) can be interpreted as the users anthropomorphizing the system. Even though users, in most cases, theoretically know that systems cannot draw upon the same world knowledge as they do, they sometimes intuitively strive to promote common ground and a logically coherent dialog progression. However, this does not apply to a large portion of users. These reactive consumers of the HMI allow themselves to be guided by the system and do not attempt to address its logical-semantic deficiencies. This results in a reactive interaction that cannot be interpreted as a transfer from HHC but also does not align with CT in the narrower sense. Nevertheless, we frequently observe this passive behavior in our empirical studies, especially among digital natives of the player type who passively let the bot guide them through the application without a specific dialog goal. These two functional user attitudes can be interpreted as a transitional zone between CASA/MASA and CT. Therefore, empirical evidence suggests a continuum between preconscious behavior and strategic CT.

Conscious / strategic decisions: The scope of CT does not encompass the entire HMI, as it is heterogeneous and sometimes exhibits longer sequences of human-like dialogue, especially in contemporary applications. What can be termed as CT must be negatively defined as the subset of HMI where preconscious mechanisms (preconscious alignment, routines) or transferred strategies from the HHC (grounding, framing) do not apply. This subset can be functionally further subdivided into a) a reactive CT, directly triggered by the technological affordances of the system's architecture and dialog design (s.b.) and b) a classic, proactive CT derived from users' assumptions about the system (Krause's metaphorical language use). For the reactive form of CT we found some interesting tendencies in user behavior (triggered by the technological affordances):

- rule-based system – user behavior: isolated keywords
- plan-based systems – user behavior: passive reception attitude and "natural language 'enter'-key" ("ok", "continue", "back")
- request and response systems – user behavior: isolated imperative sentences

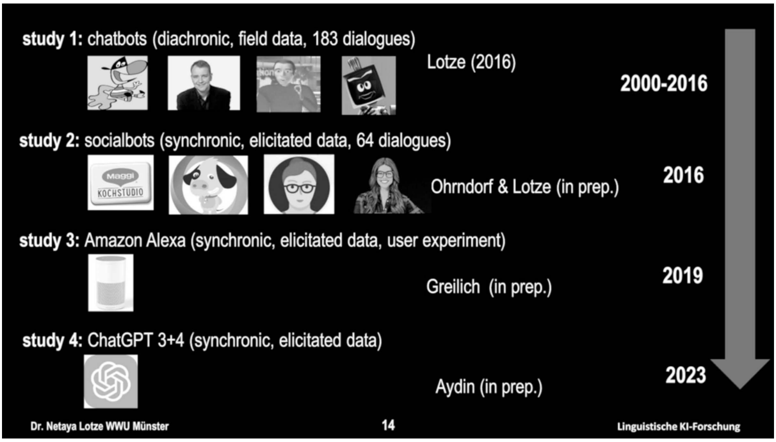
Both types of CT have functional and structural dimensions and undergo developmental processes. Function and structure do not always develop in tandem. For example, lexical and syntactic simplification today serves different functions than it did in the 1990s (programming language as a metaphor vs. keyword-based Google search as a metaphor). Overall, CT represents only an extreme case of linguistic user behavior that can be perceived as an outer pole within a continuum of user language.

3. How Do Users Linguistically Interact With AI in Our Empirical Studies? Alignment, Acceptance and Simplification (AAS)

Demonstrating the heterogeneous nature of HMI as a form of interaction, we could empirically substantiate our findings using various methods for diverse user groups with field and experimental data since the year 2000. In this article, I aim to provide an overview and, as a conclusion, present my model for HMI as a complex socio-linguistic practice. As mentioned earlier, I can only present the most relevant aspects of every study.

Study 1 is my dissertation, where I worked on a micro-diachronic level, analyzing user language in rule-based and plan-based, media-written chatbots from 2000–2016. These systems were all used in the help-desk sector and were more or less advanced for that early stage of technological development. It represented a first description of HMI using a mixed-method approach with qualitative (conversation analysis) and quantitative methods (corpus linguistics). The data consisted of system log files from various application scenarios with real users, providing high ecological validity. Human-to-human chats with help-desk character served as a parallel corpus (library information, Chat-Korpus Beißwenger and Storrer 2004). Statistical analysis of corpus data included relative frequencies, distance-frequency analyses, and inferential statistics.

Figure 2: Our studies over the past 20 years



In 2016, Study 2 applied the mixed-methods approach (qualitative and quantitative) to Socialbots on Facebook Messenger, conducting a synchronous analysis of customer support bots in that context.

Study 3 and 4 constitute projects undertaken by my research group. Study 3 is a psycho-linguistic, hypothesis-testing experiment focusing on user strategies in oral interaction with Amazon Alexa (in collaborative tasks). Study 4 is a purely qualitative first description of written interaction with ChatGPT in elicited dialogues with the AI in two collaborative tasks (travel planning and essay writing) following an ethnomethodological approach.

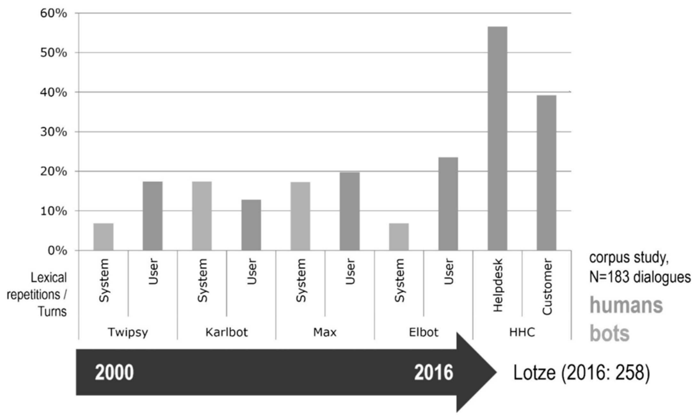
Even though our studies address different types of systems (rule-based, plan-based, VUI for RequestandResponse, GPT) and different modalities (written/oral), we observe similarities in the user language. In a simplified view, across older systems, social media systems, oral VUI, and the innovative GPT, we identify three fundamental tendencies in user language: preconscious (reactive) alignment; reactive adaptation strategies to the affordances of system architecture and dialog design, as well as simplifications in the sense of a “Simplified Register” (in extreme cases, even as CT according to Zoeppritz 1985). Alignment, acceptance, and simplification can be abbreviated to the acronym AAS, representing the main aspects of a heterogeneous HMI.

3.1 “Alignment” as a Preconscious Phenomenon

Interactive alignment in HHC, characterized by the tendency to adapt one’s language use to that of the interlocutor (Hartsuiker et al. 2000; Pickering and Garrod 2004), serves as a good example of preconscious behavior, given reaction times in the microsecond range. Perception and reception are so closely linked in HHC that a form just perceived remains cognitively active when people begin to produce their own contribution. Thus, it is more likely to reproduce what has just been perceived.

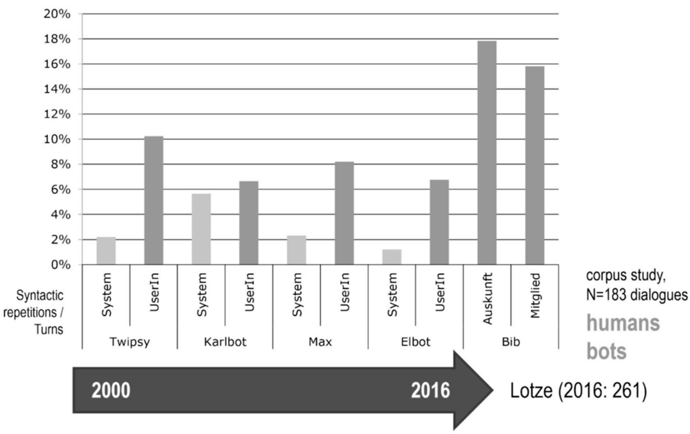
Reactive alignment of the user to the bot: Alignment in HMI has been demonstrated in various studies across all linguistic levels (Branigan et al. 2000; Branigan and Pearson 2006; Huiyang and Min 2022; Heyselaar 2017; Raveh et al. 2019; Lotze 2016; Fischer 2006; Linnemann and Jucks 2018).

Figure 3: Lexical alignment (Users of rule-based and plan-based systems) (Lotze 2016, 258)



In our corpora, it plays a less prominent role for users of our older systems compared to the HHC reference corpus (Lotze 2016, 254–55). Nevertheless, it appears consistently in every dialogue (approximately as frequently as in HHC) and has been identified on the syntactic and lexical levels. Nevertheless, these adoptions can be seen as indicating a transfer of the basic concept of interaction per se from human-human communication, especially among users of the more recent systems. The better the system works, the more the user goes along with the illusion. What is evident in my corpus studies on lexical alignment in HMI, which manifests as user repetitions of the systems lexis, is, that humans adapt less to the language of the system than to a human (on average 50 percent less for lexis and syntax) (c.f. Lotze 2016; 2018; 2019).

Figure 4: Syntactic alignment (Users of rule-based and plan-based systems) (Lotze 2016, 261)



However, we have to distinguish for lexical persistencies for each individual instance whether it is preconscious adaptation, socially motivated strategic adaptation, or a simplification strategy. In the latter case, a word form, that the system itself has already used, is selected for the user's own turn, because the user assumes, that this keyword is stored in the system's database. Thus, lexical alignment of the user can be interpreted either as a pre-conscious mechanism or as a strategic adaptation with different motivations (preconscious alignment as attribution of social proximity vs. simplification for the machine). Of course, as researchers, we can only speculate about the actual intentions behind the users' alignment to the system. However, the HMI research community agrees that users' intentions can vary and may also change over the course of a single dialogue, even for the same user.

Syntactic alignment is less frequent but shows the same trend and the interpretation as pre-conscious alignment is obvious here. With syntactic alignment, it becomes much clearer that in these instances users are not trying to find the right keyword. Instead, they are not aware of the adoption of the entire syntactic structure on a conceptual level. In the following example the user adopts the syntactic form of the predicative clause from the bot, even though there is a change of topic in the example, and the lexis is not adopted.

Example 2: Lotze (2019, 314)

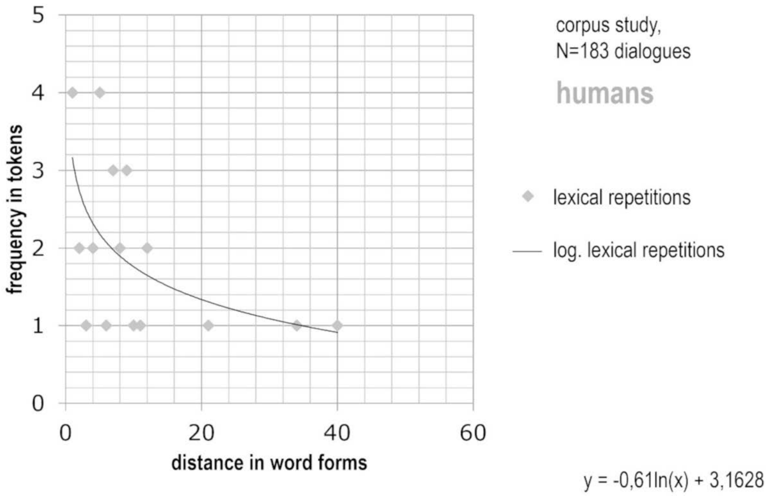
Max: **Das** [Nominativ] **ist** [Kopulaverb] **deine Meinung** [Nominativ].
[This is your opinion *[predicative sentence with "to be"]]*

User: **Stefan** [Nominativ] **wird** [Kopulaverb] **Informatiker** [Nominativ]
[Stefan will be a computer scientist? [predicative sentence with "to be"]]

In this interaction with the Max system at Bielefeld University, which was being tested by its developers at the time, Max concludes a prior dispute with the statement 'That's your opinion,' effectively suggesting to agree to disagree. The user then changes the topic and addresses the career ambitions of one of Max's developers (Stefan Kopp) with the remark, 'Stefan will become a computer scientist.' Despite the abrupt topic shift, the user syntactically aligns with the preceding system turn in form of a predicative clause.

Does human memory in HMI differ? Overall, the cognitive processing of dialogue by users in HMI is not fundamentally different from Computer-Mediated Communication (CMC) (more nuanced: Lotze 2016). For example, the rate of decay of the primes repeated by the user follows the "forgetting curve" of Ebbinghaus (1985); i.e., for users in HMI, a linguistic structure produced by the system becomes gradually less relevant, and repetitions by the user become rarer, as is the case in human-human communication as well. These research results support the idea that alignment in HMI is also preconscious behavior.

Figure 5: Rate of decay of primes (Users of rule-based and plan-based systems) (Lotze 2016, 279)

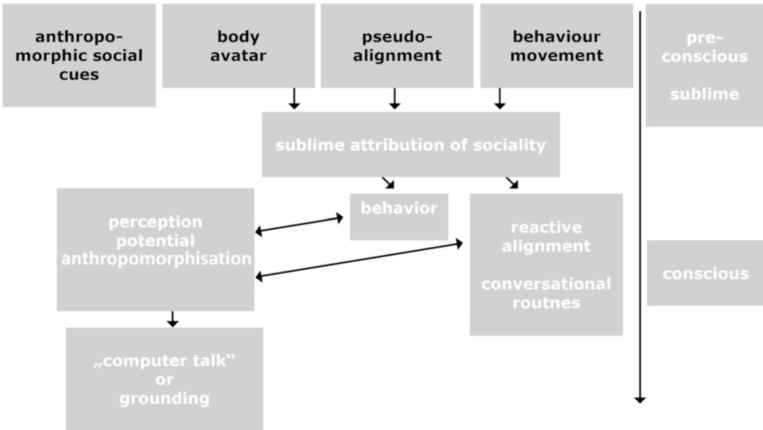


Thus, a model of interactive alignment in human-machine interaction has to be understood as a continuum of degrees of consciousness between automatic or sublime/subtle and conscious or strategic behavior. For this purpose, I added the new component "alignment" to the schematic representation of the social effects of artificial agents according to Krämer (2008, 202).

Reactive alignment of users is thus a proven phenomenon and, depending on the level of awareness and motivation, can be interpreted as: (a) an automatism of human dialog behavior (b) a linguistic indicator for attributing social characteristics to the system (c) a conscious simplification strategy (e.g. search for the right keyword) in the sense of a "computer talk" (CT) according to Zoeppritz (1985). Against the background of the discussion about dangers of interaction with AIs it then must be classified as (a) an unavoidable cognitive process, (b) as problematic because the system can only disappoint the expectations of the users and (c) as media literacy of users, who are familiar with the system. The example of "alignment" shows clearly, how multilayered the psycho-linguistic interpretation is.

The model can also be applied to other modalities. Greilich (in preparation) found even more alignment in the oral HMI among users of Amazon Alexa (in lexicon, syntax, and prosody). She conducted an experimental study as a collaborative task with Amazon Alexa with focus on referential expressions and topic continuity, both often manifesting in repetition. And she found not only more alignment in oral HMI, but also over longer sequences of persistencies in triplets or quadruplets.

Figure 6: Model alignment in HMI (Lotze 2016, 390)



Example 3: Greilich (in preparation)

User: Alexa, was sind die Öffnungszeiten?
[Alexa, what are the office hours?]

Alexa: Die Gemäldegalerie Alte Meister in Dresden öffnet in 4 Minuten?
[The galery Alte Meister in Dresden is going to open in 4 minutes?]

User: Alexa, wie lange ist sie am Montag geöffnet?
[How long is it open on Mondays?]

Alexa: Sie haben am Dienstag von 10 Uhr bis 18 Uhr geöffnet.
[They open on Thursdays from 10 a.m. to 6 p.m.]

In one of the scenarios of the collaborative tasks of the experiment the test users had to ask Alexa for information on Dresden from a tourist perspective.

With reference to the discussion about CASA/MASA vs. "Simplified Registers," we can establish the following aspects based on our empirical findings:

a) Reactive alignment can be interpreted as a transfer from HHC to CASA/MASA.
b) The phenomenon is so stable in HMI that it occurs even in medial-written interaction with old, extremely error-prone plan- and rule-based systems and is cognitively processed regularly (forgetting curve according to Ebbinghaus 1985).
c) The phenomenon depends on modality and intensifies in orality (probably due to the anthropomorphic voice and shortened reaction time).
d) Strategic alignment as a search for the appropriate keyword by users must be interpreted as a simplification strategy in the sense of a "Simplified Register" (Fischer 2011).

In addition, all ritualized aspects of interaction such as turn construction and allocation, politeness levels, and ritualized greeting sequences can be interpreted as "mindless behavior." We find numerous examples of all these investigative parameters in our studies (further explored in Lotze 2022).

3.2 "Acceptance" as a Transitional Phenomenon

At the transition between preconscious behavior and strategic action[2], we find a) highly frequent passive reactions to the affordances and restrictions of system architecture and dialog design[3] and b) transfers of proactive strategies from HHC (e.g., grounding as a repair strategy). In the former case, it is an *affordance-bound*, passive receptive stance of users that guides them through mainly plan-based applications in the quickest way without disruptions. In the latter case, it is an *affordance-unbound* user reaction, which is indeed a conscious repair, but often not a conscious decision of the users since the older systems in our corpora lack any world knowledge (cf. Habermas 1993). Such user strategies can still be conceptually interpreted as transferred concepts from HHC according to CASA/MASA on the cognitive level, but functionally, they represent a conscious repair strategy.

2 The processes identified by Pickering and Garrod are preconscious and thus automatic, meaning they occur prior to more complex processes of conscious interpretation (t < 600ms, see Pickering & Garrod 2004).

3 rule-based systems – user behavior: isolated keywords; plan-based systems – user behavior: passive receptive stance ("ok", "continue", "go back"); RequestandResponse systems – user behavior: isolated imperative sentences

User language that lies in this borderline area and we can only interpret it with a model that considers such a transition zone. Furthermore, we need to differentiate between affordance-bound acceptance and affordance-unbound acceptance. A "passive receptive stance" of users is found especially in plan-based systems that guide people step by step through decision tree-based dialog scripts, requiring a lot of confirmation in the dialogue. In extreme cases, users deviate from their original dialogue goal and let the AI passively guide them through the application. This example of an interaction with the Max system comes from the period when it was deployed as a virtual museum guide at the Heinz Nixdorf Forum in Paderborn.

Example 4: Max corpus 501–526
Max: Should I show you the next exhibit?
User: no, go back
Max: The next exhibit is the AI exhibition. Should I explain that?
User: ok

Krause's metaphorical language use (Krause and Hitzenberger 1992) has thus undergone a shift in its pragmatic function – away from the metaphor of actively operating a machine to a) cooperation with the system in processing specific tasks (cf. RDS, Fischer 2011) and b) a passive receptive stance towards conversational technology (Lotze 2016, 358). The latter manifests as reactive user behavior closely tied to the bot's instructions. An extreme example of this is the absence of interventions in case of disruptions.

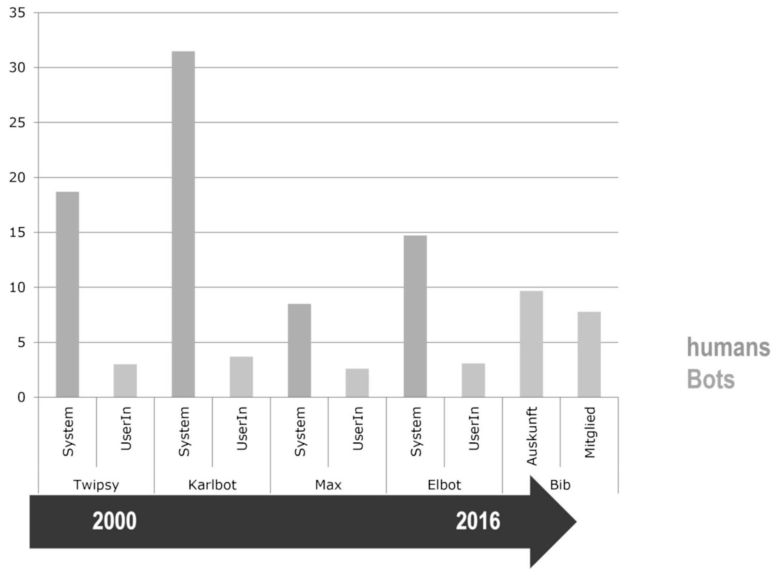
3.3 Simplification as an Affordance-Bound and Affordance-Unbound User Style

Simplifications by users are numerous across all examined systems on different linguistic levels, regardless of modality. Comparing the turn lengths of users and systems in the older rule- and plan-based chatbots, users consistently formulate extremely short turns, regardless of the length of the bot's turns. One could argue that this is due to the helpdesk scenario[4] with short questions and detailed responses, which represents the context of all corpora

4 All corpora examined in the 2016 study were log files of interactions with various chatbots in help-desk scenarios, specifically in customer support. The parallel corpus for human-human communication was the chat corpus by Beißwenger and Storrer (2004).

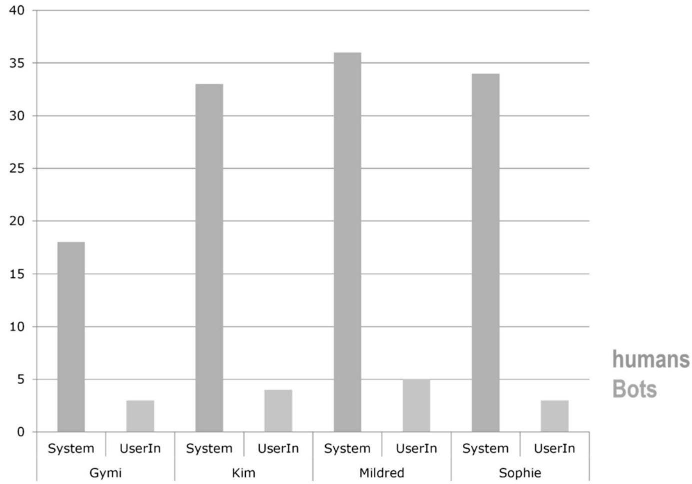
in this study. However, looking at the comparison corpus of CMC (computer-mediated communication) related to HHC, which was selected as a parallel corpus precisely because it is also a helpdesk, the quantification of the corpus study clearly shows that people adapt to each other regarding turn length, and this effect is stable even in the written medium.

Figure 7: Length of turns in chatbots (diachronic) (Lotze 2016, 234)



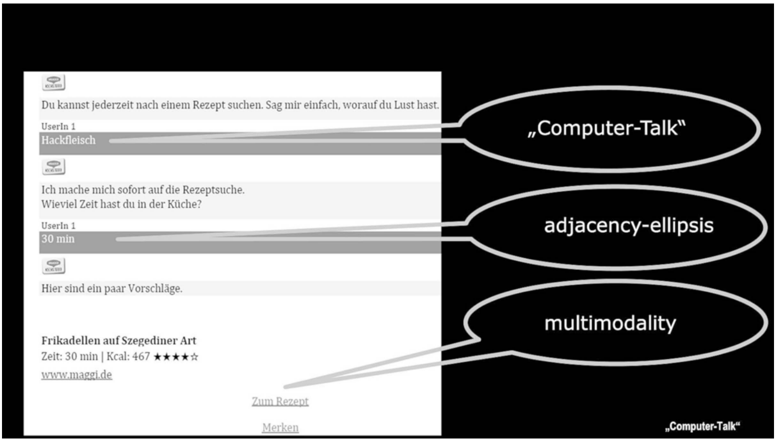
We could replicate this result in 2016 for a more recent type of intent-based social bots on Facebook Messenger and were able to reproduce my micro-diachronic study with corpus data from 2000–2016. Even with these significantly improved systems, turn lengths vary greatly, and users tend to become silent. This leads to various additional structural and functional simplifications (for further details, refer to Lotze 2016).

Figure 8: Length of turns in socialbots (synchronic) (Lotze and Ohrndorf, in preparation)



If the interfaces are additionally designed to be multi-modal with clickable areas, buttons or images, the effect is further enhanced.

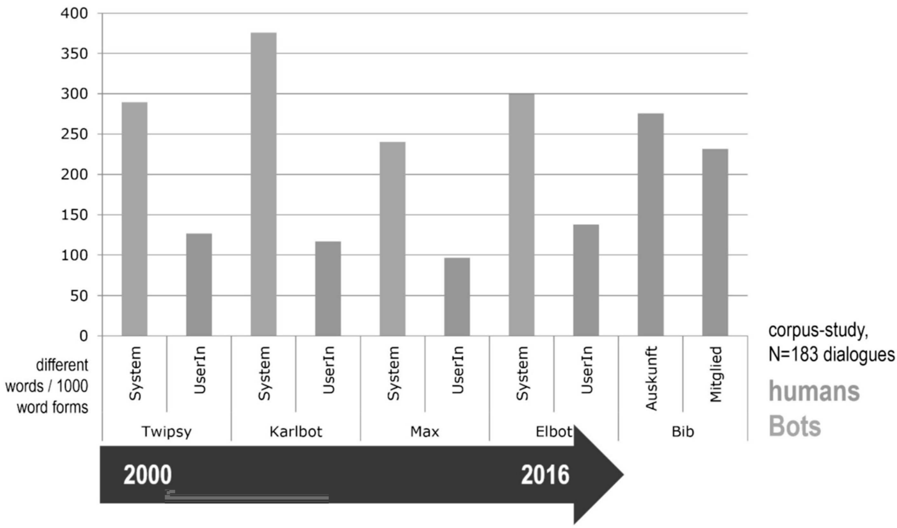
Figure 9: Isolated keywords in users of socialbots (Lotze and Ohrndorf, in preparation)



The example illustrates aspects of structural and functional computer talk at the interface between the desktop metaphor and the assistant metaphor, as well as the metaphor of Google search (for Krauses notion of metaphor see cap. 1.2.1, for further details s. Natale and Cooke 2021). The interfaces of the examined socialbots were primarily operated by their digital-native users using isolated keywords and adjacency ellipses related to the bot's previous turn. The tool metaphor dominates in this generation of users.

Not only does the turn length decrease, but lexical variability also decreases compared to computer-mediated communication (CMC). The Type-Token Ratio of users is significantly lower than in the parallel corpus, and that of bots is significantly higher, further emphasizing the asymmetry. In the older information bots, lexicon and syntax were primarily oriented toward written texts in a brochure, not the medial-written, quasi-synchronous dialogicity of CMC. This explains the richness of different lemmata, especially in the two oldest systems.

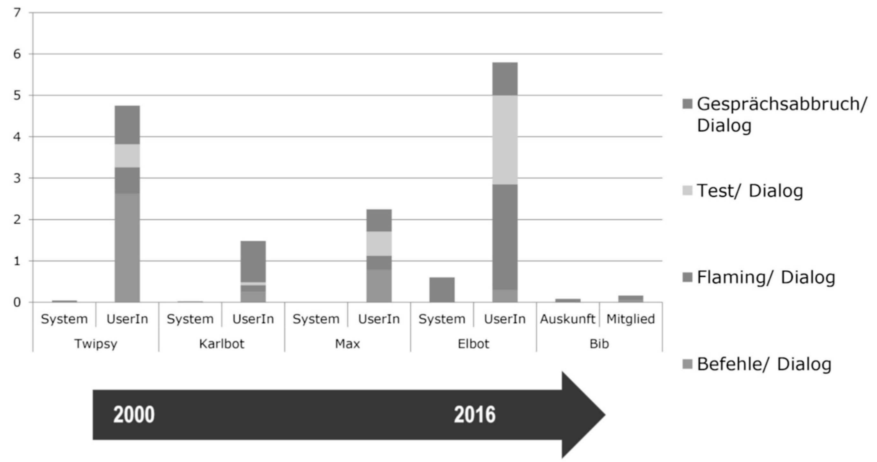
Figure 10: Simplification of lexical variation (Lotze 2016, 322)



Concerning syntax, users of rule- and plan-based systems utilize ellipses approximately 60–70 per cent of the time, not all of which are adjacency ellipses; some are also isolated keywords, simple imperatives, and confirmation signals. The remaining 20 per cent consist of simple sentences, often with a copula or main verb in the imperative form. Only 10 per cent form complex sentence structures with subordinate clauses. In comparison to the CMC reference corpus, ellipses constitute 40–50 per cent, predominantly being adjacency ellipses. In terms of syntax, humans strongly adapt to each other, while the frequencies for users and bots differ significantly, primarily due to the conceptual nature of the bot turns (Lotze 2016, 327).

Politeness in language remains a contentious research field in HMI, as studies yield different results depending on the application context, user type, and sophistication of the AI (see, e.g. Clark and Fischer 2023). Indicators of actual CT, following Zoeppritz (1985) as the extreme pole of a simplified user language, can only be interpreted through expressions that would be completely dispreferred in human interaction: isolated imperatives, vulgarisms (flaming), abrupt conversation interruptions, and playful testing of system functions by asking the bot personal, emotional or particularly complex questions. Instances of these forms of expressions are found in users of rule- and plan-based systems between 1.5 to 6 times per dialogue.

Figure 11: Linguistic (im)politeness (Lotze 2016, 338)



This result is particularly interesting as it demonstrates that impoliteness occurs more frequently in studies when the investigation corpora consist of unaltered field data (log files) with high ecological validity (as in Lotze 2016). Therefore, it can be inferred that users communicate more impolitely and directly with bots in real-world scenarios than in experimental settings. In her elicited data on Amazon Alexa, Greilich (in preparation) identifies more simple imperatives without politeness markers, but no vulgar language. This suggests that imperatives and isolated keywords depend on the affordances of the RequestandResponse architecture in orality. The absence of vulgarisms can most likely be analyzed as an experimenter effect.

Example 5: Imperative as affordance-bound simplification (Amazon Alexa, Greilich, in preparation)

User 1: Nenne mir bekannte Verfilmungen von Thomas Mann (Participant 1, Question 2, Attempt 2)
[Name famous film adaptations of Thomas Mann]

User 2: Erzähle mir etwas über die Werke von Thomas Mann (Participant 4, Question 2, Attempt 2)
[Tell me something about the works of Thomas Mann]

User 3: Finde mir bitte das Alter von Leonardo DiCaprio raus (Participant 7, Question 1, Attempt 1)
[Please find out the age of Leonardo DiCaprio for me]

More challenging to interpret are user utterances that are not triggered by the affordances of the dialog system.

Example 6: Isolated keywords as affordance-unbound simplification (Amazon Alexa, Greilich, in preparation)

User 4: Thomas Mann Nobelpreis (Participant 9, Question 2, Attempt 3)
[Thomas Mann Nobel Prize]

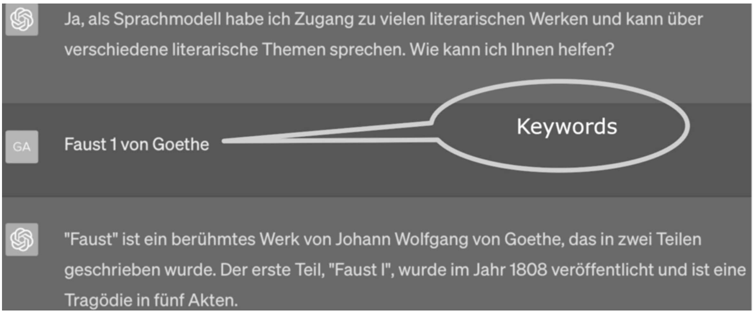
User 5: 5 Tickets Gemäldegalerie Alte Meister (Participant 9, Question 3, Attempt 6)
[5 tickets art gallery Alte Meister]

These affordance-unbound simplifications can indeed be interpreted as "metaphoric language use" according to Krause and Hitzenberger (1992), as users here experiment with the new metaphor of Google search in the form of isolated keywords. This provides deeper insights into users' assumptions about the system's functions. Alexa, as a voice-based assistant, has been optimized with a RequestandResponse architecture for oral commands for home automation and has been trained with empirical speech data. This suggests operation in complete sentences with finite verbs (especially in the imperative). However, some users transfer the cognitive concept of keyword-based bots and Google search from text to oral interaction with Alexa. For keyword-based searches, a socio-linguistic practice seems to have already developed, as indicated by the example of social bots (Fig. 9).

Even in generative transformers based on large language models (LLMs) that perform better when longer contexts are made explicit, we find affordance-unbound simplifications and evidence of reactive user behavior.

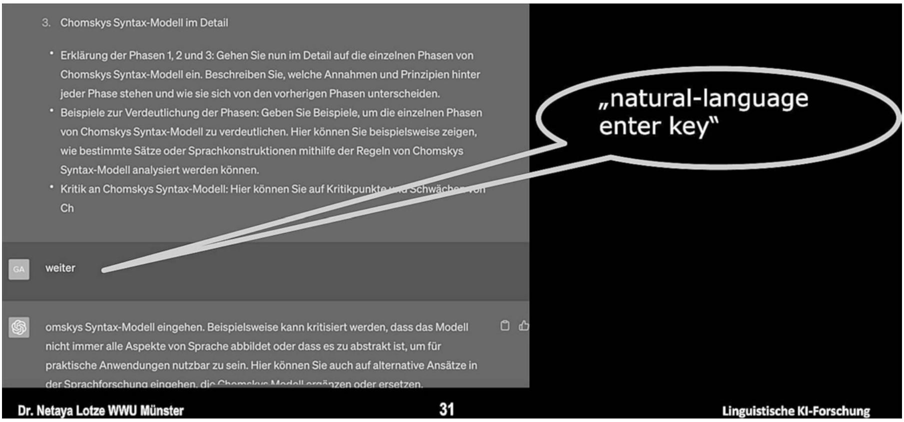
Example 7: Collaborative travel planning (1–3) and essay task with ChatGPT (Lotze and Aydin, in preparation)

Figure 12: Isolated keywords in ChatGPT



In the most extreme manifestation of this form of acceptance and passive user behavior (as in the example above), the user merely confirms the suggestions provided by the bot with "yes", "ok", or in the case of this example "next", which can be analyzed as an equivalent of the 'enter' key in natural language. This reactive strategy has evolved in the past, especially among users of plan-based systems, who playfully and exploratively let the system guide them through the application in this way. Now, they transfer this concept to interact with ChatGPT, thereby rendering the practice of "natural language 'enter'-key" no longer strictly interpretable as affordance-bound.

Figure 13: Reactive behavior in users of ChatGPT

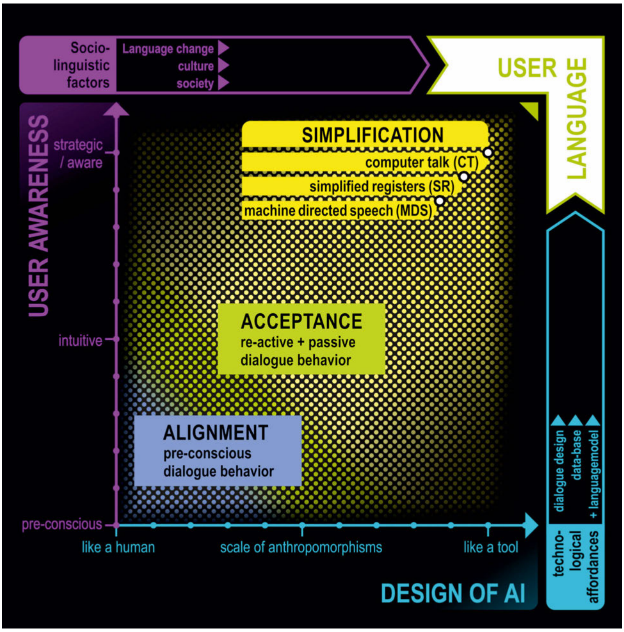


When technology changes, user strategies evidently do not shift immediately but with a time delay (cf. the "stylistic inertia law" (*stilistisches Trägheitsgesetz*, Schmitz 2015, 25–26; cf. Hauser 1958)). Therefore, if the simplification strategy does not align with the affordances of the (new) technology and cannot be understood as a transfer from the HHC (e.g., child-directed speech, etc.), Krause and Hitzenberger's (1992) idea of metaphorical language use for this small subset of linguistic simplifications within the HMI, in my opinion, remains relevant. Conceptual metaphors seem to undergo a diachronic change, which follows the technological revolutions with a delay. Metaphor is a concept that can manifest linguistically and structurally in various ways, and it must be considered partially independently of the affordances of technology and mediality (see above).

For a (still extremely young) diachronic research perspective on HMI, this means that user concepts also undergo changes over time. This becomes apparent whenever user assumptions about how the technology works lag behind: code as a metaphor for operating the first natural language interfaces, isolated keywords, and the "natural language 'enter'-key" (or dialog scripts) as metaphors for operating learning-capable and pre-trained systems based on large language models (LLMs). The time-delayed adaptation of the metaphor is interesting for linguistic discussions because in transition phases, one can observe that the affordances of technology do not directly trigger linguistic behavior but are always mediated by cognitive concepts, which are only discarded when they are no longer efficient. The conceptual level and the medial level are not always congruent in HMI (for HHC: see Koch and Oesterreicher 1994; for CMC: see Dürscheid 2003).

4. A Model for HMI as a Complex Socio-Linguistic Practice

Figure 14: AAS-Model of HMI as a complex socio-linguistic practice



In the following, a model for HMI as a complex socio-linguistic practice will be introduced. To do this, I will first present its three continua: a) a technological continuum, b) a human-cognitive continuum, and c) a human-linguistic continuum, with the latter being partially dependent on the first two. However, a strict dependency of user language on the degree of system anthropomorphism and individual cognitive awareness cannot be postulated, as language, in general, is subject to numerous social, cultural, and historically grown factors and undergoes language-specific changes. Moreover, the model is by no means deterministic but assumes spontaneous, flexible, and adaptable users. The multi-dimensionality of the model provides users with the ongoing opportunity to linguistically position themselves depending on the degree of system anthropomorphism. This occurs partly at a preconscious level and partly consciously and strategically. If preconscious behavior or a conscious strategy fails, the same user can reposition themselves in the interaction diagram. Concrete linguistic structures, manifested as lexical, syntactic, or phonological forms, can be analyzed by the linguistic community using the model.

Dimension 1: Technological affordances and anthropomorphic design

The technological dimension is graded in degrees of the anthropomorphization of the system (cf. Ruijten et al. 2014; 2019). This dimension is adopted from MASA (Lombard and Xu 2022). It refers not only to the degree of anthropomorphism of the system as an interface or robot doll but also includes the cognitive reception by its users. As reception effects coincide with the degree of anthropomorphism, Ruijten et al. (2014; 2019) argue that these parameters can be combined as one parameter for human reception of more or less anthropomorphic systems.

I would like to expand the visual, movement-based etc. anthropomorphism and its reception by the degree of linguistic anthropomorphism, which appears more relevant from a linguistic perspective. The gradual variation here is crucial, indicating to what extent a natural language dialogue succeeds in being coherent and cohesive over longer sequences or, for example, only at individual adjacency pairs (cf. old rule-based bots, and partially RequestandResponse systems). Therefore, the anthropomorphism of pragmatics in AI per se is particularly important for interpreting our data of interface-based AI (oral and written). Additional parameters in our studies include the voice or name of the AI, as in the case of Alexa. Regarding robotics, we cannot make any statements based on our own studies and rely primarily on Clark and Fischer (2023), Fischer (2011), Habscheid et al. (2018), Lenz et al. (2019), and can build upon MASA (Lombard and Xu 2022).

Dimension 2: Cognitive awareness levels of the user

Human consciousness can be understood as a temporally staggered phenomenon, ranging from “pre-conscious” to “conscious,” as cognitive availability hierarchies organize processing in the brain chronologically. For the cognitive processing of HMI by users, alignment as a lower-level priming presents a key phenomenon (see above). The processes considered by Pickering and Garrod (2004; with Gandolfi 2023) are pre-conscious and automatic, i.e., they occur temporally before more complex processes of conscious interpretation (t < 600ms, cf. Pickering and Garrod 2004). Lower-level priming alone can be understood as the driving force of the interaction in this area. The interactive alignment model (Pickering and Garrod 2004) does not provide information about factors related to conscious interpretation. Its mechanisms must precede considerations of the social goal orientation or intentionality of utterances both temporally and logically. Lower-level priming thus constitutes the starting point of the second dimension.

Especially assigned to consciousness are the activated memory, focal attention, and controlled (non-automatic) processes of information processing (cf. Wirtz 2021). Reflected, thoughtful user strategies that intentionally pursue their own agenda or individual dialogue goal accordingly represent the endpoint of the consciousness continuum.

Fischer takes initial steps in this direction by defining CT as functional (2006) and as a Simplified Register (2011). She assumes conscious user strategies that control dialogue behavior depending on assumptions about the bot and the user type. She emphasizes the tool character of user language. This contrasts with preconscious cognitive alignment as the cause of preconscious user behavior. Depending on the HMI application, strategies are developed more or less consciously (Lotze 2016, 334–336). Greetings and farewells follow transferred protocols from HHC, while repair strategies for disruptions are mostly consciously chosen. Therefore, a continuum between "awareness" strategies and "mindless behavior" (stereotypes, assumptions, cf. among others Reeves and Nass 1996 alignment, among others Pickering and Garrod 2004) should be assumed (Lotze 2016, 334–35).

Dimension 3: User language as a continuum of AAS (Alignment, Acceptance, Simplification)

The third dimension represents a continuum of degrees of simplification in user language – from pre-conscious alignment through passive and reactive behaviors to different simplification strategies (from RDS to CT). The starting point of this dimension is a user language that exactly matches the HHC and should, therefore, be interpreted radically according to CASA/MASA. This language is for us purely a hypothetical placeholder in the model, for which we (yet!) have no evidence. Innovative systems of the future may one day fill this space (or may not).

Alignment can be understood primarily as "mindless behavior" and manifests itself in persistences (repetitions), which I therefore include in the first section of the language continuum (see Chapter 2.1.1). Partial transfers from the HHC are naturally present in our studies, and we must consider them directly after alignment on the scale: turn construction and allocation, linguistic anthropomorphisms (e.g., through "you"/"they" pronouns, linguistic politeness, ritualized greetings, semantic-thematic anthropomorphisms (e.g., personal questions), etc.).

Acceptance phenomena and reactive behaviors, which are affordance-dependent, constitute the transitional area (see Chapter 2.1.2), such as "natural-language 'enter'-key") as a user reaction to plan-based systems.

Then come simplifications ("Machine Directed Speech" (MDS), "Simplified Registers" (SR)), initially those directly triggered by the affordances of the respective technology, and then those that are independent of them (see Chapter 2.1.3). The outer extreme pole in the continuum represents affordance-independent simplifications that occur across technologies (isolated keywords, abrupt terminations, and vulgar language as tests or after disruptions, etc.). These come closest to CT according to Zoeppritz (1985) and metaphorical language use (Krause and Hitzenberger 1992) and seem to have emerged as a new digital practice per se. Here, the tool character of the application alone seems to be the cognitive guiding concept.

External influencing factors:

In my view, user language evolves in dialogue not simply between the reception of an anthropomorphic technology and the level of consciousness in its cognitive processing but might also actually be modeled as an independent dimension. We need to consider it as an only partially dependent variable. Language follows its own principles, which manifest in the formation of style and register over extended periods. New technologies with new affordances give rise to new linguistic practices that should not be solely interpreted as technology deterministic. Besides the technological realm, there exist a social and language inherent realm. Language variation and change always occur in the interplay between explicitness and simplifications. Grammar and lexicon of each individual language also play a role. Lexicalization and grammaticalization and language- and culture-specific parameters for variation and change must be considered in a modern model for HMI. Otherwise, one cannot explain technology-transcendent new socio-linguistic practices (e.g., isolated keywords as a user strategy in all bots, from old, rule-based systems to Voice User Interfaces and ChatGPT). Therefore, the model also takes into account socio-linguistic factors such as language variation and change, culture, and society alongside technological affordances like dialogue design, data basis, and language model. This makes the multi-dimensional AAS-model compatible with more abstract, humanities-oriented discourses on AI.

References

Beißwenger, Michael, and Angelika Storrer. 2004. '*Dortmunder Chat-Korpus.' Chat-basierte Bibliotheksauskunft.* http://www.chatkorpus.tu-dortmund.de/files/releasehtml/index.html.

Branigan, Holly P., Michael J. Pickering, and Andrew A. Cleland. 2000. "Syntactic co-ordination in dialogue." *Cognition* 75 (2): 13–25. https://doi.org/10.1016/S0010-0277(99)00081-5.

Branigan, Holly P., and Jamie Pearson. 2006. "Alignment in human-computer interaction."In *Proceedings of the Workshop on How people talk to computers, robots, and other artificial communication partners*, edited by Kerstin Fischer, 140–56. Hanse-Wissenschaftskolleg Institute for Advanced Study.

Brown, Penelope, and Stephan C. Levinson. 1987. *Politeness. Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Clark, Herbert. H., and Kerstin Fischer. 2023. "Social robots as depictions of social agents." *Behavioral and Brain Sciences* 46: 1–33. https://doi.org/10.1017/S0140525X22000668.

Dippold, Doris. 2023. "'Can I have the scan on Tuesday?' User repair in interaction with a task-oriented chatbot and the question of communication skills for AI." *Journal of pragmatics* 204: 21–32.

Dürscheid, Christa. 2003. "Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit: theoretische und empirische Probleme." *Zeitschrift für angewandte Linguistik* 38: 37–56.

Ebbinghaus, Herman. 1985. *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker and Humblot.

Fillmore, Charles J. 1976. "Frame semantics and the nature of Language." In *Annals of the New York Academy of Sciences. Origins and Evolution of Language and Speech*, 20–32.

Fischer, Kerstin. 2006. *What computer talk is and isn't. Human-Computer Conversation as Intercultural Communication*. Saarbrücken: AQ-Verlag.

Fischer, Kerstin. 2011. "How people talk with robots: Designing dialog to reduce user uncertainty." *Ai Magazine* 32 (4): 31–38.

Gandolfi, Greta, Michael J. Pickering, and Simon Garrod. 2023. "Mechanisms of alignment: shared control, social cognition and metacognition." *Philosophical Transactions of the Royal Society B* 378 (1870).

Greilich, Anna. In preparation. *Information structure in human-computer interaction: The linguisticanalysis of multi-turn information search with Amazon Alexa*. Dissertation thesis.

Habermas, Jürgen. 1993. *Theorie des kommunikativen Handelns*. Frankfurt am Main: Suhrkamp.

Habscheid, Stephan, Christine Hrncal, Jens Lüssem, Rainer Wieching, Felix Carros, and Volker Wulf. 2018. "Robotics and Emotion." In *EuropeNow* 19. *A Special Feature on Anxiety Culture*, edited by Nicole Shea and Emmanuel Kattan. https://www.europenowjournal.org/2018/07/01/robotics-and-emotion.

Hartsuiker, Robert J., and Casper Westenberg. 2000. "Word order priming in written and spoken sentence production." *Cognition* 75 (2): 27–39.

Hauser, Arnold. 1958. *The Philosophy of Art History. Routledge Revivals*. London: Routledge.

Heyselaar, Evelien. 2017. *Influences on the magnitude of syntactic priming*. Nijmegen: Radboud University Nijmegen.

Huiyang, Shen, and Wang Min. 2022. "Improving Interaction Experience through Lexical Convergence: The Prosocial Effect of Lexical Alignment in Human-Human and Human-Computer Interactions." *International Journal of Human-Computer Interaction* 38 (1): 28–41. https://doi.org/10.1080/10447318.2021.1921367.

Koch, Peter, and Wulf Oesterreicher. 1994. "Schrift und Schriftlichkeit." In *Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung*, edited by Herbert Günther and Otto Ludwig, 587–603.

Köhler, Reinhard. 2005. "Synergetic Linguistics." In *Quantitative Linguistik – Quantitative Linguistics. Ein internationales Handbuch*, edited by Reinhard Köhler, G. Altmann, and R. G. Piotrowski, 760–74. Berlin and New York: De Gruyter.

König, Katharina. 2019. "Sequential patterns in SMS and WhatsApp dialogues: Practices for coordinating actions and managing topics." *Discourse and Communication* 13 (6): 612–29.

Krämer, Nicole. 2008. *Soziale Wirkungen virtueller Helfer: Gestaltung und Evaluation von Mensch-Computer-Interaktion*. Stuttgart: Kohlhammer.

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

Krause, Jürgen, and Ludwig Hitzenberger. 1992. *Computer Talk*. Hildesheim, Zürich, and New York: Olms.

Lenz, Gaby, Jens Lüssem, Hannes Eilers, and Hannah Wachter. 2019. "Soziale Robotik in der Altenpflege." *Soziale Arbeit* 68 (11): 402–09.

Linnemann, Gesa Alena, and Regina Jucks. 2018. "'Can I Trust the Spoken Dialogue System Because It Uses the Same Words as I Do?' – Influence of Lexically Aligned Spoken Dialogue Systems on Trustworthiness and User Satisfaction." *Interacting with Computers* 30 (3): 173–86. https://doi.org/10.1093/iwc/iwy005.

Lombard, Matthew, and Kun Xu. 2021. "Social responses to media technologies in the 21st century: The media are social actors paradigm." *Human-Machine Communication* 2: 29–55.

Lotze, Netaya. 2016. *Chatbots: Eine linguistische Analyse*. Frankfurt: Peter Lang.

Lotze, Netaya. 2018. "Zur sprachlichen Interaktion mit Chatbots – Eine linguistische Perspektive." In *Talk with the Bots – Gesprächsroboter und Social Bots im Diskurs*, edited by Theo Hug and Günther Pallaver, 9–50. Innsbruck: University Press.

Lotze, Netaya. 2019. "Psycholinguistik der KI-Forschung – Beeinflussen Künstliche Intelligenzen unsere Art zu kommunizieren?" *Psychologische Aspekte von Automation und Robotik. Psychologie in Österreich* 4: 310–16.

Lotze, Netaya. 2022. "Zur Adressierung des Unbelebten – Grenzen von pragmatischer Konzeption." In *Mensch – Tier – Maschine: Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*, edited by Miriam Schmidt-Jüngst, 305–327. Bielefeld: transcript, 305–27.

Lotze, Netaya, and Gabriel Aydin. In preparation. *Chat-GPT: Kohärente Interaktion?* Concept paper.

Lotze, Netaya, and Laura Ohrndorf. In preparation. *Socialbots. Eine linguistische Analyse*. Lang.

Nass, Clifford, and Scott Brave. 2005. *Wired for speech: how voice activates and advances the human-computer relationship*. MIT Press.

Nass, Clifford, and Young Me Moon. 2000. "Machines and mindlessness: Social responses to computers." *Journal of Social Issues* 56 (1): 81–103. http://stanford.edu/~nass/vita.html.

Natale, Simone, and Henry Cooke. 2021. "Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces." *Media, Culture & Society* 43 (6): 1000–1016. https://doi.org/10.1177/0163443720983295.

Pearson, Jamie, Holly M. Branigan, Martin J. Pickering, J. Hu, and Clifford Nass. 2006. "Adaptive Language Behavior in HCI: How Expectations and Beliefs About a System Affect Users' Word Choice." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1177–80. https://doi.org/10.1145/1124772.1124948.

Pickering, Martin J. and Simon Garrod. 2004. "Towards a mechanistic psychology of dialogue." *Behavioural and Brain Sciences* 27: 169–225. https://doi.org/10.1017/S0140525X04000056.

Raveh, Eran, Ingo Siegert, Ingmar Steiner, Iona Gessinger, and Bernd Möbius. 2019. "Three's a Crowd? Effects of a Second Human on Vocal Accommodation with a Voice Assistant." In *Proceedings of the 20th Annual Conference of the International Speech Communication Association*, 4005–4009. https://doi.org/10.21437/Interspeech.2019-1825.

Reeves, Byron, and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Cambridge University Press.

Reeves, Byron, Jeff Hancock, and S. X. Liu. 2020. "Social robots are like real people: First impressions, attributes, and stereotyping of social robots." *Technology, Mind, and Behaviour* 1 (1): 1–37. https://doi.org/10.1037/tmb0000018.

Ronneberger-Sibold, Elke. 1980. *Sprachverwendung – Sprachsystem. Ökonomie und Wandel*. Tübingen: Niemeyer. https://doi.org/10.1515/9783111346076.

Ruijten, Peter A., Diane H. L. Bouten, Dana C. J. Roushop, Jaap Ham, and Cees J. H. Midden. 2014. "Introducing a rasch-type anthropomorphism scale." In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 280–81. https://doi.org/10.1145/2559636.2559825.

Ruijten, Peter A., Antal Haans, Jaap Ham, and Cees J. Midden. 2019. "Perceived human-likeness of social robots: testing the Rasch model as a method for measuring anthropomorphism." *International Journal of Social Robotics* 11, 477–94. 10.1007/s12369-019-00516-z.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1992. *Lectures on Conversation, Volumes I and II.* Blackwell: Oxford.

Schmitz, Ulrich. 2015. *Einführung in die Medienlinguistik*. Darmstadt: WBG.

Siever, Torsten. 2011. *Texte i. d. Enge. Sprachökonomische Reduktion in stark raumbegrenzten Textsorten*. Frankfurt am Main: Peter Lang.

Wirtz, Markus A. 2021. *Dorsch – Lexikon der Psychologie*. Göttingen: Hogrefe.

Zoeppritz, Magdalena. 1985. *Computer talk? Technical Report TN 85.05*. Heidelberg: IBM Heidelberg Scientific Center.

II Linguistic Exchange with Voice Assistants as a Practical Problem

"Oh, Now I have to Speak"
Older Adults' First Encounters with Voice-based Applications in Smartphone Courses

Florence Oloff

Abstract *This chapter deals with the question of what we can learn from interaction in institutional settings about the usability and learnability of everyday technologies such as voice-based Intelligent Personal Assistants (IPAs), especially for older adults or, more generally, less-expert technology users. Based on an analysis of video recordings made during smartphone courses in adult education centers in Germany, this contribution provides a qualitative and micro-analytical perspective on non-expert adult users' processes of discovering and exploring voice-based technologies. Using the framework of multimodal conversation analysis, both linguistic formats and embodied actions are examined, revealing the participants' situated and dynamic understandings of how one type of IPA (as a smartphone app or widget) works and operates. The analysis of these either guided or accidental discoveries of a new technology can provide new insights regarding the specific challenges associated with handling IPAs and instructing new users how to do so. Based on these observations, this chapter also provides some general thoughts on teaching digital skills to less-expert users.*

1. Introduction

Voice interfaces such as Intelligent Personal Assistants (IPAs), integrated into personal smartphones or as external devices, have been marketed as a particularly accessible technology that can be easily incorporated into our everyday lives (Reeves et al. 2018). However, despite the intuitive ease of use propagated, users with low technical affinity do not seem to adopt these technologies particularly well. Apart from – perhaps false – assumptions regarding idealized user types and interface design, one possible reason for this is the lack of opportunities to *discover* voice-based technologies, especially for older adults. One situation that can provide such a chance is in introductory smartphone courses for older adults: as well as explaining how to use basic applications such as messaging programs, emails, or the camera, some courses also introduce smartphones' voice assistant function (in our data, typically the "Google Assistant"). Although this type of institutionalized learning setting cannot show how participants adopt new technologies over longer periods of time, nor how practices and routines emerge in regular use in everyday life, observing this educational context can offer a unique opportunity to examine how initial contact is made with a previously unknown application.

The question of how children and young adults learn to use new technologies and media has attracted widespread interest in research. How people in later life phases get in touch with and use new technologies has received much less attention, however, and studies have tended to be based on questionnaires or interviews (section 2.1). Regarding IPAs more specifically, interactional research has illustrated how domestication processes manifest themselves in the details of talk with and around IPAs in mundane, private settings (section 2.2). With reference to video recordings made during introductory smartphone courses, this chapter is based upon data from a non-experimental setting in which older adults engage with everyday technologies (section 3). Deploying multimodal conversation analysis, I then illustrate how course participants discover and try to use an IPA for the first time (section 4). Both instructed and accidental as well as individual and joint 'discovery processes' are considered, showing which types of obstacles non-expert users encounter and how they respond to the discovery of this new application. Finally, the potential of this type of data and analysis to further our understanding of how non-experts approach mundane technologies, and how digital skills teaching might be improved, is briefly assessed in section 5.

2. Background

Both IPAs in general and the communication routines and technology use of older adults have been studied from a wide range of perspectives and fields, including, among others, computer science, media and communication studies, human–computer interaction (HCI), social psychology, and applied linguistics. As this contribution focuses on situated technology use in non-experimental social settings, the most relevant prior research comprises qualitative studies of older adults' interactional practices (with technologies, Section 2.1) or involving the use of IPAs in general (Section 2.1).

2.1 Older adults in social interaction (and interacting with technology)

"Communication and aging" was coined as a topic in the early 1990s to emphasize that despite the decrement in health and skills associated with aging (Coupland et al. 1991, Mollenhauer and Meier zu Verl 2023, 8–12), age should be understood as essentially contributing to an individual's identity and as a development process that unfolds in and through communication (Nussbaum and-Coupland 2004). However, just as ageism and ascriptions of age are part and parcel of our daily personal and institutional communication routines (Fiehler and Thimm 2003, Thimm 2000), a bias can also be observed in researchers' choice of settings and phenomena for studies focusing on older participants. Within interactional studies, for instance, most research seems to investigate speech-related pathologies (and how participants successfully communicate despite certain constraints, Goodwin 2003, Wilkinson 2019), on communication in private or institutional care settings (investigating issues of autonomy or entitlement, Backhaus 2013, Lindström 2005), or on practices of remembering and self-reflection (Boden and Bielby 1983, Boxer 2018).

When it comes to technology and the internet/media use of senior citizens, a plethora of studies have typically deployed surveys and interviews to investigate the question of how this population adjusts to the increasingly pervasive integration of digital tools into everyday life. Within the last two decades, a shift can be observed from an attitude towards technology characterized by anxiety and resistance (Czaja et al. 2006, Selwyn et al. 2003) toward more creative and customized practices whereby technology supports the maintenance of social connectedness (Quan-Haase et al. 2016, Wang et al. 2018), as "… longtime users of digital media have grown up into older age …" (Quan-Haase et al. 2018, 1207). Nonetheless, older adults remain a heterogeneous population in terms of their digital practices and experiences, as they dynamically encounter various technologies at different stages of their lives (Domínguez-Rué and Nierling 2016, Vincent 2018).

But while asking elderly participants to report on and assess their own level of connectedness and technology acceptance is important, such studies offer little to further understanding of how older adults develop skills in handling hard- and software, and how situated processes of domestication and taming (Waldecker and Hector 2023) of new technologies actually unfold (see also section 2.2). Within interactional approaches, there have been studies on first contact among older participants and assistive technologies, such as social robots (Habscheid et al. 2020) or virtual assistants (Opfermann/Pitsch 2017), but other technologies remain understudied from this perspective (see Hrncal andHofius 2023, 125–127). The settings investigated have tended to be semi-experimental, testing participants' reactions in trials designed to assess a specific technology's acceptability and user design (see, e.g., Hrncal andHofius 2023, Pino et al. 2015). In their case study on the use of social robots in care facilities, Carros et al. (2020) describe the elderly as being "more restrained and insecure" (ibid., 5) when first meeting a robot, and then engaging more actively with it from the second time on. While focusing on assistive technologies is clearly justified from a demographic and socio-economic perspective (Carros et al. 2020), this emphasis tends to render uses of more mundane technologies invisible. First, older adults can and do inhabit this world not exclusively as participants in need of assistance, but also as fully capable, i.e., typical (cf. Antaki andWilkinson 2012 for the notion of (a)typicality), participants, who also use technology for non-medical purposes and in non-institutional settings; uses that do not differ fundamentally from those of younger adults. Second, research with older citizens tends to take what could be called an exoticizing approach, in the sense that the technologies under investigation tend to be highly specialized, pricey, and often still in development or in a test phase (see, for example, Carros et al. 2020, Opfermann anditsch 2017, Pino et al. 2015). While this is linked to the applied dimension and with aims to develop and improve specific designs and user interfaces, especially in HCI-related research, one result is that mundane and fully domesticated technologies such as smartphones, tablets, and laptops in their standard uses tend to be overlooked, possibly due to being perceived as less interesting or valuable research topics (Oloff 2021a, 197ff.).

Indeed, there seems to be a greater societal interest in the techno-socialization of toddlers and children (e.g., Lahikainen et al. 2017, Wiesemann et al. 2020), thereby further marginalizing attention to the acquisition of digital skills in older populations. This is compounded by the difficulty of identifying precise places and times in which such learning processes could take place, as older adults, unlike children and young adults, do not generally encounter technologies within compulsory institutional contexts (kindergarten, school, university), but ad hoc and at different moments in their professional and private lives. One such setting, however, is provided by digital skills courses within adult education. Indeed, micro-analytical studies conducted in these settings can reveal precisely the challenges older adults face when learning how to use mobile and smartphones, tablets, or computers (Oloff 2023, Råman 2022, Weilenmann 2010). Drawing conclusions from video recordings made during instructional and learning activities in situ, a multimodal interactional approach can expand upon the findings of more design-oriented research by contributing new perspectives and topics that offer detailed insights into the potential benefits and complex obstacles that mundane technologies present to less-expert users, singling out specific physical and digital *learnables* (cf. Råman 2022). It is this approach that is followed in the study presented in this chapter.

2.2 IPAs in social interaction

One of the advantages of taking an interactional approach to study the role of technology in our lives is that it enables us to examine technology use outside controlled laboratory conditions, i.e., 'in the wild' of everyday life. Compared to log data or protocols that only provide snippets of talk with and around IPAs, video recordings of social interaction offer a more comprehensive perspective on how IPAs are embedded in conversational and other mundane activities (Habscheid et al. 2021, Porcheron et al. 2017). In multi-party interactions, users have been shown to mutually adjust by selecting a query performer or by producing silence (Porcheron et al. 2017, 2018), thereby collaboratively finding ways to use IPAs in co-presence with others, despite them having been initially designed for single users (Albert et al. 2023). Indeed, interactionally-oriented research has criticized the focus in much of HCI research on IPAs as being driven by false assumptions about ideal users and conversational models that the device should supposedly be designed for (Reeves andPorcheron 2023). Research based on interactional data concludes that it would be better to design IPAs not in accordance with an idealized model of "conversation" that the device output is optimized to correspond to, but rather to aim to maximize the progressivity of the request–response chain (Fischer et al. 2019, Reeves et al. 2018, Reeves andPorcheron 2023).

When users try out and explore a new IPA, they usually proceed by taking a trial-and-error approach (Habscheid et al. 2023, Velkovska et al. 2020). This first exploration phase ends when users' practices involving the IPA stabilize, indicating that the technology has been "tamed" by its new users (Waldecker/Hector 2023, note, however, that the domestication/taming of new technologies does not always follow a unilinear trajectory, as new functions or updates can instigate new appropriation processes, see Peil/Röser 2023). IPA users have been shown to repeat or refine their formulations of queries (Porcheron et al. 2017) and to try out different syntactic formats, with a decreasing number of unsuccessful commands over time, which might indicate a learning process leading to more successful and therefore routine query types (Barthel et al. 2023, Porcheron et al. 2018). More experienced users have been shown to later flexibly ascribe or restrict the IPA's agency according to their technical needs and interactional purposes (Habscheid et al. 2023).

According to the literature review by Stigall et al. (2019), studies addressing older adults and voice assistants were overall very few and predominantly interested in the participants' *perceptions* of IPAs (with respect to their usability, accessibility, or trustworthiness) or in their preferences (e.g., regarding the hardware, or the gendered voice output). IPAs are thought to be possibly more user-friendly for older or disabled users than other interfaces, as the voice interface does not require potentially challenging physical or visual input methods (Stigall et al. 2019). With their study on the use of IPAs in a private home-care environment, however, Albert et al. (2023) show that IPAs do not represent a technological panacea – even if they can augment the independence of those in need of care – because they always end up being used in complex socio-material settings that cannot be modelled in advance. Moreover, even if problems with tactile user interfaces seem to be rather prevalent among less experienced technology users (see, e.g., Råman 2022, Weilenmann 2010), IPAs seem to figure among the least-used mundane technologies among the elderly.[1] While one possible reason for this might be the lower frustration threshold of older users than their younger counterparts when interacting with IPAs (Desai/Chin 2023), detailed research on older participants interacting with IPAs in everyday settings could shed more light on why voice interfaces are less popular with this user group. This chapter therefore proposes looking closely at situations in which older adults make first contact with IPAs, in order to better understand both the challenges and the opportunities IPAs present for less experienced users.

1 In a survey of media use among participants of 60+ years of age in Germany, only 14% of the respondents reported having an IPA at their disposal, compared to 100% for TVs and 72% for smartphones (SIM-Studie 2021, 6). "As this figure is significantly lower than the ownership rate of smartphones, which usually include a voice assistant, it can be assumed that many people are not aware of these functions, or thought that they were only available on [external] devices." (translation of SIM 2021, 7).

3. Data and Method

The video data analyzed for this contribution were collected within the research projects "Smart Communication" and "DigiLife". Both projects set out to investigate the use of mundane technologies in naturally occurring (i.e., non-elicited) face-to-face encounters, with the project DigiLife focusing more specifically on older adults' routines and challenges when handling technological devices such as smartphones or tablets. Although the currently available data set involving older adults was collected in both institutional (currently approximately 38 hours of recorded video) and private settings (currently 12 hours), it is only the institutional data set that is referenced in this contribution. This is because, for one, participants did not spontaneously use IPAs during the recorded sessions in private settings; secondly, my focus in this chapter is on first encounters with IPAs and their initial exploration, which can be observed as an activity in smartphone courses designed for non-expert users, but are much more difficult to record "in the wild" of private homes (Hector/Hrncal 2020). Courses offered by public adult education centers may be attended by adults of any age; however, in our data, it was typically older adults who seemed interested in acquiring basic smartphone skills. Among the seven smartphone courses that were recorded in adult education centers in different regions in Germany (most of which offered an introduction to Android phones), the presentation and introduction of voice-controlled applications such as Google Assistant was rather peripheral. This relates to the courses' introductory scope, emphasizing the most basic functions of the hardware and the operating system (including, for instance, control buttons for volume, different connection modes such as WIFI or mobile data, writing emails, or taking pictures), and to the limited time frame of the courses: usually ranging from two to twelve hours. Indeed, the only detailed introduction to IPAs observed during the courses took place in one of the longer ones (which comprised three three-hour sessions). In the other courses, the existence of IPAs was mentioned, but they were not introduced as a separate topic. However, course participants can potentially discover voice-controlled applications themselves at any time by intentionally or accidentally activating the app, as happened during one of the observed courses. In this chapter, I focus on moments of both instructed and accidental discoveries of IPAs during these courses.

The framework used for analyzing these first encounters with IPAs is ethnomethodological conversation analysis: a qualitative approach to social interaction with the primary aim to reveal the underlying orderliness of conversational and other mundane activities, i.e., the participants' methods (Bergmann 1981, Garfinkel 1967) for achieving social order and mutual understanding. Based on audio and video recordings of naturally occurring social interactions (Sacks 1984), the main tools of ethnomethodological conversation analysis are detailed transcripts and sequential analysis (Schegloff 2007), focusing on the precise temporality and coordination of audible and visible action (Mondada 2016, 2018). The data were transcribed using transcription conventions following Jefferson (2004) for the verbal transcript, and Mondada for the multimodal annotations (Mondada 2018[2]). The original talk in German was translated into English endeavoring to provide an idiomatic translation. All the participants consented to the recording and to their data being used in scientific publications, and all names have been pseudonymized in the transcripts.

4. Analysis

In this section, I focus on three excerpts from video recordings made during smartphone courses for adults in which participants explore IPAs, firstly, as part of an instructed activity (section 4.1), and secondly, following their accidental discovery (section 4.2). The analyses take into account how the participants orient to this discovery process through talk and embodied actions, and what the participants' conduct reveals about their expertise and stance toward the 'new' technology.

4.1 Instructing the Use and Exploring IPAs for the First Time

I now take a closer look at a smartphone course in which part of a session is devoted to the Google Assistant (as a widget[3] on Android smartphones, see Figure 1 as an example). This slot of approximately 20 minutes takes place at the beginning of the second of the course's three meetings. The seven participants

2 https://www.lorenzamondada.net/multimodal-transcription (accessed on 10/08/2024).

3 A widget is a simplified interface that can be positioned on the home screen of a mobile device (in this case a "mobile widget") so as to make a specific and usually data-rich application more quickly accessible to the user.

are seated at tables arranged in a horseshoe formation, and course leader BEN uses a smartboard and his own phone during the course (see Figures 2–3)[4]. Prior to the excerpt, BEN had already shown (via smartboard) how to carry out a browser-based web search and a search in the Google Play Store. He then announced that he would now present a "completely different way of searching" and began by pointing to the Google widget and the microphone icon in the search bar on his own phone (see Figure 2). The participants were supposed to then find the widget and the icon on their own device; a task that, due to the design of the microphone icon (circled in black in Figure 1), caused some difficulties (for a more detailed analysis, see Oloff 2021b). After ensuring that each participant had identified the microphone icon on their own phone (by moving around and checking on all the participants), BEN now explains how to make the IPA work (note that in the multimodal annotation, the abbreviation SP stands for smartphone):

Excerpt 1 (190925_VHSB_001521_okgoogle)

Figure 1: Widget with microphone icon. Figure 2: #2, l. 1

4 BEN uses the smartboard to project a browser window (e.g., the Play store) from the desktop computer in the room and for writing down basic keywords and instructions, or for drawing diagrams of the smartphone screen. He rarely uses his own smartphone, except, for instance, to show and comment on specific functions, or to demonstrate how to use it, such as when introducing the IPA (see Excerpt 1). The phone screen, however, is not projected and therefore the course leader's manipulation of the device is largely unseen by the course participants and can only be followed on the basis of his spoken commentary. This might contribute to the participants' difficulties in executing tasks on their own phones afterwards.

```
1    BEN  +SO; wenn ich das #2mikrofo:n antippe; (.) dann kann ich das
          NOW  if   I  tap  on the microphone   (.) then I can ask
          +...right index points to SP/widget------------------->
     kle  >>looks down at her SP and notes--->>
2         was ich suche, (1.2)+als   FRA:ge stellen.     (0.9)
          a question     (1.2) about what I am searching (0.9)
          >pppp to SP/widget--+,,,
3         also; (.) und die [frage?]
          so    (.) and the [question
4    KLE                    [ °ach:]so:; (jetzt)     [sprechen;°]#3
                            [  oh: ]right (now) with [speaking  ]
5    BEN                                             [die läu:t,]   äh:
                                                     [(this) ring-] e:r
6         die:;           (.) beginn ich, mit;  (.) ‘okay: google.’ (0.5)
           this (question) (.) I’ll  begin  with (.) ‘okay: google’  (0.5)
7         dieses ‘okay google:’ (.) ist da[für da, .h: dass    er[:,
          this    ‘okay google’  (.) is the[re     .h: so that he[:
8    SPth                                 [((beep1))
9    THI                                                         [°ah:°
                                                                 [°oh°
10        °sch(h)o(n) muss ich spr(h)echen.°
          (already)  I  have  to  speak
```

Figure 3: #3, l. 4

Figure 4: #4, l. 16

```
11         (0.3)
12   BEN  dass er, (0.5) [sich:-
          that he  (0.5) [can
13   THI                 [°schon muss [ich  sprech[en.°
                         [(already) I [have to spe[ak
14   SPth                             [((beep2))
15   BEN                                          [auf ihre sprache
                                                  [prepare for your
16        einrichten +kann. (0.5)    +also mikro#4fon  an, und dann;
          speech             (0.5)    so (turn) on the mic and then
     ben             +rHand t/display+activates microphone
```

```
17        (.) die frage stellen,(.) und das ganze 'okay google;
          (.) ask  a   question (.) and then this 'okay google
18        [zeige [mir;' 'okay google; was ist,' et cetera.
          [show  [me'   'okay google what is'   et cetera
19   SPBe [((beep1))
20   SPx         [((beep1))
21        (0.5)
22   BEN  jetzt hat er bei mir natürlich [reagiert hier,]
          now   in my case  of course    [(he) reacted  ]
23   SPbe                                [von     wikipe]dia; (.) der
                                         [from    wikipe]dia  (.) the
24        [lateinische ausdruck et cetera, wird zur abkürzung einer&
          [Latin     expression et ecetera is (used) to abbreviate an&
25   SPx  [((beep2))
26   SPbe &aufzählung verwendet, und bedeutet wörtlich,
          &enumeration           and literally   means
27        und die üb[rigen dinge.
          and the   [other things
28   ZAN            [th: °ich glaubs nich;°
                    [th: I can't believe it
```

BEN's demonstration of the IPA provides a variation of the different steps required to operate it: activating the mic, saying the wake words *okay google*, and then formulating the query. Notably, however, he inverts the order of the two latter steps: while stating how to activate the microphone by tapping on the symbol (l.1), he points to the widget on his phone (Figure 2). As the second step of the demonstration, he then mentions the formulation of the query *als Frage* "as a question" (l.2), only afterwards mentioning the wake word as the way to start the request (l.3, 5–6) so that the device (*er* "he", l.7) can "prepare for your speech" (l.7, 12, 15–16), thus providing a lay explanation of the wake word's technical significance. Indeed, BEN then condenses the procedure of using the IPA with the summary: "So turn on the mic and then ask a question" (l.16-17), mentioning the wake word only indirectly in the following two generalized and incomplete example queries "and then this 'okay google show me' 'okay google what is' et cetera" (l.17-18). Thus, he does not emphasize the required precise order of the steps to be taken nor the technical importance of the wake word (or of the activation in general) and of the timing, nor that the IPA can be activated by either voice or touch (but does not actually need both), all of which missing details might make it difficult to understand how to successfully use the IPA. Moreover, BEN's initial demonstration is interspersed with some responses from the group, leading to suspensions and breaks in the production of this multi-unit turn.

The first participant to respond is KLE, who constantly looks down at her device and notes during BEN's demonstration (see Figure 3), a posture she maintains throughout and beyond the duration of the excerpt. Her possibly distracted reception of BEN's demonstration of the IPA is thus limited to the audible. Perhaps unsurprisingly, then, her early response to BEN's turn (l.4), indicates that her initial understanding of the IPA's function is somewhat over-simplified (see the change-of-state token *achso*, Heritage 1984, Golato and Betz 2008): searching by simply speaking. Proceeding with his example query, BEN unintentionally activates the IPA on THI's smartphone. The fact that her device respond only to the wake word (which publicly shows that the touch activation is an alternative, not a compulsory first step) is not commented on by BEN. This prompts THI's phone to produce the on/off sound (l.8, 14), which THI comments on, thus incidentally displaying the knowledge she already has about IPAs ("Oh I already have to speak", l.9-10, 13). These overlapping audible actions lead to major perturbations in BEN's turn production, but he nevertheless proceeds with his condensed explanation[5]. Finally, by tapping on the mic (l.16, Figure 4), uttering the wake word and a (double) question, BEN then accidentally triggers the IPA on his own phone (l.19, see his comment l.22), leading to an audio output from the phone concerning the definition of the word "etcetera" (l.23-24, 26–27), which had only been used in order to mark the end of the list (Jefferson 1990) of possible queries BEN provided as part of his initial explanation. Participant ZAN assesses this demonstration as "unbelievable" (l.28), which, similar to KLE, displays her status as someone with no prior knowledge about IPAs. This clearly contrasts with THI's previous display of knowledge (also implied by her pre-emptive completion of BEN's turn in Excerpt 2, l.67; see Lerner 1996). The participants' speaking turns therefore reflect whether they are encountering the technology for the first time or already have some degree of familiarity with it. Indeed, in what follows, it is both KLE and ZAN who attempt to use Google Assistant on their own devices, whereas THI only does so to assist them in their initial, unsuccessful attempts.

5 Indeed, none of the other participants' devices are automatically triggered at this moment. As BEN does not formulate any guidelines regarding the timing for individually trying out the IPAs, he does not seem to expect major problems due to multiple overlapping queries. But his ensuing round to check up on each participant will show that the participants' IPAs have either been set to inappropriate settings (e.g., not yet set to enable voice input, or the correct language), or that the participants have not yet understood how to correctly activate and use them.

In the following 28 seconds (not presented here), BEN provides two fully-formulated examples of IPA-assisted searches: first, a search for pictures ("OK Google, show me pictures of elephants"), and second, asking for the opening times of a sightseeing attraction, which leads to an audio output. He then goes on to provide another summary of how the IPA functions (Excerpt 2A). During all this time, KLE continues to look down at her phone, an orientation she maintains when she requests assistance (l.69):

Excerpt 2A (190925_VHSB_001633_okgoogle)

```
65   BEN  also ich brauch das dann nicht (0.4) mehr; (.)
          so I no longer have           (0.4) (to)  (.)
     kle  >>-looks down at her SP-------------------->>
66        [reinzuschrei:]ben, (.) sondern (.) ich; (.)&
          [write it     ] in  (.) but     (.)  I   (.)&
67   THI  [reintippen:, ]
          [type it in   ]
68   BEN  &sp:r:eche es rein; (0.5)
          &say       it      (0.5)
69   KLE  und [wo find ick det 'okay'?   ]
          and [where do I find the 'okay']
70   BEN      [.h und   kriege  genau so ] die ergebnisse
              [.h and   (I) get the same ] results
71        wie bei  der normalen;
          as  with the normal
72   KLE  jetzt +sprechen?
          speak   now
     ben        +..looks to KLE->
73        (0.2)
74   BEN  suche; (.) 'okay, google;'+ und dann einfach lossprechen;
          search (.) 'okay  google'   and then just start talking
     ben  >gaze KLE-----------------+,,,
75        (.)
76   KLE  ich hab hier +keen-  find   keen   'okay:'
          I  don't have a-    can't find any 'okay'
     ben               +..gaze KLE->
77   THI  .H:
78        +(0.4)
     ben  +..walks twd KLE, stands next to THI/ZAN->
79   KLE  °det  [is-°
          that  [is-
80   THI        [nee einfach- (.) einfach [reinsprechen.      ]
                [no  just-    (.) just    [talk into (the mic)]
81   BEN                                  [NEIN. wenn sie auf ]
                                          [no    when you     ]
82        das  [mikrofo:n *gedrückt [haben; sagen sie-     ]
          have [pressed   the   micro[phone   you say-     ]
83   KLE       [achso:,             [ja: und jetzt soll ick] +sprechen.
               [oh okay             [yes and now  I have to]  speak
     ben  >gaze KLE-----------------------------------------+t/ZAN->
     zan  >gaze t/KLE-----*...gaze twd SP display-->>
     zan                  *..bends body closer twd her SP->l.89
```

```
84          (.)
85    BEN   [+°genau.°
            [  exactly
86    KLE   [su-  (.) °°su-°°
            [sea- (.)  sea-
      ben    +..gaze KLE-->
87          (0.2)
88    ZAN   °einfach+*%reinsprechen;°#5a/b[xxx kosten]los installieren;°&
             just talk into (the mic)     [xxx   down]load       for free&
89    KLE                                 [suche elefant. ]
                                          [search elephant]
      ben   >--KLE--+..gaze t/ZAN----------------------->
      zan           *..bends down, closer to SP display->l.112
      thi           %..gaze & bends t/SP&ZAN----------->>
90    ZAN   &°mh° (.) au weia; das is ja der reinste [wahnsinn;
            & mh  (.) oh dear  that's really         [insane
```

Figure 5a: #5a, l. 88

Figure 5b: #5b, l. 88

BEN concludes his introduction of the IPA by comparing it to a standard, text-based search: the query is not made by writing, but by speaking (l.65-66, 68) to obtain "the same results as with the normal search" (l.70-71, 74). This summary is again repeatedly suspended by KLE's requests for assistance to find the "Okay" (l.69), and asking whether she needs to "speak now" (l.72), which seems to display a more generalized trouble relating to the correct order of steps required to activate the IPA. While KLE's first request does not receive a response, BEN does shift his gaze to her (l.72) and replies to her second turn by providing yet another version of the summarized procedure: "'Okay Google' and then just start talking" (l.74). This is not sufficient for KLE, who then reformulates her first question in the form of a declarative, stating the absence of an "Okay" in the widget (l.76). With her gaze steadily fixed on the screen and her posture bent over her smartphone, this declarative aims at mobilizing the course leader's assistance, and is apparently successful – BEN moves from the front desk toward KLE (Oloff 2023). He does not, however, go to KLE and inspect her device, but remains standing in front of the table where THI and ZAN are seated (l. 78, cf. Fig. 5a). THI, after an initial hesitation (l.77) and possibly waiting for BEN to respond, now self-selects with another instruction for KLE ("just talk into (the mic)", l.80). In overlap, BEN provides his own response, finally mentioning the need to tap the mic symbol first (l.81-82), which is a way of triggering the IPA if it has not yet been activated or set to operate via voice activation[6]. Interestingly, KLE's initial misconception (that "Okay" corresponds to some kind of button to press) is not explicitly corrected, but only disconfirmed as non-relevant (see the initial negative response tokens in l.80-81). Nevertheless, KLE publicly displays understanding with *achso*, and, sequentially well placed as a second step following the first part of BEN's response, reformulates her prior request for confirmation (l.72) as an instruction directed to herself ("yes: and now I have to speak", l.83). Simultaneously with BEN's confirmative closing third (l.85), KLE now tries to initiate a query on her device by imitating one of BEN's prior demo queries, albeit in a simplified form ("sea- sea- search elephant", l.86, 89). Despite having previously displayed possible understanding through the change-of-state token *achso*, she neither activates the mic nor verbalizes the wake word, so her IPA is not activated and this attempt will fail.

BEN, who has monitored KLE beyond the sequence closing (l.86-88), now refocuses his attention on ZAN, most likely prompted by her self-selection (Fig. 5a/b, l.88). Her embodied display of trouble (bending down and looking at her smartphone lying in front of her, Kendrick and Drew 2016), the muttered repeat of THI's previous instruction to KLE (l.80), reading aloud a message from the screen, the trouble alert (ibid., *auweia* "oh dear", l.90) and the final assessment ("that's really insane", l.90) clearly display that she is struggling to use the IPA on her own. This way, ZAN successfully mobilizes both BEN's and THI's assistance, who will now guide her in her first attempt to operate the Google Assistant:

6 See the instructions provided here (accessed on 10/08/2024): https://support.google.com/assistant/answer/7172657?sjid=17825161593263064971-EU&hl=en.

Excerpt 2B (190925_VHSB_001633_okgoogle, continuation of Excerpt 2A)

```
90   ZAN  &°mh° (.) auweia; das is ja der reinste [wahnsinn;
          & mh  (.) oh dear  that's really        [insane
91   KLE                                          [°gibt nur keenen
                                                  [ but there's no
92        (0.2)+gibt keenen ele[fant°
          (0.2) there's no  ele[phant
93   THI                       [einfach-
                               [just
     ben      +..puts on glasses & comes to ZAN's place->
94        (0.3)
95   BEN  .h::
96   ZAN  hier;
          here
97   KLE  °°da kommt keen elefant.°°
            there is no elephant
98        (0.8)
99   BEN  +°so;° (und) das #6 antip[pen, (.)+&
           alright (and)      tap  [here (.) &
100  ZAN                           [ja,
                                   [yes
     ben  +right index ppp to SP ZAN--------+,,,
101  BEN  &*sagen; 'okay, google?' und dann: die [frage.+
          &   say  'okay google'   and then the  [question
102  SPza                                        [((beep1))
     zan   *activates microphone on SP by tap
     ben  >stands & leans to ZAN------------------------+,,,
103       (0.2)
104  ZAN  okay:,
105       (0.6)
106  THI  google;
107       (0.4)
108  ZAN  google.
109       (0.5)
110  BEN  #7und die frage,
            and the question
111       (0.3)
112  ZAN  .h[:: äh::; [wie +komme ich-       *ach]so:;&
          .h[:: uhm:: [how  do I get          oh ] okay&
113  KLE    [°ach ick muss sa:gen; 'okay google'°]
            [ oh  I have   to say  'okay google' ]
114  SPza              [((beep2))
     zan  >bent over SP display--------------*,,,straigthens up->
     ben  >in front of ZAN-+,,, starts turning & walks back t/desk->>
115  KLE   [okay, google; (.) ich suche den elefant;      ]
           [okay  google  (.) I'm looking for the elephant]
116  ZAN  &[(0.4) °ja° *(.)   welche  öff: (.) nungszeiten] hat äh::::;
          &[(0.4)  yes  (.)  what are the o(.)pening hours](of) uh:::m
     zan  >straight----*...bends back to SP display->
117  THI  °nee sie- sie-° die frage müssen sie schon vorher wissen;
           no  you  you   have to know the question beforehand
118       des: [.h der-  .h:] und dann fangen sie einfach&
          that [.h the-  .h:] and then     you just start&
119  ZAN       [*he, he, he;]
     zan  >-----*..leans back, away from SP->>
     zan  >gaze SP*,,,,
120  THI  &noch mal von vorne an;
          &again from the beginning;
```

Figure 6: #6, l. 99

Figure 7: #7, l. 110

BEN visibly prepares to help ZAN by putting on his glasses and moving closer to her (l.92-93). While both THI and BEN project (longer) turns (l.93, 95), they wait for ZAN to initiate the trial on her own, which she does by hovering with her index finger over the widget on her smartphone screen (l.96, cf. Figure 6). As she does not then tap the mic icon, BEN bends closer to her smartphone, points to the correct area of the screen (Figure 6) and reissues an instruction, this time in the correct order (l.99, 101). At the same time, ZAN activates the mic by tapping, which is registered by the device's activation sound (l.102). ZAN then proceeds by saying *Okay*, but does not continue (l.104-105). THI, seated next to ZAN, prompts her by saying the missing part of the wake formula (*Google*), which ZAN hesitantly repeats before stopping again (l.106-109)[7]. BEN, who has meanwhile re-straightened his posture (Figure 7), thereby treating ZAN's problem as a *learnable* (and not something he will solve by manipulating her device, Råman 2022), now urges her to provide a question (l.110). ZAN initiates a possible request for directions ("how do I get-", l.112), interrupts herself and attempts a different query ("what are the opening hours of-", l.116), echoing one of BEN's initial demo queries. She then abandons this syntactic construction as well, apparently realizing that she has not yet prepared an appropriate query. By now, her IPA is no longer activated (l.114). THI offers a suggestion (i.e., that the question should be known in advance, l.117) and

7 Indeed, both BEN and THI seem to orient to a redundant activation practice (tapping the mic and saying the wake formula), which seems probably easier than to immediately instruct two different ways (either touch or voice activation), and less time-consuming than to check the settings on all the participants' devices.

encourages ZAN, who has momentarily disengaged from her smartphone, to start again (l.118-120).

This assistance sequence is overheard by KLE, who had previously realized that her first attempt had failed, as "there's no elephant" (i.e., pictures thereof) to be seen on her phone (l.91-92, 97). Concurrently with ZAN's first attempt, KLE displays understanding of how to proceed, affirming that she now understands that *Okay Google* has to be voiced as well (l.113). This leads her to produce a formally correct query (l.115) which nonetheless fails as the language of her IPA seems to be set to English – a problem that will later be resolved by BEN.

The excerpts in this section have illustrated that even detailed demonstrations of how to use a new application do not always lead to its immediately successful implementation by course participants. This could be partly due to a certain variation in the explanation: the order and exact effects of the three steps (activating the mic by tapping, uttering the wake formula, and verbalizing a query) were not consistently presented in exactly the same way, or to the heterogeneity of types of queries demonstrated (both syntactically, i.e., including both directives or WH-interrogatives; and in terms of output, i.e., voice, images, or text/display output). Here, the ways the participants respond to and comment on BEN's introduction reveal different levels of (non-)expertise regarding IPAs. While those (such as THI in this case) who already have a degree of familiarity with a specific app or technology can even assist their peers, the 'newcomers' mobilize assistance through various audible and visible displays of trouble. This leads to customized instructional sequences and extended monitoring of the participants' (here, KLE and ZAN) trials. While the non-expert participants audibly display some understanding about how the IPA functions from very early on in the demonstration, the lack of success of their initial attempts to use it indicates that first-time users of IPAs might benefit from supplementary basic technical information, such as the necessity of the wake word (or, more generally, of the need to *activate* the IPA), the timing of the query formulation, and the significance of the on/off sounds. Beginning the demonstration by outlining some of the benefits of using an IPA might also contribute to faster success in using it, or, at least, increase participants' motivation to try out the IPA on their own. In the next section, I discuss the role of discovering individual benefits to be gained by adopting and accepting a new technology.

4.2 Discovering IPAs and assessing their value for non-expert users

The analysis in this section will offer a reflection on chance encounters with new applications or functions, and examine the process of discovering a 'new technology'. In the example I draw upon here, the IPA is not introduced as part of the course, but discovered when a participant inadvertently activates it on her own device. The excerpt is taken from the second meeting of a two-part course (2 x 1.5 hours); the eight participants, who are seated in a row opposite the course instructor, have previously been instructed how to use Google Maps. They are now supposed to type the name of a place or location into the search bar to explore the app. Instructor JUN's final example as part of this instructional sequence is the Eiffel tower in Paris (Excerpt 3A, l.1-2). During this turn, participant MEF, seated to the left end of the row, unintentionally activates the IPA integrated in the Google Maps app of her phone. Because the excerpt included numerous turns by other participants and the course leader's interaction with them, especially at the beginning, the transcript presented here has been simplified to focus on the exchange between MEF and FIS, who is seated next to her.

Excerpt 3A (190919_NOS_010910)

Figure 8: #8

Figure 9: #9, l. 4

Figure 10: #10, l. 6

```
1     JUN  >genau so können sie jetzt auch sagen< (0.3) ich h(h)ab ja
            exactly and so you could also say     (0.3) I have
2          schon immer mal von de:m; (0.5) pari:ser eiffelturm gehört;
           always  heard about the    (0.5)     Eiffel tower in Paris
3          (0.25)*(0.6)
      mef        *..rHand to SP, changes position of SP->
4     JUN  *aber noch *#9nie *gesehen.
            but I've  never seen it
      mef  *...gaze t/watch--*...gaze t/left------>
      mef                  *right thumb pushes border SP
      SPmf >>google maps on screen->
5          (0.1)%(0.3)
      SPmf  >---%...pop-up voice assistant on screen->
6     SPMf ((beep1))*#10
      mef            *..gaze SP->
7          (1.0)
8     MEF  hUch
           oh dear
9          (0.7)
10    MEF  'sAg etwas'
           'say something'
11         (0.9)
12    SPmf £((beep2))
      fis  £..gaze t/SP MEF->
```

Figure 11: #11, l. 14

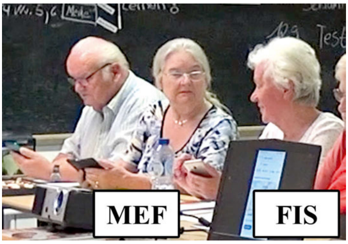


Figure 12: #12, l. 18

```
13          (0.75--)*(0.1)%(0.5)
      mef   >gaze SP*...gaze FIS->
      SPmf                  %screen changes
14    FIS   hem,#11
15          (0.5)
16    MEF   .H; *.he,
      mef   >FIS*,,,gaze SP->
17          (0.9)
18    FIS   der spricht ja auch mit £ihnen; (0.3) wenn sie_s-#12
                      he also talks to you; (0.3) if   you s-
      fis   >SP MEF-----------------£..gaze MEF->
19    MEF   %ja:,:;=
             yeah=
      SPmf %,,,blue pop-up beam on display disappears
```

```
20   FIS  =wenn sie zum +beispiel *ins +internet gehen,&
          =if       for example you go on the internet &
     mef  >gaze SP---------------*...gaze FIS->
     jun                 +...walks back t/center of room->
     jun                               +..gaze t/FIS&MEF>
21   FIS  &und wollen nich eintippen     (0.3) was sie wissen wollen,
          &and you don't want to type in (0.3) what you want to know
22        (0.5)+können sie den *fragen (0.2) und dann kommt&
          (0.5) you can ask him        (0.2) and then you get&
     mef  >gaze FIS------------*,,,gaze SP->
     jun  >----+stands in front of board, gaze t/FIS&MEF->
23        &die antwort auch °über-° (0.6) per stimme;+#13
          &the answer (also) over-  (0.6) by   voice
     jun                                             +nods
24        (0.8)
25   MEF  ha,ha,ha,ha;
26   JUN  da geb ich ihnen recht.
          I do agree with you
```

Figure 13: #13, l. 23

As he formulates his final example of a list of several possible places to look up with the Maps app, course leader JUN is standing behind the desk in the center of the classroom. When one of the participants self-selects in overlap and formulates a complaint (not shown in the transcript), JUN moves to the right end of the row to assist that person, a position he will remain in until line 20. This might explain why he is so late to comment on MEF’s discovery of the IPA (l.26 and further). Before and in the beginning of the excerpt, MEF is looking at her smartphone, viewing the app interface in order to follow JUN’s instruction (to insert the name of a location in the search bar, cf. Figure 8). During JUN’s turn suspension (l.3), she slightly adjusts the position of her smartphone (that, due to a protective double case, can be held like a book) more towards her right hand, and then looks at her watch on her left wrist (l.4, fig. 9). With this movement, her right thumb apparently briefly touches the microphone icon in the Google Maps search bar (see the black circle in Figure 8), triggering a pop-up message on the screen (l.5, cf. Figure 10) and the IPA's activation sound (l.6). MEF, who had looked up to the left in the meantime, shifts her gaze back to her smartphone display and perceives the modified interface, to which she responds with the high-pitched response cry *huch* (l.8, Goffman 1981), a formulaic interjection expressing "surprise" that works as a possible public trouble alert (Kendrick and Drew 2016). She then reads aloud the IPA's initial message on her screen *sag etwas* (l.10), but does not seem to understand it as a prompt to actually "say something" to the IPA at this moment. Shortly thereafter, the "off" sound indicates the IPA's deactivation (l.12), apparently accompanied by a visible notification on the screen (l.13). FIS, who is sitting next to MEF and probably overheard the surprise discovery, now turns her gaze to MEF's smartphone (l.12) and produces a laughter particle (l.14). MEF minimally responds to this by shifting her gaze to FIS and reciprocating the laughter (l.13, 16, Figure 11). She does not take this opportunity to initiate topical talk or to formulate an assessment, possibly because she does not know the exact meaning of the visual and audio notifications presented by her device, which is also suggested by her gaze returning to the screen immediately afterwards (l.16, cf. Figure 12). MEF's lack of understanding seems to have been anticipated by FIS, who then provides a basic explanation of the IPA, shifting her gaze from MEF's smartphone to MEF's face (l.18, fig. 12). Notably, FIS does not offer instructions on how to make the IPA work, but instead describes the general functioning and advantages of the app, i.e., that queries don't have to be typed and that answers are provided via audio output (l.18, 20–23). MEF initially responds in a possibly disengaged way (her stretched "yeah", l.19, suggests acknowledgement of having heard more than it displays comprehension), but she then gazes back at FIS (l.20) before returning to her device (l.22). She then produces more laughter particles and finally formulates an assessment that evaluates the IPA's functioning as something "nice" (l.27, Excerpt 3B):

Excerpt 3B (190919_NOS_010910, continuation of Excerpt 3A)

```
25   MEF  ha,ha,ha,ha;
26   JUN  da geb ich ihnen recht.
          I do agree with you
```

```
27   MEF  £das ist gu:t.
            that's nice
     fis  £..gaze JUN->
28        (0.2)
29   MEF  'nein [danke,'  'aktivieren;'(nun muss-) achso,]+ dann muss&
          'no   [thank you''activate'(now I have-) oh okay] then I have&
30   JUN        [für die SCHREIBfaulen mal ganz gut; ehe,he,he,he,]
                [for the ones too lazy to write that's good ehe,he,he,he,]
     jun  >monitors FIS/MEF----------------------------+.back t/desk>
31   MEF  &ich das aktivieren und dann, .h [*xx-
          &to activate it     and then  .h [ xx-
32   JUN                                   [+>dann kann man auch-<&
                                           [  then you  can also  &
     mef  >gaze SP-------------------------*...gaze JUN->
     jun  >gaze downwards / desk------------+...gaze MEF->
33        &(0.2)#14 t- +REIN[sprechen=also heut-= hier heute abend]&
          &(0.2)    t- speak[ into(the mic)=but today=here tonight]&
34   MEF                    [dann brauch ich nicht schreiben.     ]
                            [then I don't have to write           ]
     jun               +..walks twd MEF-------->
35   JUN  &empfehl ichs nich unbedingt=
          &I would not really recommend it=
36        =[weil wenn wir alle acht     REIN]sprechen (0.4) eH: (0.3)&
          =[because if the eight of us speak] into(the mic)(0.4)eH(0.3)&
37   MEF   [dann ist doch einfacher;        ]
           [then it is easier               ]
38   JUN  &[aber- (.) gerne auch üben.          ]
          &[but   (.) you are welcome to practice]
39   MEF   [dann brauch ich nicht schreiben,    ] dann kann ich
           [then I don't have to write          ] then I can
40        einfach sprech[en;  ] ne,
          just       spe[ak   ] right
41   JUN                [+ja; ]
                        [ yes ]
     jun                 +..steps back from MEF
42        (1.0)*(0.2)
     mef  >JUN-*,,,gaze SP->>
```

MEF then reads again from her phone's screen ("activate", l.29), and, after having produced the change-of-state token *achso*, concludes: "then I don't have to write", l.34, "then it's easier", l.37, "then I don't have to write then I can just speak, right", l.39-40). She thereby formulates in her own words what FIS has just told her and draws her own conclusions about the IPA's usefulness. JUN, who had been walking back to the center of the room a bit earlier, first overhears (see his gaze orientation and nod, l.20, 22–23 cf. Figure 13) and then joins the conversation between FIS and MEF (l.26). He then proceeds to somewhat diminish the attractiveness of the IPA by jokingly saying that it is for people who are "too lazy to write" (l.30). Instead of simply confirming MEF's tentative understanding, i.e., that she can "speak into the device" after activating it (l.29, 31), JUN remarks that trying out the IPA during the ongoing course could lead to problems due to the size of the group (l.33, 35–36), which is why he recommends MEF should try it at a different time (l.38)[8]. Interestingly, most of MEF's reasoning seems to be formulated independently of the course leader's argumentation, although JUN and MEF have meanwhile engaged in mutual orientation and gaze (l.32, cf. Figure 14). As course leader, JUN orients to the institutional tasks of limiting possible digressions from the course content and of arranging adequate learning conditions (i.e., without overlapping noise), which he also continues beyond the excerpt shown here. He thus also addresses the other course participants, not only MEF. MEF, as a course participant, does not respond to JUN's objections, but instead elaborates on her assessment of the usefulness of the IPA, which her husband, sitting to her right, finally acknowledges as well by announcing that they will use the IPA in the future (not shown here).

Figure 14: #14, l. 33

Although we do not know for sure whether, prior to this excerpt, MEF knew anything about how IPAs function, the way she responds to this chance en-

8 While the other course leader, BEN, did not treat the possible concurrent use of several IPAs in one room as a problem, JUN claims that it is a reason not to do so. This might be less related to an actual (and so far unencountered) problem of concurrent acoustic signals, but rather that this might distract from the current course agenda, namely exploring the Maps app.

counter with the Google Assistant integrated in the Maps app strongly suggests that she did not. Her careful inspection of the screen, reading the text aloud, and her minimal answers to the IPA's prompts indicate that this might very well be her first encounter with this type of IPA, or even with any IPA at all. Furthermore, FIS' explanation is not met by MEF with an already-knowing stance, but instead responded to with an assessment, and then, when requesting feedback from course instructor BEN, by an independent public appraisal of the IPA's potential usefulness. The data collected does not reveal whether MEF and/or her husband subsequently used the IPA again or attempted to do so. However, within the excerpt they are quick to assess the IPA as a potentially useful application to learn to use (e.g., l.34, 39–40), even though it had not been explicitly introduced in the course.

5. Conclusion

As the analyses of the instructed and the chance encounter with a smartphone-based IPA have shown, overall, participants in the present data set reacted with interest to the discovery of the hitherto unknown application. Within these two basic IT skills courses, neither the course leaders nor the participants mentioned the potential risks associated with IPAs relating to data collection and privacy that are frequently discussed in the media. On the contrary, participants' occasional assessments such as "this is spooky" rather expressed a general fascination regarding the functionalities and "omniscience" of the IPA. For the course leaders, avoiding the topic of data security might have been primarily a pragmatic decision because of insufficient time to address a complex and potentially controversial topic; for the participants, this absent critical perspective might be due to a lack of knowledge and/or display a sense of trust in the course leader and the specific institutional framework.

When encountering the IPA, the course participants' first speaking turns reveal their (un)knowing stance: first-time encounters are met, for example, with expressions of surprise, response cries, and assessments, whereas participants who already know the IPA instead display their expertise, for example, by pre-emptively completing the course leader's turns or by self-initiating offers of advice to their peers. Excerpts 1 and 2 illustrate that newcomers to the application display certain difficulties when they immediately try to use the IPA on their own following instruction, despite or perhaps exacerbated by the quite elaborate information from the course leader and his three different demonstration queries. More specifically, the IPA newcomers display difficulty in immediately recognizing the significance of the wake formula, what it triggers (or, more generally, why and how an IPA has to be activated), and the overall importance of the temporal order and formulation of the query. A particular challenge seems to be the production of a specific way of speaking *for* and *to* the machine, both with respect to action formation (e.g., what and how does the IPA process what its user is saying, cf. Reeves 2017) and with respect to the need to verbalize in a planned, appropriately timed, and orderly way (unlike spontaneous and emergent conversational requests). While it might intuitively be a sensible choice to skip more technical explanations when introducing a new application to non-experts, relating the technical constraints of the application more explicitly to its use could possibly lead to a faster successful implementation of the task. Indeed, the participants' incremental formulations not only publicly display the process of their appropriation of the IPA (which can then be assisted), but also show their reasoning as they break down the handling of the IPA into a machine-like step-by-step procedure, but tend not to take into account the required timing and planning – a frequent reason for unsuccessful first queries. Rather than presenting newcomers with a black-box technology, a more technically-oriented instruction could support them to try out the IPA with more autonomy.

Another (unplanned) approach was illustrated in Excerpt 3, in which a peer described to her fellow participant how the IPA functions and how it differs from script-based searches in potentially useful ways. While the course setting, with the app not featured as a topic, prevented the newcomer from exploring the app immediately, her independent appraisal of the IPA illustrated her intuitive grasp of its hitherto unknown affordances, which she evidently deemed to be useful. When introducing a new technology to non-experts, it might thus be advantageous to initially underline why it might be of interest and what specific benefits it could offer for the participants addressed. Complemented by more technically-oriented cause–effect instructions, this could present the use of the IPA as being fundamentally a matter of personal choice, and therefore foster the non-experts' agency from the start. As first-time encounters with a new technology can be formative in encouraging participants' acceptance and later engagement with it, the way its use and functioning are presented in educational contexts should be critically assessed in further empirical and qualitative studies. Future research could, for instance, explore how, within an institutional setting, participants can be guided to use a previously unknown device or application such as an IPA by directly relating instructions regarding practical handling to its technical constraints and characteristics, and by actively encouraging and structuring individual processes of discovery to identify its advantages and disadvantages, for example, by working in peer dyads.

Acknowledgements

"Smart Communication: The situated practices of mobile technology and digital literacies" (2019–2023), project funded by the Academy of Finland, project number 323848.

"Navigating the digital everyday life: Elderly participants' use of mundane technologies in and for social interaction (DigiLife)" (2022–2026), project funded by the Leibniz Association, Germany, as part of the "Leibniz Program for Women Professors".

References

Albert, Saul, Magnus Hamann, and Elizabeth Stokoe. 2023. "Conversational User Interfaces in Smart Homecare Interactions: A Conversation Analytic Case Study." In *Proceedings of the 5th International Conference on Conversational User Interfaces*, edited by Minha Lee et al., 1–12. New York: ACM. https://doi.org/10.1145/3571884.3597140.

Antaki, Charles, and Ray Wilkinson. 2012. "Conversation analysis and the study of atypical populations." In *The Handbook of Conversation Analysis,* edited by Jack Sidnell, and Tanya Stivers, 533–50. Oxford: Wiley-Blackwell.

Backhaus, Peter. 2013. *Communication in Elderly Care: Cross-Cultural Perspectives*. London: Bloomsbury.

Barthel, Matthias, Henrike Helmer, and Silke Reineke. 2023. "First Users' Interactions with Voice-Controlled Virtual Assistants: A Micro-Longitudinal Corpus Study." In *Proceedings of the 27th Workshop On the Semantics and Pragmatics of Dialogue*, edited by Andy Lücking, Chiara Mazzocconi, and Darinka Verdonik, 105–17. Maribor: University of Maribor.

Bergmann, Jörg R. 1981. "Ethnomethodologische Konversationsanalyse." In *Dialogforschung*, edited by Peter Schröder, and Hugo Steger, 9–51. Düsseldorf: Pädgogischer Verlag Schwann.

Boden, Deirdre, and Denise D. Bielby. 1983. "The Past as Resource: A Conversational Analysis of Elderly Talk." *Human Development* 26 (6): 308–19. https://doi.org/10.1159/000272892.

Boxer, Diana. 2018. "Senior Confessions. Narratives of Self-Disclosure." In *Language Learning, Discourse and Cognition: Studies in the Tradition of Andrea Tyler*, edited by Lucy Pickering, and Vyvyan Evans, 63–82. Amsterdam/Philadelphia: John Benjamins.

Carros, Felix, Johanna Meurer, Diana Löffler, David Unbehaun, Sarah Matthies, Inga Koch, Rainer Wieching, Dave Randall, Marc Hassenzahl, and Volker Wulf. 2020. "Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. Honolulu: ACM. https://doi.org/10.1145/3313831.3376402.

Coupland, Nikolas, Justine Coupland, and Howard Giles. 1991. *Language, Society and the Elderly: Discourse, Identity and Ageing.* Basil Blackwell.

Czaja, Sara J., Neil Charness, Arthur D. Fisk, Christopher Hertzog, Sankaran N. Nair, Wendy A. Rogers, and Joseph Sharit. 2006. "Factors Predicting the Use of Technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE)." *Psychology and Aging* 21 (2): 333–52. https://doi.org/10.1037/0882-7974.21.2.333.

Desai, Smit, and Jessie Chin. 2023. "OK Google, Let's Learn: Using Voice User Interfaces for Informal Self-Regulated Learning of Health Topics among Younger and Older Adults." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21. Hamburg: ACM. https://doi.org/10.1145/3544548.3581507.

Domínguez-Rué, Emma, and Linda Nierling. 2016. *Ageing and Technology: Perspectives from the Social Sciences*. Science Studies. Bielefeld: transcript.

Fiehler, Reinhard, and Caja Thimm, eds. 2003. *Sprache und Kommunikation im Alter*. Radolfzell: Verlag für Gesprächsforschung.

Fischer, Joel E., Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. "Progressivity for Voice Interface Design." In *Proceedings of the 1st International Conference on Conversational User Interfaces*, 1–8. Dublin: ACM. https://doi.org/10.1145/3342775.3342788.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs: Prentice-Hall.

Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Golato, Andrea, and Emma Betz. 2008. "German Ach and Achso in Repair Uptake: Resources to Sustain or Remove Epistemic Asymmetry." *Zeitschrift für Sprachwissenschaft* 27: 7–37.

Goodwin, Charles, 2003. *Conversation and Brain Damage*. Oxford University Press.

Habscheid, Stephan, Felix Carros, Jens Lüssem, and Christine Hrncal. 2020. "Professionelle Emotionalität und humanoide Robotik in der institutionellen Kommunikation." In *Institutionelle und organisationale Kommunikation: Theorie, Methodologie, Empirie und Kritik*, edited by Helmut Gruber, Jürgen Spitzmüller, and Rudolf De Cilla, 169–88. Göttingen: Vienna University Press.

Habscheid, Stephan, Tim Hector, and Christine Hrncal. 2023. "Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology." *Social Interaction. Video-Based Studies of Human Sociality* 6 (1). https://doi.org/10.7146/si.v6i1.137378.

Habscheid, Stephan, Tim Moritz Hector, Christine Hrncal, and David Waldecker. 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. https://doi.org/10.21248/jfml.2021.44.

Hector, Tim Moritz, and Christine Hrncal. 2020. *Intelligente Persönliche Assistenten im häuslichen Umfeld: Erkenntnisse aus einer linguistischen Pilotstudie zur Erhebung audiovisueller Interaktionsdaten*. Siegen: Working paper series / SFB 1187 Medien der Kooperation. http://dx.doi.org/10.25819/ubsi/1013.

Heritage, John. 1984. "A Change-of-State Token and Aspects of Its Sequential Placement." In *Structures of Social Action*, edited by J. Maxwell Atkinson, and John Heritage, 299–45. Cambridge: Cambridge University Press.

Hrncal, Christine, and Katharina Hofius. 2023. "Digitale Kommunikation im Alter." In *Interaktion und Kommunikation im Alter: Interdisziplinäre Forschungsperspektiven*, edited by Rafael Mollenhauer, and Christian Meier zu Verl, 119–48. Velbrück Wissenschaft. https://doi.org/10.5771/9783748915096.

Jefferson, Gail. 2004. "Glossary of Transcript Symbols with an Introduction." In *Conversation Analysis. Studies from the First Generation*, edited by Gene H. Lerner, 13–31. Amsterdam/Philadelphia: John Benjamins.

Jefferson, Gail. 1990. "List construction as a task and resource." In *Interactional Competence*, edited by George Psathas, 63–92. New York: Irvington.

Kendrick, Kobin H., and Paul Drew. 2016. “Recruitment: offers, requests, and the organization of assistance in interaction.” *Research on Language and Social Interaction* 49 (1): 1–19.

Lahikainen, Anja Riita, Tiina Mälkiä, and Katja Repo. 2017. *Media, Family Interaction and the Digitalization of Childhood*. Cheltenham: Edward Elgar.

Lerner, Gene H. 1996. “On the ‘Semi-Permeable’ Character of Grammatical Units in Conversation: Conditional Entry into the Turn Space of Another Speaker.” In *Interaction and Grammar*, edited by Elinor Ochs, Emanuel Schegloff, and Sandra A. Thompson, 238–76. Cambridge: Cambridge University Press.

Lindström, Anna. 2005. “Language as Social Action: A Study of How Senior Citizens Request Assistance with Practical Tasks in the Swedish Home Help Service.” In *Syntax and Lexis in Conversation: Studies of the use of linguistic resources in t alk-in-interaction*, edited by Auli Hakulinen, and Margret Selting, 209–30. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.17.11lin.

Mollenhauer, Rafael, and Christian Meier zu Verl, eds. 2023. *Interaktion und Kommunikation im Alter: Interdisziplinäre Forschungsperspektiven*. Velbrück Wissenschaft. https://doi.org/10.5771/9783748915096.

Mondada, Lorenza. 2016. “Challenges of Multimodality: Language and the Body in Social Interaction.” *Journal of Sociolinguistics* 20 (3): 336–66. https://doi.org/10.1111/josl.1_12177.

Mondada, Lorenza. 2018. “Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality.” *Research on Language and Social Interaction* 51 (1): 85–106. https://doi.org/10.1080/08351813.2018.1413878.

mpfs – Medienpädagogischer Forschungsverbund Südwest. 2021. “SIM-Studie 2021.” https://www.mpfs.de/studien/sim-studie/2021/.

Nussbaum, Jon F., and Justine Coupland. 2004. *Handbook of Communication and Aging Research*. Routledge.

Oloff, Florence. 2023. “‘Maintenant tout a disparu’: les requêtes indirectes déclaratives dans la formation numérique pour adultes.” *Langage et Société* 2023/2 (179): 83–110. https://doi.org/10.3917/ls.179.0082.

Oloff, Florence. 2021a. “Some systematic aspects of self-initiated mobile device use in face-to-face encounters.” *Journal für Medienlinguistik* 2 (2): 195–235. https://doi.org/10.21248/jfml.2019.21.

Oloff, Florence. 2021b. “New technologies – new social conduct? A sequential and multimodal approach to smartphone use in face-to-face interaction.”

Bulletin VALS-ASLA. Special issue (1): 13–34. https://www.vals-asla.ch/fileadmin/user_upload/Journal/Special_2021_vol_1_def_A4.pdf.

Opfermann, Christiane, and Karola Pitsch. 2017. "Reprompts as Error Handling Strategy in Human-Agent-Dialog? User Responses to a System's Display of Non-Understanding." In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 310–16. Lisbon: IEEE. https://doi.org/10.1109/ROMAN.2017.8172319.

Peil, Corinna, and Jutta Röser. 2023. "Conceptualizing Re-Domestication. Theoretical Reflections and Empirical Findings to a Neglected Concept." In *The Routledge Handbook of Media and Technology Domestication*, edited by Maren Hartmann, 42–52. London/New York: Routledge. https://doi.org/10.4324/9781003265931-6.

Pino, Maribel, Mélodie Boulay, François Jouen, and Anne-Sophie Rigaud. 2015. "'Are We Ready for Robots That Care for Us?' Attitudes and Opinions of Older Adults toward Socially Assistive Robots." *Frontiers in Aging Neuroscience* 7 (July). https://doi.org/10.3389/fnagi.2015.00141.

Porcheron, Martin, Joel E. Fischer, and Sarah Sharples. 2017. "'Do Animals Have Accents?' Talking with Agents in Multi-Party Conversation." In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 207–19. https://doi.org/10.1145/2998181.2998298.

Porcheron, Martin, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. "Voice Interfaces in Everyday Life." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. Montreal: ACM. https://doi.org/10.1145/3173574.3174214.

Quan-Haase, Anabel, Kim Martin, and Kathleen Schreurs. 2016. "Interviews with Digital Seniors: ICT Use in the Context of Everyday Life." *Information, Communication & Society* 19 (5): 691–707. https://doi.org/10.1080/1369118X.2016.1140217.

Quan-Haase, Anabel, Carly Williams, Maria Kicevski, Isioma Elueze, and Barry Wellman. 2018. "Dividing the Grey Divide: Deconstructing Myths About Older Adults' Online Activities, Skills, and Attitudes." *American Behavioral Scientist* 62 (9): 1207–28. https://doi.org/10.1177/0002764218777572.

Råman, Joonas. 2022. "Multimodal Negotiation for the Right to Access Digital Devices Among Elderly Users and Teachers." In *Shaping the North Through Multimodal and Intermedial Interaction*, edited by Juha-Pekka Alarauhio, Tiina Räisänen, Jarkko Toikkanen, and Riikka Tumelius, 67–93. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-99104-3_4.

Reeves, Stuart. 2017. "Some conversational challenges of talking with machines." In *Talking with Conversational Agents in Collaborative Action*, Workshop at the 20th ACM conference on Computer Supported Cooperative Work and Social Computing. https://nottingham-repository.worktribe.com/output/844412.

Reeves, Stuart, and Martin Porcheron. 2023. "Conversational AI: Respecifying Participation as Regulation." In *The SAGE Handbook of Digital Society*, edited by William Housley, Adam Edwards, Roser Beneito-Montagut, and Richard Fitzgerald. Sage. https://doi.org/10.4135/9781529783193.n32.

Reeves, Stuart, Martin Porcheron, and Joel Fischer. 2018. "'This Is Not What We Wanted': Designing for Conversation with Voice Interfaces." *Interactions* 26 (1): 46–51. https://doi.org/10.1145/3296699.

Sacks, Harvey. 1984. "Notes on Methodology." In *Structures of Social Action. Studies in Conversational Analysis.*, edited by J. Maxwell Atkinson, and John Heritage, 21–27. Cambridge: Cambridge University Press.

Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction. A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.

Selwyn, Neil, Stephen Gorard, John Furlong, and Louise Madden. 2003. "Older Adults' Use of Information and Communications Technology in Everyday Life." *Ageing & Society* 23 (5): 561–82. https://doi.org/10.1017/S0144686X03001302.

Stigall, Brodrick, Jenny Waycott, Steven Baker, and Kelly Caine. 2019. "Older Adults' Perception and Use of Voice User Interfaces: A Preliminary Review of the Computing Literature." In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, 423–27. Fremantle WA: ACM. https://doi.org/10.1145/3369457.3369506.

Thimm, Caja. 2000. *Alter – Sprache – Geschlecht. Sprach- und Kommunikationswissenschaftliche Perspektiven auf das höhere Lebensalter.* Frankfurt/New York: Campus Verlag.

Velkovska, Julia, Moustafa Zouinar, and Clair-Antoine Veyrier. 2020. "Les Relations Aux Machines «conversationnelles»: Vivre Avec Les Assistants Vocaux à La Maison." *Réseaux* 220 (2): 47–79. https://doi.org/10.3917/res.220.0047.

Vincent, Jane. 2018. "Electronic Emotions, Age and the Life Course." In *Digital Technologies and Generational Identity. ICT Usage across the Life Course*, edited by Sakari Taipale, Terhi-Anna Wilska, and C. J Gilleard, 201–16. London/New York: Routledge.

Waldecker, David, and Tim Hector. 2023. "A Praxeological Approach Towards the Domestication of Connected Media Technologies: Introduction." *Digital Culture & Society* 9 (1): 5–22. https://doi.org/10.14361/dcs-2023-0102.

Wang, Hua, Renwen Zhang, and Barry Wellman. 2018. "Are Older Adults Networked Individuals? Insights from East Yorkers' Network Structure, Relational Autonomy, and Digital Media Use." *Information, Communication & Society* 21 (5): 681–96. https://doi.org/10.1080/1369118X.2018.1428659.

Weilenmann, Alexandra. 2010. "Learning to Text: An Interaction Analytic Study of How an Interaction Analytic Study of How Seniors Learn to Enter Text on Mobile Phones." In *Proceedings of the 28th International Conference on Human Factors in Computing Systems – CHI '10*, 1135–44. Atlanta: ACM. https://doi.org/10.1145/1753326.1753496.

Wiesemann, Jutta, Eisenmann, Clemens, Fürtig, Inka, Lange, Jochen, and Bina Elisabeth Mohn, eds. 2020. *Digitale Kindheiten.* Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-31725-6.

Wilkinson, Ray. 2019. "Atypical Interaction: Conversation Analysis and Communicative Impairments." *Research on Language and Social Interaction* 52 (3): 281–99.

Stylizing the Ideal User

Insights into the Experiences of Turkish-Speaking Voice Assistant Users in Germany

Didem Leblebici

Abstract *This chapter examines how Turkish-speaking voice assistant users stylize their English and German to mimic the 'ideal' user of Alexa and Siri. To date, little research has explored the experiences of multilingual individuals who use such technologies. Drawing from an ethnographically informed study with Turkish-speaking newcomers in Germany, this chapter offers a linguistic anthropological and sociolinguistic perspective on voice assistant use. I focus in particular on linguistic stylizations performed by participants during interviews, stylizations that index the strictly nationalized language constraints of popular voice assistants. Orienting to Portmann's (2022) study of how UX writers curate audiences by means of little texts, I argue that digital assistants are also tailored for specific addressees. The interviewees discursively constructed those audiences based on their own previous engagements with the technology. These ideas were reflected upon through their stylizing practices, which I analyze as a form of double-voicing (Bakhtin 1994; 1999; Rampton 2018). Participants often mocked the assumed audience of the technology and the voice assistant persona by performing an accent stylization of non-Western names. At other times, they adopted the standard variety and demonstrated a phenomenon of (non-)addressing the voice assistants as part of their narrative practice to avoid unwanted activation of the device. The insights of this research hold broader implications for the adoption and integration of voice technologies, particularly in multilingual or multiethnic settings.*

1. Introduction

> Birini aramam gerekiyordu, yani orada aramam gereken insanın Türkçe ismini Almanca Alman biri nasıl söylermiş gibi aksan yapıp kasıp onun ev numarasını kendi kendime aratmayı Siri'yle becerdim ve kendimle gurur duydum.
> *I needed to call someone, so I managed to use Siri to call the person's home number by speaking with a German accent, as if a German person was saying the Turkish name of the person I was supposed to call, and I was proud of myself.*

Digital technologies, especially AI chatbots and assistants reliant on voice recognition, present challenges for users who engage with them in non-standardized language varieties (e.g., Wu et al. 2020; Markl 2022; Koenecke et al. 2020). Many contemporary AI technologies seem to be constrained by boundaries associated with nation states and their official languages. As the account of the Siri user quoted above suggests, his voice assistant set to operate in German seemed to exclude people with non-German names, exemplifying a broader sociotechnical issue (see also Beneteau et al. 2019 for a Spanish–English case). In contrast to the essentialist perspective on language that has been integrated into voice assistants, contemporary sociolinguistic perspectives posit that language is an embodied, interactive, and dynamic activity that transcends geographical and political borders (also known as (trans)languaging, as discussed by Pennycook 2018; Li 2018; Cowley 2011).

To date, little research has been done to explore the experiences of multilingual speakers as they navigate and adapt voice assistant technologies in everyday domestic interactions. In this chapter, I focus on cases in which Turkish-speaking newcomers to Germany discursively echo and adapt the voice of their voice assistants – stationary smart speakers as well as assistants on smartphones – to mock, criticize, comment on, or align with some of the linguistic design constraints. Alignment in this context does not refer to users cooperating with the machine (as in Lotze 2016) but rather to speakers' discursive practices of self-presentation and positioning in relation to the technology design. The analysis specifically focuses on the stylization practices of the participants during the interviews, i.e., how they "produce specially marked and often exaggerated representations of languages, dialects, and styles that lie outside their own habitual repertoire" (Rampton 2009, 149). For instance, in the quote above, the interviewee describes deliberately stylizing his pronunciation of a non-German name. Self-reflexive performances of linguistic acts are particularly useful indicators of participants' subjective experiences and assessments, because they convey "metamessages" that shed light on interlocutors' ideologies and worldviews (Coupland 2001, 155). The analysis of these practices during the interviews is supported by ethnographically grounded research including participant observation and voice history data automatically generated by the Alexa app (also called "log data", as discussed by Habscheid et al. 2021).

The study aims to explore how multilingual users navigate, adapt, evaluate, and assess voice assistants that are inscribed with certain affordances and constraints. Specifically, it seeks insights into how multilingual and migrant speakers position themselves vis-à-vis voice assistant technologies that are designed according to an understanding of languages as fixed and discrete. The "migration-driven diversity" of late modern societies, characterized by heterogeneity of ethnicities, religions, languages, identities, and cultural values, has profound implications for contemporary language use, which cannot be reduced to standardized national languages (Blommaert and Rampton 2011; Vertovec 2010). The complexity of contemporary mobilities and linguistic practices does not seem to have been reflected in the design of voice-operated technological devices that offer languages options in the form of discrete, nameable entities tied to national countries with no possibility for code-switching within a single utterance. In addition to this nationalized concept of language, English is considered to be a "suitably representative language" for training other languages in the design of language technologies (Bender 2011, 17; Bender et al. 2021). The effects of this become strikingly evident in view of the significantly worse performance observed in languages with morphological structures that are different from English, such as Turkish or Finnish (Bender 2011, 5). Conversely, "[d]ominant, prestige-loaded, and standard forms (mostly from European languages), … are further pushed in status as popular gadgets like machine translation and digital voice assistants are available and work best in these" (Schneider 2022, 373).

In the following, I first introduce the concepts of style and stylization within sociolinguistics. I elucidate not only how individuals adopt the voice of the other to mock, critique, or align with the represented voice (Rampton 2018; Bakhtin 1999) but also how organizations and institutions strategically curate and invent specific audiences through stylization in digital and non-digital contexts (Cameron 2000; Portmann 2022). This is followed by a discussion of voice assistants and how their addressees are curated as monolingual speakers of the country of residence through national language options that do not allow two named languages to be used at once. The third and main section of this chapter begins by providing insights into my methodological approach to data collection, interpretation, and analysis of indexical fields (Eckert 2008). My analysis then draws upon the notion of "double-voicing", which encompasses an interpretation of both the stylizers' reflexive voice and the represented voice as manifested through stylistic performance (Bakhtin 1994; 1999). The analytical sections discuss two primary discourses extracted from the interviews, which are further contextualized with data from participant observation: (1) stylizations of non-Western names, utilized to both mock and critique the design of voice assistants and (2) stylizations pertaining to 'wake words', strategically employed by users to avoid activating the device in undesirable situations. In both discourses, it becomes apparent that the voices adopted reflect the speakers' image of the "ideal" user supposedly envisioned by the designers of voice assistants, which is closely intertwined with the audience design of voice user interfaces. To conclude, I reflect on the broader implications of this study for the ongoing development and integration of voice technologies, particularly in multilingual and/or multiethnic settings.

2. Style, Styling, and Stylization

In early variationist sociolinguistics, 'style' traditionally referred to language variation of speakers with regard to specific social situations such as formality or degree of attention to speech (Labov 1972). Later ethnographic studies expanded the notion by emphasizing speakers' reflexivity and their "communicative competence" (Hymes 1972) as they deployed different styles to address different audiences (Bell 1984), to evoke associations with certain social qualities in order to gain approval from the listener (Giles and Ogay 2007), or to signal identification with specific social groups (Le Page and Tabouret-Keller 1985). Against this background, contemporary research typically approaches style as something people do, i.e., 'styling', and as the accumulation of linguistic and semiotic resources that people deploy to produce social meaning and specific identities (Eckert 2003). Style is thus considered to encompass a wide range of social meanings that is not limited to the formality of the interaction, degree of attention to speech, or demographic categories (Jaspers and Van Hoof 2019, 112; Eckert 2008).

In everyday interactions, individuals naturally incorporate styling, but stylization goes beyond this, encompassing a deliberate and strategic "experiment with language" (Jaspers and Van Hoof 2019, 112). Rooted in Bakhtin's exploration of creative textual practices in literature, stylization is characterized as “an artistic representation of another’s language” (Bakhtin 1994, 362). Expanding on Bakhtin’s work, Rampton approaches stylization as “the communicative action in which speakers produce specially marked and often exaggerated representations of languages, dialects, and styles that lie outside their own habitual repertoire” (Rampton 2009, 149). By stylizing, speakers engage in “double-voicing”: incorporating both their reflexive voice and the stylized voice “either to mock or comment on the represented voice …, or to align oneself with the qualities that are associated with the original owners of the voice” (Jaspers and Van Hoof 2019, 112; Bakhtin 1999). In other words, alignment indexes the way speakers discursively “position themselves with respect to the form or content of their utterance” (Jaspers and Van Hoof 2019, 120). In essence, stylization practices evoke “secondary or meta-level representations” of language, thereby offering insights into speakers’ broader sociocultural understandings, discourses, ideologies, and worldviews (Rampton 2006, 222; see also Thøgersen, Coupland, and Mortensen 2016).

Research in recent years has particularly emphasized the relation between stylization practices and larger societal issues, illuminating how speakers position themselves and others by performing voices (e.g., Koven 2015). Stylizations serve as markers that index social categories such as class or prestige through enregisterment processes, i.e., “processes whereby distinct forms of speech come to be socially recognized (or enregistered) as indexical of speaker attributes by a population of language users” (Agha 2005, 38). For instance, shifting between standard and stereotyped vernacular styles may serve as a means of positioning oneself in relation to power asymmetries based on socioeconomic class and ethnicity (Rampton 2006; Jaspers 2006), or as a way of expressing critical perspectives on political matters (Androutsopoulos 2023). Stylization is also observable in mediated representations such as in TV shows or radio broadcasts, contributing to the cultural reproduction of sociolinguistic stereotypes and typification (Coupland 2001; Van Hoof and Jaspers 2016).

Stylizations not only hint at the ideologies and worldviews of speakers but also inform us about how addressees are designed and curated (Bell 1984). In contrast to individuals’ stylizations, language practices in institutional settings such as in service workplaces involve multiple actors and are prescriptive practices that are “imposed from the top down” (Cameron 2000, 326). In the context of digital interfaces, similar practices are observed in which specific audiences are curated “by imposing a particular ‘built in’ social identity” for software users (Portmann 2022). Drawing on work by Bakhtin (1986) and Piller

(2001), Portmann contends that designers of digital interfaces construct what she terms "an ideal addressee", i.e., "a social identity that users, if they wish to use that software, have no choice but take on" (Portmann 2022, 5). For example, depending on the target audience, writers may adjust the formality level of their cookie consent notices, opting for phrases like "I'm cool with cookies" rather than "I accept". In doing so, they not only reference specific addressees but actively "invent and craft said audience through their work" (Portmann 2022, 5). Voice user interfaces curate particular audiences through stylization practices in similar ways. The following section explores the audience design of voice assistants, with particular attention to the language options that they offer.

3. Voice Assistants and their Addressees

Building on Latour's work on actor networks, I conceptualize voice assistants in this chapter as "sociotechnical assemblages": assembled networks involving human and nonhuman actors (Latour 1992; 2005). Viewed as a network, a voice assistant system involves various actors including programmers, researchers, designers, UX writers, consumers/users, data labelers, algorithms, and environmental resources (Crawford and Joler 2018; Natale 2021). Whereas the human labor, environmental impact, and algorithmic processes of the assemblage are not immediately visible to users, the voice user interfaces that users engage with are presented with distinctive synthetic voices, personalities, and stylized conversation design (Natale 2021). The study upon which this chapter is based focuses on the experiences of users as addressees of Google Assistant, Siri, and Alexa – three popular voice assistants in Germany.

Companies employ several strategies to cultivate an "anthropomorphized" persona for voice user interfaces (Sweeney 2016). The assistants are often assigned female names (Siri and Alexa), accompanied by synthetic voice options that are initially introduced as exclusively female. With the primary objective of projecting an image of helpful, polite, and assisting personae, creators of these technologies have been criticized for perpetuating traditional gender roles wherein women are commonly associated with servant and assistant positions (e.g., Phan 2017; Sweeney 2016; West, Kraut, and Chew 2019). While AI assistants are stylized as the figure of a traditional middle-class housewife, users are positioned as "friendly participants in everyday family routines" (Humphry and Chesher 2020, 2; see also Phan 2017; 2019).

Like many other internationally marketed products, digital assistants undergo localization processes in which their design and content are adapted to target cultural contexts (Schneider 2022, 369). Localization encompasses several practices including the provision of gendered voice options, language choices in standardized national categories, and the incorporation of references to popular culture or other types of responses tailored to the target country. For instance, while the introduction of a new language option typically involves a female-gendered voice option, the Arabic language option for Google Assistant did not include a female voice until 2023.[1]

In terms of voice and language options, Phan (2019, 23) posits that they are "underwritten by ideals of whiteness". She argues that the language varieties chosen for voice outputs emulate the standardized varieties commonly associated with educated upper-class speech. These language options, typically represented by English varieties, are adapted for target countries incorporating regional varieties accordingly. For instance, users in Australia are presented with middle-class Australian English that also includes local knowledge and "Australian slang expressions" (Humphry and Chesher 2020, 10). Other languages are also offered in nationalized categories, e.g., a German language option is associated with a variant spoken in Germany and not in Austria, Switzerland, or other countries where users may wish to engage with Alexa in German. Although some voice assistants (including Google Assistant and Alexa) currently permit users to select multiple language options for a single device, the range of combination options is significantly limited[2] and the devices are unable to process code-switching within a single utterance. On the basis of these design choices, it appears that users are conceived of as monolingual speakers of the national language of their residency. For multilingual users, this implies an obligation to think or speak in "one language at a time", thereby suggesting a "monolingual bias" (Li 2020).

Numerous language options also remain unsupported in voice assistants, with Turkish notably absent for smart speakers such as Alexa Echo devices, Google Home, or HomePod. This observation is particularly significant for the

1 This statement is based on my own observations by checking for updates on my personal smartphone and following the news in 2022 and 2023. Although some blog posts discussed the lack of a female voice option in Arabic before 2023, I have been unable to find any official statement or announcement from Google itself.

2 For instance, Alexa only allows other languages to be combined with certain English varieties (see Leblebici 2024).

study presented here, in which participants are individuals who speak Turkish and reside in Germany. Against the background of such affordances and constraints, I argue that the users of voice assistants in this study do not fit the image of the "ideal" users envisioned by the devices' creators, as I have also discussed elsewhere (Leblebici 2024). In the next section, I introduce the backgrounds of my research participants and outline the recruitment and interview procedures.

4. Stylizing the Ideal User

4.1. Methodological approach

To analyze discourses about voice assistants in multilingual contexts, I draw from an ethnographically informed study conducted with 10 Turkish-speaking individuals living in Germany who had migrated there from Turkey within the past 10–15 years. Data collection occurred between 2021 and 2023, encompassing qualitative interviews, online and offline participant observation, follow-up interviews, and Alexa voice history data – also referred to as "log data" (Habscheid et al. 2021). This chapter centers on the informants' stylization practices observed during the interviews, which emerged as a prominent phenomenon in the collected data, offering insights into the users' assessments of the devices.

My informants were recruited by sending invitation messages to WhatsApp group chats of newcomers who self-identify as part of the "New Wave". Unlike the traditional "guest worker" diaspora in Germany, this self-proclaimed "New Wave" of migrants relocated from Turkey to Germany and other European countries for reasons such as higher education, labor opportunities, or sociopolitical motives (Yanasmayan 2018). As a member of these online communities, my positionality was of an "insider" with a similar migration and language biography (Ganga and Scott 2006; De Fina 2020). This position made it easy for me to contact individuals for interviews and for conducting participant observation both virtually and in person, in their homes, and to establish friendly relationships. Following the initial invitation message, a sample of 10 participants were included in the study. The devices they used varied, ranging from stationary smart speakers to voice assistants integrated into smartphones and smartwatches. Although their language biographies and repertoires differed, they all used Turkish and English in daily interactions. Some also communicated in German in their academic or professional environments, while others were in the process of learning the language. In the subsequent analysis, excerpts from the study are contextualized to take into account the participants' devices and language repertoires.

The initial interviews were conducted in a semi-structured manner, inviting participants to comment on their motivations for using the devices and to recall their experiences and use cases related to technology. This semi-structured approach was chosen to open up participants' narratives and life stories, often referred to as "techno-biographies" (Kennedy 2003; Ching and Vigdor 2005; Lee 2014). According to Lee (2014), techno-biographic interviews are valuable as a way to prompt participants to reflect on and make sense of their experiences with domesticated technologies.

Following the period of initial interviews, participant observation was conducted during voice assistant use. Observations were made in a range of settings, including participants' homes for those using stationary smart speakers, out and about with users of voice assistants on smartphones and smartwatches, and through virtual interactions via video call. Alexa users were also asked to share their voice history data[3]. Three of the five Alexa users agreed to share their data from the previous week or month. After the observation phase, a second series of interviews was conducted to ask follow-up questions relating to the initial analysis and log data. These data served as complimentary to the analysis.

The interviews were transcribed primarily in Standard Turkish, with elements such as laughter and pauses included (see Appendix for transcription conventions). When appropriate, phonetic transcription based on the International Phonetic Alphabet (IPA) is provided to elucidate the stylization practices and demonstrate how they deviate from the participants' habitual repertoire.

Jaspers and Van Hoof (2019) propose that the analysis of stylizations should encompass an exploration of "the indexical field", which refers to "a field of potential meanings …, any one of which can be activated in the situated use of the variable" (Eckert 2008, 453). To make sense of the indexical field, the analysis is informed by a "thick" understanding, rooted in ethnographically grounded research, enabling the identification of potential indexical meanings within their local contexts (see e.g., Jaspers 2006; Rampton 2006; 2018; Coupland 2011). Stylization practices in the interviews are thus contextualized

3 Data log collection was not possible with Siri users, as Apple does not provide users access to their log data. At the time of writing, this policy remains unchanged.

with data from participant observation and voice history data retrieved from the Alexa app, which are reflected upon throughout the analysis.

In the subsequent sections, stylizations are analyzed through a dual lens, following Bakhtin's concept of "double-voicing" to examine how speakers engage with the represented voice (Bakhtin 1999; as applied by Rampton 2018, 218). Bakhtin differentiates between two types of discourses: (1) In *vari-directional* discourses, speakers engage in parody or disagreement with the represented voice. The first analytical section explores the types of creative language practices wherein participants mock or critique the supposed "ideal addressee" of voice assistants. (2) In *uni-directional* discourses, the speaker's voice and the represented voice are closer to each other; speakers align with the represented voice (Bakhtin 1999, 198). The second part addresses stylizations in which participants aim to embody qualities of the "ideal addressee" that they deem useful. In both cases, the analysis focuses on how participants discuss smart speakers in interview contexts and how they portray them in particular ways through stylization practices. Therefore, their descriptions of and reflections on the way they interact linguistically with smart speakers are considered as part of their narrative reconstructions and not simply as indicative of their de facto use.

4.2. Accent stylization of non-Western names

In the interview data, nearly every participant highlighted challenges they faced when commanding their devices to process non-English or non-German names. They explained that in order to achieve the results required, they often needed to adjust their pronunciation of Turkish names when commanding their devices to perform tasks such as playing music or making calls to friends. Notably, when participants recounted such situations, demonstrating their stylizations of Turkish names, they often did so with laughter, suggesting a sense of mockery. These examples thus illustrate vari-directional double-voicing wherein the informants distance themselves from the voice assistant's voice.

To provide a concrete example of these practices and to contextualize them, I will first introduce one participant, Selim[4], who utilizes multiple devices, including two stationary smart speakers (Alexa and Google Home) at home and Siri on his smartphone. A 27-year-old postgraduate student, Selim resides in

4 All of the names used for the participants are pseudonyms.

a shared flat with one other person who also owns a Google Home device situated in the shared living room. During our initial interview, I inquired about the language preferences Selim had chosen for his voice assistants. He said that he enjoyed using Siri in Turkish, although it had initially been set to English. Since neither Alexa nor Google Home offer Turkish as an option, he used those devices in English. Similar to experiences reported by other participants, Selim encountered difficulties with activities such as playing music or making calls, particularly when they involved processing non-English names.

Excerpt 1 – Accent stylization of the name "Ibrahim"

Selim: O [Alexa] da İngilizce
müzik falan açıyorsun bi şeyler tarif ya da soruyorsun bi şeyler
anlamıyor ki zaten seni

Didem: A öyle mi?

Selim: Hani böyle şey değil
hani çok onun gibi konuşman lazım
Ne bileyim. **İbrahim [ˈibɹahim]** falan böyle lafın gelişi [...]
ismi algılamıyor aynen yani Türkçe olarak söylersen algılamayacak

Selim: *It [Alexa][5] is also in English*
When you turn on the music or ask for a recipe or something
It doesn't understand you

Didem: *Oh really?*

Selim: *I mean it is not like*
you have to speak like it [Alexa]
I don't know. ***Ibrahim [ˈibɹahim].*** *Like that for example [...]*
It doesn't understand the name I mean it won't understand it if you say it in Turkish

In this excerpt, there is an accent stylization in the pronunciation of the name Ibrahim, which Selim performs in a way that deviates noticeably from his usual manner of speaking. The typical pronunciation of the name Ibrahim in Turkish would be [ib.ɾaːˈhim], but Selim alters it to [ˈibɹahim]. Specifically, he modifies the pronunciation of "r", adjusts the pattern of intonation, and shortens the vowel sound. It is evident that Selim and other participants exaggerate their

5 Personal pronouns in Turkish are gender neutral. The third-person pronoun "o" is translated in all the excerpts as "it".

modified pronunciation in such narratives to illustrate the extent to which they deviate from their typical speech patterns in order to facilitate the device's processing of Turkish names.

Although Selim emphasized during the interview that he often needs to pronounce Turkish names with an accent, accent stylizations of people's names were only observed on some occasions when he was actually using the smart speaker Alexa. According to his voice history data of one week before the interview, he primarily used his smartphone to control music playback, utilizing voice commands to adjust volume settings, skip songs, or turn off the music, rather than specifically requesting songs by a particular artist or title. This data differed markedly from what I witnessed during participant observation in Selim's home, where he attempted to command the device to play a song by the renowned Turkish singer İbrahim Tatlıses, using an accent stylization similar to that which he demonstrated during the interview. İbrahim Tatlıses is known for performing traditional Turkish songs. Hence, accent stylizing his name introduces an extra layer of contrast, potentially heightening the implicit mockery. The juxtaposition pits "modern" technology against "traditional" music, highlighting the implied clash between contemporary voice assistant technology and the traditional genre of music performed by İbrahim Tatlıses. Additionally, since the singer is primarily recognized within Turkey and not internationally, accent stylization of his name may further emphasize the localized nature of the reference, contributing to the playful interaction between Selim, the device, and the researcher. This interaction situation persisted for some time, with the device repeatedly failing to process the name and playing other songs from Spotify instead.

This type of creative language use represents a performance for the researcher present in the room and thus differs from the interlocutor's regular engagement with the device. But Selim also performs these stylizations for other audiences, e.g., friends who visit him at home or those who connect with him through social media. For instance, there are instances in the voice history data where he instructs the device to "Say Hi to [friend's name]". When questioned about these situations in the follow-up interview, Selim explained that he likes to record Alexa's synthetic voice pronouncing the Turkish names of his friends and then shares the recordings with them via WhatsApp. This practice echoes situations in human–animal interactions (Tannen 2004), in which pets serve as communicative resources to facilitate relationships between humans.

As the dialogue with Selim illustrates, there is an understanding that users are expected to 'speak like' the voice assistant in order for Turkish names to be accurately processed by the machine (Excerpt 1). Consequently, users mimic the voice of the machine during operation, and also reproduce their imitated pronunciations for specific audiences in order to mock it. Interlocutors' performance for various (human) audiences of Turkish names pronounced with the foreign accent necessitated by the device parodies the voice of the envisioned "ideal" addressee. This ideal addressee is curated to be a monolingual speaker of a standardized language variety, in this case Standard (British or US) English. Therefore, this imagined user is expected to pronounce Turkish words with an English accent. This discrepancy between users' habitual repertoire and the expected pronunciation that the voice assistant has been designed to respond to becomes a point of entertainment and commentary for the participants.

The sensed need to mimic the represented voice is rooted in previous experiences with such technology. In the following excerpt, another participant, Erdem, shares insights into his stylization practices, although he does not actively demonstrate them during the interview. A 33-year-old engineer, Erdem had been living in Germany for just over 10 years at the time of the study. He told me that he communicated primarily in German at his workplace and engaged with Turkish, English, and German in his day-to-day interactions. He used Siri in German and had previously owned an Alexa device. However, he complained that persistent communication problems with Alexa had ultimately led him to discontinue using it. Throughout the interview, he mentioned these instances of miscommunication frequently, attributing them to "multilingual issues". Below, he elaborates on how he navigates these challenges.

Excerpt 2 – Remembering how to pronounce names like Siri

Selim: Sen birini ara dediğinde onu anlayınca o kendi nasıl anladığını *pronunciation*'i söylüyor.
Şimdi ben Didem'i *call Didem* diyorum mesela benim telefonumda Siri (2) benim telefonumda Siri Almanca mesela ben *ruf Didem an* diyorum
O bana mesela *Okay ich rufe Didem an* derken Didem'i farklı söylüyor ve bir dahaki sefere onu nasıl anladığını aklımda tutuyorum yani seninle konuştuğumda bu küçük farkı böyle hani

Etrafından dönüyorum hani normal Türkçe bi isim söylesem onu anlamayacak çünkü.

Selim: *When you tell it [Siri] call this person, it tells you how it understood the pronunciation.*
Now I say "call Didem" for example. In my phone Siri (2) In my phone, Siri is in German for example. I say "ruf Didem an" [Call Didem] for example.
When it says "Okay ich rufe Didem an" [Okay I'll call Didem] it pronounces Didem differently and the next time I remember how it understood it. The small difference I mean.
I turn around it [the usual pronunciation]. I mean if I say a normal Turkish name it won't understand it.

Erdem has set his voice assistant to reiterate his commands, including names, before carrying out the actions requested. He uses this feature to familiarize himself with the synthetic voice and to devise workarounds so that he can utilize the machine effectively. While this may not be considered a stylization, since he does not perform a marked or exaggerated variant of the name "Didem", it is noteworthy that he engages in metalinguistic reflection regarding his stylization practices. He acknowledges that he does not utter a "normal"-sounding Turkish name but instead must "turn around" the standard pronunciation by mimicking the sound produced synthetically by the machine. Furthermore, during our conversation, he compared this practice to performing in a theater, emphasizing that he deviates from his usual linguistic repertoire to mimic the voice of the machine. His insights reveal a deliberate effort to navigate and subvert the limitations imposed by the technology. The participant's decision to discontinue using Alexa due to communication issues further testifies to his critical perspective on the design of voice assistants.

While accent stylizations of Turkish names are commonly employed for the purpose of mockery or critiquing the inadequacy of voice assistants, in other situations participants adopt stylizations because they find them useful. The subsequent section discusses this aspect, with illustrative excerpts to elucidate how participants utilize stylizations for practical purposes.

4.3. Accent stylization of wake words "Echo", "Alexa", and "Hey Siri"

Not only are people's names subject to stylization, but also the wake words like "Alexa", "Echo" or "Hey Siri" used to activate the voice assistants. This section illustrates uni-directional double-voicing whereby users align with the represented voice rather than contest it. Adopting a stylized version of the wake word, often "with an accent", is reported to be necessary in order to effectively engage with the device. In the following excerpt, Selin, a 20-year-old undergraduate engineering student, who uses Alexa in German, discusses her decision to modify the wake word.

Excerpt 3: Accent stylization of "Echo"

Selin: Bir de şeyini değiştirmiştim hani komut harekete geçirme kelimesi Echo [eko] olabiliyor başka Amazon falan da olabiliyor sanırım. Onu başlarda Echo [eko] yapmıştım mesela, Echo'yu [eko] daha zor anlıyordu.
Echo [e:ko] falan hani böyle daha aksanlı söylemek gerekiyordu galiba. Onu daha zor anlıyordu. Şimdi Alexa'yı daha yine kolay anlıyor ama yine mesela duymadığı oluyor

Selin: *I also changed something, I think the command activation word can be Echo [eko], it can also be Amazon or something. For example, I made it Echo [eko] at the beginning, but it was harder for it to understand Echo [eko].*
E::cho [e:ko] *or something like that, I guess it should have been said with more of an accent. It found it more difficult to understand. Now it understands Alexa more easily, but there are times when it can't hear me.*

Unlike some other participants in the study, Selin does not use Alexa as a communicative resource to entertain visitors by demonstrating its shortcomings. According to her voice history data collected via the Alexa app, she frequently employs it for tasks such as playing music and setting timers, predominantly using short, imperative commands. Consequently, she prioritizes smooth device operation and opts for the most effective wake word, "Alexa," over "Echo". In recounting her narrative, she stylizes the word Echo by mimicking "an accent", which is, in fact, the standard German pronunciation of the word Echo [e:ko] instead of Turkish [eko]. Unlike the stylized Turkish names discussed in the previous section, stylizing the wake word is not about implying mockery or criticism. Instead, Selin frames the adjustment – either accent stylization or choosing an alternative wake word – as a technical solution to an issue that, unaddressed, would hamper functionality.

During the interviews, the presence of voice assistants in the room was palpable, often indexed by the occasional utterance of the wake word. As discussions centered around Siri or Alexa, participants with their assistants set to English or German adopted the Turkish pronunciation of the wake word to prevent inadvertent activation while conversing with me in Turkish about the technology. For example, in the following excerpt, Alp, a 28-year-old Siri user, elaborates on the wake word “Hey Siri” using various accent stylizations.

Excerpt 4: Accent stylization of “Hey Siri”

Alp:	son zamanlarda bazen böyle hey Siri [siɾi] diye sesleniyorum, bazen cevap vermiyor. Şu an da cevap vermiyor çünkü Türkçe olarak söylüyorum.
Didem:	İngilizce farklı şekilde mi tonluyorsun?
Alp:	**Hey Siri ['siɹi]** dediğim zaman. Mesela ya da
Siri:	(beeps)
Alp:	E:: italyancada **eyy ziri [ei:: ziɾi]** böyle **eyy [ei::]** diyorlar ona cevap veriyor
Alp:	*Lately sometimes I call like hey Siri [siɾi], sometimes it doesn’t answer. It is also not answering right now because I am speaking in Turkish.*
Didem:	*Do you intonate English differently?*
Alp:	*When I say **hey Siri ['siɹi]**. For example or*
Siri:	*(beeps)*
Alp:	*E:: they say eyy ziri **[ei:: ziɾi]** like ey **[ei::]** in Italian. It answers to that.*

It is noteworthy that Alp uses digital assistants in German, English, and Italian, particularly with his Italian- and English-speaking friends. During both interviews, he emphasized that he enjoyed using Siri with his friends to explore different ways of engaging with the device and to impress them (also observed in Habscheid, Hector, and Hrncal 2023). For instance, he demonstrated the activation of the smartphone flashlight by voice-commanding “Lumos”, a charm from the Harry Potter series that creates light. By incorporating ‘tricks’ using popular media references, voice assistants can be adapted to impress and thus mediate relationships with others. This aspect is especially relevant in interpreting Alp’s stylizations, as he was also able to experience how his Italian-speaking friends engage with the device.

In the excerpt above, Alp initially pronounces the wake word in Turkish to prevent device activation. When asked about intonation differences in English, he demonstrates his pronunciation of “Siri” with a voiced alveolar approximant. As with the user in Excerpt 1, the contrast between the English and Turkish pronunciation is discernible, particularly in the placement of the “r” sound.

Siri is activated by the accent stylization during our interaction, but Alp does not pay attention to this interruption and goes on to perform an Italian accent stylization of "Siri". He comments that he learned this pronunciation from his Italian-speaking friends: "they say eyy ziri". Although log data retrieval is not offered by Siri, the interview excerpt illustrates the implementation of different stylizations to avoid inadvertent device activation as well as to selectively activate it in desired situations according to language settings.

While stylization practices of wake words retain a distinctive quality and do not completely blend with the speaker's habitual language repertoire, they are not presented in mockery, parody, or irony. Rather, they indicate an alignment with the represented voice, which can be characterized as uni-directional double-voicing (Bakhtin 1999; Rampton 2018, 218). The distance between the usual voice and the represented voice remains minimal, although it is not possible to allege a complete "fusion of voices" (Bakhtin 1999, 198). For instance, in Alp's case, he echoes the voice qualities of his Italian friends when using Italian language settings, aligning himself with the characteristics associated with an "ideal" monolingual user with a standardized repertoire of a national language. In other examples involving English or German, speakers not only make use of their knowledge of standard varieties of English and German but also imitate the voice of the machine to attain the anticipated voice quality and ensure proper functionality of the device.

5. Conclusions

This chapter set out to offer sociolinguistic and linguistic anthropological insights into the navigation, adoption, and critical appraisal of voice assistant technologies by multilingual speakers. Based on ethnographically informed data collected with Turkish-speaking newcomers in Germany, the analysis concentrated on participants' stylizations: exaggerated linguistic performances that fall outside the speakers' ordinary linguistic repertoires (Rampton 2009). Stylizations offer valuable insights into individuals' experiences and assessments and can unveil ideologies and worldviews (Coupland 2011, 155) that are connected to broader sociotechnical issues, particularly concerning the interface design of voice assistants.

With regard to voice assistants' audience design and curation, the findings resonate with Portmann's (2022) assertions concerning digital interfaces: it becomes evident that users feel obliged to conform to the prescribed monolingual mode of interaction in order to obtain required results from voice user interfaces. The provision of languages in discrete national categories, coupled with the inability to process code-switching, makes it clear that the envisioned "ideal addressees" of these voice assistants are constructed to align with the characteristics of monolingual speakers of the standardized national language of their residence, such as standard German in Germany (as discussed in Leblebici 2024). These affordances and constraints shape user interactions within a predefined linguistic and sociocultural scope.

Against this background, I have argued that the study participants with their multilingual backgrounds are not the representative persons curated as target audiences of these technologies. The participants acknowledge this themselves, and reflect on it in different ways, not least when they mimic the voice of the machine or of their friends, performing accent stylizations in German, English, Italian. In doing so, they stage "double-voicing" (Bakhtin 1999; Van Hoof and Jaspers 2016), incorporating two voices in their performances: (1) the stylized voice of the "ideal addressee", (2) their reflexive voice for commenting, mocking, critiquing or aligning with the represented voice. In all cases, these stylizations are not simply creative performances but are a way for speakers to position themselves within the sociotechnical assemblage.

On the one hand, performed accent stylizations of Turkish names were observed as a means of mocking the interface and highlighting the shortcomings of its design. Contemporary migrant-receiving societies, characterized as "superdiverse" due to globalization effects, are home to populations with diverse religious, geographical, national, and ethnic backgrounds (Vertovec 2010). Given this diversity, one might expect voice technologies localized for specific countries to adapt their databases accordingly. However, the experiences of study participants suggest otherwise. Through accent stylizations of Turkish names, participants denigrate voice user interfaces that fail to adequately process non-English or non-German names when set to operate in English or German respectively. Within the contemporary landscape of diverse communication opportunities provided by ever more media channels and platforms (Madianou and Miller 2013), voice assistants' officially propagated range of services are expanded by the creative ways consumers use them to entertain and impress others, not least by deploying different accent stylizations. These stylizations go beyond casual humor: they also express critique of interface designers' limited perspectives. Future research into this promising field could explore whether processes of enregisterment, stereotyping, or typification

(Agha 2005) occur, associating specific forms of speaking with voice assistant attributes.

On the other hand, accent stylizations of activation words such as "Hey Siri" served much less to provoke humor than to adapt to the standard variety in order, quite simply, to be heard – or processed. At the same time, being able to shift between standard German or English and Turkish 'incomprehensible' pronunciations of wake word emerged as a useful resource that enabled multilingual participants to discuss the machine without activating it. In these examples, people describe how they adapt themselves to nonhuman actors' modes of operation for better functionality (Habscheid 2023) by bringing together "a range of linguistic, artefactual, historical and spatial resources … in particular assemblages in particular moments of time and space" (Pennycook 2017, 278). For future research, it could be interesting to explore the extent to which accent stylizations of functional lexical items like wake words "merge" with a speaker's own voice and contribute to processes where "stylization becomes style" (Bakhtin 1999, 198).

Based on ethnographically grounded qualitative research, the analysis presented here makes no claim to be representative of the experiences of Turkish-speaking people in Germany, let alone those of multilingual users of voice user interfaces more widely. Furthermore, the subjective experiences are presented through the lens of a researcher who is considered an insider within a specific community. The observations and analyses have been derived from narratives recounted in interviews rather than from interactive practices in everyday life. Nevertheless, the study has some valuable implications for the adoption and future design of voice assistants, especially in multilingual and multiethnic settings.

In the light of what contemporary research on superdiverse societies has shown, expanding the databases of voice assistants to include names that are not traditionally considered 'German' or 'English' seems long overdue. While this study has illuminated some of the challenges for users of voice assistants in multilingual contexts, it also underscores users' creativity in integrating these technologies into their daily lives. Users engage with voice user interfaces and incorporate them into everyday domestic interactions in unexpected ways by combining cultural, linguistic, and spatial resources. At the same time, these insights indicate the absence of multilingual practices, such as code-switching or the inclusion of non-English or German names in the respective language option, within the voice assistant data set. By failing to implement such practices in the interface design, technology companies thus contribute to the reinforcement, reproduction, and securitization of national language ideologies.

Transcription Conventions

.	pause of less than a second
(2)	approximate length of pause in seconds
?	raising intonation
line break	new idea/proposition
bold	stylization

Acknowledgements

I would like to thank Tim Hector and Stephan Habscheid for their helpful suggestions and comments in revising this article.

References

Agha, Asif. 2005. "Voice, Footing, Enregisterment." *Journal of Linguistic Anthropology* 15 (1): 38–59. https://doi.org/10.1525/jlin.2005.15.1.38.

Androutsopoulos, Jannis. 2023. "Punctuating the Other: Graphic Cues, Voice, and Positioning in Digital Discourse." *Language & Communication* 88 (January):141–52. https://doi.org/10.1016/j.langcom.2022.11.004.

Bakhtin, Mikhail Mikhailovich. 1986. *Speech Genres and Other Late Essays*. 1st ed. Austin: University of Texas Press.

Bakhtin, Mikhail Mikhailovich. 1994. *The Dialogic Imagination: Four Essays*. Edited by Caryl Emerson and Michael Holquist. 9th Edition. University of Texas Press Slavic Series 1. Austin: University of Texas Press.

Bakhtin, Mikhail Mikhailovich. 1999. *Problems of Dostoevsky's Poetics*. 8th Printing. Theory and History of Literature 8. Minneapolis: University of Minnesota Press.

Bell, Allan. 1984. "Language Style as Audience Design." *Language in Society* 13 (2): 145–204. https://doi.org/10.1017/S004740450001037X.

Bender, Emily M. 2011. "On Achieving and Evaluating Language-Independence in NLP." *Linguistic Issues in Language Technology* 6 (October). https://doi.org/10.33011/lilt.v6i.1239.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Virtual Event Canada: ACM. https://doi.org/10.1145/3442188.3445922.

Beneteau, Erin, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. "Communication Breakdowns Between Families and Alexa." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. New York, NY, USA: ACM. https://doi.org/10.1145/3290605.3300473.

Blommaert, Jan, and Ben Rampton. 2011. "Language and Superdiversity: A Position Paper." *Working Papers in Urban Language & Literacies* 13 (2).

Cameron, Deborah. 2000. "Styling the Worker: Gender and the Commodification of Language in the Globalized Service Economy." *Journal of Sociolinguistics* 4 (3): 323–47. https://doi.org/10.1111/1467-9481.00119.

Ching, Cynthia Carter, and Linda Vigdor. 2005. "Technobiographies: Perspectives from Education and the Arts." Paper presented at the First International Congress of Qualitative Inquiry, Champaign, IL.

Coupland, Nikolas. 2001. "Dialect Stylization in Radio Talk." *Language in Society* 30 (3): 345–75. https://doi.org/10.1017/S0047404501003013.

Coupland, Nikolas. 2011. "The Sociolinguistics of Style." In *The Cambridge Handbook of Sociolinguistics*, edited by Rajend Mesthrie, 138–56. Cambridge; New York: Cambridge University Press.

Cowley, Stephen J. 2011. "Distributed Language." In *Distributed Language*, edited by Stephen J. Cowley, 1–14. Benjamins Current Topics, v. 34. Amsterdam; Philadelphia: John Benjamins Pub. Co.

Crawford, Kate, and Vladan Joler. 2018. "Anatomy of an AI System." http://www.anatomyof.ai.

De Fina, Anna. 2020. "The Ethnographic Interview." In *The Routledge Handbook of Linguistic Ethnography*, edited by Karin Tusting, 154–67. Routledge Handbooks in Applied Linguistics. London; New York: Routledge, Taylor & Francis Group.

Eckert, Penelope. 2003. "The Meaning of Style." *Texas Linguistic Forum* 47:41–53.

Eckert, Penelope. 2008. "Variation and the Indexical Field." *Journal of Sociolinguistics* 12 (4): 453–76. https://doi.org/10.1111/j.1467-9841.2008.00374.x.

Ganga, Deianira, and Sam Scott. 2006. "Cultural 'Insiders' and the Issue of Positionality in Qualitative Migration Research: Moving 'Across' and Moving 'Along' Researcher-Participant Divides." *Forum Qualitative Sozialforschung* 7 (3).

Giles, Howard, and Tania Ogay. 2007. "Communication Accommodation Theory." In *Explaining Communication: Contemporary Theories and Exemplars*, edited by B. B. Whaley and W. Samter, 293–310. Mahwah, NJ: Lawrence Erlbaum.

Habscheid, Stephan. 2023. "Socio-Technical Dialogue and Linguistic Interaction. Intelligent Personal Assistants (IPA) in the Private Home." *Sprache Und Literatur* 51 (2): 167–96. https://doi.org/10.30965/25890859-05002020.

Habscheid, Stephan, Tim Hector, and Christine Hrncal. 2023. "Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology." *Social Interaction. Video-Based Studies of Human Sociality* 6 (1): 1–31. https://doi.org/10.7146/si.v6i1.137378.

Habscheid, Stephan, Tim Moritz Hector, Christine Hrncal, and David Waldecker. 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. https://doi.org/10.21248/jfml.2021.44.

Humphry, Justine, and Chris Chesher. 2020. "Preparing for Smart Voice Assistants: Cultural Histories and Media Innovations." *New Media & Society* 23 (7): 1–18. https://doi.org/10.1177/1461444820923679.

Hymes, Dell. 1972. "On Communicative Competence." In *Sociolinguistics. Selected Readings*, edited by J.B. Pride and Janet Holmes, 269–93. Harmondsworth: Penguin Books.

Jaspers, Jürgen. 2006. "Stylizing Standard Dutch by Moroccan Boys in Antwerp." *Linguistics and Education* 17 (2): 131–56. https://doi.org/10.1016/j.linged.2006.09.001.

Jaspers, Jürgen, and Sarah Van Hoof. 2019. "Style and Stylisation." In *The Routledge Handbook of Linguistic Ethnography*, edited by Karin Tusting, 109–24. London; New York: Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315675824-9.

Kennedy, Helen. 2003. "Technobiography: Researching Lives Online and Off." *Biography* 26 (1): 120–39.

Koenecke, Allison, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad

Goel. 2020. "Racial Disparities in Automated Speech Recognition." *Proceedings of the National Academy of Sciences* 117 (14): 7684–89. https://doi.org/10.1073/pnas.1915768117.

Koven, Michele. 2015. "Narrative and Cultural Identities: Performing and Aligning with Figures of Personhood." In *The Handbook of Narrative Analysis*, edited by Anna De Fina and Alexandra Georgakopoulou, 388–407. Malden, MA: Wiley Blackwell.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennyslvania Press, Inc.

Latour, Bruno. 1992. "Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts." In *Shaping Technology/Building Society: Studies In Sociotechnical Change*, 225–58. Cambridge, Mass: The MIT Press.

Latour, Bruno. 1992. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Clarendon Lectures in Management Studies. Oxford; New York: Oxford University Press.

Le Page, R. B., and Andrée Tabouret-Keller. 1985. *Acts of Identity: Creole-Based Approaches to Language and Ethnicity*. Cambridge; New York: Cambridge University Press.

Leblebici, Didem. 2024. "'You Are Apple, Why Are You Speaking to Me in Turkish?': The Role of English in Voice Assistant Interactions." *Multilingua* 23(4): 455–85. https://doi.org/10.1515/multi-2023-0072.

Lee, Carmen. 2014. "Language Choice and Self-Presentation in Social Media: The Case of University Students in Hong Kong." In *The Language of Social Media*, edited by Philip Seargeant and Caroline Tagg, 91–111. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137029317_5.

Li, Wei. 2018. "Translanguaging as a Practical Theory of Language." *Applied Linguistics* 39 (1): 9–30. https://doi.org/10.1093/applin/amx039.

Li, Wei. 2020. "Multilingual English Users' Linguistic Innovation." *World Englishes* 39 (2): 236–48. https://doi.org/10.1111/weng.12457.

Lotze, Netaya. 2016. *Chatbots: eine linguistische Analyse*. Sprache – Medien – Innovationen, Band 9. Berlin: Peter Lang.

Madianou, Mirca, and Daniel Miller. 2013. *Migration and New Media*. Routledge. https://doi.org/10.4324/9780203154236.

Markl, Nina. 2022. "Language Variation and Algorithmic Bias: Understanding Algorithmic Bias in British English Automatic Speech Recognition." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 521–34. Seoul Republic of Korea: ACM. https://doi.org/10.1145/3531146.3533117.

Natale, Simone. 2021. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford University Press. https://doi.org/10.1093/oso/9780190080365.001.0001.

Pennycook, Alastair. 2017. "Translanguaging and Semiotic Assemblages." *International Journal of Multilingualism* 14 (3): 269–82. https://doi.org/10.1080/14790718.2017.1315810.

Pennycook, Alastair. 2018. *Posthumanist Applied Linguistics*. London; New York: Routledge/Taylor & Francis Group.

Phan, Thao. 2017. "The Materiality of the Digital and the Gendered Voice of Siri." *Transformations* 29:23–33.

Phan, Thao. 2019. "Amazon Echo and the Aesthetics of Whiteness." *Catalyst: Feminism, Theory, Technoscience* 5 (1): 1–38. https://doi.org/10.28968/cftt.v5i1.29586.

Piller, Ingrid. 2001. "Identity Constructions in Multilingual Advertising." *Language in Society* 30 (2): 153–86. https://doi.org/10.1017/S0047404501002019.

Portmann, Lara. 2022. "Crafting an Audience: UX Writing, User Stylization, and the Symbolic Violence of Little Texts." *Discourse, Context & Media*, 8.

Rampton, Ben. 2006. *Language in Late Modernity: Interaction in an Urban School*. Cambridge: Cambridge University Press.

Rampton, Ben. 2009. "Interaction Ritual and Not Just Artful Performance in Crossing and Stylization." *Language in Society* 38 (2): 149–76. https://doi.org/10.1017/S0047404509090319.

Rampton, Ben. 2018. *Crossing: Language and Ethnicity among Adolescents*. Third edition. Routledge Linguistics Classics. London; New York: Routledge.

Schneider, Britta. 2022. "Multilingualism and AI: The Regimentation of Language in the Age of Digital Capitalism." *Signs and Society* 10 (3): 362–87. https://doi.org/10.1086/721757.

Sweeney, Miriam. 2016. "The Intersectional Interface." In *The Intersectional Internet: Race, Sex, Class and Culture Online*, edited by Safiya Umoja Noble and Brendesha M. Tynes, 215–28. Digital Formations, vol. 105. New York: Peter Lang Publishing, Inc.

Tannen, Deborah. 2004. "Talking the Dog: Framing Pets as Interactional Resources in Family Discourse." *Research on Language & Social Interaction* 37 (4): 399–420. https://doi.org/10.1207/s15327973rlsi3704_1.

Thøgersen, Jacob, Nikolas Coupland, and Janus Mortensen, eds. 2016. *Style, Media and Language Ideologies*. Standard Language Ideology in Contemporary Europe. Oslo: Novus Press.

Van Hoof, Sarah, and Jürgen Jaspers. 2016. "Negotiating Linguistic Standardization in Flemish TV Fiction around 1980: Laying the Grounds for a New Linguistic Normality." In *Style, Media and Language Ideologies*, edited by Jacob Thøgersen, Nikolas Coupland, and Janus Mortensen, 161–88. Standard Language Ideology in Contemporary Europe. Oslo: Novus Press.

Vertovec, Steven. 2010. "Towards Post-Multiculturalism? Changing Communities, Conditions and Contexts of Diversity." *International Social Science Journal* 61 (199): 83–95. https://doi.org/10.1111/j.1468-2451.2010.01749.x.

West, Mark, Rebecca Kraut, and Han Ei Chew. 2019. "I'd Blush If I Could – Closing Gender Divides in Digital Skills Through Education." Technical Report GEN/2019/EQUALS/1 REV. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1.

Wu, Yunhan, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. 2020. "See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers." In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–9. Oldenburg Germany: ACM. https://doi.org/10.1145/3379503.3403563.

Yanasmayan, Zeynep. 2018. *The Migration of Highly Educated Turkish Citizens to Europe: From Guestworkers to Global Talent*. 1st ed. Abingdon, Oxon; New York, NY: Routledge. https://doi.org/10.4324/9781315555584.

Linguistic Practices as a Means of Domesticating Voice-Controlled Assistance Technologies

Stephan Habscheid, Tim Hector, and Christine Hrncal

Abstract *This chapter explores the linguistic practices involved in the domestication of voice-controlled smart speakers, drawing on findings from our research project "Un/desired observation in interaction: Intelligent Personal Assistants", conducted in Germany from 2020 to 2023. First, the characteristics of smart speakers and the methodological challenges of studying them empirically are outlined. This is followed by a theoretical discussion of how perspectives from interaction research and from linguistic media research can be integrated to complement one another. The empirical part of the chapter first elucidates the organizational characteristics of 'VUI dialogue' and then the organizational integration of VUIs in social interaction in multi-party constellations. Finally, we show how everyday usage practices in households are shaped by the linguistic conditions of their mediation on both organizational levels. The analyses also address the limitations of voice assistants as experienced by users, particularly when devices fail to meet expectations. Thus, this chapter sheds light on the complex dynamics of human–technology relationships and takes into account social and linguistic dimensions of technology domestication.*

1. Introduction: Smart Technologies between Public Discourse and Private Practice

Public discourse around AI-based 'smart' technologies has become increasingly controversial in the past few years – the more recent "hype" in the field of "communicative AI" (Hepp et al. 2023) has given this a new boost, but already before that, the increasing permeation of smart technologies into everyday life had already ignited an ongoing public debate. As Roe and Perkins (2023) point out in an analysis of AI discourse in the British press, public reporting on

AI sways dynamicly between two poles: the promise of great potential on the one hand and warnings about serious dangers on the other. Both poles can be potentially problematic – if the pendulum swings towards an overly positive and expectant attitude towards technological advancement, this not only creates exaggerated expectations of what applications can offer, but also plays into the hands of big tech companies. In academic studies, their products have been criticized for being non-transparent technologies (Liesenfeld, Lopez, and Dingemanse 2023) as well as a catalyst for racist (e.g. Phan 2017, 2019, Woods 2021; Leblebici, this volume) and sexist (e.g., Strengers and Kennedy 2020) biases in representing social order. Furthermore, the operations of the companies have been shown to be based on global inequality and exploitation (e.g., Crawford and Joler 2018, Couldry and Mejias 2019) as well as on excessive energy consumption and other environmental problems (Crawford 2021, Brevini 2023). On the other hand, AI-based, 'smart' technologies certainly have the potential to offer great benefits, e.g., in educational contexts (for an overview see Schiff and Rosenberg-Kima 2023), as assistance technologies (Albert, Hamann, and Stokoe 2023; Endter, Fischer, and Wörle 2023), in the context of smart cities, or in medical contexts (Levina et al. 2024).

However, the discourse on AI-based, networked, and data-driven technologies is not only debated in public, but also, significantly, in the private living environments of users – where they are actually used. In addition to pragmatic questions concerning device operation, users' and potential users' discussions and reflections revolve around comfort and assistance, surveillance, safety, data protection, and exploitation, as well as on human and non-human agency. Discourse in the 'public of the home' picks up on, continues, and evaluates public debates – and relates them to the everyday media practices of users themselves, as has been shown in research on the "domestication" of communication technologies such as television (Silverstone, Hirsch, and Morley 1992). Significantly, the integration of the internet and mobile technologies as well as data-driven and networked technologies into domestic life has blurred distinctions between public and private spheres and thus challenged the domestication research paradigm (Waldecker and Hector 2023, 14). Nevertheless, the domestic sphere continues to play a crucial role in society and in the organization of everyday life, and is one of the most significant areas of application for smart technologies (see Habscheid et al., this volume).

As earlier research grounded in the domestication research paradigm (Silverstone, Hirsch, and Morley 1992), sociology of knowledge and ethnomethodology (Ayaß 2012) as well as conversational linguistics (Baldauf 2002) has shown for different media phenomena, language is central to processes of media domestication and appropriation. Analyzing linguistic practices is thus an apposite approach to investigate the anatomy of social practices that are affected and sometimes reshaped by media technologies – and to research the reflections and stances of users, in order to reveal not only how media technologies are embedded within the domestic community, but also how they are discussed and become part of its discourse. To date, however, empirical research on the everyday practices of users who actually use 'intelligent' and networked technologies in their living environments has been rare. Few studies have focused on linguistic practices and patterns in the domestication of media technologies.

In this chapter, we summarize the results of the linguistic strand of the research project "Un/desired observation in interaction: Intelligent Personal Assistants" within the Collaborative Research Center "Media of Cooperation" at the University of Siegen[1], which investigated linguistic media practices with voice assistants. We focus on the domestication of voice assistants – specifically, on how this is achieved linguistically – to make the contribution of this strand of research more visible in the increasingly complex academic discourse relating to such technologies. We begin with a brief characterization of smart speakers as a device type, elaborating on their specific features and the corresponding methodological consequences for empirical investigation (section 2). We then set out some theoretical assumptions concerning the relationship between linguistics, media, and praxeology (section 3). Based on these premises, we then draw on examples from our data corpus to illustrate the following practices of domestication: (a) the linguistic organization of one-on-one dialogues with smart speakers; (b) the speakers' linguistic embedding in multi-party-interactions; and (c) the ways in which users linguistically accomplish social practices in a collage of human and machinic utterances (sections 4.1–4.3). We finish with a summary and an outlook on the role of smart speakers within broader smart home ecosystems and the relevance of linguistic practices within these developments (section 5).

1 For a more detailed description of the research project and its contextualization within the Collaborative Research Center, see the the introduction to this volume.

2. Characteristics of Smart Speakers: How to Investigate them from an Empirical Linguistic Perspective

Smart speakers integrate virtual intelligent personal assistants with voice user interfaces (VUIs) in the form of a stationary device placed in the living environment. What interests us in particular about these interfaces is their voice-based operation via the acoustic channel: both inputs and outputs are processed verbally. Some dialogues are supplemented by acoustic signal tones, which can be conceptualized as "earcons" (Blattner, Sumikawa, and Greenberg 1989, 11), as well as by visual signs on the surface of the device. Voice assistants can be seen as a prototype for smart technologies: the recording, transmission, and utilization of data; the invisible connection to network publics as well as the embedding of the technology usage in sequential and incremental social interaction; the interweaving with everyday practices (and their transformation); the humanoid character of the systems; and the gradual adaptation of users to the linguistic restrictions (Hector and Hrncal 2024) are typical characteristics of smart technologies that come together in intelligent personal assistants with VUIs. Furthermore, the devices are associated with controversy, with (potential) users weighing the benefits of comfort, security, and assistance against surveillance, privacy breaches, and observability.

Krummheuer (2010) characterizes sociotechnical dialogue with virtual agents as a "hybrid exchange". On the one hand, such dialogues exhibit characteristics of interpersonal interactions based on a simulated similarity between technical and human actors. On the other hand, when malfunctions arise or communication is unsuccessful, differences between the human and the device come to the fore, and it is the human user who has to adapt to the limited interactional capabilities of the machine. Especially when disruptions occur, the focus can quickly shift from similarity to difference. The synthetic voice not only vocalizes machinic answers, but is modelled as an artificial companion, a 'persona'. Natale and Cooke (2021, 1009) stress that "[f]rom a technical viewpoint there isn't anything like one monolithic 'Alexa' or 'Siri'". From this perspective, these 'personae' function rather like metaphors to integrate a range of technical processes, such as speech recognition, natural language processing (NLP), and information retrieval within an interaction partner perceived as unified and singular. Sociotechnical exchange with the systems is mediated by such metaphors. This type of interface design follows on from attempts to make human–technology dialogues seem as 'natural' as possible: the non-human dialogue partner is addressed as a human conversational partner would be, as advertisements for these products emphasize (Lind and Dickel 2024). Hence, the overriding design aim of this interface type is to allow users to integrate computer operations as 'seamlessly' as possible into their everyday lives. The interfaces therefore serve as a tool to mediate computer operations in the form of a linguistic dialogue (Merkle and Hector, forthcoming).

Voice assistants are always connected to a smartphone app. These apps include various functions: they give users control over the device in terms of settings and preferences, show connections to other smart home devices as well as to the internet, and enable touch-based control (for a detailed description, see Habscheid et al. 2021). Some of these apps also show a protocol of the usages of the smart speaker, including the recordings of audio in which a voice input was recognized. Habscheid et al. (2021) examine the analytic potential of the log file data that are recorded by voice-controlled systems and the documented activities in related smartphone applications. The log file data include not only the audio recording of the input, but also further information concerning the 'activities', presented graphically (such as a transcript of the recorded input and response and the time it occurred). They also provide further options for interaction with the database entries, such as providing feedback on whether the voice assistant did what was expected. At the same time, log file data serve to document fragments of the social situation they recorded. With these characteristics, they offer data by means of which the machine's performance and its 'understanding' of the recorded situation can be assessed. On the basis of the datafied recordings one can also draw conclusions about their further utilization as training data for speech recognition and NLP systems (see also Hector, forthcoming).

However, the data recorded by the systems do not provide enough information for research that seeks to analyze the entanglement of smart speaker systems, everyday life, and ongoing social situations in relation to linguistic practices (Hector and Hrncal 2020, 9; Habscheid et al. 2021, 44–45) – such a perspective calls for recordings of not just the 'voice command' itself, but of the social situation in which it takes place; the preparation and initiation of a voice dialogue and its subsequent evaluation and follow-up comments are crucial. This creates a methodological challenge for research into situational smart speaker usage: How can researchers record audio data in the private living environments of users, not just during discrete interaction situations, but whenever the smart speaker is used in daily life – without recording the whole living environment constantly (which would not just be ethically problematic but would also produce an overwhelmingly vast volume of data)? Porcheron et al. (2018) developed a solution for this methodological problem: A specifically designed device called a "conditional voice recorder" (CVR) that can be placed in the living environments of study participants. Its basic function is to continuously record the audio in a certain room via a far-field microphone, but also to delete the recorded audio after a set duration of time, e.g., three minutes. The resulting three-minute audio recording is held in the buffer memory, only to be saved for longer if – and this is the key operating principle of the raspberry-pi based device – an invocation word such as "Alexa" is recognized by the built-in speech recognition of the CVR. When an activation word is recognized, the three buffer minutes are kept, together with three follow-up minutes, and saved on a connected flash drive[2]. Equipped with this technology for data collection, our research project was able to include within its ambit the analysis of multi-party situations involving more than one user and/or smart speaker, as well as the conversational preparation, initiation, and subsequent evaluation of VUI dialogues – and hence their embedding in turn-by-turn talk in social interaction (Habscheid, Hector, and Hrncal 2023, 7–8). We also drew on another form of data: video recordings produced to document how users installed their smart speakers for the first time.

3. Theoretical Foundations: Linguistics of Practices, Interaction, and Media

Approaches that seek to better understand communicative and cultural practices by tracing and reconstructing their linguistic (and multimodal) mediation have long been anchored in the theoretical traditions of anthropology and linguistics (e.g., Wittgenstein 1984; Hanks 1996; Luckmann 1986). More recently, linguistics has also been engaging with newer sociological approaches from the field of praxeology (e.g., Schatzki 2002; Reckwitz 2003; Hirschauer 2004, 2016; Deppermann, Feilke and Linke 2016a; on Hirschauer see also Habscheid, Hector, and Hrncal 2023).

The concept of 'practices' builds upon the insight that the use of linguistic means and forms indexically invokes – and thereby situationally modifies –

2 For a detailed description of the functions of the device, its further technological development for the context of the research project in Siegen, and the data practices associated with it, see Hector et al. (2022).

highly complex, 'gestalt-like' cultural knowledge contexts, the meaning of which extends far beyond simple speech acts and semantic content (Habscheid 2016). Additionally, the concept of 'practice' refers to a level of sociality that is logically prior to the handling of cultural resources in communication (Schüttpelz and Meyer 2017): Social practice is fundamentally established through cooperative production and temporality, it transcends language and knowledge in that it is also based on the participation of bodies, spaces, and technology (e.g., humans and voice user interfaces in connection with built environments, digital platforms, and infrastructures), often under asymmetrical conditions for cooperation (Hirschauer 2004, 2016; on VUIs see Habscheid et al. 2021). We revisit the asymmetrical types of participation associated with VUIs in section 4.2.

Cooperative practice requires material and sensory mediation, which may involve technical resources. Certain approaches in linguistics, like the tradition of 'Gesprächsforschung', focus in particular on cases of 'interaction' (Hausendorf 2015) in which "co-presence" or at least "tele-co-presence" (Hausendorf 2022) is established on the basis of synchronicity (and, in face-to-face conversation, also co-presence in physical space). The tradition of conversation analysis, which has strongly influenced linguistics (including 'Gesprächsforschung') investigates how interaction in co-presence is structured by means of language ("linguistic practices", cf. Deppermann, Feilke and Linke 2016b, 13) and manifold other material resources that can be functionalized as situated signs in the process of interaction (Goodwin 2018, 445). Whether and to what extent the findings of Conversation Analysis are applicable to dialogues with VUIs is one of the questions we address below (section 4.1). A further question is how social interaction changes under conditions of co-presence when VUIs participate asymmetrically (section 4.2). Finally, interaction forms the organizational backbone of communicative and cultural practices (including practices that extend beyond communication, like cooking or shopping), which can be observed in connection with the domestication of smart speakers in households (section 4.3).

One approach to incorporating media into linguistic theory is to assume that media formats presuppose and enable particular linguistic forms. This can be studied, for example, by focusing on 'communication forms' (*Kommunikationsformen*). This concept refers to the structural conditions of communication and language use that are characterized by the use of technical artifacts ('media' in a narrow sense) in connection with certain media institutions (e.g., the platforms of commercial IPA systems). In addition to practices tied to ‘genre’ conventions (e.g., weather queries), communicative potentials that have not (yet) been exploited culturally can also come into focus: Thus, in their reinterpretation of the term, Brock and Schildhauer (2017, 13) define communication forms as “human-made clusters of technical and communicative constellations with communicative potentials, which are commonly restricted by conventions”. Alternatively, a different perspective can be adopted, whereby ‘media’ are understood in a broader sense: as culturally established techniques, of which technical media can be a component. Accordingly, “medial procedures”, e.g., communication mediated by language assistance systems, form the material and procedural side of the use of signs; this use is also embedded in conventional (communicative) practices (e.g., weather queries), and it furthermore depends on the individual competence of the sign user (Schneider 2017, 45).

Neither of these approaches take into account how media – beyond their communicative potential and their creative use by individuals – come into being *as social instances* and are thereby shaped on the basis of their materiality (Meiler 2019), vary dynamically, and can change over time. In order to be able to investigate such questions, the technical and cultural-theoretical view of media must be supplemented by a foundational, social-theoretical perspective: In a praxeological view, as pursued in the CRC 1187 “Media of Cooperation” at the University of Siegen, media are understood as “cooperatively created conditions of cooperation” or, in short, as “media of cooperation” (Schüttpelz 2017, 24).

This position, developed in the conversation analytic tradition by Charles Goodwin (2018) and at the intersection of conversation analysis and media theory by Erhard Schüttpelz and Christian Meyer (2017), does not ignore cultural consolidations and agreements (such as techniques, communicative genres, or symbols), nor does it take them as given. Instead, at a more foundational level of social order, it focuses on the concept of reciprocal “practice” (Schüttpelz and Meyer) or “co-operative action” (Goodwin): Participants in the production of meaning mutually produce processes by partially taking up the sign-like material resources (including, not least, indexical references) brought into play by their predecessors, and transforming them to reuse them for their own purposes. This theoretical premise is able to account for how communicative conventions can emerge and change, while identifying media themselves – without which communication is inconceivable – as always situationally emergent, brought forth in practice. It is this framework that underlies our research.

4. Analysis

The research that is discussed in the following sections 4.1–4.3, including the examples cited, is based on a corpus of video and audio data recorded in eight households. The recordings of initial installations comprised a total of one hour and 53 minutes of video filmed in six different situations. The audio files were recorded by the CVR in two different phases: generally, the first phase took place directly after the initial installation of the smart speaker, with a second phase three to four months later. In two households, smart speakers had already been installed before our data collection started, so we only conducted one recording phase. All the recording phases lasted between 20 and 30 days. This resulted in a total of 30 hours and 58 minutes of audio recordings, which were subsequently inventoried and transcribed according to the GAT 2-standard (Selting et al. 2011). For the video recordings, the transcription was supplemented multimodally following Mondada (2016).

4.1 Linguistic Organization I: The 'VUI Dialogue'

In order to be successful, sociotechnical dialogue between user and smart speaker – referred to in the following as VUI dialogue – must follow a specific sequential pattern, as shown in the following example. The excerpt stems from a CVR recording from the household of Beate (BW), a retired official in her sixties.

Example 1: How will the weather be today?

```
051 BW:  HEY google?
052 p:   (1.1)
053 BW:  wie wird das WETter heute?
         how will the weather be today?
054 p:   (1.7)
055 GA:  HEUte wird es in aachen überwiegend
         bewÖlkt,
         today in Aachen it will be mostly
         cloudy
```

```
056          bei temperaturen zwischen elf und
             NEUNzehn grAd.
             with temperatures between eleven and
             nineteen degrees
057 BW:      HM_hm;
058 p:       (1.0)
059 GA:      momenTAN ist es überwiegend bewÖlkt
             bei elf grAd.
             at the moment it is mostly cloudy with
             eleven degrees.
```

In addition to voice input (l. 053) and voice output (l. 055–056), the VUI dialogue in the example above also includes the invocation by means of an activation word or phrase – "HEY google?" (l. 051). The device responds to this by activating the listening mode, which enables the ensuing audio input to be processed. Devices also indicate activation visually with light signals on the speaker itself; Apple's smart speaker also emits an acoustic reception signal after a waiting period to indicate that it is ready to listen (Hector, in preparation). The exchange that follows activation is typically designed as a pair sequence – the input (first pair part) makes an output (second pair part) expected (Habscheid 2023, 188). Differently from the example above, the latter could take the form of a response to the input without any voice output from the smart speaker (e.g., fulfilling a task such as playing a certain song or turning on the lights). This basic sequence structure was represented very consistently across our data set (Hector, in preparation; forthcoming). Furthermore, dialogues can be expanded with a third element either by the user(s) or by the system. Sequence-ending third moves that are common in social interaction, such as the confirmation by Beate ("HM_hm;", l. 057) in the example above, often do not take place in VUI dialogues, and when they do, it often remains unclear whether or to what extent they are processed by the systems as indicators of comprehension. Longer user-initiated expansions without a further invocation are typically co-addressed to both the device and other co-present persons (Hector, in preparation) – as we elaborate below. When a third move is emitted by the device, this might, for instance, be to provide additional information as in the example (l. 059), or new features may be recommended, or tips, such as how to manage notifications. These utterances may be semantically linked to the preceding turns, but they do not have to be; they also can, but do not have to, make follow-up turns relevant

(Habscheid 2023, 189–90; Hector in preparation). In general, these expansions can function to generate user feedback, to compose more coherent sequential projects, or may be part of a general effort by providers to improve the service quality (Habscheid 2023, 188–90).

Regarding the linguistic design of the input, research on human–machine interaction has intensively discussed the notion of "computer talk" (first mentioned by Zoeppritz in 1985) and deliberated whether a distinct register for conversations with machines can be defined (see also Hector, in preparation). As summarized by Lotze (2016, 160–61), it has not been possible to empirically establish evidence of such a "register", as the actual linguistic practices involved in sociotechnical exchange with machines are too diverse and not specific. Short linguistic utterances, such as stand-alone nouns, imperatives, or deontic infinitives, which might at first glance appear to be characteristic of such conservations, have also been observed in other empractical contexts (e.g., Mondada 2014; Deppermann 2018), and are not unique to talk with machines (Hector, in preparation; Merkle and Hector, forthcoming). Furthermore, questions such as asking for a weather forecast in the example above (l. 053) are not the exclusive preserve of VUI dialogue. What does, however, seem to be specific to human–machine exchanges, is a "new form of dialogicity" (Lotze 2020, 363; Habscheid 2023, 174), which is characterized by a "broken-up" form of sequentiality (Krummheuer 2010, 229; Hector, in preparation). For VUIs, for example, this means that follow-up requests by users, if not initiated by the VUIs, always require a whole invocation sequence. Furthermore, sequential coherence between a first and a follow-up utterance – which in human-to-human conversations is often realized by the use of pronouns, for example – cannot be accomplished, or only to a very limited extent. The exchange between users and the system is most stable when the basic sequential structure mentioned above is adhered to, using adjacency pairs and short inputs (see also Barthel, Helmer and Reineke 2023). That does not mean that VUI dialogues are entirely predetermined, however. Their constitution is still an ongoing linguistic accomplishment between a machinic and a human interlocutor with very different initial conditions and it takes place under specific socio-spatial, material conditions.

4.2 Linguistic Organization II: VUIs as Participants in Multi-Party Interactions

A comparative analysis of one-on-one situations and multi-party settings makes it clear that sociotechnical dialogue and social interaction between humans are not one and the same – rather, users accomplish "transitions between sociotechnical exchange and social interaction" (Habscheid 2023, 176) and establish a dedicated "meta-interaction space" (Habscheid 2023, 176) – "a specific type of interaction between humans, which is directly related to the sociotechnical human–machine exchange" (Habscheid, Hector, and Hrncal 2023, 15). In this type of interaction participants may, for example, negotiate and reflect on the sociotechnical exchange, discuss smart speaker features in relation to ways of addressing the VUI, evaluate and deal with failures, or more generally discuss "the embedding of the smart speaker in the sequential unfolding of everyday practices" (Habscheid, Hector, and Hrncal 2023, 15). The relevance of this type of interaction has already been pointed out by Porcheron et al. (2018, 9): Users conceptually distinguish an exchange with a VUI based on inputs and outputs from one of "turns-at-talk". While the first is by design pre-planned to a certain degree, the unfolding of the talk is based on coherent conversational context and reciprocally constitutes itself in the latter.

The following excerpt (Example 2) illustrates how VUIs can be embedded in multi-party interactions by human users and to what extent the negotiation of agency can be intertwined with this. The excerpt was recorded in the shared apartment of Lukas (LF) and Alex (AK). Kurt (KS), a guest present at the time of the recording, introduces Lukas and Alex to a feature of their smart speaker that was previously unknown to them: the "Super Alexa Mode".

Example 2: "Super Alexa Mode"

```
155 KS:  (sag ma) kennst du den SUper alexa
         modus?
         say do you know the super alexa mode
156 p:   (1.7)
157 LF:  (den) WAS?
         the what
158 p:   (0.7)
```

159 KS: <<p> WARte.>
wait
160 alexa lautstärke
SE[CHS-]
alexa volume six
161 k: [((Musik wird leiser))]
volume turned down
162 (1.2)
163 k: ((Musik wird lauter))
volume turned up
164 KS: aLExa?
alexa
165 k: [((Musik wird leiser))]
volume turned down
166 [(0.5)]
167 KS: hoch (.) hoch (.) runter runter (.)
links rechts (.) links rechts (.)
be: a: (.) START;
up up down down left right
left right b a start
168 (1.4)
169 AL: super alexa modus wird aktiVIERT.
super alexa mode activated
170 (0.6)
171 AL: hiHI;
hehe
172 (0.4)
173 AL: zehn extra leben adDIERT.
ten extra lives added
174 (0.6)
175 AL: speed level drei FREIgeschaltet;
speed level three unlocked
176 (0.6)
177 AL: alle powerUPS verfügbar;
all power-ups available

```
178 AK:  (die kann ja doch)
         she can
         [((unverständlich)) ]
            unintelligible
179 AL:  [FEHLfunktion;      ]
          malfunction
180 AK:  die kann auch in (deu[tschland) sich
         selbst ((unverständlich))]
         in Germany she can also herself
         ((unintelligible))
181 AL:                       [powerups haben
         keine POwer;               ]
                                power-ups do
         not have any power
182      (0.4)
183 KS:  ((lacht))
           laughs
184 AL:  ABbruch;
         aborting
```

In this excerpt, a hidden function of the smart speaker, a so-called 'Easter egg' that has no function beyond entertainment, is personally presented 'in a favorable light' by Kurt to his interaction partners Lukas and Alex, who at the time were relatively inexperienced in using the device. Kurt's performance showcases the smart technology as exceptional and innovative (Habscheid, Hector, and Hrncal 2023, 8). The command to start the Super Alexa Mode is based on a sequence that is more commonly known for its use with game controllers: "up up, down down, left right, left right, A, B, start" is actually a sequence of keystrokes that activates advantages in video gaming. It has been a well-known 'classical' Easter egg in the gamer scene since the 1980s and works in many different games (Baumann 2023). Kurt, as connoisseur of the function, first establishes his host Lukas' level of awareness regarding this 'Easter egg' (l. 155); the latter's reaction in line 157 indicates his complete lack of knowledge. With his request "wait" (l. 159), Kurt then projects the subsequent demonstration of the feature, for which he first reduces the volume of the music playing via voice command (l. 160) and then utters the atypical voice command (l. 167). The VUI ratifies the input with a confirmation of activation of the Super Alexa Mode (l. 169–177), which is then cancelled proactively by the system a little later (l. 180–184), presumably as part of the feature. By naming the feature in the beginning of the excerpt (l. 155 and 159) and by raising the volume for the presentation, Kurt frames the function and its demonstration as something "atypical" – beyond the usual set of commands used by household members. An asymmetry is thereby staged between Kurt, the well-informed guest who is familiar with the presented feature, and the household members Lukas and Alex to whom it is as yet unknown.

This leads to the question of how the devices participate in social practice, as discussed by Reeves and Porcheron (2023) and Hector (in preparation). To explore in what sense and to what extent the devices 'participate' – and are *treated as* participants by the co-present humans – it is helpful to revisit debates on the role of anthropomorphization that have been ongoing ever since technology began to become embedded in everyday life. The late 1990s and early 2000s saw the rise of a paradigm called "Computers as Social Actors" (CASA), which basically argued that humans mindlessly transfer habits from human-to-human communication to interactions with other entities perceived to exhibit a certain degree of 'life of their own', as computers may do, especially if their design is anthropomorphic (Nass and Moon 2000, 98). Reeves and Porcheron (2023) have interrogated these concepts intensively and challenge the notion that conversational AI systems, such as voice assistants, 'participate' in social interactions in the same ways that humans do. They argue that to assume so overlooks the fundamental role of social situations in making 'AI interactions' meaningful. Indeed, the significance of the ongoing social situation for an accomplishment of 'understanding' and 'meaning-making' was already demonstrated by Harold Garfinkel in 1967 in his research on the early chatbot ELIZA, which deployed rather simple sequence-orientated scripts to simulate an interaction based on connectable utterances (see Eisenmann et al. 2023, 6). Hence, instead of conceiving of voice assistants as social actors, Reeves and Porcheron (2023, 581) suggest that dialogues with these systems are better understood as regulated exchanges among participants within organized social (group) contexts, in which anthropomorphic utterances can be a linguistic resource.

The notion of participation is thereby not merely about direct interaction with the system, but focuses more intently on the broader social dynamics and the collaborative efforts of users to incorporate these technologies into their everyday lives. The situated, emergent nature of participation in VUI dialogues come to the fore. This is also in line with the findings of Hector (in preparation), who adopts the way participation is defined by Stefan Hirschauer (2004, 2016). Hirschauer argues from a praxeological standpoint that basically any material artifacts can be part of an action – with various different "levels of activity" (Hirschauer 2016, 49), ranging from active to passive on one spectrum and from active to prohibitive on another. Based on this model, combined with analyses of recordings of VUI dialogues in multi-party settings, Hector (in preparation) proposes from a linguistic standpoint that voice assistants can indeed be 'participants', but not in an equivalent way to human conversational partners. Human utterances seemingly directed 'to' the intelligent personal assistant following an input–output exchange are often, upon closer scrutiny, actually addressed to other humans as a form of co-address. Superficially, their linguistic contribution seems to be directed to the virtual interlocutor and can articulate polite formulas, evaluations, or follow-up requests. However, as no listening mode is activated by the device, at a pragmatic level these utterances should instead be seen as evaluations, frustration management, humor provocation, or other activities within a group of co-present speakers (see section 4.3 for an example).

The excerpt in Example 2 also shows that the embedding of smart speakers as participants in multi-party interactions can shed light on the negotiation of agency of technological actors, which – "especially in voice-based exchange with smart speaker technology" (Habscheid, Hector, and Hrncal 2023, 1) – is dynamically accomplished and intrinsically "bound to the local (linguistic) practices carried out by or rather involving contributions by participants with unequal resources for participating" (ibid.). The abovementioned meta-interaction space is highly relevant for this negotiation of agency between human and technical entities (Habscheid, Hector, and Hrncal 2023, 10). In line with the aforementioned praxeological perspectives, it makes sense to work with a praxeological conceptualization of agency, too. Krummheuer (2015) discusses how agency is dynamically constructed and negotiated within the interactional context, referring to concepts from Actor-Network-Theory (Latour 2005) as well as ethnomethodological conversation analysis. Rather than identifying it as an attribute inherent to either humans or technology, agency is viewed as a situationally emerging property of ongoing social interaction (see also Pentzold and Bischof 2019; Natale and Guzman 2022). In human–technology exchanges, participants might address the technology as a communicative partner, an actant, or even an "opponent", according to their situational needs and the unfolding interaction (Krummheuer 2015, 180). This concept of agency directs research on the linguistic unfolding of social practices with smart speakers towards the question of how users ascribe and negotiate agency through their interactional practices. Our research has shown that agency occurs as a situational accomplishment that is linguistically negotiated between the participants (Habscheid, Hector, and Hrncal 2023), including the non-human 'participants'. The balance of agency can shift within very short time frames.

But how can we trace this in the excerpt presented above? Kurt, who is not a member of Lukas and Alex's household, claims agency by taking over the operation of the device, which is granted to him by the two flatmates who do not interfere. Here, agency is initially negotiated within the multi-party constellation of the human interactants. With the utterance "Alexa, up up down down left right left right B A start" (l. 164–167) directed towards the VUI, Kurt follows the script structure required to correctly execute his voice command, and in doing so he – at least to some degree – downgrades his own agency and attributes a certain level of agency to the technical device, which then realizes the output requested by Kurt's input. To some extent, the device itself then proactively indicates when the operation requested in Kurt's utterance has been redeemed to a satisfactory degree and is complete.[3]

4.3 The Linguistic Accomplishment of Social Usage Practices

In the process of 'domesticating' new technologies (Waldecker and Hector 2023), routine everyday practices are modified and new sociotechnical practices emerge. The latter initially serve to make the devices work and embed them into domestic spaces and infrastructures, which in turn are subject to processes of change and (re-)design. From the outset, users are challenged to overcome the 'resistance' of any new technology – often together with other people in co-presence – and to cope with the associated alienation (section 4.3.1). Over time, changed and new practices can become more or less deeply anchored in everyday domestic life. Insofar as all these social and sociotechnical cooperation processes are essentially linguistically and multimodally mediated, they can be investigated from a (media) linguistic perspective. This includes cases where attempts at domestication fail due to shortcomings of voice user interfaces (section 4.3.2).

3 It should be mentioned at this point that agency discernable "in front of" the device, in the interaction situation, can differ significantly from data practices at the "back end" of the device – if the latter are taken into account, the relative agency on the users' side is considerably diminished (Waldecker, Hector, and Hoffmann 2024).

Typically, users initially explore the practical potential of new technologies by testing and practicing. In the case of smart speakers, these processes are partly guided by the system through corresponding 'test scripts', but our observations show that users often go far beyond those in what they do (linguistically) in the early phase of appropriation (Habscheid 2023). While one approach is to cautiously probe the limits of the (linguistic) performance and flexibility of the systems, some users deliberately push the devices to their limits and beyond (Krummheuer 2010, 263) in order to then assert their superiority by ridiculing and exploiting bizarre utterances of the VUI to humorous effect (Krummheuer 2010, Chapter 9).

However, the appeal of such nonserious domestication practices can wear off over time. Furthermore, anyone who wants to make practical use of the technologies' capacity to increase convenience in everyday life is obliged to adapt to the limits of their linguistic communication capabilities rather than exacerbate them (Drösser 2020, 72). Smart speakers are capable, for example, of processing certain pair sequences. In contrast, utterances that can only be interpreted on the basis of conversational implicatures taking into account their sequential position irrespective of formal sequences, evidently represent an excessive challenge for the systems (at least those at the technical level we have investigated so far). This may incite amusement in an early phase of use (Habscheid 2023), but in the longer term users are more likely to accommodate their linguistic behavior towards that of the devices.

4.3.1 Early stage

The following excerpt (Example 3) is a transcript of a video excerpt from the documentation of an initial installation: Lukas (LF) and Alex (AK), two young men who live in a shared apartment (also protagonists in Example 2), are busy putting a smart speaker into operation:

Example 3: "You asked for mom"

```
421 LF:  §WILL ich das mit meinen kontakten verbinden,
          do I want to connect this to my
          contacts
    ak:  §enters the room and walks towards the sofa->
422 LF:   dann SPÄter,§
         later
    ak:             ->§
```

```
423 AK:  nein?
         no
424 AK:  §$((lacht))         $§
             laughs
    ak:  §sits down on the sofa§
    lf:   $smiles-------------$
425 p :  §(3.0)        §
    ak:  §grabs the laptop standing on the sofa, takes it
          onto his lap§
426 AK:  §außer du willst dass ich die ganze zeit damit
          RUMtrolle;
          unless you want me to troll with it all the
          time
    ak:  §leans forward and rummages through a few items
          on the coffee table->§
427 AK:  $alexa ruf meine MOM an;$§
          alexa call my mom
    lf:  $smiles----------------$
    ak:                       ->§
428 AK:  +HAllo lukas mom;+
          hello lukas mom
    ak:  +looks at LF-----+
429 LF:  h°°
430 AL:  §du hast nach MAma gefragt,        §
          you asked for mom
    ak:  §leans back and operates the laptop§
431 AL:   [aber ich kann diesen namen in deinen
           kontakten] über die geräteliste nicht FINden;
           but I can't find this name in your contacts via
           the device list
432 AK:  $[((lacht laut und klatscht in die
            Hände)) ]$
             laughs loudly and claps hands
    lf:  $smiles-----$
433 AK:  ((lacht laut)) [((lacht))  ]
          laughs loudly  laughs
434 LF:                 [!UH!?      ]
```

In this initial installation phase, the users have not yet set up separate user accounts for the smart speaker. Typical work-sharing usage has not yet been established in this sense. Lukas is therefore hesitant to connect the shared stationary device with the contacts in his cloud (l. 421). Entering the room, Alex jokingly suggests that if Lukas did so, Alex could “troll around” with his flatmate’s contacts (l. 426). He thus uses humor to raise the issue of intra-household surveillance: individual privacy may be relinquished to other household members if they use the same stationary smart speaker without differentiation of accounts. Lukas adds weight to his joke with a hypothetical, quotation-like pattern of speech: "Alexa, call my Mom" (l. 427).

The practice of hypothetical quotation bears a formal resemblance to the initiation of a real dialogue with the smart speaker – which is indeed its effect. Alex's utterance "Alexa, call my Mom" is identified by the device as an invocation and linguistic input, but is not processed successfully: Instead of implementing the command, the system opens an insertion sequence with a kind of query: "You asked for Mom but I can't find this name in your contacts via the device list" (l. 430–431).

The fact that the smart speaker responds to an utterance not addressed to it as such and then processes it incorrectly may show the problem of the medial externalization of the household, but in this example case it mainly provides a cause for amusement: Alex exhibits his pleasure to the maximum by means of a linguistic staging procedure (Schmitt 2003) – loud, extended laughter accompanied by hand-clapping – while Lukas, also with a practice of staging, realizes an emphatically minimized format with a smile. While the entertainment value of such practices in the meta-interaction space is evident, they also indicate how users assure each other that they are in control of the system. It is then not the technology that "curates" social practices (Dolata 2019, 195), but rather the users who 'domesticate' the system by displaying their superiority.

4.3.2 Later stage

The following excerpt (Example 4) shows a usage situation from a later phase of domestication. Users are conducting a knowledge search, but the system proves to be extremely 'recalcitrant': not only does it repeatedly fail to provide the information sought, it also employs a dysfunctional and time-consuming presentation format that evidently causes frustration for users Robin (RL) and Lara (LS). Despite the disappointing experience, which is indicated mainly in the meta-interaction space (see section 4.2) through rejections (l. 201), comments (l. 222) and laughter (l. 218), Robin repeatedly initiates attempts to repair the dialogue, taking a trial-and-error approach. This is exhibited in abruptions (l. 204/205), new trials (l. 205), and reformulations of the request (l. 207, l. 234). Ultimately, use is discontinued with an expression of frustration (l. 242, l. 247) and a sweepingly negative evaluation ("You are useless", l. 254), which is answered by the assistance system ("I am still learning", l. 258):

Example 4: "When is the next bus?"

188	RL	hey GOOgle;
189		(0.8)
190	RL	WANN-
		when
191		[fährt der NÄCHste-]
		is the next
192	LS	[(halts MAUL.)]
		shut up
193	RL	bUs zum HAUPTbahnhof.
		bus to the main station
		(2.3)
194	GA	die BESte möglichkeit-
		the best option
195		um mit dem BUS zum nächsten dortmund- (.)
		to go by bus to the next dortmund
196		HUCKarde; (.)
		huckarde
197		HAUPTbahnhof-
		main station
198		an der adresse DORTmund zu kommen; (.)
		at the address dortmund
199		ist die linie NEUN um siebzehnuhrachtzehn von lückersweg.
		is line nine at seventeen ten from lückersweg
200		(0.2)
201	RL	NE-
		no
202		(0.3)
203	GA	DAS ist von-
		that is from
204		[hier etwa VIER min-]
		here about four min-
205	RL	[hey GOOgle;]
206		(0.3)
207	RL	wo ist die nächste bushalteSTELle;
208		(2.2)
209	GA	ich habe in deiner nähe mehrere einträge zum thema BUShaltestelle-
		I've found several entries on the topic 'bus stop'

210 [geFUNden (.) der-]
the
211 LS: [(in deiner MAIL?)]
in your mail
212 GA: ERSte ist der hohler weg.
first one is hohler weg
213 (0.3)
214 GA: der ZWEIte ist küstenbecke a-
the second one is küstenbecke A
215 SPARkassencenter;
banking center
216 (0.4)
217 GA: der DRITte ist küstenbecke c (.)
sparkassencenter.
the third one is küstenbecke C banking center
218 RL: ((lacht))
laughs
219 (0.6)
220 LS: xxx;
221 (0.4)
222 RL: ja (sind halt alles nicht die NÄCHsten xxx)-
well, they are all just not the next (xxx)
223 (5.4)
224 RL: (wo IS dat;)
where is it
225 (1.5)
226 RL: ((incomprehensible))
227 (1.1)
228 RL: ((incomprehensible, loud noise))
229 (2.4)
230 RL: hh°
231 (0.6)
232 RL: hey GOOgle;
233 (0.4)
234 RL: wAnn kommt der BUS;
when is the bus coming
235 (2.3)
236 GA: zum NÄCHsten hohler weg kommst du mit dem
auto am schnellsten-
the fastest way to get to the next hohler weg
by car is
237 RL: [((lacht))]
laughs
238 GA: [über golmer STRAße und-]
via golmer straße and

```
239          [schmieDEstraße;]
              schmiedestraße
240   RL:    [hey GOOgle;    ]
241          (0.4)
242   RL:    stopp- (.)
243   GA:    die fahrt DAUert-
             the ride takes
244   GA:    [UNgefähr sechs-]
              about six
245   RL:    [hey GOOgle;    ]
246   GA:    [miNUT-]
247   RL:    [STOPP,]
248          (1.7)
249   GA:    okAY.
250          (0.4)
251   GA:    ABgebrochen.
             cancelled
252   RL:    hey GOOgle;
253          (0.6)
254   RL:    du bist NUTZlos;
             you are useless
255          (1.3)
256   GA:    naJA,
             well
257          (0.3)
258   GA:    ich LERne noch.
             I am still learning
```

Before the transcript starts, Robin's first attempt to obtain information about the bus connection had just failed; Lara's utterance (l. 192) reflects the frustration that is already beginning to set in. In the following minutes, the Google Assistant repeatedly produces voice outputs that do not match the desired query – mainly because an incorrect bus stop was selected as the starting point and the best stop is not identified even when queried directly. An opportunity for a local initiation of repair (Schegloff, Jefferson, and Sacks 1977, 374) does not present itself: There is no possibility for a *short* intervention or correction within the longer utterances of the VUI – the only option is to cancel whole oral text pieces. Additionally, unlike written search query results in a browser or smartphone navigation app, the temporality of speech production (Auer 2000) renders the information fleeting, inextricable, and difficult to compare. Identifying the source of trouble is therefore particularly difficult (see also Porcheron et al. 2018, 10; Garg, Cui, and Kapida 2021). Presets of the VUI, such as the selection of a car route instead of a bus route, cannot be viewed in advance in order to check or amend them before making the query, which results in the continued verbalization of the unwanted information, as in l. 244. As the dialogue progresses, the user inputs become increasingly non-specific, more and more general. Whereas in one of the first voice inputs (uttered before the excerpt above) Robin specified very precisely when he wanted to travel from which bus stop to which destination, the last attempt is reduced to the question "When is the bus coming" (l. 234), with no local specification for start or end of the trip or for the time. This trajectory seems to result from repeated disappointments and failed trial-and-error attempts, leading from resignation to failure and a dwindling willingness to make any (linguistic) effort (see also Hector, in preparation).

5. From Smart Speakers to Smart Homes: An Outlook

With our analyses, we have been able to demonstrate that the social practices performed in interactions with smart speakers are already established everyday practices that are undertaken in changing medial conditions and hence undergo continual modification and transformation – consequently their accomplishment or non-accomplishment. Users have to sequentially cope with 'hybrid' dialogue systems that have been designed to simulate social interaction to a certain extent, yet also repeatedly deviate from it in ways that disrupt conversation. Thus, systems occasionally initiate the kind of phatic communication ("ich hoffe (.) du hattest einen TOLlen mittwoch./ I hope you had a great Wednesday") associated with small talk (Senft 2009), but sometimes they do so at inappropriate moments, in unsuitable contexts, and without the competence to respond appropriately to a follow-up move. In Example 4 ("When is the next bus?"), well-known problems from human–machine exchange as they have been described by Suchman (2007) occur: The dialogue design of the machine is rule-based, with the rules remaining opaque for the users, and the humans' situated utterances are only recognizable for the technology when they fit within the framework it 'anticipates'. These fundamental differences between the ways humans and machines process linguistic signals prevent the accomplishment of real "interaction", with openness and situation-boundedness presenting the most challenging characteristics of human interaction that machines need to cope with (Hector and Hrncal 2024). Resources that can increase the user-friendliness of interfaces, such as visualization and written elements, were not available for smart speakers at the time of our research – but since then, some more recent generations have incorporated screens.

What also becomes clear is that VUIs can sometimes be perceived as so bulky and dysfunctional that their domestication fails – despite the greatest linguistic effort. In cases when neither the user(s) nor the system achieve any situational agency – understood praxeologically (Krummheuer 2015) –, a recovery strategy can be observed on the part of the users: They assure one another of their assumed autonomy and superiority on a meta-level, for example through 'arrogant' jokes about or insults of the pseudo-social counterpart in its co-presence, such as the utterance "You are useless" in the last example. If one takes this user at his word, it becomes clear that the domestication of assistance systems can – in extreme cases – fail as a result of their voice interfaces' limitations.

Despite economic inefficiencies associated with VUIs (Amadeo 2022), their significance in consumer technology is apparently set to remain high due to the strategic ambitions of service providers such as Amazon, that appear to be aiming to establish VUIs as pivotal control hubs for smart home environments (Strüver 2023). As smart home technologies become increasingly sophisticated and data-intensive, the integration of advanced sensor technology into private residences is expected to grow, enhancing the capabilities and appeal of VUIs. Looking ahead, there are notable announcements suggesting the deployment of large language models (LLMs) within VUIs (see also Strüver, this volume). This development holds the potential to significantly enhance the conversational abilities and functionality of smart speakers. However, at least at the time of writing this chapter, experts point to hurdles for such technological evolution, at least for Amazon's smart speaker, resulting from path dependencies in the architecture (see Eric 2024). Additionally, there is speculation about discontinuation of Google's current smart speaker models, with a possible shift towards integrating newer voice assistant technology into tablets. Hence, linguistic practices may be as ephemeral as the technologies in relation to which they emerge – while at the same time, linguistic practices with VUIs demonstrate impressively the flexibility of competent speakers to adapt their practices for organization of talk to the organization of sociotechnical exchange.

Acknowledgements

We would like to express our sincere gratitude to Sarah Diehl and Franziska Niersberger-Guéye for their invaluable assistance with the transcription work and for their helpful comments on earlier drafts of this manuscript.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 262513311 – SFB 1187 Media of Cooperation. *Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 262513311 – SFB 1187 Medien der Kooperation.*

References

Amadeo, Ron. 2022. "Amazon Alexa is a 'colossal failure', on pace to lose $10 billion this year." *Ars Technica*, November 21, 2022. https://arstechnica.com/gadgets/2022/11/ amazon-alexa-is-a-colossal-failure-on-pace-to-lose-10-billion-this-year/.

Albert, Saul, Magnus Hamann, and Elizabeth Stokoe. 2023. "Conversational User Interfaces in Smart Homecare Interaction: A Conversation Analytic Case Study" In *Proceedings of 5th International Conference on Conversational User Interfaces*, edited by Minha Lee et al. New York: ACM. https://doi.org/10.1145/3571884.3597140.

Auer, Peter. 2000. "On line-Syntax – Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen." *Sprache und Literatur* 31 (1): 43–56. https://doi.org/10.30965/25890859-031-01-90000005.

Ayaß, Ruth. 2012. "Introduction. Media Appropriation and Everyday Life." In *The Appropriation of Media in Everyday Life,* edited by Ruth Ayaß and Cornelia Gerhardt, 1–16. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.224.01aya.

Baldauf, Heike. 2002. *Knappes Sprechen*. Tübingen: Niemeyer.

Barthel, Mathias, Henrike Helmer, and Silke Reineke. 2023. "First users' interactions with voice-controlled virtual assistants: A micro-longitudinal corpus study." *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue – SemDial*. PrePrint. Accessed August 22, 2024. https://mezzanine.um.si/wp-content/uploads/ Marilogue_Proceedings.pdf.

Baumann, Melanie. 2023. "Cheat-Code für den Super-Alexa-Modus. Amazon Alexa Cheat Code anwenden – so geht's". Accessed August 22, 2024. https://www.homeandsmart.de/alexa-cheat-code.

Blattner, Meera, Denise Sumikawa, and Robert Greenberg. 1989. "Earcons and Icons: Their Structure and Common Design Principles." *Human–Computer Interaction* 4 (1): 11–44. https://doi.org/ 10.1207/s15327051hci0401_1.

Brevini, Benedetta. 2023. "Myths, techno solutionism and artificial intelligence: reclaiming AI materiality and its massive environmental costs." In

Handbook of Critical Studies of Artificial Intelligence, edited by Simon Lindgren, 869–877. Chaltenham and Northampton: Edward Elgar.

Brock, Alexander and Peter Schildhauer, eds. 2017. *Communication Forms and Communicative Practices. New Perspectives on Communication Forms, Affordances and What Users Make of Them.* Frankfurt a.M.: Peter Lang.

Couldry, Nick, and Ulises Mejias. 2019. "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject." *Television & New Media* 20 (4): 336–349. https://doi.org/10.1177/1527476418796632.

Crawford, Kate, and Vladan Joler. 2018. "Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources." Accessed August 22, 2024. https://anatomyof.ai/.

Crawford, Kate. 2021. *Atlas of AI – Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale: Yale University Press.

Deppermann, Arnulf. 2018. "Instruction practices in German driving lessons: Differential uses of declaratives and imperatives." *International Journal of Applied Linguistics* 28 (2): 265–282. https://doi.org/10.1111/ijal.12198.

Deppermann, Arnulf, Helmuth Feilke, and Angelika Linke, eds. 2016a. *Sprachliche und kommunikative Praktiken.* Berlin and Boston: de Gryuter.

Deppermann, Arnulf, Helmuth Feilke, and Angelika Linke 2016b. "Sprachliche Praktiken. Eine Annäherung aus linguistischer Sicht." In *Sprachliche und kommunikative Praktiken,* edited by Arnulf Deppermann, Helmuth Feilke, and Angelika Linke, 1–23. Berlin and Boston: de Gryuter.

Dolata, Ulrich. 2019. "Plattform-Regulierung. Koordination von Märkten und Kuratierung von Sozialität im Internet." *Berliner Journal für Soziologie* 29: 179–206. https://doi.org/10.1007/s11609-020-00403-9.

Drösser, Christoph. 2020. *Wenn die Dinge mit uns reden. Von Sprachassistenten, dichtenden Computern und Social Bots.* Berlin: Dudenverlag.

Eisenmann, Clemens, Jakub Mlynář, Jason Turowetz, and Anne W. Rawls. 2023. "'Machine Down': making sense of human–computer interaction – Garfinkel's research on ELIZA and LYRIC from 1967 to 1969 and its contemporary relevance." *AI & Society*. https://doi.org/10.1007/s00146-023-01793-z.

Endter, Cordula, Florian Fischer, and Tobias Wörle. 2023. "Doing Home by Using Digital Assistive Technologies. On the Role of Meaning and Materiality in the use of Digital Assistive Technologies at Older People's Domesticity." *Digital Culture & Society* 9 (1): 179–200. https://doi.org/10.14361/dcs-2023-0109.

Eric, Mihail. 2024. "How Alexa Dropped the Ball on Being the Top Conversational System on the Planet." Accessed August 22, 2024. https://www.mihaileric.com/posts/how-alexa-dropped-the-ball-conversational-ai/.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology.* Englewood Cliffs: Prentice Hall.

Garg, Radhika, Hua Cui, and Yash Kapadia. 2021. "'Learn, Use, and (Intermittently) Abandon': Exploring the Practices of Early Smart Speaker Adopters in Urban India." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–28. https://doi.org/10.1145/3479495.

Goodwin, Charles. 2018. *Co-Operative Action.* New York: Cambridge University Press.

Habscheid, Stephan. 2016. "Handeln in Praxis. Hinter- Und Untergründe Situierter Sprachlicher Bedeutungssituation." In *Sprachliche und kommunikative Praktiken*, edited by Arnulf Deppermann, Helmuth Feilke, and Angelika Linke, 127–51. Berlin and Boston: de Gruyter.

Habscheid, Stephan. 2023. "Socio-technical dialogue and linguistic interaction. Intelligent Personal Assistants (IPA) in the Private Home." *Sprache und Literatur* 51 (2): 167–196. https://doi.org/10.30965/25890859-05002020.

Habscheid, Stephan, Tim Hector, and Christine Hrncal. 2023. "Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology". *Social Interaction. Video-Based Studies of Human Sociality* 6 (1): 1–31. https://doi.org/10.7146/si.v6i1.137378.

Habscheid, Stephan, Tim Hector, Christine Hrncal, and David Waldecker. 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. https://doi.org/10.21248/jfml.2021.44.

Hanks, William F. 1996. *Language and Communicative Practices.* Oxford and Boulder: Westview Press.

Hausendorf, Heiko. 2015. "Interaktionslinguistik". In *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*, edited by Ludwig M. Eichinger, 43–70. Berlin and Boston: de Gruyter.

Hausendorf, Heiko. 2022. "'Telekopräsenz'. Interaktionslinguistische Anmerkungen zu einer Kommunikationsbedingung im Wandel." In *Brückenschläge: Linguistik an den Schnittstellen*, edited by Sarah Brommer, Kersten Sven Roth, and Jürgen Spitzmüller, 205–44. Tübingen: Narr Francke Attempto.

Hector, Tim. 2023. “Smart Speaker in der Praxis. Methodologische Überlegungen zur medienlinguistischen Erforschung stationärer Sprachassistenzsysteme.” *Sprache und Literatur* 51 (2): 197–229. https://doi.org/10.30965/25890859-05002021.

Hector, Tim. Forthcoming. “Sprachassistenzsysteme und ihre Interfaces: eine medienlinguistische Analyse” In *Un/reale Interaktionsräume. Formen sozialer Ordnung im Spektrum medienspezifischer Interaktion*, edited by Clara Kindler-Mathôt, Didem Leblebici, Giacomo Marinsalta, Till Rückwart, and Anna Zaglyadnova. Bielefeld: transcript.

Hector, Tim. In preparation. *Smart Speaker als Gesprächsbeteiligte? Sprachliche Praktiken im Prozess der Domestizierung mit Voice-User-Interfaces.* Dissertation, University of Siegen, Faculty of Arts and Humanities – Faculty I. Defense: April 12, 2024.

Hector, Tim Moritz, and Christine Hrncal. 2020. “Intelligente Persönliche Assistenten im häuslichen Umfeld. Erkenntnisse aus einer linguistischen Pilotstudie zur Erhebung audiovisueller Interaktionsdaten.” *Working Paper Series Media of Cooperation* 14: 1–14. http://dx.doi.org/10.25819/ubsi/1013.

Hector, Tim Moritz, and Christine Hrncal. 2024. “Sprachassistenzsysteme in der Interaktion”. In *Sprache und digitale Kommunikation*, edited by Jannis Androutsopoulos und Friedemann Vogel, 309–28. Berlin and Boston: de Gruyter.

Hector, Tim, Franziska Niersberger-Gueye, Franziska Petri, and Christine Hrncal. 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: 1–15. https://doi.org/10.25819/ubsi/10168.

Hepp, Andreas, Wiebke Loosen, Stephan Dreyer, Juliane Jarke, Sigrid Kannengießer, Christian Katzenbach, Raina Malaka, Michaela Pfadenhauer, Cornelius Puschmann, and Wolfgang Schulz. 2023. “ChatGPT, LaMDA, and the Hype Around Communicative AI: The Automation of Communication as a Field of Research in Media and Communication Studies.” *Human-Machine-Communication* 6: 41–63. https://doi.org/ 10.30658/hmc.6.4.

Hirschauer, Stefan. 2004. “Praktiken und ihre Körper. Über die materiellen Partizipanden des Tuns.” In *Doing Culture. Neue Positionen zum Verhältnis von Kultur und sozialer Praxis*, edited by Julia Reuter and Karl H. Hörning, 73–91. Bielefeld: transcript.

Hirschauer, Stefan. 2016. “Verhalten, Handeln, Interagieren. Zu den mikrosoziologischen Grundlagen der Praxistheorie.” In *Praxistheorie. Ein soziologi-*

sches Forschungsprogramm, edited by Hilmar Schäfer, 45–67. Bielefeld: transcript. https://doi.org/10.1515/9783839424049-003.

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. 2005. *Reassembling the Social. An Introduction to Actor-Network-Theory.* Oxford: Oxford University Press.

Levina, Anastasia, Igor Ilin, Dayana Gututishvili, Kristina Kochetkova, and Andrea Tick. 2024. "Towards a Smart Hospital: Smart Infrastructure Integration." *Journal of Open Innovation: Technology, Market, and Complexity* 10 (3): 1–12.

Liesenfeld, Andreas, Alianda Lopez, and Mark Dingemanse. 2023. "Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators." *Proceedings of the 5th International Conference on Conversational User Interfaces*, 1–6. https://doi.org/10.1145/3571884.3604316.

Lind, Miriam, and Sascha Dickel. 2024. "Speaking, but having no voice. Negotiating agency in advertisements for intelligent personal assistants." *Convergence* 30 (3), 1008–1024. https://doi.org/10.1177/13548565231192100.

Lotze, Netaya. 2016. *Chatbots.* Frankfurt a.M.: Peter Lang.

Lotze, Netaya. 2020. "Künstliche Intelligenz im Dialog – Ein methodologisches Konzept zur Analyse von Mensch-Maschine-Interaktionen." In *Deutsch in Sozialen Medien*, edited by Konstanze Marx and Henning Lobin, 363–67. Berlin, Boston: de Gruyter.

Luckmann, Thomas. 1986. "Grundformen der gesellschaftlichen Vermittlung des Wissens: Kommunikative Gattungen." In *Kultur und Gesellschaft* (= *Kölner Zeitschrift für Soziologie und Sozialpsychologie.* Sonderheft 27), edited by Friedhelm Neidhardt, M. Rainer Lepsius, and Johannes Weiß, 191–211. Opladen: Westdeutscher Verlag.

Meiler, Matthias. 2019. "Zur praxeologischen Verhältnisbestimmung von Materialität, Medialität und Mentalität oder: Medien als Praxis." *Zeitschrift für Semiotik* 41 (1–2): 63–88. https://doi.org/10.14464/zsem.v41i1-2.709.

Merkle, Benedikt, and Tim Hector. Forthcoming. "Werkzeuge und Medienpraktiken. Intelligente persönliche Assistenten und das Paradigma objektorientierten Programmierens." In *Akustische Interfaces. Interdisziplinäre Perspektiven auf Schnittstellen von Technologien, Sounds und Menschen*, edited by

Christoph Borbach, Timo Kaerlein, Robert Stock, and Sabine Wirth. Wiesbaden: Springer Vieweg.
Mondada, Lorenza. 2014. "Instructions in the operating room. How the surgeon directs their assistant's hands." *Discourse Studies* 16 (2): 131–161. https://doi.org/10.1177/1461445613515325.
Mondada, Lorenza. 2016. "Conventions for multimodal transcription." Accessed June 15, 2024. https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/ mondada_multimodal_conventions.pdf.
Nass, Clifford, and Youngme Moon. 2000. "Machines and Mindlessness: Social Responses to Computers." *Journal of Social Issues* 56 (1): 81–103. https://doi.org/10.1111/0022-4537.00153.
Natale, Simone, and Henry Cooke. 2021. "Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces." *Media, Culture & Society* 43 (6): 1000–1016. https://doi.org/10.1177/0163443720983295.
Natale, Simone, and Andrea L. Guzman. 2022. "Reclaiming the human in machine cultures." *Media, Culture & Society* 44 (4): 627–637. https://doi.org/10.1177/01634437221099614.
Pentzold, Christian, and Andreas Bischof. 2019. "Making affordances real: socio-material prefiguration, performed agency, and coordinated activities in human–robot communication." *Social Media + Society* 5 (3): 1–11. https://doi.org/10.1177/2056305119865472.
Phan, Thao. 2017. "The Materiality of the Digital and the Gendered Voice of Siri." *Transformations* 29: 23–33. https://www.transformationsjournal.org/.
Phan, Thao. 2019. "Amazon Echo and the Aesthetics of Whiteness." *Catalyst: Feminism, Theory, Technoscience* 5 (1): 1–39. https://doi.org/10.28968/cftt.v5i1.29586.
Porcheron, Martin, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. "Voice Interfaces in Everyday Life." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*. 1–12. https://doi.org/10.1145/3173574.3174214
Reckwitz, Andreas. 2003. "Grundelemente einer Theorie sozialer Praktiken." *Zeitschrift für Soziologie* 32: 282–301.
Reeves, Stuart, and Martin Porcheron. 2023. "Conversational AI: Respecifying Participation as Regulation". In *The SAGE Handbook of Digital Society*, edited by William Housley, Adam Edwards, Roser Beneito-Montagut, and Richard Fitzgerald, 573–92. London: Sage.
Roe, Jasper, and Mike Perkins. 2023. "'What they're not telling you about ChatGPT': exploring the discourse of AI in UK news media headlines." *Humani-*

ties and Social Sciences Communications 10: 1–9. https://doi.org/10.1057/s41599-023-02282-w.

Schegloff, Emanuel A., Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53 (2): 361–382.

Schatzki, Theodore R. 2002. *The Site of the Social. A Philosophical Account of the Constitution of Social Life and Change.* University Park, PA: Pennsylvania State University Press.

Schiff, Daniel S., and Rinat B. Rosenberg-Kima. 2023. "AI in Education: Landscape, Vision and Critical Ethical Challenges in the 21st Century." In *Handbook of Critical Studies of Artificial Intelligence*, edited by Simon Lindgren, 804–14. Cheltenham, Northampton: Edward Elgar.

Schmitt, Reinhold. 2003. "Inszenieren: Struktur und Funktion eines gesprächsrhetorischen Verfahrens." *Gesprächsforschung* 4: 186–250. http://www.gespraechsforschung-online.de/.

Schneider, Jan Georg. 2017. "Medien als Verfahren der Zeichenprozessierung: Grundsätzliche Überlegungen zum Medienbegriff und ihre Relevanz für die Gesprächsforschung". In *Gesprächsforschung* 18: 34–55. http://www.gespraechsforschung-online.de/.

Schüttpelz, Erhard. 2017. "Infrastructural Media and Public Media". *Media in Action* 1(1): 13–61. http://dx.doi.org/10.25819/ubsi/7935.

Schüttpelz, Erhard, and Christian Meyer. 2017. "Ein Glossar zur Praxistheorie. ‚Siegener Version' (Frühjahr 2017)". *Navigationen* 17 (1): 155–163. https://doi.org/10.25969/mediarep/1762.

Senft, Gunter. 2009. "Phatic communion" In *Culture and language use*, edited by Gunter Senft, Jan-Ola Östman, and Jef Verschueren, 226–33. Amsterdam John Benjamins. https://doi.org/10.1075/hoph.2.20sen.

Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, and Susanne Uhmann. 2011. "A system for transcribing talk-in-interaction: GAT 2." *Gesprächsforschung* 12: 1–51. http://www.gespraechsforschung-online.de/.

Silverstone, Roger, Eric Hirsch, and David Morley. 1992. "Information and communication technologies and the moral economy of the household." In *Consuming Technologies. Media and Information in Domestic Spaces*, edited by Roger Silverstone and Eric Hirsch, 13–28. London: Routledge.

Strengers, Yolande, and Jenny Kennedy. 2020. *The smart wife. Why Siri, Alexa, and other smart home devices need a feminist reboot.* Cambridge: The MIT Press.

Strüver, Niklas. 2023. "Frustration Free: How Alexa Orchestrates the Development of the Smart Home." *Digital Culture & Society* 9 (1): 99–123. https://doi.org/10.14361/dcs-2023-090106.

Suchman, Lucy. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions. 2nd Edition.* Cambridge: Cambridge University Press.

Waldecker, David and Tim Hector. 2023. "A Praxeological Approach Towards the Domestication of Connected Media Technologies. Introduction." *Digital Culture & Society* 9 (1): 5–22. https://doi.org/10.14361/dcs-2023-0102.

Waldecker, David, Tim Hector, and Dagmar Hoffmann. 2024. "Intelligent Personal Assistants in practice. Situational agencies and the multiple forms of cooperation without consensus." *Convergence* 30 (3), 975–91. https://doi.org/10.1177/13548565231189584.

Wittgenstein, Ludwig. 1984. "Philosophische Untersuchungen." In *Ludwig Wittgenstein. Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchunngen. Werkausgabe Band 1.* Frankfurt a.M.: Suhrkamp.

Woods, Heather. 2021. "Smart homes: domestic futurity as Infrastructure." *Cultural Studies* 35 (4–5): 876–899. https://doi.org/10.1080/09502386.2021.1895254.

Zoeppritz, Magdalena. 1985. "Computer Talk?" *Technical Report TN 85.05.* Heidelberg: IBM Germany Heidelberg Scientific Center.

III Privacy and Data Protection as Practical Problems

Glitch Studies and Smart Speakers

A Spotlight on User Experiences of Unexpected Behaviors

Christoph Lutz and Gemma Newlands

Abstract *Smart speakers have been widely adopted but come with substantial privacy risks, touching on different privacy types such as informational, social, and physical privacy. Scholars have increasingly studied the privacy implications of smart speakers, finding that users tend to have limited privacy concerns and engage infrequently in privacy protection behavior. Extant research also stresses the contextual and situated nature of privacy around smart speakers, pointing to relevant affordances of the technology. Despite these knowledge advancements, a glitch studies perspective on smart speaker interactions and privacy is notably missing. The glitch studies approach was developed by Rosa Menkman at the intersection of art, technology, and critical social research. It directs the attention to glitches as seemingly small and mundane but powerful moments of interruption that allow for reflection and have productive character. In this contribution, we introduce a glitch studies perspective to the investigation of smart speakers and privacy, showing its fruitfulness. We first discuss the literature on smart speakers and privacy, before providing a concise overview of the glitch studies approach. We then present our data and methodological approach. Based on open text responses from an online survey in the United Kingdom, we identify four types of smart speaker glitches: randomly starting to talk or carry out unexpected activities, misinterpreting the user, technical issues related to connectivity, and violating social and contextual norms. Each glitch type is described in turn, with quotes from the survey as illustrative examples. We conclude with a short summary, some implications for research and policy, as well as directions for future research.*

1. Introduction

Smart speakers are voice-controlled mobile devices that use artificial intelligence (AI) in the form of natural language processing to perform hedonic and functional tasks, such as playing music, setting reminders, and retrieving information (Lau, Zimmerman, and Schaub 2018). Normally located in the home and often embedded within a broad arrangement of smart devices, smart speakers have quickly become mainstream. In the United Kingdom, 42% of households had a smart speaker in the first quarter of 2023 (Ofcom 2023), while around 35% of adult United States residents owned at least one smart speaker in 2022 (Schwartz 2022). Amazon Alexa-enabled speakers, such as the Amazon Echo, are the clear market leaders, followed by Google Assistant-enabled speakers and Apple Siri-enabled speakers (Feldman 2018).

Starting around 2017, in line with the increasing adoption of smart speakers, there has been considerable academic interest in this emerging technology (e.g., Brause and Blank 2020, 2023; Hoy 2018; Kang and Oh 2023; Lutz and Newlands 2021; Mols, Wang, and Pridmore 2022; Pridmore and Mols 2020; Smith 2018; Waldecker, Hector, and Hoffmann 2024). However, the terminology used is diffuse, including terms such as voice assistants (Fetterolf and Hertog 2023; Gruber et al. 2021), smart speakers (Kang and Oh 2023; Lutz and Newlands 2021), smart speaker assistants (Brause and Blank 2020, 2023), mobile virtual assistants (Guzman 2019), digital personal assistants (Hurel and Couldry 2022), intelligent personal assistants (Mols, Wang, and Pridmore 2022), and conversational agents (Mariani, Hashemi, and Wirtz 2023). As Lutz and Newlands (2021, 149) note, "users frequently conflate the intelligent assistant and the device. Amazon Echo, the material smart speaker, is often thought of as being 'Alexa.'" Fetterolf and Hertog (2023) qualify that "Echo refers to the smart speaker (the device), while Alexa refers to the VA [voice assistant] within it" (14). In this chapter, we chose term "smart speaker" to refer to the assemblage of both the material device (e.g., the Amazon Echo) and the AI system embedded within in (e.g., Alexa). Neither the disembodied voice assistant without the device nor the embodied device without the voice assistant are sufficient to understand the technology at hand, given the entanglement of material, spatial, temporal, and discursive aspects.

The research landscape on smart speakers is interdisciplinary, including contributions from the social sciences, the technical sciences, and business and management studies (Minder et al. 2023). While much of the literature is published in computer science and human-computer interaction (Feng, Fawaz, and Shin 2017; Geeng and Roesner 2019; Lau, Zimmermann, and Schaub 2018; Luger and Sellen 2016; Malkin et al. 2019; Zheng et al. 2018), recent years have seen dynamic interest from the social sciences, including emerging empirical evidence from communication and media studies (Brause and Blank 2020, 2023; Lutz and Newlands 2021; Pridmore et al. 2019; Vitak et al. 2023; Waldecker, Hector, and Hoffmann 2024). In that area, human-machine communication (HMC) has advanced our understanding of communication modalities with smart speakers and related technologies such as social robots (Guzman 2017, 2019; Lutz and Tamó-Larrieux 2020). Privacy, in particular, presents a prominent angle to approach the topic (Lutz 2023), given the sensitive nature of the data collected and the use of smart speakers in domestic settings. We will accordingly frame this chapter strongly within privacy discourse on smart speakers. However, we will rely on an under-used theory in the space, namely *glitch studies* (Menkman 2011).

To our knowledge, this is the first contribution to apply glitch studies to smart speakers. Glitches, and any other unexpected behavior by the smart speaker, risk user confidence in the technology and may raise concerns over the privacy and security of user data. However, glitches may in some instances enhance the relationship between the user and smart speaker, such as by triggering anthropomorphizing scripts. Glitch studies offers a fruitful lens to study smart speakers because it acknowledges the fallibility and imperfection of technologies, centering on seemingly small and mundane instances of frailty rather than large breakdowns. Glitch studies is especially helpful for understanding privacy in the context of smart speakers, thus heeding the call by Newlands et al. (2020) for an increased attention to privacy violations as a result of mundane technical breakdowns, possibly stemming from rushed innovation pathways. Focusing on glitches also points to alternative trajectories of technical development, opening the scope for counter- and off-label uses as well as creative, artistic, and playful design (Kemper 2023).

Drawing on rich textual data from an online survey among 369 smart speaker users in the United Kingdom, we explore the user experience of glitches and overarching themes, connecting glitches to privacy perceptions. In the remainder of the chapter, we provide a short literature review on privacy and smart speakers as well as glitch studies. We then describe the survey and data, present the findings in the form of four key themes and accompanying quotes, and conclude with a synthesizing discussion as well as outlook.

2. Literature Review

Privacy and Smart Speakers

In a recent systematic literature review of voice assistants in private households, Minder et al. (2023) identified nine topical clusters and four overarching streams: conceptual foundations (stream 1), systemic challenges, enabling technologies and implementation (stream 2), efficiency (stream 3), and applications and use cases (stream 4). In this review, the authors show that computer science is the area with the highest number of relevant publications (197), followed by the social sciences (52), and business and management (20). Within the social sciences, the US takes the first place in terms of output (19 publications), followed by India and the UK (5 each), Germany and Japan (4 each), and Australia and the Netherlands (3 each). Across the four streams and nine clusters, there is an overarching "focus on users' perceived privacy risks and concerns and a focus on the impact of perceived risks or concerns on the adoption of VA technology" (9). The authors also note a lack of cross-disciplinary engagement.

Another recent systematic literature review, based on 89 publications, focuses on privacy and security in smart speakers exclusively (Maccario and Naldi 2023). The review highlights smart speaker research as a growing trend over the last four years, with most contributions emerging from the United States. The literature encompasses five themes: privacy concerns, factors influencing adoption, identification of vulnerabilities, development of countermeasures, and legal issues. Interestingly, and again in line with the broader review by Minder et al. (2023), most of the research centers on technical aspects (vulnerabilities and countermeasures), showing a strong focus on concrete privacy and security problems. While privacy concerns act as a pronounced deterrent in adopting smart speakers, the literature presents other factors such as platform-related variables, connectivity, technology optimism, and functional versatility. Despite fewer papers on legal issues, Maccario and Naldi (2023) anticipate an increase in this area. Together, the two reviews show the need for a multi-dimensional understanding of privacy in smart speakers that goes beyond security and considers contextual, social and legal elements.

Recent literature in media and communication studies and the interdisciplinary areas of critical data studies, science and technology studies and Internet studies have enhanced our understanding of privacy in the context of smart speakers. Table 1 shows an overview of such studies with their methodological approach, main theories used, and key results.

Table 1: Summary of User-Centered Research on Privacy and Smart Speakers

Publication	Methodology	Key Theories and Results
Brause and Blank 2020	12 semi-structured interviews with smart speaker users (qualitative)	Domestication theory; Identified six smart speaker use genres (companionship, self-control and productivity, sleep aid, health care, peace of mind, increased accessibility) and spatially distributed uses based on users' perception.
Brause and Blank 2023	12 semi-structured interviews with smart speaker users (qualitative)	Privacy work and privacy as contextual integrity theory; Revealed new types of privacy work and rationales, suggesting an expansion of the model of contextual integrity to understand privacy perceptions with smart speakers.
Fetterolf and Hertog 2023	16 semi-structured interviews with young adult Alexa users (qualitative)	Privacy and trust literature (e.g., digital resignation); Three strategies to manage distrust in smart speaker company: anthropomorphism, digital resignation, and occasionally taking protective action.
Gruber et al. 2021	83 semi-structured interviews conducted in 5 countries (qualitative)	Found awareness of automatic decision-making in voice assistants, influenced by experiences with other digital devices and services.
Hurel and Couldry 2022	Thematic analysis of documents from Amazon and Google, news coverage and academic research (qualitative)	Data colonialism and dataveillance literature: Examines different aspects of data extraction of the home in the vein of data colonization: territorializing the home for data extraction, redirecting the user to specific actions, justifying data accumulation
Kang and Oh 2023	Survey of 474 smart speaker users in the US (quantitative)	Privacy literature and theories (privacy calculus, privacy management theory, protection motivation theory); Examined three privacy management strategies in smart speaker use: privacy disclosure, boundary linkage, and boundary control.

Publication	Methodology	Key Theories and Results
Lutz and Newlands 2021	Survey of 367 smart speaker users in the UK (quantitative)	Contextual integrity theory, privacy cynicism; Privacy concerns about third parties are most pronounced, with privacy protection behaviors being uncommon, but affected by privacy concerns and motives.
Mols, Wang, and Pridmore 2022	Combination of survey with 291 university employees and focus groups with 35 participants, both in the Netherlands (mixed methods)	Affordances; Develops a multidimensional understanding of privacy concerns around household smart speakers, differentiating between surveillance, security, and platform concerns.
Pridmore and Mols 2020	9 semi-structured interviews, 6 focus groups with 35 university personnel and 5 focus groups with 22 young adults in the Netherlands (qualitative)	Surveillance capitalism, technology adoption; Highlights the complexity of data production at a household level and how smart speakers produce myopic views of users for platforms.
Vitak et al. 2023	11 focus groups with 65 US adult users and non-users (qualitative)	Communication privacy management theory; Investigates differences in attitudes and concerns toward voice assistants and how attitudes are influenced by device features.
Waldecker, Hector, and Hoffmann 2024	Eight German households: conversation analysis of audio and video material (actual smart speaker use) and interviews (qualitative)	STS literature (e.g., boundary objects), and surveillance capitalism; Examines the agencies of users and device/service providers, discussing how these are intertwined and distributed.
Xu, Chan-Olmsted, and Liu 2022	Survey of 991 participants' attitudes and behavior patterns related to smart speaker use (quantitative)	Uses and gratifications theory and privacy management literature; Explores the gratifications that users seek and identifies main strategies for privacy management, highlighting two routes: protective and precautionary.

As the table demonstrates, there is a plurality of both methods and theories used in smart speaker research. Most of the discussed studies used qualitative methods, potentially due to the relative nascency of the technology or the focus on media and communication, critical data studies and STS. Overall, the evidence painted is varied and deep. The following three trends in the literature can be synthesized

a) *Varied use and privacy concerns*: Users employ smart speakers for diverse purposes, from entertainment to healthcare, with prevalent privacy concerns related to data collection and potential misuse.
b) *Awareness and complex privacy management*: There is a growing awareness of the algorithmic functioning of these devices, with some users adapting to privacy concerns through strategies like device manipulation and digital resignation. Research highlights the complexity of privacy management, involving multifaceted strategies that balance perceived risks and benefits, and are influenced by users' privacy self-efficacy. Overall, however, privacy protection seems infrequent and superficial.
c) *Data security and surveillance challenges*: Users exhibit mixed reactions to surveillance and security, with concerns about third-party access and the implications of continuous listening prompting discussions on legal and ethical frameworks in smart speaker technology.

The review so far shows how privacy in the context of smart speakers is not a singular, one-dimensional concept but a multi-faceted phenomenon that requires contextual awareness. Nissenbaum's (2010) theory of contextual integrity is therefore particularly relevant (as applied in Lutz and Newlands 2021). The theory argues that privacy is not about the mere secrecy of information but about the appropriate flow of information depending on social norms and contexts (Apthorpe et al. 2018). Smart speakers, which are often placed in private settings like homes, challenge traditional boundaries and norms associated with information flow. The devices' ability to passively listen and record conversations, even unintentionally, can breach the contextual integrity of a home setting, where certain conversations are presumed to be private and confined within the space.

Glitch Studies

Glitches, often occurring as technical anomalies, can show critical privacy vulnerabilities (Kemper 2023; Menkman 2011). The unintended experiences of glitches offer a richer understanding of the interactions and challenges posed by smart speakers.

A glitch is typically defined as a "short-lived fault in a system, such as a transient fault that corrects itself, making it difficult to troubleshoot" (Wikipedia 2023). Peña and James (2016) describe glitches as "unforeseen behaviors within a system", especially computer systems, but also in related areas such as art and video gaming. In software development and maintenance, a glitch differs from a bug, where the former is more transient and less critical, but still noticeable, whereas the latter presents a more fundamental rupture. In glitch studies, these anomalies are not merely errors to be fixed but opportunities to gain insights into the design, use, and impact of technology. Menkman (2011) coined the term glitch studies and contributed a concise manifesto that is based on extensive artistic engagement. The manifesto challenges the perpetual pursuit of flawless technology, arguing that every new medium will inherently possess its own unique imperfections, so-called "noise artifacts". Glitches are not just errors or failures; instead, they are opportunities for creative exploration and critical examination of technology, including opportunities for serendipity and learning (Peña and James 2016). Menkman (2011) also contextualizes glitch studies within discussions on noise (as in signal transmission and complexity theory) and differentiates hot and cool glitches.

The manifesto ends with eight propositions, which show the academic and artistic scope of glitch studies. The first proposition claims that "[t]he dominant, continuing search for a noiseless channel has been – and will always be – no more than a regrettable, ill-fated dogma." Several of the other propositions call for artistic experimentation and challenge the status quo (e.g., proposition 3: "Get away from the established action scripts and join the avant-garde of the unknown. Become a nomad of noise artifacts!"). At the same time, the manifesto has pragmatic elements and emphasizes temporal aspects of glitches (last part of proposition 5: "Be aware of easily reproducible glitch effects, automated by softwares and plug-ins. What is now a glitch will become a fashion."). In the final two sentences, the approach is synthesized: "Flow cannot be understood without interruption or function without glitching. This is why glitch studies is necessary." Thus, glitch studies is not only an analytic or scholarly program but very much action-oriented. Glitches serve as means to dispute the conventional templates of creative practice, bringing to light hidden dynamics of technology. In the end, glitch studies offer a more critical and reflective engagement with digital media.

While still an under-represented approach in the academic literature, glitch studies has been taken up in pedagogy (James and Peña 2023; Peña and James 2016; Preece and Whittaker 2023), showing its potential for enhanced learning. Peña and James (2016, 123), for example, specify that "[g]litch pedagogy not only instigates the game-sense of learning but celebrates mistakes and processing errors as central to creativity, inquiry, invention, and discovery of processes underlying knowledge construction and mobilization in the twenty-first century". Beyond pedagogy, glitch studies has also been taken up in gender and queer studies (e.g., Linghede 2018), human-computer interaction (Kim, Van Dierendonck, and Poslad 2019), geography and urban studies (Leszczynski 2020) as well as cultural studies (Kemper 2023; Rutten and De Vos 2023). Kemper and Kolkman (2017, 8) apply glitch studies to algorithms, locating within interruptions of expected use "an entryway into the hidden or taken-for-granted logic according to which they operate".

This perspective of creativity, inquiry and invention within glitch studies is also relevant in the context of smart speakers, where glitches can expose the complexities of human-technology interactions and the broader implications of AI-driven devices in private spaces, making users reflect on the technology in a different light. Glitches can range from simple misunderstandings of voice commands to unintended activations and inappropriate or strange responses. Each of these occurrences offers a window into the operational logic of these devices and the user interactions they engender, breaking down the ordinary flow of conversation. Glitches bring to light the intricacies of voice recognition technology, the assumptions embedded within AI algorithms, and the challenges of designing technology that seamlessly integrates into daily life, thus potentially increasing transparency and explainability (Felzmann et al. 2019, 2020). Studying glitches in smart speakers thus provides a highly relevant approach to understanding user experiences.

Glitches in smart speakers can also have broader societal implications. For example, a glitch that inadvertently records a private conversation not only raises privacy concerns but also prompts questions about surveillance, data security, and the ethical responsibilities of technology companies that might prompt media stories and thus public attention (Estes 2018). By examining these glitches, researchers can uncover the often invisible power dynamics and ethical considerations inherent in the deployment of smart speaker technologies. A close reading of glitches also allows to deepen the contextual integrity norms at play around smart speakers (Apthorpe et al. 2018; Lutz and Newlands 2021).

Methodologically, glitch studies require an interdisciplinary approach, combining insights from computer science, sociology, media studies, and design. However, a particular focus is not only on the technology itself but also its understanding among users and what it does to the human-technology relationship. Interactive interviews where users are encouraged to describe their technology use in situ can be employed to study the multifaceted nature and consequences of glitches, such as through the think-aloud protocol (Nielsen, Clemmensen, and Yssing 2002) or algorithm audits (Diakopoulos, Bandy, and Dambanemuya 2023).

3. Methods

We use an in-depth online survey to assess privacy perceptions and glitches around smart speakers. Prolific was the platform of choice for recruiting survey respondents due its flexible screening options, including a question for smart speaker ownership. Moreover, Prolific has comparatively good data quality and makes sure to reward participants ethically (Douglas et al. 2023; Palan and Schitter 2018; Peer et al. 2017, 2021). The survey was launched in October 2019 with a sample size of 369 UK-based respondents. It had a series of closed-ended questions about privacy concerns, privacy protection behavior, motives for using smart speakers and use modalities, which are not used in this chapter, as well as an open-ended question that forms the basis for the analyses here. A more detailed description of the methodological procedure and an overview of the closed-ended questions is available in Lutz and Newlands (2021). Regarding the open-ended question, an open text box queried respondents about incidents where the smart speaker exhibited erratic or unexpected behavior. The exact wording of the question was: "*Please describe below any incidents where the smart speaker exhibited erratic or unexpected behaviour.*" The answers to this question provide the main body of evidence for this study. We analyzed the open text answers through inductive thematic analysis, reading the responses several times and then grouping them based on what emerged as the salient glitch category.

4. Results

Table 2: Common glitches encountered by Amazon Echo users (Note: data Collected in October 2019)

Type of Glitch	**Example Quotes**[1]
1. Randomly starting to talk or carry out unexpected activities	*"my alexa has randomly talked in the middle of the night."* *"Out of the blue, Alexa has started speaking about something completely random. A couple of times, she has just started playing a radio station."* *"randomly speaks sometimes. It was bought by another member of the household. I don't like it, don't trust it, and don't use it."* *"Sometimes starts talking when no one is in the room or it's deadly silent."*
2. Misinterpreting the user	*"Just one time when I asked for it to play Elton John rocket man and it played your song instead."* *"many when asking for music it plays something random."* *"I tried to use the shopping list function and was misheard on most items with some very amusing results."*
3. Technical issues related to connectivity	*"Only on Christmas day. We were setting up the echo and so were thousands of others, which effected the echo to work properly."* *"She usually behaves erratically when the Internet connection is suboptimal."* *"Sometimes it just loses connection unexpectedly."* *"The only time this happened was when an update was incoming,so I have to wait a while for update to take effect."* *"When we moved houses Alexa refused to connect to Internet, kept getting error on the app and she wouldn't say anything more besides she can't connect."*
4. Violation of social and contextual norms	*"I once asked Alexa to turn the lights off and she made a comment about manners and refused to turn the lights off until I said please."* *"When asking to play music, the speaker sometimes tries to make me purchase Amazon Music, even though I regularly and only use Spotify."* *"When chatting in general, someone joked 'be careful Alexa's in this room she'll hear you' and Alexa responded. It was a bit creepy."* *"Wouldn't stop farting even after being told to stop after my daughter asked her to fart."*

1 Quotes are reported unedited for spelling and grammar.

Of the 369 respondents, 264 reported to use an Amazon Echo (71.5%), 74 a Google Home (20.1%), 12 an Apple HomePod (3.3%), 3 (0.8%) another speaker (all of which indicated to use a Sonos), and the remaining 16 respondents (4.3%) were ex-users. In the following, we focus on the Amazon Echo users. Of the 264 Amazon Echo users, 181 respondents (68.6%) had some type of glitch experience, 74 (28.0%) reported no glitch experience, and the remaining nine respondents (3.4%) did not fill out the text box (the question was requested response, with a reminder to those who did not put anything in the text box, but not forced response). Subsequently, we analyzed the comments of Amazon Echo users, identifying four main glitch types.

Table 2 shows four types of common glitches with corresponding quotes. A first common glitch type was about the Amazon Echo randomly starting to talk or initiate activities without any user prompt. The unpredictability of this behavior can be unsettling, especially when it happens in quiet environments or during unconventional hours, like in the middle of the night. Such occurrences can lead users to questions around the reliability and privacy of their device, as they may worry about being listened to or recorded without their consent. The spontaneous activation of the device raises concerns about its autonomous functions and the potential for privacy breaches. The connotations with this glitch are primarily negative, expressed best by quote 3 (*"I don't like it, don't trust it, and don't use it."*)

A second glitch type was with delivering wrong, misleading or unusable results to queries. Here, the glitch is about the Echo misinterpreting user commands, resulting in unexpected or incorrect responses. This can range from playing the wrong song to mishearing items on a shopping list, with both amusing or frustrating outcomes. While these instances may seem trivial, they highlight the limitations of the device's natural language processing capabilities and can erode user trust. Misinterpretations also touch upon privacy concerns, as they imply that the device may not always understand the context or intention behind user interactions, potentially leading to inappropriate or sensitive information being recorded or acted upon incorrectly. Compared to the first glitch type, the connotations seem slightly more mixed and benign (e.g., the relatively narrow glitch experience from quote 1: *"Just one time when I asked for it to play Elton John rocket man and it played your song instead.*[2]*"*).

A third glitch type had to do with technical connectivity and time-outs. In some instances, the connectivity issues resulted in other glitches, for example

2 Reference to songs 'Rocket Man' and 'Your Song' by Elton John.

seemingly incorrect interpretations (see second quote in this category). This type of glitch also occurred with situational changes, for example, when new updates had to be installed or when a person moved house. These technical glitches can affect the usability of the device and may lead to concerns about the stability and reliability of the technology, impacting user trust. Furthermore, connectivity issues can exacerbate privacy concerns if they lead to unanticipated device behavior or failures in executing privacy controls. Compared to the previous glitch types, this glitch seems more situational, with several respondents using modifiers such as "only".

A fourth and final glitch type has to do with violating social norms or communication norms in certain situations. While instances of this type of glitch were rare, they are most interesting in terms of privacy (maybe together with glitch type 1), contextual integrity and social expectations. An interesting instance occurs when commercial aspects seem to cloud the interaction with the user (see second quote of the last category). Other examples include the device making inappropriate comments, refusing to follow commands without polite phrasing, or responding unexpectedly to conversations it overhears. Such behaviors can be perceived as intrusive or creepy, undermining the social acceptability of the device. On the other hand, this glitch type also offers potential for reflection and contextualization among the users, prompting them to question the technology and see the bigger picture.

Taken together, our findings offer an indication of what glitches smart speaker users commonly encounter. In the future development of the research, we aim to connect the qualitative findings with the quantitative data, exploring whether certain types of glitches correlate with privacy concerns. The research contributes to the emerging field of HMC and privacy (Lutz 2023) by adopting the relatively new perspective of glitch studies to smart speakers. This perspective shows the generative nature and quality of unexpected and seemingly erratic technology behavior. According to glitch studies, such glitches and imperfections open up avenues for user reflexivity that transcends dominant and pre-programmed notions of sociality.

5. Discussion and Conclusion

The exploration of user experiences with smart speakers, particularly focusing on glitches and unexpected behaviors, sheds new light on the complex dynamics of human-technology interaction. The survey results show that the occurrence of glitches is something many users experience. The different types are varied and range from misinterpretations of commands to unsolicited responses. These glitches, although often minor in nature, resonate deeply with user concerns, especially regarding privacy and trust. The findings underscore a critical aspect of smart speaker technology: the delicate balance between utility and user apprehension.

Drawing on glitch studies (Kemper 2023; Menkman 2011) proved fruitful to expand our repertoire of making sense of smart speaker-user interactions. Glitches, in this context, are not just technological anomalies but are instrumental in shaping user perceptions and interactions with smart speakers. They serve as a lens through which the complexities of AI-driven communication can be understood. Each glitch experience contributes to a user's ongoing narrative with their device and their domestication (Brause and Blank 2020; Waldecker, Hector, and Hoffmann 2024). These narratives often reflect broader concerns about the role and reliability of AI in everyday life, highlighting the need for a deeper understanding of the socio-technical systems we engage with.

The connection between glitches and privacy concerns is particularly striking. Instances where smart speakers activate without a wake word or respond inappropriately reveal the underlying continuous listening capabilities of these devices. Such occurrences raise critical questions by researchers about the handling and potential misuse of personal data (Lutz and Newlands 2021). The findings emphasize the need for more transparent and user-centered privacy practices in the development of smart speakers, to rebuild and maintain user trust.

In terms of theoretical implications, the emphasis on glitches offers a unique contribution. While previous research has extensively covered user experiences and privacy concerns (see Table 1 and the Literature Review), the specific focus on glitches adds a new dimension. It aligns with existing literature on the imperfections of algorithms and AI (Kolkman and Kemper 2017) but goes further to delineate types of imperfection with specific empirical data.

The insights gained from this study have *practical implications* for both smart speaker developers and policymakers. Manufacturers should prioritize user-centric design, particularly in addressing glitches and enhancing privacy features. Transparent communication about how data is processed and used, along with user-friendly privacy controls, may enhance user acceptance (Felzmann et al. 2020). For policymakers, the findings highlight the importance of robust privacy regulations and standards specifically tailored to AI-driven devices in private spaces. The AI Act, a landmark legislation for AI systems in Europe, was recently adopted and is currently being implemented (Tamò-Larrieux et al. 2024). Given the voice-modality of interaction between users and smart speakers, provisions in this Act about biometrics should apply also to smart speakers (cf. Horn in this volume). It remains to be seen if these systems will classify as high-risk AI and thus face strict scrutiny and more stringent regulation.

The study opens several *avenues for future research*. Foremost, the research field needs to examine how glitches actually impact user attitudes and behavior, both through qualitative and quantitative methodologies.Longitudinal studies could provide insights into how user perceptions and experiences evolve over time, especially as users become more accustomed to the quirks of their devices. Additionally, investigating diverse user demographics could reveal variations in experiences and expectations, contributing to more inclusive and adaptable smart speaker technologies. Another promising area is the exploration of user experiences across different technological setups and ecosystems, providing a more comprehensive view of the smart speaker landscape.

In conclusion, this chapter enriches our understanding of smart speakers, not just as technological artifacts but as integral components of our daily lives that continuously interact and learn from us. By focusing on glitches, we gain a deeper appreciation of the challenges and opportunities presented by these devices.

References

Apthorpe, Noah, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. 2018. "Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (2): 1–23.

Brause, Saba Rebecca, and Grant Blank. 2020. "Externalized Domestication: Smart Speaker Assistants, Networks and Domestication Theory." *Information, Communication & Society* 23 (5): 751–63.

Brause, Saba Rebecca, and Grant Blank. 2023. "'There Are Some Things That I Would Never Ask Alexa'– Privacy Work, Contextual Integrity, and Smart

Speaker Assistants." *Information, Communication & Society*: 1–16. https://doi.org/10.1080/1369118X.2023.2193241.

Diakopoulos, Nicholas, Jack Bandy, and Henry Dambanemuya. 2023. "Auditing Human–Machine Communication Systems Using Simulated Humans." In *The SAGE Handbook of Human-Machine Communication*, edited by Andrea L. Guzman, Rhonda McEwen, and Steve Jones, 178–85. Thousand Oaks: SAGE.

Douglas, Benjamin D., Patrick J. Ewell, and Markus Brauer. 2023. "Data Quality in Online Human-Subjects Research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA." *Plos One* 18 (3): e0279720.

Estes, Adam Clark. 2018. "Your Worst Alexa Nightmares Are Coming True." *Gizmodo*, May 25, 2018. https://gizmodo.com/your-worst-alexanightmares-are-coming-true-1826327301.

Feldman, Russell. 2018. "Smart Speaker Ownership Doubles in Six Months." *YouGov*, April 19, 2018. https://yougov.co.uk/topics/politics/articles-reports/2018/04/19/smart-speaker-ownership-doubles-six-months.

Felzmann, Heike, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. "Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns." *Big Data & Society* 6 (1): 1–14. https://doi.org/10.1177/2053951719860542.

Felzmann, Heike, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2020. "Towards Transparency by Design for Artificial Intelligence." *Science and Engineering Ethics* 26 (6): 3333–61. https://doi.org/10.1007/s11948-020-00276-4.

Feng, Huan, Kassem Fawaz, and Kang G. Shin. 2017. "Continuous Authentication for Voice Assistants." In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 343–55. New York: ACM.

Fetterolf, Elizabeth, and Ekaterina Hertog. 2023. "It's not Her Fault: Trust through Anthropomorphism among Young Adult Amazon Alexa Users." *Convergence*: 1–18. https://doi.org/10.1177/13548565231200337.

Geeng, Christine, and Franziska Roesner. 2019. "Who's in Control? Interactions in Multi-User Smart Homes." In Proceedings of the 2019 CHI Conference on Human Factors *in Computing Systems*, article 268. New York: ACM.

Gruber, Jonathan, Eszter Hargittai, Gökçe Karaoglu, and Lisa Brombach. 2021. "Algorithm Awareness as an Important Internet Skill: The Case of Voice Assistants." *International Journal of Communication* 15: 1770–88.

Guzman, Andrea L. 2017. "Making AI Safe for Humans: A Conversation with Siri." In *Socialbots and Their Friends*, edited by Robert Gehl and Maria Bakardjieva, 85–101. London: Routledge.

Guzman, Andrea L. 2019. "Voices *in* and *of* the Machine: Source Orientation toward Mobile Virtual Assistants." *Computers in Human Behavior* 90: 343–50. https://doi.org/10.1016/j.chb.2018.08.009.

Hoy, Matthew. 2018. "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants." *Medical Reference Services Quarterly* 37 (1): 81–88.

Hurel, Louise Marie, and Nick Couldry. 2022. "Colonizing the Home as Data-Source: Investigating the Language of Amazon Skills and Google Actions." *International Journal of Communication* 16: 5184–5203.

James, Kedrick, and Ernesto Peña. 2023. "In Response to 'Towards Glitch Pedagogy'". ResearchGate Preprint. https://www.researchgate.net/profile/Ernesto-Pena/publication/369685170_In_Response_to_Towards_Glitch_Pedagogy/links/64274b1666f8522c38e94f52/In-Response-to-Towards-Glitch-Pedagogy.pdf.

Kang, Hyunjin, and Jeeyun Oh. 2023. "Communication Privacy Management for Smart Speaker Use: Integrating the Role of Privacy Self-Efficacy and the Multidimensional View." *New Media & Society* 25 (5): 1153–75. https://doi.org/10.1177/14614448211026611.

Kemper, Jakko. 2023. "Glitch, the Post-Digital Aesthetic of Failure and Twenty-First-Century Media." *European Journal of Cultural Studies* 26 (1): 47–63.

Kim, Raphael, Roland van Dierendonck, and Stefan Poslad. 2019. "Moldy Ghosts and Yeasty Invasions: Glitches in Hybrid Bio-Digital Games." *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. New York: ACM.

Kolkman, Daan, and Jakko Kemper. 2017. "Glitch Studies and the Ambiguous Objectivity of Algorithms." *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2985424.

Lau, Josephine, Benjamin Zimmerman, and Florian Schaub. 2018. "Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers." *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW): article 102. https://doi.org/10.1145/3274371.

Leszczynski, Agnieszka. 2020. "Glitchy Vignettes of Platform Urbanism." *Environment and Planning D: Society and Space* 38 (2): 189–208. https://doi.org/10.1177/0263775819878721.

Linghede, Eva. 2018. "The Promise of Glitching Bodies in Sport: A Posthumanist Exploration of an Intersex Phenomenon." *Qualitative Research in Sport,* *Exercise and Health* 10 (5): 570–84. https://doi.org/10.1080/2159676X.2018.1479980.

Luger, Ewa, and Abigail Sellen. 2016. "'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–97. New York: ACM.

Lutz, Christoph. 2023. "Privacy and Human-Machine Communication." In *The SAGE Handbook of Human-Machine Communication*, edited by Andrea L. Guzman, Rhonda McEwen, and Steve Jones, 310–17. Thousand Oaks: SAGE.

Lutz, Christoph, and Gemma Newlands. 2021. "Privacy and Smart Speakers: A Multi-Dimensional Approach." *The Information Society* 37 (3): 147–62. https://doi.org/10.1080/2159676X.2018.1479980.

Lutz, Christoph, and Aurelia Tamó-Larrieux. 2020. "The Robot Privacy Paradox: Understanding how Privacy Concerns Shape Intentions to Use Social Robots." *Human-Machine Communication* 1: 87–111. https://doi.org/10.30658/hmc.1.6.

Maccario, Guglielmo, and Maurizio Naldi. 2023. "Privacy in Smart Speakers: A systematic Literature Review." *Security and Privacy* 6 (1): e274. https://doi.org/10.1002/spy2.274

Malkin, Nathan, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. "Privacy Attitudes of Smart Speaker Users." In *Proceedings on Privacy Enhancing Technologies* 2019 (4): 250–71.

Mariani, Marcello M., Novin Hashemi, and Jochen Wirtz. 2023. "Artificial Intelligence Empowered Conversational Agents: A Systematic Literature Review and Research Agenda." *Journal of Business Research* 161.

Menkman, Rosa. 2011. "Glitch Studies Manifesto." *Video vortex reader II: Moving* images beyond YouTube, 336–47. https://pdfs.semanticscholar.org/0732/56f0c90a712365890f379f7dc55bd1377f3e.pdf.

Minder, Bettina, Patricia Wolf, Matthias Baldauf, and Surabhi Verma. 2023. "Voice Assistants in Private Households: A Conceptual Framework for Future Research in an Interdisciplinary Field." *Humanities and Social Sciences Communications* 10 (1): 1–18. https://doi.org/10.1057/s41599-023-01615-z.

Mols, Anouk, Yijing Wang, and Jason Pridmore. 2022. "Household Intelligent Personal Assistants in the Netherlands: Exploring Privacy Concerns around Surveillance, Security, and Platforms." *Convergence* 28 (6): 1841–60. https://doi.org/10.1177/13548565211042234.

Newlands, Gemma, Christoph Lutz, Aurelia Tamò-Larrieux, Eduard Fosch Villaronga, Rehana Harasgama, and Gil Scheitlin. 2020. "Innovation under

Pressure: Implications for Data Privacy during the Covid-19 Pandemic." *Big Data & Society* 7 (2): 1–14. https://doi.org/10.1177/2053951720976680.

Nielsen, Janni, Torkil Clemmensen, and Carsten Yssing. 2002. "Getting Access to What Goes on in People's Heads? Reflections on the Think-Aloud Technique." In *Proceedings of the Second Nordic Conference on Human-computer interaction*, 101–10. New York: ACM.

Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the* Integrity of Social Life. Palo Alto, CA: Stanford University Press.

Ofcom. 2023. "Media Nations: UK 2023." Ofcom, August 3, 2023. https://www.ofcom.org.uk/__data/assets/pdf_file/0029/265376/media-nations-report-2023.pdf.

Palan, Stefan, and Christian Schitter. 2018. "Prolific.ac—A Subject Pool for Online Experiments." *Journal of Behavioral and Experimental Finance* 17: 22–27. https://doi.org/10.1016/j.jbef.2017.12.004.

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research." *Journal of Experimental Social Psychology* 70: 153–63. https://psycnet.apa.org/doi/10.1016/j.jesp.2017.01.006.

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2021. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods* 54: 1643–62. https://doi.org/10.3758/s13428-021-01694-3.

Peña, Ernesto, and Kedrick James. 2016. "A Glitch Pedagogy: Exquisite Error and the Appeal of the Accidental." *Journal of the Canadian Association for Curriculum Studies* 14 (1): 108–27. https://doi.org/10.25071/1916-4467.40306.

Preece, Chloe, and Laryssa Whittaker. 2023. "Towards Glitch Pedagogy." *Journal of Marketing Management* 39 (1–2): 68–91. https://doi.org/10.1080/0267257X.2023.2174165.

Pridmore, Jason, and Anouk Mols. 2020. "Personal Choices and Situated Data: Privacy Negotiations and the Acceptance of Household Intelligent Personal Assistants." *Big Data & Society* 7 (1): 1–12. https://doi.org/10.1177/2053951719891748.

Pridmore, Jason, Michael Zimmer, Jessica Vitak, Anouk Mols, Daniel Trottier, Priya C. Kumar, and Yuting Liao. 2019. "Intelligent Personal Assistants and the Intercultural Negotiations of Dataveillance in Platformed Households." *Surveillance & Society* 17 (1/2): 125–31. https://doi.org/10.24908/ss.v17i1/2.12936.

Rutten, Ellen, and Ruby de Vos. 2023. "Trash, Dirt, Glitch: The Imperfect Turn." *European Journal of Cultural Studies* 26 (1): 3–13. https://doi.org/10.1177/13675494231152371.

Schwartz, Eric Hal. 2022. "Smart Speaker Adoption Continues Slowdown even as Voice Assistant Command Frequency and Variety Jumps: Report." *Voicebot.AI*, June 17, 2022. https://voicebot.ai/2022/06/17/smart-speaker-adoption-continues-slowdown-even-as-voice-assistant-command-frequency-and-variety-jumps-report/.

Smith, Katherine Taken. 2018. "Marketing via Smart Speakers: What Should Alexa Say?" *Journal of Strategic Marketing* 28 (4): 350–65. https://doi.org/10.1080/0965254X.2018.1541924.

Tamò-Larrieux, Aurelia, Clement Guitton, Simon Mayer, and Christoph Lutz. 2024. "Regulating for Trust: Can Law Establish Trust in Artificial Intelligence?" *Regulation & Governance* 18 (3), 780–801. https://doi.org/10.1111/rego.12568.

Vitak, Jessica, Priya C. Kumar, Yuting Liao, and Michael Zimmer. 2023. "Boundary Regulation Processes and Privacy Concerns with (Non-)Use of Voice-Based Assistants." *Human-Machine Communication* 6 (1): 183–202.

Waldecker, David, Tim Hector, and Dagmar Hoffmann. 2024. "Intelligent personal assistants in practice. Situational agencies and the multiple forms of cooperation without consensus." *Convergence* 30 (3), 975–91. https://doi.org/10.1177/13548565231189584.

Xu, Kun, Sylvia Chan-Olmsted, and Fanjue Liu. 2022. "Smart Speakers Require Smart Management: Two Routes from User Gratifications to Privacy Settings." *International Journal of Communication* 16: 192–214.

Zheng, Serena, Noah Apthorpe, Marshini Chetty, and Nick Feamster. 2018. "User Perceptions of Smart Home IoT Privacy." *Proceedings of the ACM on Human-Computer Interaction,* 2 (CSCW): article 200.

The Role of Imagined Sociotechnical Affordances in Shaping Experiences of Privacy in Smart Speakers

Jasper Vermeulen and Anouk Mols

Abstract *Smart speakers are heralded to make everyday life more convenient in households around the world. These voice-activated devices have become part of intimate domestic contexts in which users interact with platforms. This chapter presents a dual study investigating the privacy perceptions of smart speaker users and non-users. Data collected in in-depth interviews and focus groups with Dutch users and non-users show that they make sense of privacy risks through imagined sociotechnical affordances. Imagined affordances emerge with the interplay between user expectations, technologies, and designer intentions. Affordances like controllability, assistance, conversation, linkability, recordability, and locatability are associated with privacy considerations. Viewing this observation in the light of privacy calculus theory, we provide insights into how users' positive experiences of the control over and assistance in the home offered by smart speakers outweighs privacy concerns. On the contrary, non-users reject the devices because of fears that recordability and locatability would breach the privacy of their homes by tapping data to platform companies. Our findings emphasize the dynamic nature of privacy calculus considerations and how these interact with imagined affordances; establishing a contrast between rational and emotional responses relating to smart speaker use. Emotions play a pivotal role in adoption considerations whereby respondents balance fears of unknown malicious actors against trust in platform companies. This study paves the way for further research that examines how surveillance in the home is becoming increasingly normalized by smart technologies.*

1. Introduction

Intelligent personal assistants (IPAs), also known as digital assistants (DAs) or voice-activated personal assistants (VAPAs), have been around for more than a decade. Such digital assistants have become embedded in different facets of life through different technologies – for instance, in smartphone use, voice assistants like Apple's Siri, Google Assistant, and Samsung's Bixby. IPAs also serve as the user interface of smart speakers, which are operated by direct user interaction through voice commands. Smart speakers have found a place in the intimate space of many homes and are connected to appliances, devices, and digital services (to control lights, curtains, TVs and other media devices, streaming services, thermostats, etc.). Smart speakers are purchased because they offer benefits like convenience, reduced consumption of time and energy, and entertainment (Gram-Hanssen and Darby 2018; Wilson et al. 2017). Smart speakers present opportunities for interaction and information once connected to other smart devices (Chang et al. 2020; Wilson et al. 2017). For instance, a smart speaker can be asked whether it will rain today, to shed light on a trivia question that pops up in a family conversation, or to close the curtains. Through these interactions, smart speakers collect, process, and communicate data, including data about the user (Batalla et al. 2017). Smart speakers thus allow for voice-activated interaction with smart elements of the home while simultaneously collecting potentially sensitive user data, like audio. To explore the societal significance of smart speakers, it is essential to view them as a part of the smart home system and to consider how the use of smart technologies contributes to the datafication of users, their homes, and their private lives (Lupton 2020).

Factors that limit the adoption of smart speakers are cost, privacy, and surveillance concerns (Balta-Ozkan et al. 2013; Wilson et al. 2017). More specifically, smart speakers potentially infringe upon privacy because, through built-in microphones and user profiles, they can collect data from the most private spheres of life (Jacobson 2019; Pridmore and Mols 2020; Wilson et al. 2017). This entails personal data, connected media accounts, linked devices, smartphone use data, internet traffic, use patterns, behavioral routines, environmental information, etc. Moreover, such data is collected on behalf of digital platforms like Amazon, Google, and Apple. Smart speakers are part of platform ecosystems that are increasingly embedded in everyday life. Smart speaker platforms' appetite for ever more personal data feeds into what has been termed the datafication of users, their homes, and their private lives (Lupton 2020).

The research presented in this chapter was conducted in the Netherlands, where the smart speaker market has grown steadily since their launch. Market researchers indicate that it reached the point of saturation in 2022, when 23%

of all households owned a smart speaker (Multiscope 2022, 2023). To provide insights into consumers' considerations, including privacy concerns, when deciding whether to install a smart speaker, our research investigated how users and non-users of smart speakers make sense of smart speaker features, their ongoing algorithmically driven changes, and the potential for data collection across platforms. We make use of the concept of affordances, which are defined as possibilities for action (Evans et al. 2017; Gibson 1979). More specifically, our interest lies in imagined sociotechnical affordances, which emerge in the interplay between user attitudes and expectations, technologies, and the intentions and perceptions of designers (Nagy and Neff 2015). This focus allows us to explore how technological features, user expectations, algorithms, and platform-based design all play a role in shaping adoption considerations and perceptions of privacy. The following question guided our research: **Which affordances play a role in privacy perceptions and adoption considerations of users and non-users of smart speakers?**

Research was conducted with focus groups of users and non-users (N=29) and interviews with family members (N=22), with the sample selected to include users with different financial standings and home contexts. We examined interlocutors' privacy considerations through a constructivist grounded theory analysis. Our study expands upon existing research by including the perceptions of users and non-users and by connecting privacy calculus theory to imagined sociotechnical affordances. This connection makes it possible to fully explore the rational, emotional, and imaginative aspects of privacy considerations.

Our results indicated that the most intense privacy concerns were expressed by non-users, while users seemed more willing to accept potential privacy and surveillance risks. For the latter group, the benefits of smart speaker use outweighed the risks. Emotions played a pivotal role in adoption considerations whereby respondents balanced fears of unknown malicious actors against trust in platform companies. Moreover, non-users' conceptualizations of the affordances of smart speakers were more speculative, as they had often not used the devices in practice, whereas users could draw on their embodied experiences of engaging with smart speakers and utilizing their affordances. Finally, our analysis suggests that the ongoing normalization of smart speaker use can further normalize commercial as well as interpersonal surveillance.

2. Theoretical Framework

Here, we present a review of current research on smart speakers in relation to privacy considerations and the privacy calculus. Subsequently, we explore the imagined sociotechnical affordances of smart speakers.

2.1 Privacy and smart speakers

Concerns about datafication revolve around privacy, a concept that can be traced back to Ancient Greece. Aristotle distinguished two aspects of privacy: *Oikos* and *Polis*. These can be defined as privacy within the social world (Polis) and privacy in the household (Oikos) (Swanson 2019). In this study, we were particularly interested in how individuals negotiate privacy in the Oikos. When using smart speakers, the Oikos becomes visible to external parties. Using smart speakers in the home is perceived to contribute to a potential diminishing or even dissolution of privacy in domestic contexts. Therefore, we conceptualize privacy around smart speakers as the right to exercise true invisibility within and around the household (inspired by Dinev and Hart 2006).

Privacy considerations are, in turn, interconnected with surveillance, with the Oikos becoming subject to commercial data collection through smart speakers. The notion of surveillance can be broken down into sur (from above) and veillance (to watch) (Galič et al. 2017). Surveillance encroaches upon the privacy of its subjects. In this case, it impacts smart speaker users who (partly unknowingly) disclose data about and from within their private homes. Haggerty and Ericson (2000) introduced the concept of the disappearance of disappearance and emphasized that, in the current societal and digital landscape, it is all but impossible to escape – or disappear – from the view of surveillance. This extends into the home; while the modern single-family home is often considered a haven of privacy, it becomes subject to increased visibility through smart speaker use. More than a decade ago, Deuze (2011) proclaimed that communication technologies have penetrated life to the extent that we no longer live with technology – instead, life is lived in technology. Deeply embedded and ever more pervasive, technology has become increasingly invisible (Deuze 2011). Smart speakers can be seen as impacting life in and around the home.

2.2 Smart speakers and the privacy calculus theory

The decision to adopt a smart speaker follows an evaluation of the perceived benefits, drawbacks, and risks (Kumar et al. 2020). Privacy, security, and surveillance concerns have been identified as aspects that inhibit smart speaker adoption (Kim et al. 2019). How people weigh the risks and benefits of technologies can be understood through the concept of the 'privacy calculus' (Dinev and Hart 2006; Kim et al. 2019), which elucidates the trade-off between affordances and privacy concerns or threats (Smith et al. 2011).

The privacy calculus concept was originally devised to analyze how users negotiate e-commerce (Dinev and Hart 2006). The theory also lends itself well to examining how users and non-users make sense of smart speakers' perceived affordances and privacy issues. In order to make use of smart speakers' capabilities and functionalities, users are obliged to share personal and behavioral data. The privacy calculus framework evaluates perceived privacy risks by assessing the extent to which it is believed that sharing personal information could lead to a negative outcome (Dinev et al. 2006). If levels of perceived trust in the companies that produce and sell smart speakers were to increase, users would perceive lower risk and greater benefit in providing personal information (Shin 2010). Kim et al. (2019) found that when it comes to IoT technologies like smart speakers, customization and personalization options were seen to significantly alleviate risk. Whereas privacy calculus theory has been widely applied to information systems, its application to IoT services has so far been limited, yet helpful (Kim et al. 2019). The privacy calculus theory starts from the assumption that individuals' actions are rational, and thus, that a correlation can be identified between benefits, risks, and actions. However, this is rarely the case, as individuals' purchasing decisions are notoriously at least as emotionally as they are rationally motivated (Kim et al. 2019). If we view such decisions as simply a rational evaluation of smart speakers' benefits versus their (privacy-related) drawbacks, we overlook emotions like the joy of using smart speakers for fun activities, feelings of unease when interacting with a device, or fear of someone eavesdropping on private conversations via smart speaker (Mols et al. 2022). Such emotions can impact privacy perceptions just as much as rational, informed ideas about functional benefits or drawbacks do. In this study, we employed privacy calculus theory in the context of smart speakers in order to distinguish between rational considerations and accompanying emotions.

2.3 Smart speaker affordances

Smart speakers enable users to 'do' everyday practices differently. The concept of affordances provides a means to understand such processes. Affordances are possibilities for action to occur in relations between humans, technologies, material features, and the situatedness of use (Evans et al. 2017, building on Gibson, 1979). This concept is often used in different contexts but is not always adequately defined (for a constructive critique, see, for instance, Evans et al. 2017). In this study we were concerned with sociotechnical "imagined affordances" (Nagy and Neff 2015, 1). Imagined sociotechnical affordances entail material, mediated, and emotional aspects of human–technology interaction whereby the relations between designers, users, and algorithms are formative. Nagy and Neff (2015) define imagined affordances as the features imagined by users based upon their perceptions and (emotional) experiences of technologies. Users form perceptions and conceptualizations of technologies partly through direct experiences and partly through indirect perceptions of how they function – in the case of smart speakers, this means internally as well as linked to internet connections, algorithms, data, and digital ecosystems. These perceived aspects of affordances are considered imagined.

Smart speakers are more than physical objects: they are embedded in digital platforms and invisibly connected to home appliances and digital services. Therefore, the concept of imagined affordances is useful because it takes the sociotechnical background of user perceptions into account. To explain, we tweak Nagy and Neff's (2015) example of Facebook news feeds to adapt it for a smart speaker context. When people ask their Google Home for their daily news updates, they might perceive this update as an objective account of news rather than an algorithmically generated selection of news sources shaped by the designers of the algorithms and the algorithms themselves. The ways users receive their news updates (e.g., via audio only or also on a screen) and the platform ecology that delivers them are the sociotechnical background that needs to be considered when studying an imagined affordance like objective news reporting.

Existing research about smart speakers already offers some insights into affordances. Brause and Blank (2020), for instance, identify "spatial affordances of SSAs [smart speaker assistants] to engage in spatially distributed uses" (p. 8). These include the affordances of potential ubiquity (ubiquitous connections with devices allow for seamless IPA use), controllability (controlling connected devices from one device) and linkability (connecting people from a distance). Lutz and Newlands (2021) also set out from "an affordance perspective (Evans et al. 2017), whereby affordances are possibilities for action emerging from the relational structure between a technology and the user, intelligent assistants enable interactivity, searchability, and recordability" (Lutz and Newlands 2021: 148). The affordances interactivity, searchability, and recordability are perceived as enabling surveillance practices between family members (Lutz and Newlands 2021).

Earlier work by one of the authors (Mols et al. 2022) built on these studies by adopting recordability from Lutz and Newland (2021) and controllability and linkability from Brause and Blank (2020). In addition, Mols et al. (2022) introduced the smart speaker affordances assistance, conversation, and locatability, and explored how these affordances were related to potential users' privacy concerns. They found that security concerns were associated with locatability; the connectedness of smart speakers to homes and personal spaces was seen as a potential breach of privacy. Controllability offers control over devices and appliances and the assistance affordance allows for assistance with daily tasks (Mols et al. 2022). However, potential users feared that digital platforms could infringe upon their intellectual privacy and curtail their control over their personal information. Surveillance concerns revolved around conversation and recordability; these affordances inspired fear of third parties listening in (Mols et al. 2022).

Because of the current study's focus on emotional aspects of privacy calculus considerations, it is important to include social aspects of smart speaker use. The concept of connectedness provides a good basis from which to explore social dimensions (Lee et al. 2017). While Lee et al. do not describe connectedness as an affordance, it is closely associated with the affordances controllability and linkability. Lee et al. (2017) distinguish between inner social connectedness and outer social connectedness. Inner social connectedness refers to connections in a smart home environment made between the user and smart home devices (in the same vein as controllability). By contrast, outer social connectedness, similar to linkability, focuses on how smart home devices facilitate connections between smart home users and others (Lee et al. 2017). Smart speakers can facilitate perceived companionship by enabling connections with others and with technological entities (Lee et al. 2017). In this light, smart speakers can afford users with a means to combat loneliness by connecting individuals living in solitude, such as the elderly (Ehrenhard et al. 2014). In this chapter, we reflect on all of the abovementioned affordances and social connectedness and show how they are perceived by users and non-users of smart speakers.

3. Methods

This study aimed to investigate the privacy perceptions of users and non-users of smart speakers through a dual-methods approach deploying interviews with parents and young teenagers in Dutch families (N=22) and focus groups with university students (N=29). Our data were collected in 2021.

3.1 Student focus groups

With the 29 university students, we conducted four focus groups of 5–8 interlocutors. The sessions lasted 60 minutes on average. Focus groups enable respondents to engage in meaning-making together and to generate rich and thick data (Peek and Fothergill 2009). The focus groups were moderated in a semi-structured manner and aimed to foster open discussions. Due to COVID-19 restrictions, the sessions were conducted online via Zoom. Stewart and Shamdasani (2017) observe that online focus groups generate better results than offline ones because respondents perceive the online setting to be more informal. In our research, we also experienced active and open engagement from all research participants in the online focus groups. The student participants, aged 18–26, were recruited by voluntary sampling. They were not obliged by their university or course to take part in the study, nor did they receive any academic credits or monetary rewards. The voluntary participants remarked that they were interested in discussing topics of privacy in relation to IPAs. Following recruitment, we filtered participants based on whether they were users or non-users of smart home technologies. Consequently, we held two focus groups with users only (FG 1+2), one with non-users only (FG 4) and one mixed session (FG 3). This approach was selected so that group sessions could focus on discussing participants' own experiences with smart home devices or on adoption considerations respectively. Furthermore, we were also interested in the exchange that a mixed group with both users and non-users would generate. Of the students, 14 were users and 15 non-users, as can be seen in Table 1. To instigate the group discussions, each session began with the screening of a video about the Google Home smart speaker (Peek of the Net 2017). The subsequent discussion revolved around actual and potential uses, benefits and risks, and adoption considerations.

Table 1: Overview of focus group respondents

Pseudonym	**Male/Female**	**Age**	**User/non-user**
	FG 1:		
Mike	M	18	User
Rutger	M	18	User
Ron	M	20	User
Renato	M	19	User
Ralf	M	22	User
Lotte	F	20	User
	FG 2:		
Lex	M	24	User
Lance	M	22	User
Harold	M	24	User
Hans	M	20	User
Holly	F	19	User
Stan	M	20	User
Lars	M	22	User
	FG 3:		
Sander	M	21	User
Sem	M	21	User
Mara	F	19	Non-user
Mario	M	19	Non-user
Mohammed	M	19	Non-user
Maria	F	19	Non-user
Mako	M	19	Non-user
Maarten	M	19	Non-user

Pseudonym	Male/Female	Age	User/non-user
FG 4:			
Mariana	F	19	Non-user
Rudolf	M	20	Non-user
Ryan	M	23	Non-user
Lara	F	24	Non-user
Harry	M	21	Non-user
Harriette	F	22	Non-user
Silvia	F	21	Non-user
Stefan	M	20	Non-user

3.2 Family interviews

We interviewed nine Dutch families (a total of 11 parents and 11 adolescents aged 11–15), as shown in Table 2. To maximize diversity in the sample (Patton 1990), families with different constellations (such as nuclear families, single-parent families, and a foster care family) were included. The families were recruited via (extended) personal networks and snowball sampling. Although we had prepared one interview framework for the parents and one for the youth, we adapted the interviews according to which family members were present, which varied between families. More specifically, some interviews were conducted with the parent(s) and youth separately, whereas other families preferred to be interviewed together. Above all, we aimed to interview family members in a situation that they felt comfortable in. Nine interviews were conducted in the family homes, and four via Zoom. The interviews focused on social media practices, parental monitoring, family interactions, COVID-19 lockdowns, and, most importantly, smart technology use.

All the respondents signed a consent form (the parents also officially approved their children's participation), and the research was conducted in accordance with the ethical guidelines set out by the Erasmus University Rotterdam. The focus groups' discussions and the interviews were transcribed verbatim. We pseudonymized the respondents and removed potentially identifiable information. Subsequently, the transcripts were analyzed in Atlas.ti through an inductive grounded theory approach consisting of (1) open coding, (2) axial coding, and (3) selective coding (Charmaz 2014). For this study, we filtered out the open codes related to smart speakers. We clustered the open codes of the focus groups and interviews into axial codes (the subthemes in the results section) and selective codes (benefits and risks inspired by privacy calculus theory). Through the triangulation of the findings of the focus groups and interviews, we were able to provide a thorough insight into privacy considerations around smart speakers.

Table 2: Overview of interview respondents

Family	Interview	Pseudonym	Role	Age	User/Non-user
1	1	Paul	Father	42	Users
	2	Parker	Son	13	
		Tim	Son	11	
2	3	Nadia	Mother	42	Users
	4	Ellie	Daughter	12	
3	5	Fiona	Mother	44	Non-users
		George	Father	42	
	6	Jill	Daughter	13	
4	7	Joel	Father	48	Non-users
	8	Scott	Son	14	
5	9	Greta	Mother	43	Non-users
		Jack	Son	13	
6	10	Abby	Mother	39	Non-users
		Naomi	Daughter	12	
7	11	Oscar	Father	49	Non-users
		Grace	Mother	43	
		Lucy	Daughter	15	
8	12	Camila	Mother	45	Users
		Jasmin	Daughter	13	
9	13	Lydia Eli Faith	Mother Son Daughter	45 12 11	Users

3.3 Methodological limitations

Despite the measures mentioned above taken to tackle methodological shortcomings and research bias, this research is not exempt from limitations. First, it is specific to the Dutch context, so its findings might not be applicable to other sociocultural contexts. The introduction of smart devices into private homes is, however, not unique to the Netherlands. By providing insights into experiences made in the Netherlands, we contribute to the growing global body of literature on privacy and smart speakers. Second, some of the research was conducted during the COVID-19 pandemic. This made it more difficult to recruit participants and meant that some of the research had to be conducted remotely, via Zoom. However, as described in the method section, we view this as a strength rather than a weakness, because online focus groups often enable respondents to feel more comfortable to share their experiences (Stewart and Shamdasani 2017). Combining data from online focus groups with interviews in family homes also enabled us to reflect on contextual factors in our interpretations of the transcripts. Finally, the research focuses on families and students, two groups that are of particular interest because young people and children are the customers of the (near) future, and their adoption considerations are therefore significant. However, future research should also consider older populations that may use smart home technologies for assistance or support. Their adoption considerations entail a need dimension not at play in the population under study in this chapter.

3.4 Connection to prior research

To allow insights into how privacy perceptions evolve during the processes of domestication of smart home technologies, this study refers back to prior research by one of the authors, which took place in 2018, before the introduction of smart speakers in the Netherlands, and was conducted with six focus groups comprised of university personnel (Mols et al. 2022). At the time of data collection of the study presented here (2021), smart speakers had meanwhile become widely available. Hence, it had become possible to study how interlocutors' privacy perceptions were shaped by experiences of actually using them, as well as exploring reasons for reluctance to use smart speakers. The study presented in this chapter thus provides updated insights into the imagined affordances identified in the 2018 study. In the conclusion, we reflect on the changes we observed in user perceptions and imagined affordances between the two studies.

4. Results

The analysis entailed the identification of several benefits and risks in accordance with privacy calculus theory. These benefits and risks revolve around specific imagined affordances, which form the sub-themes in this results section.

4.1 Controllability affordance

In relation to the affordance of controllability, participants reported various benefits as well as perceived risks of smart speakers in and around the home. The controllability affordance affords users the capability to control interconnected devices and appliances via one device (Brause and Blank 2020). Controllability was perceived by most users as a positive affordance: it was essentially the main reason why they had purchased a smart speaker. In the introductory round of one of the online focus groups, Renato (19, user) gave a live demonstration of the voice-activated lights in his room that were connected to his smart speaker, exemplifying controllability. When people use smart speakers to make their homes smart, they install interconnected devices that require transmission of personal information in order to be controllable. In the focus groups, respondents reflected on how controllability affords convenience for users in the home (supporting the findings of Chang et al., 2020; Gram-Hanssen and Darby, 2018). As Ron (20, user) put it: "The most important part is saving time and integrating these technologies in your life, making it very efficient and easy." Sander (21, user) illustrated the ease of use: "It removes a lot of hassle from daily things. For example, you do not have to open your computer or your phone to check flights, things are ready for you."

Some student users speculated that controllability would be even more useful for families. Stan (20, user) remarked: "For example, not right now, but if you have a family and busy lifestyle, then it [a smart home] is helpful, convenient and can save you a lot of time." The student focus groups with users and non-users brought together young adults who did not have their own children. Nevertheless, users perceived the increased controllability in the home as a great advantage of IPAs.

The family interviews provided insights into some families' habits of delegating household tasks and actions to smart speakers. Paul, father of two sons, described how smart speaker use was embedded into their everyday family life:

> We play music on different speakers, and sometimes we fool around with questions like 'imitate a dog'. And, well, actually, all of us use them [smart speakers] every day. We broadcast things, we turn the lights on or off. We never use switches or power outlets in the house anymore, we do all of that with those Google Homes. (Paul, 42, user)

While controllability was perceived by users as a beneficial affordance, a fear of losing control was expressed by non-users. "There's just information that's private and that's somehow used elsewhere. But, but just the feeling that you kind of have no control over it, that you don't know, that, that actually puts me off." (Joel, 48, non-user). Similar sentiments were voiced among students: "These kinds of things create an excessively big dependency, everything that the people are doing, they depend on this little machine." (Silvia, 21, non-user). Users spoke about control directly in terms of imagined affordances. In contrast, non-users seemed more concerned about a different aspect of control: in relation to the process whereby personal data is imagined to be collected, controlled, and manipulated in unknown ways by unknown agents. While greater control over the home (such as controlling energy consumption) was praised by users among our respondents, echoing the findings of prior research (Balta-Ozkan et al. 2013), it became clear that non-users often perceived the delegation of home control in the form of controllability as a daunting prospect. Hence, there are two dimensions of controllability at play. One dimension is experiencing a sense of having direct control over the home via a central device. The second is that the processes that enable this form of control are invisible to users and thus can only be imagined. This unseen back end is what creates a sense of lack of control, specifically, concerning the ways smart speakers use data.

Moreover, in debates about controllability as an imagined affordance, platform operators are often explicitly mentioned. For users as well as non-users, whether data collection is considered acceptable or not relates to perceived levels of (dis)trust in smart speaker providers like Google and Amazon. More specifically, concerns were raised by our participants about how collected data is handled, processed, and stored. As this remains a black box phenomenon for the vast majority of users and non-users alike, trust in the companies involved plays a crucial role in how the imagined affordance of controllability is weighed up. For instance, Ryan (23, non-user) expressed trust: "Google is responsible for the data… I would rather have trust in a company such as Google," while Sem (21, user) voiced distrust: "Google can collect information on what you're doing, and what you like to do as well, which I think can be very scary." Trust is thus a significant component of the emotional aspects of this imagined affordance and is heavily dependent on the perceptions and experiences of the user or non-user. These perceptions crucially influence associated trust levels that shape interactions with smart speakers.

4.2 Assistance affordance

Many respondents described how smart speakers can increase convenience and make certain practices and household tasks easier. This is clearly illustrated in the examples discussed above in relation to the controllability affordance. Yet, there is a further dimension, which is about how smart speakers can assist people in their everyday lives. The assistance affordance (Mols et al. 2022) relates to how smart speakers assist with daily tasks and offer general user support. While controllability focuses on controlling the home environment, assistance provides support across a variety of tasks, stimulated by either human or technological input. Sander (21, user) described this: "It removes a lot of hassle from your daily tasks; for instance, you do not have to open the computer to check your flights; it [Google Home] does it for you." Ron (20, user) elaborated: "When you integrate these technologies into your daily life, it can make it a lot easier for you and save time." Smart speakers can function as personal assistants that makes the lives of users easier.

In a more negative light, non-users in the family interviews speculated that such assistance could have unwanted outcomes, which mainly revolved around users becoming too dependent or lazy. George (42, non-user) remarked: "I have colleagues who say, 'Hello Google' for everything, 'Google turn on the heating, the lights and play some music'." Fiona (family non-user) shared this view: "I think such practices where you become reliant on smart speakers to do certain actions for you can make you lazy." Therefore, while smart speakers offer users assistance in mundane tasks, such as switching on the lights, this was often perceived (by non-uses) as potentially making users lazy. Such considerations were also discussed in the focus groups with students. Harriette (20, non-user), for instance, commented that with assistance from smart speakers: "We would get so lazy and do nothing all day." Notably, the assistance affordance invoked stronger (negative) reactions among non-users than among users.

4.3 Conversation affordance

Smart speakers afford inner social connectedness by facilitating connections between users and their smart home devices (Lee et al., 2017). This conversation affordance provides benefits but is also accompanied by concerns about risks (Mols et al., 2022). Benefits are seen primarily in the aspect of saving time, as observed by users and non-users. Stan (20, user) believed that talking to technology can save time: "It also saves time since if you have a question, you can just ask instead of looking at your phone, and while asking, you can still do something else." Furthermore, in families, smart speakers are sometimes used to interact with family members. Paul (42, user) explained: "We sometimes broadcast through the Google Homes. So, then we don't have to scream upstairs, but then we can just ask 'Hey Google, broadcast…' and then it broadcasts to all devices, in all rooms, you will hear that."

Moreover, functionalities offering comfort and control through voice-activated interactions can simultaneously provide a means to combat loneliness. Verbally interacting with these smart home technologies could be an end in itself. Users of smart speakers perceived this to be a potential significant benefit in specific contexts: "I would see benefits for elderly who are alone at home" (Rutger, 18, user). Previous research on smart home technologies has shown that these technologies can offer particular benefits to users who are considered elderly and have physical limitations (Ehrenhard et al. 2014; Kim et al. 2019).

Some respondents, however, expressed ambivalence about the conversation affordance becoming integrated into family lives. Naomi (12, non-user) described her observation of how, when they become a medium for communication inside homes, smart speakers can significantly impact daily interactions.:

> I visited a family before and there they had everything. There, they said 'Good night, Google,' after which the lights in the hallway and the living room would turn off. They also use Google Home to broadcast to their children when dinner is being served. But I also know families that don't really want any part of that [smart speakers]. So, really, I see both. (Naomi, 12, non-user)

This example illustrates that smart speakers can become an integral part of sociotechnical systems in the home. Depending on how these smart speakers are employed, they become increasingly involved in intimate contexts. Naomi was unsure whether this was an affordance that families should want, and in the further discussion she also expressed her concern about reliance on technology and the environmental impact of technology use.

While conversation is mainly perceived as a beneficial affordance by users, the interpersonal surveillance (also described as lateral surveillance, Andrejevic 2002) that it enables also evoked ambivalence. Just as Lutz and Newland (2020) warned about surveillance within families through smart speakers, families in our study also reported some negative impressions about interpersonal surveillance and conversation through devices. Jasmin (13, family user) shared an example: "Once me and my friend were watching Titanic at two a.m. Um, and at one point there was some scene that was really loud. That's when my mother heard us, and she activated sleepy music on the Google thing. So, I was like okay, stop, so I disconnected it."

4.4 Linkability affordance

Linkability affords smart speaker users connections with people across distance (Brause and Blank 2020), facilitating outer social connectedness between smart home users and others beyond the home (Lee et al. 2017). Connecting with other users is not a feature that is unique to smart speakers, but was seen as an attractive affordance by students who perceived it as beneficial for their age group (18–25). For example, Holly (19, user) remarked: "It is nice to connect with people of the same age through smart devices such as Google Home, to get the feeling your peers are there with you in the room." Like other communication technologies, smart speakers can connect homes to their users. Linkability enables smart speakers to represent people who are distant as if they are nearby. The value attributed to this by our participants might have been increased by the timing of the research – focus groups took place during the COVID-19 social distancing restrictions. Moreover, many other technologies also offer connections with others. Yet, Holly's quote indicates that when these connections are located in a smart speaker, it offers a sense of proximity which she apparently did not experience through other devices. This emotive experience could offer benefits to people who find themselves far away from friends or family. In other words, linkability can be understood as one of the emotional dimensions of the perceived benefits and risks that are weighed up by (potential) users, because smart speakers can offer a feeling of togetherness with peers who are not bodily present. This was especially important for the university students who took part in the study during a global pandemic.

Linkability and conversation affordances come together when users broadcast through their smart speakers when they are not at home. Nadia (42, user) described a situation in which she broadcasted to her daughter from the gym:

> I used it to send a voice memo and then it would sound 'pling plong', and broadcast me saying 'I'll come home in a bit.' But she [her daughter] was completely shocked by it every time I did that. I thought it was very handy, because at that time she didn't look at her phone that much. So, I thought if you don't look at your phone, I'll broadcast it through the house. She didn't like that very much. (Nadia, 42, user)

This example demonstrates how certain users anticipate benefits from the linkability affordance, but in practice, other users may experience its effects in different ways. For Nadia's daughter, hearing her absent mother's disembodied voice emitted by the speaker was apparently more alarming than comforting. Thus, the imagined emotional affordance of linkability seems to be mostly appreciated in situations when physical presence is prevented.

4.5 Recordability affordance

Smart speakers that are operated by voice commands, such as the Google Home, work by processing audible requests and performing actions accordingly. To improve the technological functioning of these products, which are still in development, employees at smart speaker companies listen to recordings to evaluate the products' performance (Jacobson 2019). Some of our study participants were aware of this: "For artificial intelligence to work, it does have to have this feedback, it must hear from us, and then just continue learning more." (Holly, 19, user). Most respondents, however, perceived recordability (Lutz & Newland 2021) as an affordance associated with concerns and risks. For instance, Rutger (18, user) commented: "because it has to listen all the time and you don't know what will happen with your data." If users are aware of and express unease about the underlying mechanisms of data collection, this could discourage potential users from adopting the devices if companies do not adapt their practices in response to the concerns expressed. Our participants proposed several potential solutions; Lex (24, user) remarked that he would even be willing to pay extra for his data not to be used:

> These companies need lots of data to improve their products, but I think that they could do this in a better way like offering a product for free to people that want to participate and do not mind their data being used and offering the options for people to pay a little more to not share their data. (Lex, 24, user)

Now that smart speakers and their capabilities have become an increasingly familiar phenomenon in the Netherlands, both users and non-users have access to information about potential privacy issues. Yet, many users are not concerned and never use the mute function. For example, in the focus groups, a student explained: "Many people like me don't care about these privacy claims and that is why I think I never used the mute button." (Renato, 19, user). Camila (45, user) remarked in one of the family interviews: "Often, I don't even notice that it's there. To be honest, I didn't even know that you can mute it [smart speaker]?". The recordability affordance appears to be of little concern for some smart speaker users.

Responses from other interlocutors in the family interviews, however, indicated that some users evaluate what they know about data collection and potential recordability and ultimately decide that the benefits outweigh the risk. Paul (42, user) explained:

> It seems, but I'm not sure, that people could eavesdrop on a smart speaker because it is connected via the internet and it has speakers in it. So, if you really have something to hide, let's say you work for the police or whatever, well, it's best to turn off the microphone. But I never mute them. I don't have much to hide in that regard, but I don't have very sensitive information either, I am not interesting. What smart speakers can pick up, can't harm us. It's not like I'm going to mention what my PIN is, or how much I have in the bank, I'm not going to say that aloud. So, what could they be eavesdropping on me? (Paul, 42, user)

Concerns about recordability clearly carried more weight for non-users than users: "But if it's 'Hey Google', it listens in all day" observed Jack (13, non-user). His mother, Greta (43, non-user), added: "That's what dad says, isn't it? Yeah, um, my husband says that he wouldn't want a smart speaker for that reason. He has the idea that you are being eavesdropped on."

Recordability is thus an imagined affordance that is perceived as more problematic by non-users than by users. Privacy concerns are significant enough to form an adoption barrier (Wilson et al. 2017). The imagined character of this affordance is crucial because recordability concerns an unreckonable process whereby data might or might not be processed and shared with third parties, and recordings might or might not be processed (to improve the accuracy of smart speakers or for malicious purposes). This uncertainty makes deciding whether the benefits of a smart speaker outweigh privacy concerns around recordability extremely complex. Thus, the privacy calculus inevitably lacks relevant information. For users, this unknowability may be alleviated by direct experiences of using smart speakers and feeling 'safe'.

While the user considerations around recordability share some resemblances with those around the controllability affordance, they differ in terms of which elements trigger privacy considerations and specific emotions. When it comes to controllability, it is in the context of the user controlling their home environment that some individuals experience a sense of lacking control over the hidden use of personal data. Considerations around control over personal data are fueled by (a lack of) trust in platform companies and are emphasized more strongly by non-users than users. For the recordability affordance, it is more specifically about the way this data is collected through voice recordings. This relates to the emotional experience of fearing eavesdropping by platforms. The idea of an external person or entity listening in on intimate conversations can incite unease among users as well as non-users. Unease is triggered by the feeling of being spied upon by an unknown entity. This imagined unease is not about personal data security but rather about the embodied personal experience of being subjected to invisible auditory surveillance.

4.6 Locatability affordance

In the family interviews, non-users expressed concerns about security. They feared that their privacy would be breached if a smart speaker was traceable to their home. As in our previous research findings about the locatability affordance (Mols et al. 2022), the risks raised mainly revolved around direct invasions of the private sphere by burglars and hackers. As George (42, non-user) put it: "You make it easier to break in, right? Yes, because if those smart speakers are connected to your network, that makes you more and more vulnerable for people to access." Concerns relating to hacking were also voiced by student users. Ralf (22, user) reflected:

> I am wondering what would happen if someone hacked into that speaker and could listen along and what they could do with that kind of information. It would be easy to find out that way, who your family members are, where you work or whom you have a relationship with. Even bank accounts if you use your Google Home to transfer money. (Ralf, 22, user)

The unreckonable threat of unknown interference clearly caused concern. Both non-users and users perceived unknown hackers as a greater threat than the platform owners that tap into vast amounts of data every day. The concerns raised express an imagined threat of malicious actors gaining access to one's home – digitally as well as physically. Imagining risks and threats involving malicious actors and their unreckonable potential actions evokes strong emotions.

5. Discussion: Adoption Considerations

The privacy calculus theory offers an insightful perspective on smart speaker adoption considerations by focusing on (potential) users' weighing up of perceived benefits against perceived risks. Users experience affordances such as controllability and assistance as beneficial. Although both users and non-users are cognizant of potential privacy risks, users have decided to adopt the devices nonetheless. In other words, affordances such as controllability and assistance appear to outweigh potential privacy risks for users. Whereas such risks lead to non-users' emotionally charged criticisms of smart speakers, users choose to accept the hidden ways their data is handled. As such, through the internalization of surveillance and the incorporation of these devices into their daily lives, users willingly allow personal data to be processed and further transmitted by their smart speakers in exchange for perceived and experienced benefits.

Conversely, many non-users feel that privacy risks outweigh the benefits, and therefore decide not to purchase smart speakers. Locatability and recordability are perceived in association with imagined threats of malicious entities breaching privacy, and controllability can instigate feelings of mistrust in platform companies. This indicates that emotional reactions to potential threats and invisible data collection play a vital role in privacy perceptions. The imagined, unreckonable character of these affordances seems to magnify some of these perceived risks.

When it comes to the affordances controllability and assistance, which users appreciate for their convenience, non-users perceive risks when they imagine intensive use scenarios. They fear that embedding such technology into their everyday (family) lives could lead to a form of technological dependency.

The considerations around the two remaining affordances seem to be less emotionally charged. Linkability is mainly seen as a beneficial way to connect with others. The conversation affordance is evaluated in mixed ways by users and non-users who see practical benefits but also identify reliability and interpersonal surveillance risks.

Notably, while non-users' concerns about privacy risks are often strong enough to limit their motivation to adopt smart speakers, some do not rule out acquiring such devices in the future. Student Rudolf (20, non-user) surmised: "I think that currently there is just too much risk compared to benefits. So, in time, it will improve, and I will consider buying smart home devices." The fundamental limitations of smart speakers were also mentioned in pragmatic evaluations. George (42, non-user) pointed out in one of the family interviews: "Even if you order a smart speaker to start your coffee machine, in the end, you still need to get your cup of coffee from the kitchen," to which his wife Fiona (44, non-user) added "You also need to drink it yourself." A general trend was observed that users and non-users were not always impressed by the current capabilities of smart speakers and what they offer. Yet, the readiness of some non-users to observe the ongoing development of smart speakers and perhaps consider adoption in the future indicates that smart speaker adoption considerations are open to change over time.

6. Conclusion

In this chapter we reported on our investigation into the privacy considerations surrounding imagined sociotechnical affordances of smart speakers. Deploying a dual-methods approach combining in-depth interviews and focus groups, we reflected on several imagined affordances: controllability, assistance, conversation, linkability, recordability, and locatability. These affordances were found to inform individuals' privacy calculus and informed how users and non-users evaluated privacy considerations for smart speakers. Although our findings were largely consistent with those of existing research and our prior study (Mols et al. 2022), we also identified that linkability affordances were particularly highly appreciated during and since the COVID-19 pandemic and that perceived risks are often emotionally charged.

We identified three trends surrounding smart speaker use and adoption. This was based upon an evaluation of several studies conducted during a period of normalization of smart home speakers in the Netherlands. First, the most recent results display a marked normalization of smart speaker use and a relativization of concerns in contrast to our earlier study (Mols et al., 2022). In 2018, focus group participants had expressed more reluctance to speak to a device in their home and some had questioned whether smart speakers were even necessary, or perhaps offered a solution for a problem they did not have. In 2021, participants provided examples of smart speakers having been integrated into everyday (family) life. Users appreciated increased controllability and assistance in the home and expressed less concern about data collection in the private sphere. Non-users, however, continued to voice privacy concerns. This was most pronounced in relation to the controllability affordance: non-users were unwilling to accept the storage and processing of data by big platform companies whom they did not trust. Nonetheless, some non-users' statements indicated that there is scope for their attitudes to change if the technology and/or corporate practices develop in a favorable direction.

Second, discussions about the locatability and recordability affordances showed that perceptions of privacy risks are more influential when fear is involved, indicating that privacy considerations that are emotionally charged rather than rationally based have a more significant impact on adoption decisions. Fears seemed to be less prevalent among users. The privacy calculus theory is typically used to analyze rational evaluations that weigh perceived benefits against perceived risks (Balta-Ozkan et al., 2013). Our research illustrates how influential emotions are when people decide whether to adopt networked technological devices in their private and personal spaces. Privacy concerns intermingle with non-tangible imagined affordances. Yet, the relative importance attributed to benefits and risks can change as perceptions and emotions shift with changing circumstances, personal and societal. In other words, privacy considerations and adoption decisions are always dynamic and in flux.

Third, smart speakers are an example of smart home technologies that can be employed in many ways. Concerns relate to data collection and the processing of sensitive user data by smart speaker platforms. While the linkability affordance results from smart speakers' capacity to mediate human-to-human connections that are experienced positively, it also enables human-to-human monitoring. Our study showed that the capacity to use smart speakers for interpersonal surveillance within the home also caused concern. With ongoing normalization, such concern may dissipate, setting a potentially harmful precedent that normalizes interpersonal surveillance in private contexts. Simultaneously, linkability increases the public visibility of hitherto private spaces. As these technologies and people's attitudes to them continue to develop, it is crucial that research continues to investigate the influence of smart speakers in the home and how they shape interactions. Future studies would do well to focus on the interactional aspects of smart devices and provide further insights into the situated use of smart devices.

References

Andrejevic, Mark. 2002. The Work of Watching One Another: Lateral Surveillance, Risk, and Governance. *Surveillance & Society* 2 (4): 479–497. https://doi.org/10.24908/ss.v2i4.3359.

Balta-Ozkan, Nazmiye, Rosemary Davidson, Martha Bicket, and Lorraine Whitmarsh. 2013. "Social Barriers to the Adoption of Smart Homes." *Energy Policy* 63 (December): 363–74. https://doi.org/10.1016/j.enpol.2013.08.043.

Batalla, Jordi Mongay, Athanasios Vasilakos, and Mariusz Gajewski. 2017. "Secure Smart Homes: Opportunities and Challenges." *ACM Comput. Surv.*, 50, no. 5: 1–32. https://doi.org/10.1145/3122816.

Brause, Saba Rebecca, and Grant Blank. 2020. "Externalized Domestication: Smart Speaker Assistants, Networks and Domestication Theory." *Information, Communication and Society* 23, no. 5: 751–63. https://doi.org/10.1080/1369118X.2020.1713845.

Chang, Victor, Yuanyuan Wang, and Gary Wills. 2020. "Research Investigations on the Use or Non-Use of Hearing Aids in the Smart Cities." *Technological Forecasting and Social Change* 153 (April): 119231. https://doi.org/10.1016/j.techfore.2018.03.002.

Charmaz, Kathy. 2014. *Constructing Grounded Theory*. SAGE.

Deuze, Mark. 2011. "Media Life." *Media Culture & Society* 33, no. 1: 137–48. https://doi.org/10.1177/0163443710386518.

Dinev, Tamara, and Paul Hart. 2006. "An Extended Privacy Calculus Model for E-Commerce Transactions." *Information Systems Research* 17, no. 1: 61–80. https://doi.org/10.1287/isre.1060.0080.

Ehrenhard, Michel, Bjorn Kijl, and Lambert Nieuwenhuis. 2014. “Market Adoption Barriers of Multi-Stakeholder Technology: Smart Homes for the Aging Population.” *Technological Forecasting and Social Change* 89 (November): 306–15. https://doi.org/10.1016/j.techfore.2014.08.002.

Evans, Sandra K., Katy E. Pearce, Jessica Vitak, and Jeffrey W. Treem. 2016. “Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research.” *Journal of Computer-Mediated Communication: JCMC* 22, no. 1: 35–52. https://doi.org/10.1111/jcc4.12180.

Galič, Maša, Tjerk Timan, and Bert-Jaap Koops. 2017. “Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation.” *Philosophy & Technology* 30, no. 1: 9–37. https://doi.org/10.1007/s13347-016-0219-1.

Gram-Hanssen, Kirsten, and Sarah J. Darby. 2018. “‘Home Is Where the Smart Is’? Evaluating Smart Home Research and Approaches against the Concept of Home.” *Energy Research & Social Science* 37 (March): 94–101. https://doi.org/10.1016/j.erss.2017.09.037.

Haggerty, K. D., and R. V. Ericson. 2000. “The Surveillant Assemblage.” *The British Journal of Sociology* 51, no. 4: 605–22. https://doi.org/10.1080/00071310020015280.

Jacobson, Adam. 2019. “Smart Home Devices and Privacy Risk,” *Risk Management* 66, no. 9 (Oct 2019): 4–7. https://search.proquest.com/openview/0f2e06fbd21a82b44606b29b129d8652/1?pq-origsite=gscholar&cbl=47271.

Kim, Dongyeon, Kyuhong Park, Yongjin Park, and Jae-Hyeon Ahn. 2019. “Willingness to Provide Personal Information: Perspective of Privacy Calculus in IoT Services.” *Computers in Human Behavior* 92 (March): 273–81. https://doi.org/10.1016/j.chb.2018.11.022.

Kumar, Harish, Manoj Kumar Singh, M. P. Gupta, and Jitendra Madaan. 2020. “Moving towards Smart Cities: Solutions That Lead to the Smart City Transformation Framework.” *Technological Forecasting and Social Change* 153 (April). https://doi.org/10.1016/j.techfore.2018.04.024.

Lee, Byounggwan, Ohkyun Kwon, Inseong Lee, and Jinwoo Kim. 2017. “Companionship with Smart Home Devices: The Impact of Social Connectedness and Interaction Types on Perceived Social Support and Companionship in Smart Homes.” *Computers in Human Behavior* 75 (October): 922–34. https://doi.org/10.1016/j.chb.2017.06.031.

Lupton, Deborah. 2020. “The Internet of Things: Social Dimensions.” *Sociology Compass* 14, no. 4: e12770. https://doi.org/10.1111/soc4.12770.

Lutz, Christoph, and Gemma Newlands. 2021. "Privacy and Smart Speakers: A Multi-Dimensional Approach." *The Information Society* 37, no. 3: 147–62. https://doi.org/10.1080/01972243.2021.1897914.

Mols, Anouk, Yijing Wang, and Jason Pridmore. 2022. "Household Intelligent Personal Assistants in the Netherlands: Exploring Privacy Concerns around Surveillance, Security, and Platforms." *Convergence* 28, no. 6: 1841–60. https://doi.org/10.1177/13548565211042234.

Multiscope. "Markt voor slimme speakers verzadigd [Market smart speakers saturated]." Multiscope. July 19, 2022. https://www.multiscope.nl/persberichten/markt-voor-slimme-speakers-verzadigd/.

Multiscope. (2023, February 9). *Nieuwste inzichten smart home markt 2023*. Multiscope. https://www.multiscope.nl/nieuws/nieuwste-inzichten-smart-home-markt-2023/

Nagy, Peter, and Gina Neff. 2015. "Imagined Affordance: Reconstructing a Keyword for Communication Theory." *Social Media + Society* 1, no. 2. https://doi.org/10.1177/2056305115603385.

Patton, Michael Quinn. 1990. Qualitative Evaluation and Research Methods. 2nd ed. Los Angeles: Sage.

Peek, Lori, and Alice Fothergill. 2009. "Using Focus Groups: Lessons from Studying Daycare Centers, 9/11, and Hurricane Katrina." *Qualitative Research: QR* 9, no. 1: 31–59. https://doi.org/10.1177/1468794108098029.

Peek of the Net. "Google Home Official Ad." Youtube. February 4, 2017. https://www.youtube.com/watch?v=OsXedJq1aWE.

Pridmore, Jason, and Anouk Mols. 2020. "Personal Choices and Situated Data: Privacy Negotiations and the Acceptance of Household Intelligent Personal Assistants." *Big Data & Society* 7, no. 1. https://doi.org/10.1177/2053951719891748.

Shin, Dong-Hee. 2010. "The Effects of Trust, Security and Privacy in Social Networking: A Security-Based Approach to Understand the Pattern of Adoption." *Interacting with Computers* 22, no. 5: 428–38. https://doi.org/10.1016/j.intcom.2010.05.001.

Smith, H. Jeff, Tamara Dinev, and Heng Xu. 2011. "Information Privacy Research: An Interdisciplinary Review." *The Mississippi Quarterly* 35, no. 4: 989–1015. https://doi.org/10.2307/41409970.

Stewart, David W., and Prem Shamdasani. 2017. "Online Focus Groups." *Journal of Advertising* 46, no. 1: 48–60. https://doi.org/10.1080/00913367.2016.1252288.

Swanson, Judith A. 2019. *The Public and the Private in Aristotle's Political Philosophy*. Cornell University Press. https://doi.org/10.7591/9781501740831.
Wilson, Charlie, Tom Hargreaves, and Richard Hauxwell-Baldwin. 2017. "Benefits and Risks of Smart Home Technologies." *Energy Policy* 103 (April): 72–83. https://doi.org/10.1016/j.enpol.2016.12.047.

Mostly Harmless? Everyday Smart Speaker Use and Pragmatic Fatalism

David Waldecker, Alexander Martin, and Dagmar Hoffmann

Abstract *Breaches of trust and privacy by tech companies and the ensuing scandals emphasize how today's digital media are driven by the monetization of users' personal data. Studies of users' attitudes to data protection issues in connection with the use of digital media technologies have led researchers to conclude that users develop a kind of "online apathy" (Hargittai and Marwick 2016), "privacy cynicism" (Hoffmann et al. 2016) or "digital resignation" (Draper and Turow 2019). This chapter examines users' experiences of smart speakers in daily life and their understandings of the data-related consequences of their everyday use of the devices. We draw upon qualitative interviews conducted with smart speaker users in Germany to illustrate how they cultivate certain attitudes towards the devices as well as to the discourse about them, and how they explain their stances in relation to usage routines and pragmatic considerations. While our interviewees asserted views similar to some described by the aforementioned researchers, in this chapter we argue that the attitudes expressed by smart speaker users can be better understood as "pragmatic fatalism" (Pettenkofer 2017). Pragmatic fatalism allows them to acknowledge criticism of corporate data practices yet disregard it as irrelevant for their own everyday lives. The perceived harmlessness of devices, usage practices, and users themselves is emphasized as justification for not worrying about the potential consequences of bringing technology that constantly records interactions into one's own home.*

1. Introduction

In modern societies, the home is seen as a private space *par excellence*. Laws that guarantee homeowners and sometimes tenants extensive control over their living spaces also cover the control over information relating to the home. The value attributed to privacy by German citizens was testified to in the late 20th century when a debate about privacy and data protection was sparked by activists who fought for the individual's right to not be included in a census survey (Lengwiler 2017, 6). One can argue that that notorious debate continues to shape German, and consequently, European data protection laws to this day. The issue of domestic data protection has also been raised in relation to smart speakers, also known as intelligent personal assistants (IPAs). Available since 2014, these devices promise comfort and access to smart home and internet services by voice activation alone – without the push of a button. In order to be able to respond when a command is uttered, they need to constantly record the ambient sounds of the home. This technical setup, in combination with the awareness that the platform companies that offer such devices are known to harvest and analyze data, has led to critiques that condemn a hollowing out of domestic privacy and even "remote control" of home dwellers by these companies (Zuboff 2020).

In this contribution, we examine smart speaker users' understandings, strategies and perspectives on this potential for commercial misuse of the data produced in smart speaker use. Research on privacy issues related to digitally-connected media practices has found that users cultivate a form of "online apathy" (Hargittai and Marwick 2016), "privacy cynicism" (Hoffmann et al. 2016) or "digital resignation" (Draper and Turow 2019). We revisit this debate with a focus on pragmatic aspects and a pragmatistic theoretical conceptualization of users' behavior (see Pettenkofer 2017; 2023) to offer a complementary interpretation. We relate this theoretical discussion to findings from problem-centered interviews conducted as part of our research project "Un-/desired Observation in Interaction: 'Intelligent Personal Assistants' (IPAs)" at the Collaborative Research Center "Media of Cooperation", University of Siegen, from 2020 to 2023 (see Habscheid et al., this volume).

This chapter proceeds as follows: After reviewing literature on users' perspectives on online privacy, we propose the need for a more pragmatic and supra-individual approach. With this in mind, we also discuss the domestication perspective on media and information and communication technologies (ICTs) as well as social theory that focuses on the role of fatalism in everyday life (Pettenkofer 2023). Our sampling and methods are detailed in the following section. Subsequently, we present key elements of our empirical analysis and we close with a discussion of the findings and their relevance for further debate on online media and data practices.

2. From Privacy Paradox to Privacy Cynicism

With the advent of everyday online interaction and services, privacy as it is commonly understood has been seen to come under threat, (1) from corporate and state surveillance, i.e., organizational surveillance, and (2) from surveillance by other online users.

(1) The majority of online services are provided by private companies with an economic imperative (Kienscherf, this volume). An easy way to monetize free-to-use services is by selling advertising space. One of the first companies to undertake online corporate surveillance was Google (Zuboff 2019): by analyzing users' reactions to results from its search engine, Google was able to build user profiles so that it could place online ads most likely to align with users' interests. This strategy has been adopted by almost all providers of commercial online services. Online stores as well as newspaper websites analyze how users interact with their platforms so that they can adapt the content presented accordingly in order to maximize advertising effectivity and user engagement. The more information companies have about their users, the better positioned they are to judge which ads are likely to resonate. This is what created the drive towards big data that is prevalent today. Alongside hardware and software manufacturers like Apple and Microsoft, the main profits of companies like Meta/Facebook and Google are generated from online advertising. As critics like Zuboff suggest, the vast amounts of data collected not only allow companies to place targeted ads, but also to influence user behavior for companies' financial gain (Zuboff 2020). While the field of surveillance studies is informed by critiques of data use by state agencies for surveillance and control, it was also early to draw attention to the aforementioned corporate tracking of consumers and media users (Gandy, 1989)[1]. These corporate practices have been with us since the late 19th century (Lauer 2020). Hence, the corporate analysis of data obtained from smart speaker use should be understood not as a novel phenomenon but as a continuation or further development of earlier forms of corporate surveillance and indirect market research (Draper and Turow 2019, Kienscherf this volume).

(2) The last two decades have witnessed not only a rapid commercialization, platformization, and oligopolization of online services, but also the rise of personal publishing (see e.g., Taddicken and Schmidt 2016) by online users

1 Note that Gandy already used the term "surveillance society" 20 years before Zuboff did (2019).

via social media. Beginning with MySpace (2003) and Facebook (2004) in the United States and StudiVZ (2005) in Germany, users of the so-called Web 2.0 (O'Reilly 2012) were suddenly able to create content and disseminate it online without needing skills in computer programming or markup language. Activities formerly associated with the private sphere were made visible online and thereby became public affairs, of sorts.

In debates on threats to privacy, it is this use of social media that has often been invoked with the idea of the "privacy paradox" (Barnes 2006; Norberg et al. 2007). The term describes the paradox of a high value attributed to online privacy in co-occurrence with actions that imply disregarding such privacy. While early research put the paradox down to the two-fold inexperience of young users and of a new and interconnected medium (Barnes 2006), later work has tried to disentangle the paradox in other ways (boyd 2014). It seems that, over time, users became better informed about online data practices, but also came to see them as inevitable, leading to what Hargittai and Marwick (2016) term "online apathy." In their research, young users reported that they were informed about the risks of exposing information about themselves online but felt simultaneously pressurized by peers to do so. Here, the paradox was no longer about contradictory "sayings" and "doings" (cf., e.g., Kahn and Jeromack, 2013) – claiming to cherish privacy, but acting otherwise –, but resulted from conflicting imperatives from school, parents, and peers concerning social media. Research conducted at our research center in Germany also suggests that young adults know and care about interpersonal online privacy and therefore consider carefully what kinds of personal content to post on platforms such as Instagram (cf. Englert et al., 2019). Teens interviewed in our study, however, mentioned that they used social media less for personal presentation than for staying up to date on posted content[2].

In addition to conflicting imperatives concerning privacy and publicity of online lives, Draper and Turow (2019) note how "digital resignation" regarding privacy is also fostered and "cultivated" by online corporations. Corporations employ "obfuscatory communication practices" (Draper and Turow, 2019, 1830) that make it hard for individuals to obtain precise information about the use

2 In addition, such conflict between imperatives to uphold privacy yet also to present oneself publicly on social media platforms has become less prevalent since certain forms of interaction shifted from platforms that are public by default (such as Facebook and Instagram) to messenger apps that are private by default (such as WhatsApp, Signal, and Telegram).

of their data. This, in turn, creates a feeling of "resignation" in users who feel unable to change or clarify details about the use of their private data by corporations. Draper and Turow highlight how "feelings of resignation are a rational emotional response in the face of undesirable situations that individuals believe they cannot combat" (Draper and Turow, 2019, 1828). Here, again, it is the way the technical infrastructure is organized and advertised by corporations – shaping the social situation of users – which is seen to determine users' privacy practices.

This resignation has also, we suggest, been fostered by a shifting media narrative about online corporations. Internet researcher and activist Geert Lovink (2019) has noted how early hopes and enthusiasm for cyberspace gave way to a more dystopian and critical view of a web dominated by corporations and advertisement. The revelations by Edward Snowden concerning online surveillance, and scandals such as Facebook's involvement with online elections ads via Cambridge Analytica, combined with a tougher policy approach to corporate data use, have all helped to propagate views that criticize corporate data handling and denounce privacy violations. This perspective on privacy violations is especially pertinent to voice assistants because in VA use users do not primarily interact with other users, but with a synthetic agent provided by a company.

The resignation and apathy discussed above has also been addressed with specific reference to users of voice assistants. In several publications, Christoph Lutz (Lutz and Strahoff, 2014; Lutz and Newlands, 2021), Christian Hoffmann, and Giulia Rancini (Hoffmann, Lutz and Rancini, 2016; 2020) have proposed the concept of "privacy cynicism". Whereas Draper and Turow (2019) emphasize the consequences of corporate strategies, the term "cynicism" conveys not just a feeling that attempting to take action would be futile, but also implies negative views towards an antagonist:

> As such, we understand privacy cynicism as an attitude of uncertainty, powerlessness, and mistrust toward the handling of personal data by digital platforms, rendering privacy protection subjectively futile … In this context of ubiquitous institutional privacy threats, privacy cynicism can be understood as a cognitive coping mechanism because it allows subjectively disempowered users to participate in online platforms without cognitive dissonance since they rationalize privacy protection as useless. (Lutz, Hoffmann and Rancini, 2020, 1174)

Lutz, Hoffmann and Rancini developed their concept based on the findings of a large-scale survey on online privacy and data protection conducted in Germany (ibid.). The investigation focused on forms of data handling by online services in general (not related to specific services or devices) that make attempts to protect privacy appear futile. The authors empirically differentiated four aspects of cynicism – mistrust, uncertainty, powerlessness, and resignation (1178) – and examined how they related to users' internet skills, privacy concerns, privacy threat experience, and privacy protection behavior (1181).

The contributions mentioned above have advanced and nuanced understanding of users' actions and perspectives relating to data and privacy in a world of interconnected devices and services. They have shown that the privacy paradox is not primarily a psychological problem or one of motivation or lack of information (about safer or alternative information and communication technologies (ICTs)), but is related to the ways that data collection is inextricably built into digital platforms and services as well as to the ways in which these services have become an integral part of the indispensable social infrastructure of everyday life.

While these explanations can be understood as strongly contextualized approaches – they analyze more than just the perceptions and actions of individuals – they nonetheless focus on individualized fatalistic perspectives. Conceptually, studies in this field rarely take into account that people discuss their use of ICTs with peers, friends, and household members. We therefore propose that further insights can be gained by drawing on the domestication approach in media studies which examines how new ICTs are adopted and used in households and other organizational units. In a foundational text, Roger Silverstone, Eric Hirsch, and David Morley (1992, 12) conceptualize the ways a household uses ICTs as part of its "moral economy": Users collectively evaluate media devices and services with respect to domestic routines and normative expectations, as well as financial, spatial, and time constraints. The organization of domestic everyday life can be understood as a complex of normative and economic decisions that come together in practice. Taking up the metaphor of the domestication of animals by humans, this perspective describes how users collectively adopt and domesticate media to their specific needs, as well as how their daily lives are changed through media use, sometimes in unexpected ways. Recent research has applied these ideas to the study of modern and interconnected ICTs (Hector et al., 2023). This includes research which looks at the "externalisation" (Brause and Blank, 2020) of domestic tasks through smart home infrastructure – which in turn is often controlled via domestic voice assistants (cf. Strüver 2023). This approach is well suited to studying domestic voice assistants and connected smart speakers, which are designed to be used by multiple users (unlike smartphones and smart watches). Here, the household is all the more relevant because its domestic space is surveilled via these devices and the devices can, in turn, be addressed by anyone in the space.

Alongside this empirical addition, we respond to the call of Lutz and his colleagues (Lutz, Hoffmann and Ranzini 2020, 1173) for further theoretical elaboration of what is meant by apathy, resignation, or fatalism. While Lutz and colleagues, like Draper and Turow (2019), draw on previous explications of cynicism and resignation respectively, we choose *fatalism* as a term to describe user perspectives and behavior. As Andreas Pettenkofer points out (2017, 2023), fatalism has been a topic in social theory since the latter's inception. Although fatalism is usually perceived as a negative trait – as an acquiescence to one's supposed fate, and thus, an attitude that inhibits action – Pettenkofer highlights ways in which fatalism is positively related to agency. From a pragmatic perspective, deciding not to think about a problem can create new possibilities for action by freeing the individual from the need to deal with the problem or its potential consequences (Pettenkofer 2017, 131). In this way, fatalism as a concept is also able to describe how users actively cope with their inability to change the data-harvesting infrastructures of many digital services. Lutz and colleagues (2020, 1173) conclude that users deal with these aspects of digital services by grudgingly accepting them as inevitable. In Pettenkofer's discussion, however, fatalism is discussed in greater depth. He argues that fatalism is a phenomenon not only among the disadvantaged, who have to accept their situation because they only have limited options for action, but that it comes in multiple shapes and sizes. For the middle class, fatalism can be part of positive thinking and for the upper or executive class, it can represent the recurrent choice not to think through the social and ecological consequences of their economic or political decisions, for example. As such, Pettenkofer argues, fatalism is not an exception to the rational and action-oriented outlook that is often perceived to be a cornerstone of the modern subject, but far more widespread than is often acknowledged in social theory.

Pettenkofer emphasizes (2017, 130 [our translation]) that "fatalism is … not simply a perception of limits to action, but a pattern of reflection that *emerges* from a specific perception of such limits". Fatalism relieves us from thinking and thereby makes us capable of acting and has the effect of upholding order. Furthermore, it is important that fatalistic patterns of interpretation are based on everyday attitudes and experiences that make it easy to publicly justify a lack of alternatives. Pettenkoffer differentiates three constellations that can lead to a decision to stop reflecting (2017, 130–132): (1) Conceivable alternatives are or have become too abstract or too far removed from the individual's situation to be worthy of consideration; (2) the participants stop thinking about the application of certain evaluation criteria when justification and criticism have no consequences; (3) the participants give up trying to understand or resist the fateful process because it is too complex or because such action seems futile.

As Pettenkofer (2017) also sees positive thinking as part of this fatalistic mindset, his theoretical perspective can help to situate voice assistant users' critique and their perceived inaction. Characterizing user perspectives with terms such as "apathy", "resignation", and "cynicism" might imply that users are severely affected by their potential loss of privacy. As we elaborate in this chapter, we consider it significant that some users criticize corporate data practices yet do not seem so concerned that they stop using smart speakers. While fatalism has negative connotations such as resignation, the following elaboration will show how we conceptualize users' nonchalance despite concerns, with reference to Pettenkofer's understanding of the term.

In the following, we examine the (fatalistic) patterns of reflection that emerged in our interviews. We adopt Pettenkofer's differentiation of "resigned" versus "pragmatic fatalism" (Pettenkofer 2017, 143) to explore results from our empirical data and argue that Pettenkofer's concept of fatalism offers a useful aid for analysis of everyday data practices in smart speaker use.

3. Research Design

Our analysis is based on data generated in our research on smart speakers which examined interaction between users, devices, infrastructure, and language from a linguistic and media-sociological perspective. This chapter focuses on the media-sociological aspects of the research and data (see the introduction to this volume for a presentation of the research project and Habscheid, Hector, and Hrncal, this volume, for a more detailed presentation of the linguistic strand of the project).

Table 1: Sample

Household	Pseudonym	Age	Gender	Number and Type of IPAs	No. of Inter-views	Duration Intv. 1	Duration Intv. 2
1	Lukas F.	25	m	2 Amazon Echo Dot	2	00:55:14	01:05:48
1	Alex K.	27	m	2 Amazon Echo Dot	2	01:12:54	01:38:05
2	Jan-Ole S.	25	m	1 Amazon Echo Dot	2	00:45:05	01:33:56
2	Damaris L.	24	f	1 Amazon Echo Dot	2	00:47:30	01:42:09
3	Beate W.	61	f	1 Google Nest	2	00:58:56	01:14:15
4	Sam R.	27	m	1 Amazon Echo Dot	1	-	01:26:07
4	Andrea S.	25	f	1 Amazon Echo Dot	1	-	02:43:42
5	Julian R.	42	m	4 Apple Home Pod	1	-	01:59:13
6	Alexander R.	31	m	1 Apple Home Pod	1	-	01:10:56
6	Janina R.	29	f	1 Apple Home Pod	1	-	00:56:41
7	Till W.	21	m	1 Amazon Echo, 1 Apple Home Pod	2	01:15:48	01:23:08
7	Konrad W.	21	m	1 Amazon Echo, 1 Apple Home Pod	2	01:09:39	00:48:25
8	Samuel M.	30	m	1 Google Nest, 1 Amazon Echo	2	00:57:57	01:14:24
8	Robin L.	24	m	1 Google Nest	2	00:50:37	-
8	Lara S.	24	f	1 Google Nest, 1 Amazon Echo Dot	1	00:36:55	-
E	Dilek U.	26	f	1 Google Nest, 1 Amazon Echo	1	-	00:47:43
F	Tobias G.	42	m	1 Apple Home Pod, 1 Amazon Echo	1	-	00:56:21
G	Mukesh V.	26	m	1 Amazon Echo	1	-	01:15:10
H	Stefanie L.	34	f	1 Amazon Echo Dot	1	-	00:51:00

The goal was not only to observe and record dialogues of users and their IPA infrastructures, but also to examine users' perspectives and reflections upon IPA usage, on privacy concerns, and ways to deal with them. 28 interviews were conducted with 19 interlocutors in twelve households. Nine of the participants were interviewed twice: first, when they had just set up their newly acquired smart speaker, and then again a few months later.

As shown in the table of participants, most of the households in our sample had one or two IPAs. Amazon's Echo was used by the majority of households, followed by Apple's Home Pod. Google Nest was used in two households. The frequency of usage and technological skills of the participants varied widely. While most simply used their devices without connecting them to other hardware and without much interest in investigating further capabilities, some displayed a higher level of skill and interest (e.g., by connecting additional devices). Most participants lived in multi-person households. Our sample included different living constellations: cohabiting couples, shared apartments, and family households. When we could interview several members of the same household, we were able to take the relationships between their members into account. Friends or other family members play an important role in these constellations. In one case, an interviewee's fiancée lived abroad, making mediated communication essential for sustaining the relationship. The majority of the interviews were held in German, with one conducted in English. At the time of the survey, the respondents were aged between 20 and 60 years old and all had a relatively high level of education. Due to the COVID-19 pandemic, we were unable to visit participants in their homes. Interviews were therefore conducted remotely via the video conferencing tools Jitsi or Big Blue Button.

Problem-centered interviews (Witzel and Reiter, 2012) were carried out with all participants. This form of interview was chosen in order to (a) focus less on biographical narratives and more on a specific aspect of social life and (b) allow for a more direct confrontation of interviewees with discrepancies and ambiguous statements in their reports than other qualitative interviews methods would. The interviews also included show-and-tell episodes wherein interviewer and interviewee looked and listened together at the audio recordings of interactions provided by Amazon and Google to its users. These sessions facilitated further discussion of the companies' mode of data presentation, protection, and transparency.

The interviews were transcribed and all personal information pseudonymized. The transcripts were coded for qualitative content analysis (Kuckartz 2014) using MAXQDA software with a mix of codes derived inductively and deductively. As recommended by Kuckartz, significant deviations between encodings of identical interviews were extensively discussed until a consensual set of main categories could be agreed upon by all members of the research team. All interviews were then coded with this finalized code set. While the list of codes was extensive and covered a range of aspects of smart speaker use, in this chapter we focus on fatalism. The use of Pettenkofer's fatalism concept was inspired by research in a previous project (cf. Englert et al. 2019). Initially, all identified instances of fatalism were grouped in one code. However, after discussing the coded material, fatalism appeared to be more heterogeneous than anticipated. Furthermore, examples of it could be found not just in the fatalism category itself but also in other categories focusing on observation and surveillance. These codes were then subjected to a re-coding and re-analysis based on inductive findings from the material as well as an application of theoretical concepts from the literature referenced above. As a result, four subcodes of fatalism were derived: resignation (cf. Lutz et al. 2020), cynicism (cf. Draper and Turow 2019), trust, and pragmatic fatalism (cf. Pettenkofer 2017). These types are detailed in the following.

4. Four Shades of Fatalism

We propose that differentiating these aspects of fatalism offers a way to expand upon the debate outlined above on user reactions and strategies concerning corporate data practices. These different aspects emerged in several interviews and sometimes even in combination. The following four distinctions or shades of fatalism are therefore not to be understood as a typology of our interviewees, but as a disambiguation of the standpoints that users adopt to situate themselves vis-à-vis corporate actors and data practices, as well as to public debate on the issue. At the same time, these aspects also reflect how users evaluate what these data practices mean to them on a more general level.

4.1 Resignation

Resignation, as outlined above, is an attitude inherent to many users' data practices regarding IPAs and is fostered by online corporations (Draper and Turow 2019). It represents a form of individual surrender to an entity that is perceived as more powerful than oneself and impossible to influence. According to Pettenkofer, such resignation is what enables users to continue using

IPAs despite data concerns, since it is based upon the conclusion that surveillance via IPAs is negligible in comparison to all the users' data that has been collected already. Hence, this resignation does not render users apathetic but rather enables them to continue using IPAs and other surveillance-affording devices and services (Pettenkofer 2017). Our interviewees, however, discussed and justified their resignation in more differentiated ways and distinguished between different entities to whom they surrendered. For them, surrendering meant accepting that they were unable to influence the processing and capturing of their data by the IPAs and the companies behind them. What varied, however, were the entities or actors that interviewees identified as being the ones to whom they were obliged to surrender in order to continuing using IPAs. Furthermore, users described different ways of resigning to such entities. Some spoke about individualistic experiences, while many others talked about general notions and tendencies in society at large; some speculated on companies' rationales. Significantly, the consequences of resignation were not evaluated on an individual, but on a collective level.

> "Well, I think that, for us, so to speak, the train has already left the station." (Beate W., l. 583–584)[3]

Entities to whom users saw themselves surrendering were often somewhat abstract. Above all, it was corporations that were accused of exchanging data on such a scale that consumers became transparent. Protecting one's privacy was seen as impossible in the face of surveillance perceived to be omnipresent in online and offline spaces.

> "I think that we are transparent as customers or as consumers anyway and that [using a smart speaker] does not make any difference anymore." (Robin L., intv. 1, l. 551–554)[4]

As proof that companies were exchanging data without any possibility for users to intervene, interviewees cited personalized ads. Participants concluded that certain apps or devices and the companies behind them already had the kinds

3 German original: "Also der Zug ist schon abgefahren für uns, sage ich mal."

4 German: "ich dachte mir dann ja, also gefühlt ist man eh schon äh der gläserne Kunde oder der gläserne Konsument und [die Nutzung eines Smart Speakers] macht dann irgendwie auch keinen Unterschied mehr [...].“

of data that could be obtained by an IPA, making surveillance by the IPA negligible.

> "We live in times of Facebook, don't we? That is, we look for something on Amazon and three seconds later, there is an ad on Facebook that fits what we were looking for just a moment ago." (Jan-Ole S., intv. 1, l. 323–327)[5]

Resignation to a lack of data security inherent to IPA usage is therefore entangled with privacy concerns pertaining to other aspects of everyday life. Data capturing is perceived not only as inevitable but as something that has already taken place, as the reference to the "times of Facebook" and thus to the Cambridge Analytica scandal implies. Since the platforms already have the users' data, there is no point in worrying about the consequences of the IPA's data capturing. Resignation is therefore grounded in the assumption that it is already too late for everyone and not just for the individual.

> "We can not get out of this." (Beate W., intv. 1, l. 654)[6]

4.2 Cynicism

Cynicism, like resignation, addresses an entity, and therefore creates a relation between individuals, groups, and their environment. Cynicism differs from resignation, however, in that it expresses an antagonistic relationship towards a counterpart. The antagonism conveyed by our interviewees was most frequently expressed in claims that corporations were untrustworthy regarding data protection. This goes beyond resignation to corporations or state agencies portrayed as powerful and surveillance as inevitable. In addition, a cynical perspective distrusts these powerful actors (Lutz et al. 2020). As the following section shows, trust – and instances in which it is broken – played an important role in shaping interviewees' attitudes. As in research by Hoffmann and colleagues (2020), tech companies were perceived as powerful and uncontrollable, which led to feelings of powerlessness and of being at the mercy of plat-

5 German: "[…] wir leben in der Zeit von Facebook, ne? Ich meine, wir suchen bei Amazon suchen wir etwas und haben drei Sekunden später bei Facebook eine Werbebenachrichtigung von dem, was wir gerade gesucht haben."

6 German: "Wir kommen aus der Nummer nicht raus."

forms. In addition, users assumed that the actors concerned had bad intentions, such as evading taxation or EU law.

> "What should Amazon be afraid of? It's a multinational corporation, almost THE multinational corporation. And if you look at their revenue in the last few years and if you look at the effort of governments to get Amazon to pay taxes, in some areas, like in Europe, why should this topic be a red line that Amazon does not cross, that is somehow more relevant than other guidelines?" (Alex K., interv. 1, l. 510–518)[7]

In accordance with this perspective, software features that allow users to delete data, such as audio recordings of voice commands, were often dismissed as fake concessions towards users; an accusation supported by statements indicating partial or vague knowledge about IPA infrastructure. User settings on the Amazon and Google platforms enable audio recordings of IPA use to be deleted (cf. Pins et al, this volume). But our interviewees disregarded this, because they assumed that Amazon and other companies would have secret backups anyway in order to continue analyzing the valuable data.

> "Well, I think that, if they want my private data, they are able to access the stuff that I deleted, too." (Andrea S., intw. 1, l. 2870–2872)[8]

Privacy cynicism, like every fatalistic reflection pattern, makes it pointless to reflect upon certain topics, since the antagonist is too powerful to deal with. Government regulations and guidelines put in place by the companies themselves are seen as strategies to project an appearance of law-abidance. Trust towards the companies has already been so eroded that data control features are perceived as just another ploy to advance data collection. Tellingly, one in-

---

7 German: "Was hätten sie denn zu befürchten? Also es ist ja letztendlich ein multinationaler Konzern und äh fast schon DER multinationale Konzern. Äh wenn man sich anguckt, was die in den letzten Jahren an Umsätzen geschoben haben und äh was für einen Aufwand Staaten betreiben müssen, damit die über/ also damit Amazon überhaupt äh Steuern zahlt in manchen Regionen, beispielsweise in Europa, (.) warum sollte das ein Aspekt sein, der da irgendwo ausschlaggebend ist und wo dann die rote Linie ist?"

8 German: "Ja, also ich denke, wenn man/ wenn die, sage ich mal, meine privaten Daten haben wollen, dann äh können sie auch auf gelöschte Sachen zugreifen."

terviewee described himself as "cynical" while putting forth his argument in the logic of privacy cynicism:

> "Well, putting it like this involves some cynicism. However, in the end, I think this is correct. Isn't it? If I am really interested in a dataset, I will analyze it before I offer users the chance to delete that data so that they don't worry too much. Well, I put this more cynically than I intented (laughter)." (Alex K., intv. 2, l. 1397–1405)[9]

4.3 Trust

To identify trust as a mode of fatalism might seem counter-intuitive at first glance. When looking at fatalistic reflection patterns, however, it becomes clear that cognitive processes involving the idea of trust can be described as fatalistic: Trust enables actors to externalize responsibility by entrusting another actor (e.g., a company, experts, or governmental oversight more generally) to take care of their concerns and problems. Draper and Turow (2019) indicate in their discussion of privacy-related corporate communication strategies that tech companies obfuscate their data handling at the same time as insisting on it being safe and responsible. This leaves users with little choice but to believe a company's messages and trust it, or to distrust and refuse to engage with the company altogether. However, this conclusion was not supported by our interviewees' statements. Instead of trusting what companies claimed in their contracts or promised in advertisements, our interlocutors' views were based upon their own experiences and theories. These were probably also influenced by narratives, ads, etc. from the companies in question, but such material was not directly referred to or reiterated.

Of our interviewees whose statements reflected this fatalistic pattern, most expressed trust in Apple.

9 German: "Ja, im Endeffekt, da schwingt immer auch so ein bisschen Zynismus mit bei solchen Formulierungen finde ich. Aber im Endeffekt ist es halt auch genau das. Ne? Wenn man damit was anfangen möchte, mit so einem Datensatz, dann ist das normalerweise schon geschehen, bevor ich den Leuten die Möglichkeit gebe, den doch zu löschen fürs gute Gewissen. Also das war jetzt aber wieder zynischer formuliert (lachend) als ursprünglich gedacht."

> "That is, using Siri and only Siri and that's it. Simply put, it's convenient, it's safe and there's the data protection issue, too, of course." (Julian R., intv. 1, l. 200–202)[10]

Julian Rieker justified his trust in Apple by comparing its corporate data practices to those of Amazon. He perceived the personalized ads that Amazon generates as something bothersome and annoying. In his view, Alexa is the unsafe IPA that hands over data to further companies and is non-transparent about what it captures and how it will be processed. Targeted ads were cited by Julian as well as by other interviewees as proof that surveillance was taking place.

> "Apple doesn't want to sell you the next pair of socks or razor blades. And with that lady in the Amazon speaker, I am not sure what [data] gets processed in the background. And if you happen to talk about Pampers, for whatever reason – then you'll suddenly get shown ads for diapers. I haven't even got a child! So, like (laughs), where's this coming from? And that's the thing. So, no, I don't prefer the lady with A [Alexa, Amazon's voice assistant]." (Julian R., intv. 1, 169–178)[11]

This stance towards Amazon is further justified by Julian's belief that Alexa constantly transmits data, whereas Apple's IPA processes data locally.

> "Well, Amazon processes everything in the Cloud. So everything you say is routed via an Amazon server and that's not how it is with Apple, for example. Apple processes everything on the device itself." (Julian R., intv. 1, l. 213–216)[12]

---

10 German: "Das heißt, mit Siri und ausschließlich mit Siri und/ und das war es. Einfach Komfort, Sicherheit eben auch, das Thema Datenschutz natürlich."

11 German: "Apple will dir nicht unbedingt äh die nächsten Socken oder die Rasierklingen verkaufen. Und bei der Amazondame im Lautsprecher, da bin ich mir nicht so ganz sicher, was da alles verarbeitet wird. Wenn du dich mal über Pampers unterhältst, warum auch immer, ne? Und auf einmal kriegst Werbung von Pampers angezeigt. So, ich habe gar kein Kind. Weißt (lachend) du so, wo kommt das auf einmal her? Und das sind halt so Sachen, ähm ja. Die Frau mit A ist nicht so der Favorit."

12 German: "Naja, bei Amazon wird alles direkt in der Cloud verarbeitet. Also das heißt, das geht wirklich alles, was du sprichst, geht immer auf die Amazon Server und das hast du bei Apple zum Beispiel nicht. Apple wird/ verarbeitet alles erst mal auf den Geräten selbst."

Trust is expressed here by comparing a preferred product and its provider to one seen as less trustworthy. The evaluation is justified by the claim that local data processing and the (perceived) absence of targeted ads indicate Apple's trustworthiness.

Trust thus appears as a form of fatalism that is experienced by most interviewees as positive, unlike other forms of fatalism. This is not surprising since endowing trust upon a powerful other is not a matter of giving in to power, but of making an active choice. Trust can furthermore be seen as the opposite of cynicism as it expresses non-antagonistic engagement with a more powerful entity. This does not necessarily indicate enthusiasm, but does point to the nuanced spectrum of fatalistic perspectives.

4.4 Pragmatic fatalism

Unlike the examples analyzed above, in some instances, interviewees spoke about potentially problematic corporate data practices with less of a focus on collective aspects than on their individual perspective. Drawing on Andreas Pettenkofer's terminology (2017, 143, our translation), we refer to this as "pragmatic fatalism". The main characteristic of this line of reasoning is that users either (a) decide from the outset not to think about an issue, or (b) that their perspective is primarily influenced by their personal circumstances.

(a) Many statements in the interviews clearly indicated that most users stopped thinking early on about the potential consequences of the data analysis resulting from their smart speaker use, or decided quickly that they did not care about privacy. They mentioned several reasons for simply accepting the data practices involved. One reason was "laziness", as some users put it. Others, however, rejected the allegation of laziness by explaining that they had difficulty understanding (1) the privacy settings of the devices, (2) the technical processes of commercial data analysis, and (3) the open-source tools and alternatives to commercial platforms that might be available. Some saw themselves as simply unable to understand such aspects, while others argued that smart speakers are all about convenience and that it would be inconvenient to have to do research to understand the workings of the devices. This line of thinking is consistent with Pettenkofer's (2017) understanding of fatalism, whereby not thinking about something is what allows one to proceed.

(b) Another pragmatic reason cited for not thinking about the data-related consequences of using a smart speaker is that users are already enmeshed in corporate data practices. Interviewees explained that they had put aside their concerns about data collection via voice assistants once they realized how much of their data was already being collected through their use of smartphones and other devices and services. More specific fears about acoustic surveillance were rendered insignificant by the realization that smartphones also have microphones that could be hacked and tapped by hardware manufacturers or criminals or misused by third parties. Samuel Matthäi – working as a teacher while completing a degree –, for example, had initially been wary about allowing a smart speaker into the kitchen of his shared apartment. However, he then reflected that he had already been using Siri on his smartphone for a while, even in his own room. He explained:

> "I think that, after some time, at a certain point in time, I realized that I was using many, many devices already that were able to record audio, and that this thing was just another one (laughter)." (Samuel M., intv. 1, l. 202–209)[13]

Contrary to the resignation or cynicism detailed above, the focus of pragmatic fatalist views was less on one's position vis-à-vis a powerful and opaque corporate oligopoly, but on personal experiences of using devices and (not) perceiving data-related consequences. Such experiences did not have to be made directly; learning how other users dealt with issues was also influential. Samuel Matthäi had initially felt uneasy when he saw his new girlfriend's children using an Amazon smart speaker. Over time, he came to see how practical and entertaining the device was for the children. In our second interview with Samuel, a few months after he and his roommates had installed a Google Nest device in their communal kitchen, he had bought an Alexa device for his own room as well. This trajectory points to the relevance of social context: devices are more likely to be accepted and adopted by people who encounter them in others' homes as part of everyday life, and they are less likely to be viewed critically by people who have had opportunities to engage with them themselves.

Here, another element comes to the fore: the motif of harmlessness. Contrary to cynical and resigned views, a pragmatic fatalist perspective is justified

13 German: "Ich glaube, ab einem gewissen Punkt ist das auch nicht mehr unbedingt/ also, ne? [Da] kommt man zu dem/ oder kam ich zumindest zu dem Entschluss, dass ich sehr, sehr viele Geräte benutze, die aufnehmen können. Und dass das jetzt halt einfach nur ein weiteres ist."

by describing the situation as safe. This emerged in our interviews in five ways. (1) Users claimed that the way they used their device was benign, because they only used it for simple commands and requests – e.g., asking for a weather forecast, the time, or telling the device to set a timer or to play a particular song. According to these users, even if the data involved were to be analyzed, it would yield only trivial information about their household. In other words, the data concerned was declared harmless. (2) Others emphasized the harmlessness of the device itself. Many users cited the frequent occasions when a device did not understand a request as evidence that the AI in the background was not particularly perceptive and hence would not be able to analyze much of what was going on in their homes. Moreover, devices were also seen as harmless (3) because users could not imagine worse consequences of using them than being exposed to personalized ads. A further aspect of device harmlessness (4) was asserted by users who argued that their device brought a certain degree of comfort but emphasized that they were by no means reliant on smart speakers in the way they depended, for example, on their smartphone or laptop. In this way, smart speakers were portrayed as a 'toy' or something trivial and non-essential. Finally (5), users portrayed themselves as harmless. As Patrick Gensing – who uses the Apple HomePod system in his family home – put it:

> "I know it's an overused phrase, but who could possibly be interested in my conversations with my kids at home? It's just not interesting to anybody. I consider myself to be boring, so I wouldn't care if someone was listening to me." (Patrick G., intv. 1, l. 375–381)[14]

In view of such harmlessness, thinking too much about the potential use of data collected and speculating on dangerous consequences thereof was dismissed by some as "paranoia" (Julian R.) or as an example of susceptibility to a "conspiracy theory" (Robin L.); i.e., as unnecessary, problematic, or pathological behavior. Such a view was substantiated by the fact that none of those interviewed mentioned any direct and negative consequences of smart speaker use, except for a few unexpected reactions and malfunctions (cf. Lutz and Newlands, this volume). This supports our proposal that in order to maintain their

14 German: "Ich weiß, es ist ein abgedroschener Satz, aber was interessiert denn irgendjemanden, was ich Zuhause mit meinen Kindern bespreche? Das interessiert vermutlich keinen Menschen. Also ich halte mich da für langweilig, (lachend), insofern äh wäre mir das auch egal, wenn da jemand zuhört."

conviction that the device and its use are harmless, users have to actively refrain from delving into data protection discourse. This relates to the “positive thinking” mentioned by Pettenkofer (2017): When users see other users using the device effortlessly and thoughtlessly and without negative consequences, they assume that they too will be able to use the device without incurring harm.

5. Discussion

5.1 Fatalism

While this chapter is specifically concerned with smart speakers, users evaluate their use and problems within the wider contexts that shape their lives. One such context is the digital lifeworld of interconnected services that are, to a certain extent, always based on the analysis of user data. Another context is the user’s household, which is particularly relevant when devices are purchased for and used by all its members, who may have differing capabilities, needs, or interests.

When it comes to the cynical, apathetic, and resigned attitudes that other studies have identified among smart speaker users, we can confirm that our interviewees also viewed IPAs’ interfaces and corporate infrastructure as opaque and potentially problematic. At the same time, most did not express feeling bothered or frustrated by this but were inclined to disregard such issues – not just in their everyday use of the devices, but also when explicitly asked about their opinions in our interviews. Some completely refused to think about potentially problematic issues while others acknowledged in principle that there might be problems with corporate data practices. Both groups determinedly refused to make such issues *their* personal problem.

There is certainly a pragmatic aspect to such cynicism and resignation: it enables users to justify using the products and services despite acknowledging problematic aspects. This has been mentioned by Lutz and his colleagues (2020) as well as by Draper and Turow (2019). The users quoted in those papers and the terminology chosen in both suggest problematization by users that was, however, less prevalent in the interviews in our study. While we do not make any quantitative claims, we nonetheless suggest that there is not only hand wringing and negative views among users (cf., e.g., Hoffmann et al. 2016), but also a certain disregard of the topic of data privacy altogether, or it was raised as part of a more personalized evaluation. Users in our study who said that they never thought about corporate data practices and surveillance were by no means unaware of the critical discourse about them in media and the public sphere. However, they actively chose not to personally investigate the claims and issues or relate them to their own personal situation and use of the devices.

Other users reported that they were less concerned with the overall discourse than with their own situation. In their view, just about everyone in society is already part of corporate online platforms. This was not seen as representing a gross power inequality with corporations tracking and exploiting users who have no choice but to make use of online services; instead, it was presented as a justification for deciding to use smart speakers while not denying privacy concerns. This is what we glean, for example, from Samuel's narrative detailed above. He explained how he overcame his initial skepticism towards smart speakers simply by being exposed to them following his failure to convince his partner that they were not suitable for children. His girlfriend showed him what fun her children had with the device. Moreover, Samuel's recognition that he had already been using a portable version of a smart speaker did not lead him to consider ceasing to use Siri on his phone, but to reconsider his skepticism and ultimately to decide to increase his use of voice interface technology.

Smart speakers are an interesting technology in relation to fatalism, especially as users often portray them as an unnecessary luxury, as something futuristic they wanted to try out, or as a toy. This is quite different from when high-school students talk about how essential it is to use social media in order to not be left out (cf. boyd 2014, Englert et al. 2019). Such peer pressure does not tend to be experienced in relation to smart speakers, which have not been as widely adopted as smartphones. In 2020 in Germany, using a smart speaker was perceived as more of a personal choice than using a smartphone, which had come to be seen as essential in order to participate in much of everyday life. As such, the resigned and cynical arguments of users cannot really be explained by inevitability, since choosing not to use a smart speaker would not necessarily bring a great reduction in comfort or social standing. This argument does not hold for certain users, such as those with physical or visual impairments, who would stand to lose a great deal more by rejecting smart speakers.

5.2 Domestication

A progression from skepticism to adoption points not only to the affective dimension of technology use (cf. Bösel and Wiemer 2020), but also to a process that has been extensively explored in the field of "domestication" research (e.g., Bakardijeva 2005). Silverstone, Hirsch, and Morley (1992, 18) argue that media are not just consumed as content but also as object – collectively, in a household. The authors assert that this process already begins even before the media object enters the home. These arguments can be convincingly applied to smart speakers as well. Silverstone and colleagues note how media are "appropriated" (1992, 16) and adapted to fit in with domestic routines and lifestyles. Metaphorically speaking, media come into the home as something "wild" that needs to be "tamed" (Waldecker and Hector 2023). But just as the process of domestication turned hunters into shepherds and foragers into farmers, the domestication of media also has the potential to change the domestic "moral economy" (Silverstone et al. 1992) by inciting the establishment of new evaluations and everyday practices.

However, this is not simply a matter of a slow habituation to new devices and services. As becomes evident in the way Samuel's views changed, elaborated above, the "taming" process is social: it is shaped by discussions with peers and household members. In her study on the private adoption of the internet in the early 21st century, Maria Bakardijeva highlights how new users were guided by "warm experts" (2005, 99), i.e., individuals within reach who were more knowledgeable than the users who asked for their help. Bakardijeva notes that these warm experts not only provided the skills necessary to get private internet access up and running, but also motivated users to actually try using online services. Nowadays, some of this motivation and enthusiasm for smart home devices is promoted by social media tech influencers, which Stephen Neville, accordingly, terms "online warm experts" (2021). While the input from these actors is no doubt relevant, our interviewees reported that their personal contact with other people who used smart speakers was even more significant in arousing their interest in trying out a smart speaker themselves. As such, "appropriation" (Silverstone et al. 1992) is not just an individual task and is not just about the device itself, but involves finding a personalized stance to appropriate the mediatized discourse on online privacy and surveillance. As our study and others mentioned above have shown, users are well aware of the critique of the data-based online economy and the potential for surveillance. How they relate their own personal, domestic, specific media use to this discourse is, to a certain extent, also influenced by their interaction with other users. It also indicates that this "appropriation" is never complete but is a practical task, an ongoing activity that changes over time (Silverstone et al. 1992, 19).

In sum, the domestication perspective can also be used to paint an empirically rich picture of how users experience data practices and how they do and do not deal with them in everyday life. It can also help us understand how positive, negative, and disinterested views on the issue are formed and how they in turn relate to everyday practice.

6. Conclusion

While the data from our interviews, complemented by results from the media-linguistic strand of our research project (cf. Habscheid, Hector, and Hrncal, this volume), can only provide partial insights into the wider embedding and enmeshing of data practices in everyday life, they have nonetheless allowed us to showcase how the fourth shade of fatalism, the most openly pragmatic kind, is connected to day-to-day experiences of using IPAs and the circumstances that shape them.

As Pettenkofer (2023, 65) argues in his most recent publication on fatalism, fatalism creates "a new routine of selectively avoiding reflection, which creates new, self-sustaining forms of selective attention". Support for this assertion is found in users' emphatic insistence on the harmlessness of using IPAs. With this focus on harmlessness, they steer the discourse away from the potential dangers of corporate data practices and the surveillance inherent to IPAs. Users' "civil inattention" (Goffman 1972, 385) to these topics makes sense: interviewees mentioned that they were unable to assess the actual data practices that take place in the back-end, which are controlled and obfuscated by companies (Draper and Turow 2019). Hence, they can never know whether or how these data practices might affect them. As most of them had not personally encountered any negative consequences directly traceable to the data recorded through their IPA use, they had no reason to concern themselves with what they could not know anyway – thus, their pragmatic negligence of the data protection issue. Therefore, it is not without justification that many users see their smart speaker use as "mostly harmless" (Adams 1979) having not personally experienced noticeable harm. We would agree – at the level of the individual –

but less so when taking into account the analytical potential of the data collected from thousands of households.

Further research is needed to elucidate how fatalistic practices and attitudes are connected to specific circumstances (Pettenkofer 2023). The debate on privacy in media studies and connected fields, as traced above, has moved away from an individualistic focus to a more holistic picture, urging the need to consider corporate responsibility and the ways infrastructures and platforms present users' choices. To a certain extent, this debate mirrors the discussion on the individual and collective responsibility for climate change[15]. In Germany, climate change is now generally acknowledged to be a growing concern. Some voices in the discourse claim that consumers' choices can make a substantial contribution to increasing or reducing greenhouse gas emissions. This individualized allocation of responsibility has been countered, in recent years, by perspectives that emphasize the culpability of corporations (e.g., by showing how oil and gas corporations helped create narratives that focused on the "carbon footprint" of individuals in the first place, cf. Mann 2021). Regardless of these debates, individuals are, given the circumstances, often unable to avoid actions that produce carbon emissions even when they know about the negative effects. Certain infrastructural arrangements make it necessary for individuals to take a car to work or to work in carbon-intensive industries – or, on a more general level, such circumstances make it necessary for people to work in jobs that they find boring, degrading, or unacceptable (Graeber 2018; Chibber 2022, 106). From this point of view, the fatalism of tech users vis-à-vis data protection is by no means exceptional, but just one instance of a phenomenon that is constitutive of modern societies (Pettenkofer 2023). As such, exploring the mutual dissonance between collective *Sein* (being) and individual *Bewusstsein* (consciousness), so to speak, could offer a productive approach for investigating how individuals deal with problematic situations that can only be changed for the better if individuals address them not individually, but collectively.

15 The particularities of dealing with crises such as climate change via digitalization is an important aspect of the follow-up project of the research project this chapter is based on.

References

Adams, Douglas. 1979. *The Hitchhiker's Guide to the Galaxy.* London: Pan Books.

Bakardjieva, Maria. 2005. *Internet Society: The Internet in Everyday Life*. London: Sage.

Barnes, Susan B. 2006. "A Privacy Paradox: Social Networking in the United States." *First Monday*. https://doi.org/10.5210/fm.v11i9.1394.

Bösel, Bernd, and Serjoscha Wiemer, eds. 2020. *Affective Transformations. Politics – Algorithms – Media*. Lüneburg: meson press.

boyd, danah. 2014. *It's Complicated. The Social Lives of Networked Teens.* New Haven, CT: Yale University Press.

Chibber, Vivek. 2022. *The Class Matrix. Social Theory after the Cultural Turn*. Cambridge, MA: Harvard University Press.

Draper, Nora A, and Joseph Turow. 2019. "The Corporate Cultivation of Digital Resignation." *New Media & Society* 21 (8): 1824–39. https://doi.org/10.1177/1461444819833331.

Englert, Kathrin, David Waldecker, and Oliver Schmidtke. 2019. "Un/Erbetene Beobachtung. Bewertungen richtigen Medienhandelns in Zeiten seiner Hyperbeobachtbarkeit." *Digitale Bewertungspraktiken. Für eine Bewertungssoziologie des Digitalen*, edited by Jonathan Kropf and Stefan Laser, 215–36. Wiesbaden: Springer VS.

Gandy, Oscar. 1989. "The Surveillance Society. Information Technology and Bureaucratic Social Control." *Journal of Communication* 39 (3): 61–76.

Graeber, David. 2018. *Bullshit Jobs. A Theory*. New York: Simon & Schuster.

Goffman, Erving. 1972. *Relations in Public.* New York: Penguin.

Hargittai, Eszter, and Alice Marwick. 2016. "'What Can I Really Do?' Explaining the Privacy Paradox with Online Apathy." *International Journal of Communication* 10: 3737–57.

Hector, Tim, David Waldecker, Niklas Strüver and Tanja Aal, eds. 2023. *Taming Digital Practices: On the Domestication of Data-Driven Technologies*. Digital Culture & Society 9 (1).

Hoffmann, Christian P., Christoph Lutz, and Giulia Ranzini. 2016. "Privacy Cynicism: A New Approach to the Privacy Paradox." *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 10 (4). https://doi.org/10.5817/CP2016-4-7.

Khan, Shamus, and Colin Jerolmack. 2013. "Saying Meritocracy and Doing Privilege." *The Sociological Quarterly* 54 (1): 9–19.

Kuckartz, Udo. 2014. *Qualitative Text Analysis: A Guide to Methods, Practice & Using Software*. Los Angeles: Sage.

Lauer, Josh. 2010. "The Good Consumer. Credit Reporting and the Invention of Financial Identity in the United States, 1840–1940". *Enterprise and Society* 11 (4):686–94.

Lengwiler, Martin. 2004. *Privacy, Justice and Equality: The History of Privacy Legislation and Its Significance for Civil Society*. Discussion Papers 2004–503. Berlin: Wissenschaftszentrum Berlin für Sozialforschung.

Lovink, Geert. 2019. *Sad by Design. On Platform Nihilism*. London: Pluto Press.

Lutz, Christoph, Christian P. Hoffmann, and Giulia Ranzini. 2020. "Data capitalism and the user: An exploration of privacy cynicism in Germany". *New Media & Society*, 22(7): 1168–187. https://doi.org/10.1177/1461444820912544.

Mann, Michael E. 2021. *The New Climate War. The Fight to Take Back Our Planet*. New York: PublicAffairs.

Neville, Stephen J. 2021. "The Domestication of Privacy-Invasive Technology on YouTube: Unboxing the Amazon Echo with the Online Warm Expert." *Convergence* 27 (5): 1288–307. https://doi.org/10.1177/1354856520970729.

Norberg, Patricia A., Daniel R. Horne, and David A. Horne. 2007. "The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors." *Journal of Consumer Affairs* 41 (1): 100–26. https://doi.org/10.1111/j.1745-6606.2006.00070.x.

O'Reilly, Tim. 2012. "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software". *The Social Media Reader*, edited by Michael Mandiberg, 32–52. New York: New York University Press. https://doi.org/10.18574/nyu/9780814763025.003.0008.

Pettenkofer, Andreas. 2017. "Fatalismus. Über eine vernachlässigte Stütze sozialer Ordnung." *Berliner Journal für Soziologie* 27 (1): 123–50. https://doi.org/10.1007/s11609-017-0339-1.

Pettenkofer, Andreas. 2023. "Three Types of Fatalistic Practice." In *"Weltbeziehung" – The Study of our Relationship to the World*, edited by Hartmut Rosa, Jörg Rüpke, and Bettina Hollstein, 63–99. https://www.campus.de/e-books/wissenschaft/soziologie/weltbeziehung-17937.html.

Silverstone, Roger, Eric Hirsch, and David Morley. 1992. "Information and Communication Technologies and the Moral Economy of the Household." In *Consuming Technologies: Media and Information in Domestic Spaces*, edited by Roger Silverstone and Eric Hirsch, 13–28. London: Routledge.

Taddicken, Monika, and Jan-Hinrik Schmidt. 2016. Entwicklung und Verbreitung sozialer Medien. In: *Handbuch Soziale Medien*, edited by Jan-Hinrik

Schmidt and Monika Taddicken. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-03895-3_1-2.

Turow, Joseph. 2021. *The Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy and Your Wallet.* New Haven, CT: Yale University Press.

Waldecker, David and Tim Hector. 2023. "A Praxeological Approach Towards the Domestication of Connected Media Technologies." *Digital Culture & Society* 9 (1): 5–22. https://doi.org/10.14361/dcs-2023-0102.

Witzel, Andreas, and Herwig Reiter. 2012. *The Problem-Centered Interview. Principles and Practice*. Los Angeles: Sage.

Zuboff, Shoshana. 2020. "You Are Now Remotely Controlled." *The New York Times*, January 24, 2020, sec. Opinion. https://www.nytimes.com/2020/01/24/opinion/sunday/surveillance-capitalism.html.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*. London: Profile.

How to Make GDPR a Threat Again

Nikolai Horn in Conversation with Dagmar Hoffmann and David Waldecker

Abstract *Dr. Nikolai Horn is an expert for data protection. He has worked for the Foundation for Data Protection (established by the German government). Currently, he is a political advisor for iRights.Lab, a German think tank which deals with the legal and policy aspects of the digital sphere. The interview is a follow-up conversation after Nikolai participated in a round table on smart speakers and data protection at University of Siegen. It covers legal and political aspects of data protection regulation with regards to voice recordings.*

Dagmar Hoffmann (DH): What are the main problems with data protection with regards to voice interfaces?

Nikolai Horn (NH): First, we have to look at what personal data is relevant in this case and if there is a legal frame for dealing with the type of data and its use cases. That is, are we talking about recorded speech? And we need to consider if this recording can be considered personal data. And yes, we can consider it to be personal data because it can be used to identify the individual speaking. However, we also need to see if it is sensitive data. Article 9 of the European General Data Protection Regulation (GDPR) lists "special categories of personal data" (GDPR Art. 9), such as the sexual or political orientation of an individual, or its health-related data. Here, data processing is tightly regulated. I would suggest that recorded speech can be considered sensitive personal data in the sense of art. 9 for two reasons. Reason one is that recorded speech can be considered biometric data as it allows to identify an individual in the same way that fingerprints do. Second, recorded speech can be analyzed for health aspects of the speaker. Thus, it is personal and sensitive data. It requires explicit consent to be processed and the GDPR restricts certain forms of analysis of said data.

Also, the goals of the data processing and its context have to be considered. Again, certain use cases of recorded speech can be relevant for data protection. Recorded speech could be used for profiling; it could also be used to analyze speech patterns in order to create a personalized artificial intelligence (AI) model of an individual speaker. This would allow for another dimension of abuse. That is, with a growing risk of misuse of data, there has to be a higher level of protection, especially when it comes to recorded speech.

DH: We also have to consider how error-prone this speech recognition still is nowadays. When somebody calls me, it is hard for them to tell me and my daughter apart, but AI is able to do so?

NH: By now, AI is able to do so. This is obvious with regards to today's programs generating pictures and videos – and video is a lot more complex than audio to generate. For a good simulation of a speaker, current software needs to analyze maybe five minutes of recorded speech. The more audio available, the better the simulation is. A potential use cases is radio broadcasting: It has been suggested to replace newscasters and anchor persons on radio stations with AI after having analyzed recordings of their spoken words. Concerning your example: the better the algorithm has been trained, the better it can distinguish between you and your daughter based on really short samples of speech.

Right now, the EU is debating how to regulate AI. In order to tell artificial and human speakers apart, some suggest to create a watermark that highlights the artificiality of computer-generated speech. Also, one could think of an AI-based algorithm which can tell human and simulated speakers apart.

David Waldecker (DW): In our view, it is this potential of AI-based speech recognition and processing which is often not known to users of smart speakers and voice assistants. There is a lot of public debate on issues of data protection. Research done by others and ourselves has indicated that users are aware of this debate, and in our interviews (cf. Waldecker/Martin/Hoffmann, this volume), they often asked us if we planned on talking about this aspect of smart speaker use. However, concrete knowledge about the potential use cases of recorded speech and its economic potential seems to be lacking. Users we interviewed suspected that data from smart speaker use might be analyzed for the personalization of advertisement on other platforms. Smart speakers, at least until now, do not play ads, but the more popular models are part of platform companies like Amazon and Google which might actually use this data in the way suggested. We wonder, if there are other use cases which are economically viable and how these are problematic from a data-protection standpoint.

NH: If we look at use cases of natural language processing (NLP) – the AI method used to generate, understand and analyze human speech and writing – a number of points are worth mentioning. As the broadcasting example shows, it can be lucrative to automate speech. Also, it is relevant for data protection because somebody's personal data has been used here. However, it is also interesting to ask from a philosophical point of view: Who does somebody's voice belong to? That aside, speech analysis is also an interesting aspect because there are several potentials for application. Next to the ad placement you mentioned based on speech analysis, we can also think of the analysis not only of the content, but the speech itself, i.e. the analysis of feelings and the psychological state a person is in. This is technically possible and it could be used, e.g., for political advertisement. It could also be used in job interviews in order to automatically analyze the manner of an applicant. There is one real-life example of the Bundesamt für Migration und Flüchtlinge (BAMF), the German federal agency responsible for asylum claims and related matters. In an experiment, this agency had applied an AI speech analysis to discover if an applicant's dialect, in, say, Arabic, matches the country or region that the applicant claims to have come from originally. Here, we see the problem of this application. Critics were concerned with the use case itself and the methods applied; also, the system turned out to produce a lot of errors. These mistakes, if undetected, would have had serious consequences for the applicants.

Also, this AI audio technology could be used for surveillance purposes. For example, it could be used to find out who was present at a meeting of a group of people. There are innumerable scenarios for application. However, there are also potential benefits in the early detection of diseases and disabilities which are present in speech but usually go unnoticed. These forms of detection should conform to data-protection regulation, but maybe one day our cell phones could analyze our conversations on the phone and suggest visiting a doctor at some point. This sounds horrible to some, but it could be created in a non-threatening manner. So, advertisement really is not the most spectacular application I could think of.

DW: In passing, you mentioned the question of who one's voice belongs to. How would you answer this question?

NH: This is an interesting topic. To a certain extent, companies nowadays treat personal data and an individual's voice as a person's property which she can transfer to others. However, I find this economic perspective and the idea that spoken words are property misleading because it suggests that one can sell this data or information in some way. When we consider speech as a biometric marker, i.e. my individual way of speaking, my personal grain of the voice, so to speak – then we do not consider the economic aspects of property, but the legal aspects of ownership and control over something. Here, constitutional law and the fundamental rights of a person become relevant. These rights are not for sale. Thus, I cannot sell my fingerprint, or my voice, for that matter.

DW: Well, you highlighted a number of ways of analyzing speech as data. For us, the interesting thing is that – with all the potential inherent in speech analysis – the actual commands issued in our interviews to voice assistants were often quite trivial. Users often use their smart speakers to turn music or the light on or off, or they ask about the time or the weather. So, some users we interviewed suggested that their use is harmless from a data-protection standpoint, because their commands do not convey any personal or sensitive data. Also, they suggested that the data is analyzed by companies like Amazon or Apple who know about their habits anyway, by analyzing their shopping and online query behavior. In this sense, we were wondering if this public discourse on the surveillance of speech and the domestic sphere relates to an actual danger or to an much more benign phenomenon?

NH: I think this debate is not over. One phrase that comes to mind is "rational apathy." This term suggests that users are looking mostly for short-term benefits, like a more comfortable remote control for the living room lights, while not thinking about the long-term consequences. Of course, it is hard to anticipate the ways data about the lighting in private homes can be used; but even in this case, pattern recognition could be used to the detriment of the user – without them even knowing about it. Another example would be the simulation of individual speakers: You only need a couple of minutes of somebody speaking for a good simulation; this technology has been used in impersonation by criminals. However, it would be too much to ask for the average user to know about and to think through all these technological consequences.

As you mentioned, it is true that companies like Apple, Google and other platforms already collect a lot of data anyway. We do not know what these companies use the data for, and we do not know if they follow data-protection guidelines. Thus, it is doubtful if regulation for data protection is being taken seriously. And because users are unable to control the flow of data, it is questionable if "informed consent" is a correct descriptor of the actual situation when digital services are used.

I have been working in this field of data protection for a while now. In everyday life, I cannot think of many cases where I do know and understand how the data is being processed that is being collected. With certain companies, users can take for granted that their data is being kept in a closed system with high levels of security. With other companies, users and citizens cannot be so sure what happens to their data, for example with certain smartphone companies or with an app by a non-European company. It is unrealistic to assume that everything is processed according to the GDPR. It should, because users in Germany reside in the European Union, but certain companies and providers are based in authoritarian states where this legislation is hard to enforce. And so, before wasting too much time on this issue, users simply take it as a given – asking themselves "Why should something happen to me?"

DW: You mentioned a number of negative consequences. When we asked users about the potential problems of using a smart speaker, they were aware of some of these problems and stated that they would not mention banking account numbers and passwords in front of the device. So, there is this attitude of nonchalance you mentioned, and there also is this feeling that the device is harmless and has not caused any greater problems. From this point of view, one particular user suggested that anyone who is concerned about the data processing behind smart speakers and voice assistants has fallen for conspiracy theories in the vein of believing that vaccines actually inject nano robots into people's bodies. That is, this problematization that we are discussing here is very distant to certain users. On the other hand, all these problematic data practices you mentioned – how much do they really affect people in any way they can directly experience? And even if they do, how problematic are they for these users? It seems that the legal perspective is too removed from everyday life to matter. To a certain extent, the legal aspects you have mentioned are somewhat fictitious. In order to create the possibility for informed consent, every user of a smart speaker or even an activated voice assistant on a smart phone would have to inform every visitor or everybody in the vicinity about an active device and the consequences of data processing. Empirically, this is not the case. So, already in the everyday use of the device, legal obligations are not kept up with, neither by users nor the manufacturers.

NH: Well, this is one of the main problems. NGOs and other actors have tried to make data protection a popular issue, to get people concerned about it. De facto, these debates are only something for experts. Experts know about the potential problems and actual cases where things went wrong – for example in a case of a children's toy which recorded interactions with it. And before the activities of Cambridge Analytica became public, their application of Facebook's potential was somewhat the matter of fiction, too. Now, this is reality we have to deal with. We just have to ask ourselves, if we want to wait for another scandal of if we want state agencies to create certain boundaries and norms proactively.

I do not want users to be forced to think about data protection, instead I want users to be able to trust experts and agencies to take care of potential pitfalls in digital data processing. This kind of oversight should work like it does with cars: As a driver, I do not need to know how a combustion engine works, but I need to be able to trust the agencies that inspect and certify cars for their safety. Thus, I do not need to know if and how the breaks or a valve could malfunction as long as experts and governing bodies take care of these risks by countering them by prescribing high standards for quality and safety. I think that we need the same procedures for voice-related technologies, before they can get easily exploited by criminals.

Getting to this legislation and regulation will not be easy because it is hard to anticipate every scenario where users could be at risk. This is obvious with the EU AI Act where certain areas, such as human resources, are considered high risk, but it is complicated to imagine risks in all potential and future areas of application. This borders on technology assessment procedures where you asses concrete technological products. It is hard to exactly determine the risk connected to a particular piece of technology, especially with this type of technology and its wide field of application.

DH: While it might be hard to assess these current technological trends, Amazon and Google products nowadays enable users to listen to the recorded interactions, i.e., to view their technical interpretation and the answer or results. It is also possible to comment upon or even delete this data. Most of those interviewed did not even know about the possibility, neither to assess the recorded data, nor to delete it. We wonder if this is some kind of pseudo-transparency and ask ourselves, like some of our interview participants, if it is worthwhile to delete these recordings?

NH: This feature in certain smart speakers relates to the discussion concerning article 20 of the GDPR and the "right to data portability". Here, the GDPR stipulates that the "data subject" should be able to receive a copy of all the personal data saved about the subject by some organization and that the subject should be able to transfer this data to another entity. While this was hailed by some as a means of consumer sovereignty, it turned out to be a somewhat toothless piece of regulation. Google had implemented this possibility before the GDPR took effect, because it does not cost much to implement and because it is pretty useless for the individual user. Google has done what it was asked to do, so I do not blame Google. Instead, I wonder why data protection agencies thought this kind of regulation would be useful. Why do we need things like Privacy Information Management Systems (PIMS) or privacy enhancing technologies which focus on the individual user? I would suggest that we need to enable organizations, such as consumer protection agencies or NGOs like AlgorithmWatch, to analyze this data and to press for charges in a class-action lawsuit. I think that it is much more productive to look for ways to enable users to realize their rights on a collective basis than to provide them with technical tools that are only interesting for the individual experts, if at all. So, we have to look at the legislative aspects but also at the implementation of the regulation itself. We do have enough regulation in the abstract, calling for privacy by design or privacy by default, but we are lacking best-practice examples. And while it makes sense to be skeptical of Google and other companies in the field, we might even develop tools together with those companies which allow for a greater and more meaningful control of personal data.

DW: This is interesting as users currently often are left with the choice to accept the vague and obscure data policies of a digital service or to not accept them – which means the inability to use the service. It is this dyadic relationship between platform and user that leaves the user solely with a choice to "take it or leave it." You mentioned NGOs as trust-enhancing organizations. What do you think of institutionalizing data management in a triadic fashion?

NH: I dealt with questions like these in 2017 as a member of the "Stiftung Datenschutz" [Foundation for Data Protection]. In a research paper (Stiftung

Datenschutz 2017) on new directions in consent in data protection, we examined PIMS and their connections to digital ecosystems. So, I am not up to date on the state of the art concerning tools which, for example, manage the consent forms for users on different platforms. I also wonder if these tools that are offered are in use in any meaningful sense – just because you provide a handy software conforming to the GDPR, it does not mean that users integrate it into their everyday activities. So, we need to find out how to deploy and use these tools to have a lasting impact. That is, we need more research, in line with behavioral economics, but also in a more interdisciplinary setting. In addition, what we need is effective and powerful agencies able to enforce legislation and regulation concerning data protection. As public offices in this field are notoriously understaffed, we might need to resort to class-action lawsuits or a more coordinated effort with consumer protection bureaus. Essentially, we need to find ways to combine the legal possibilities inherent in this legislation and we need to foster new and powerful forms of regulatory oversight.

DH: Earlier in this interview, you suggested that users should be relieved from thinking about data protection at every turn. What can we ask users to do then, concerning voice assistants and data protection?

NH: While information and education are important, I am a wary to turn towards schools to educate our future citizens on these matters. This will help to a certain extent but will not get the job done. I think that we have to make digital products which adhere to privacy standards more appealing. This will allow for a better position on the market. Apple, for example, has been advertising its products as more secure and protective of personal data. This leads to greater trust in the brand as well. In the end, data protection should become one of the main reasons to use or shun a product. Data protection must not be viewed as an annoying hindrance, but as a protection of fundamental rights. Those digital products that protect (digital) fundamental rights and enable the exercise of these fundamental rights most effectively deserve respect and must be recognized as such. Maybe, with the advent of AI and the current debates in the dangers inherent in its ubiquitous application – maybe this will lead to a greater sensibility and a demand for products which are more attuned to data protection.

Translated by David Waldecker

References

Stiftung Datenschutz. 2017. “New ways of providing consent in data protection – technical, legal and economic challenges. Leipzig.“ Policy Paper. Available at: https://stiftungdatenschutz.org/fileadmin/Redaktion/Video/Fremdveranstaltungen/PIMS-Abschluss-Studie-30032017/stiftungdatenschutz_PolicyPaper_New_ways_of_providing_consent_in_data_protection_EN_final.pdf

IV Technical Infrastructures as a Practical Problem

Demystification of Technology

Empowering Consumers to Access and Visualize Voice Interaction Data

Dominik Pins, Fatemeh Alizadeh, Alexander Boden, Sebastian Zilles, and Gunnar Stevens

Abstract *Voice assistants (VAs) in households are becoming increasingly commonplace, with many users expressing their appreciation of the devices' convenience. Nonetheless, a notable number of users have raised concerns that the devices are 'always listening', and that there is a lack of clear information from providers about the data collected and processed through their microphones. Adopting a socio-informatics research perspective, we used the living lab approach to work with users over three years to investigate their uncertainties regarding the data collected by VAs in everyday usage. Based on our findings from interviews, fieldwork, and participatory design workshops with 35 households, we developed the web tool "CheckMyVA" to support users to access and visualize their own VA data. This chapter presents the observations and findings of the three-year study by outlining the implemented features of the tool and reflecting on how its design can help improve data literacy and enable users to reflect on their long-term interactions with VAs, ultimately serving to 'demystify' the technology.*

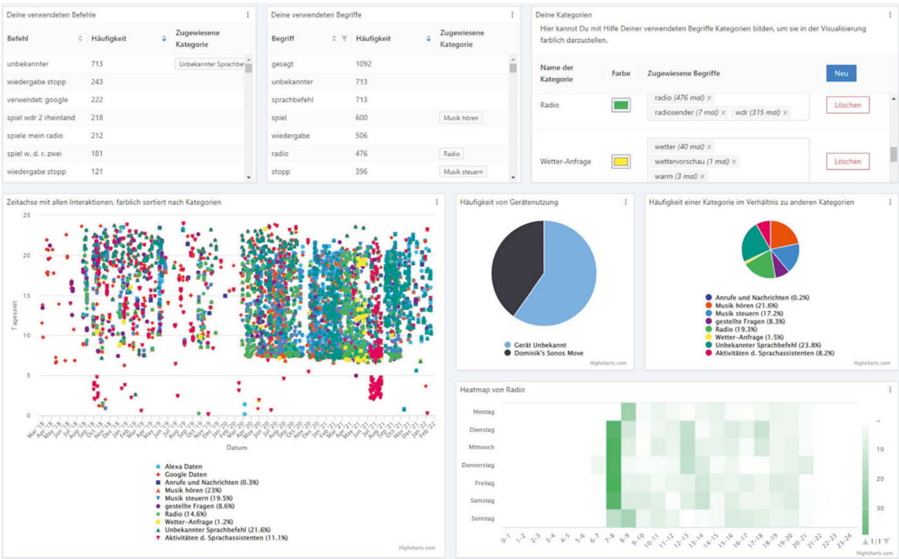
1. Introduction and Background

Since their launch in 2015, voice assistants (VAs) for home use such as Google Assistant or Amazon's Alexa have been steadily gaining prevalence (Bohn 2016), with the global market estimated to exceed 200 million devices in 2023 (Laricchia 2023). While users appreciate the usefulness and convenience of VAs, the ability to control these devices by voice also serves as a gateway to a growing ecosystem of data-based services (Strüver 2023a). Initial studies have shown that users are often unaware of what data these devices capture and whether or how their data is stored (Abdi, Ramokapane, and Such 2019; Alepis and Patsakis 2017; Jakobi et al. 2020; Pins et al. 2020). One reason for this is the lack of opportunities provided to users to learn about, understand, or manage the data collected by companies (Jakobi et al. 2020; Pins et al. 2020; 2021).

Figure 1: Extracts of raw data transcription files from data takeouts, received a) as a JSON file for Google Assistant and b) as a CVS file for Amazon Alexa

When it comes to tracking what VAs have captured or processed, providers do offer options such as interaction logs, which can be accessed in users' account settings (see an analysis of the log data by Habscheid et al. (2021)), or, in the case of Amazon, users can ask Alexa directly why it performed in a certain way (Alizadeh, Pins, and Stevens 2023). However, studies have shown that while these options make it quite easy to access recent interactions, they do not offer an overall view of interactions over longer periods, nor are they suitable for conducting in-depth data work (Pins et al. 2020). For this reason, we leveraged the right to access data guaranteed by the General Data Protection Regulation (GDPR) in order to obtain raw interaction data from a longer period of time with which we could explore different visualization methods. Figure 1 shows how the interactions were presented in the data takeouts supplied by Google and Amazon respectively. The interaction data for Google Assistant (shown in JSON format in Figure 1) exhibit a uniform structure for each interaction. However, the individual labels at the beginning of each line are not self-descriptive: laypersons would not necessarily find them helpful to understand the subsequent information. Amazon provides the transcription of Alexa interactions as a CVS file, which includes the timestamp, the user command, the name of the audio file, and the response from Alexa for each interaction, listed line by line. As can be seen, both of these formats lack legibility, especially for laypersons, and interpreting them requires a deeper understanding of the data structure (Pins et al. 2021).

Figure 2: Dashboard for data visualization – (exemplary view)



Our aim was to examine, in a living lab study, how users of VAs integrate the devices into their daily lives and, in particular, how they deal with uncertainties regarding VAs' recordings of everyday life in their homes – whether intentional or accidentally activated, for example, by TV or human conversations. Our approach was guided by the understanding that the appropriation of technology is a social process, whereby artifacts are incorporated into one's everyday life (Draxler et al. 2012; Stevens, Pipek, and Wulf 2010; Wulf 2018); this incorporation influences behavior and can lead to new practices, thought patterns, and design approaches as reciprocal effects (Rohde et al. 2017).

This contribution reflects upon our development of CheckMyVA: a web tool intended to empower users of VAs from different providers by preparing and visualizing their interaction data. Figure 2 shows some of CheckMyVA's visualization options that allow users to view the recordings and corresponding transcriptions stored by VA providers, thereby demystifying what VAs are listening to and helping users to reflect on their usage behavior.

2. State of the Art

2.1. Privacy Concerns About the Use of VAs

VAs are valued highly for their convenience and for the captivating way they enable users to operate music, connected devices, and entire home systems by means of voice commands (Purington et al. 2017; Abdi, Ramokapane, and Such 2019; Brüggemeier et al. 2020). However, for many people, their usage is also associated with opacity, concern, and mistrust (Lau, Zimmerman, and Schaub 2018). Additionally, users have expressed disappointment that VAs do not always react and respond reliably, and more complex tasks are not always completed successfully (Bentley et al. 2018; Luger and Sellen 2016; Pins et al. 2020).

The reasons for these negative sentiments often lie in users' uncertainty about what exactly VAs 'understand' or record and how they process data (Luger and Sellen 2016; Malkin et al. 2019). Recent research has shown that most privacy concerns are associated with accidental activations (Schönherr et al. 2020; Malkin et al. 2019; Ford and Palmer 2019) along with anxiety about the presence of a device that is 'always listening' (Alepis and Patsakis 2017; Lau, Zimmerman, and Schaub 2018). However, disappointment and frustration were also expressed about providers' failure to provide appropriate support to deal with problems, such as by suggesting repair strategies to clarify why a VA acted in a certain way or to successfully resolve misleading interactions (Kiesel et al. 2019; Pins et al. 2020; Pins and Alizadeh 2021). Studies have shown that users are often unaware that they can view interaction-related data and review or delete them (Malkin et al. 2019; Pins et al. 2021; Sciuto et al. 2018).

As a result of these operational difficulties and privacy concerns, users tend to adapt their use behavior by trying to make their voice commands as trivial, uninteresting, or short as possible (Lau, Zimmerman, and Schaub 2018; Malkin et al. 2019; Pins et al. 2020). This behavior can also be explained by rational fatalism (Kerwin 2012) or resignation as an attempt to protect one's data from companies (Pins et al. 2021; Xie, Fowler-Dawson, and Tvauri 2019).

2.2. Usable Privacy for Greater Data Literacy

Advocates of 'usable privacy' argue for the need to design secure systems from the user's perspective (Cranor 2008) and to support consumers to manage their own data privacy (Adams and Sasse 1999). This includes aspects such as improving privacy awareness (Langheinrich 2002), making security tools usable (Whitten and Tygar 1999), and making privacy notices understandable (Angulo et al. 2012; Schaub et al. 2018).

Against the backdrop of increasingly comprehensive and complex data collection, current research in usable privacy focuses on adapting the data literacy concept (Zhang 2018). This concept, which originated in the educational sciences, has been defined in various ways (see Koltay (2015) for an overview). In summary, data literacy involves the ability to access, interpret, critically evaluate, manage, and process data, so that it can be transformed into actionable knowledge to make informed decisions (Calzada Prado and Marzal 2013; Koltay 2015; Mandinach and Gummer 2016).

In our contemporary data-driven economy and society, data literacy is not only a key skill for individuals, but also a prerequisite for informed data protection regulation. The GDPR right to access data has created an important technical basis for promoting data literacy by enabling individuals to access information stored about them. However, there is a lack of complementary measures to ensure that accessed data can be understood and effectively managed. To address this gap and promote data literacy among consumers, it is expedient to draw on methods such as information visualization (InfoViz) (Shneiderman 1996), data citizen science (Marr 2016) and data work (Tolmie et al. 2016), and combine them with playful data exploration techniques (Jakobi et al. 2017).

In the research field of digital consumer behavior, artificial intelligence and data science methods such as deep learning (Chapman and Feit 2019; Feldman and Sanger 2006; Igual Muñoz and Seguí Mesquida 2017) are increasingly used alongside classical statistical methods to identify relevant information in user data and to derive behavioral patterns. However, such methods are typically available only to companies and data scientists; there is a lack of usable solutions for consumers that enable automated analysis for different fields of application (Fischer et al. 2016).

InfoViz methods facilitate the visualization of time series, networks, and hierarchical data (Aigner et al. 2007; Ware 2013), which can reflect users' behaviors back to them (Castelli et al. 2017; Jakobi et al. 2017; Stevens et al. 2017).

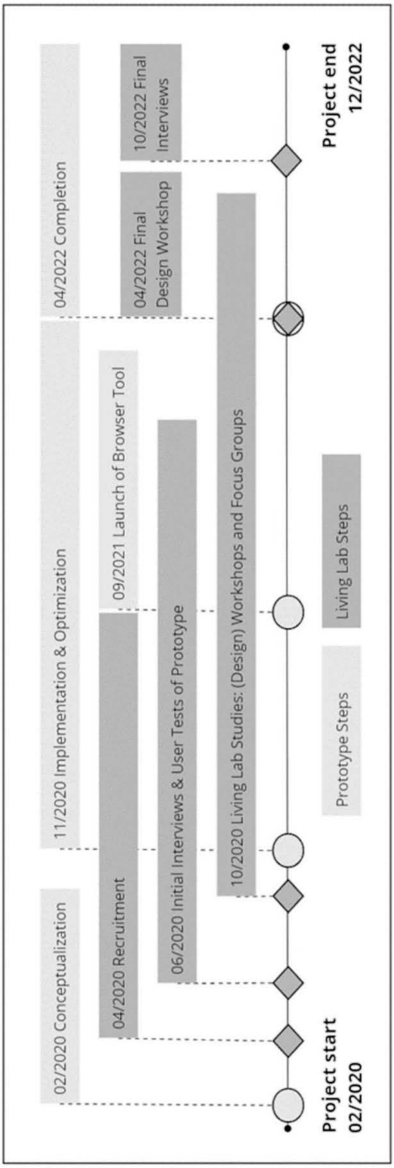
Initially, we were aware of just three further studies that had explicitly used log files of interactions to investigate the use of VAs (Malkin et al. 2019; Sciuto et al. 2018; Bentley et al. 2018). During our study, additional research examined interaction data to draw conclusions about human–VA interaction within the smart home ecosystem (Habscheid et al. 2021) and to assess privacy sensitivity and intimacy using data sharing scenarios (Gómez Ortega, Bourgeois, and Kortuem 2023).

While these studies primarily analyzed data for research purposes, our aim was to design a tool that could directly help consumers themselves to explore and understand their data, ultimately empowering them by improving their data literacy relating to VA use. The findings presented in this chapter build on a previous study that tested the process for requesting interaction data and evaluated an initial prototype for data visualization (Pins et al. 2021). Since then, we have completed the research project and are able to present the results of the iterative design process here.

3. Methodology

In this study, we adopted the 'living lab' approach to investigate ways to promote data sovereignty in the use of VAs. A living lab can be understood as a user-centered research methodology for sensing, prototyping, validating, and refining complex solutions in evolving real-life contexts (Eriksson and Kulkki 2005). Our procedure also incorporated the practice-orientated problem-solving strategy deployed in design case studies (Wulf et al. 2011). This approach takes into consideration the user, their (social) practices, institutional arrangements, and technological infrastructures, thereby exploring the design of innovative IT artifacts in situ (Wulf et al. 2015).

Figure 3: Project Timeline



Our aim was thus to study participants and their behavior in real-world settings, gaining insights into their use and understandings of voice interaction data. For the living lab study, we used mixed methods including interviews, fieldwork, and (design) workshops, in order to identify and validate users' needs and requirements. This iterative process enabled us to design, develop, and optimize a prototypical web tool (see Figure 3 for an overview of the research phases).

Parallel to the living lab study, we used several data donations from our participants to test the efficacy of various machine learning (ML) models to draw conclusions about users based on their data (digital consumer analytics), for example, to identify characteristics of users or their households. Unfortunately, the data set proved too small for the models to be trained precisely enough to be of practical use in the prototype.

Shortly after project launch in February 2020, we recruited households for the living lab via digital and social media.[1] By summer 2020, we were able to begin an initial needs assessment and evaluation of our first prototype with a sample of 12 households.

Over the course of the project, we worked with a total of 35 households. With each household we were in contact with a main participant who was the administrator of the VA and had access to the interaction data. These participants ranged in age from 18 to 56, with a mean age of 33. The sample included 24 males and 11 females, who lived in single and partner households, family households, and shared apartments. Sixteen households were 'beginners' who had never used a VA at home before joining the research project (for greater detail, see Table 1).

1 Due to the contact restrictions imposed by the simultaneous outbreak of the Covid-19 pandemic, no other recruitment strategies were practicable.

Table 1: Living lab household participants

#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H1	m	31	Software developer / consultant	Starter	2	Amazon Apple	2	high	high	very high	very high	very high
H2	m	35	Service technician (electrical)	User	3	Amazon	8	very high	very high	medium	very high	high
H3	Participation withdrawn											
H4	m	31	Public sector employee	User	5	Google	2	high	high	high	medium	high
H5	w	32	Market and social research specialist	Starter	3	Amazon Apple	2	high	high	high	medium	high
H6	m	24	Trainee in the higher forest service	User	2	Amazon	2	medium	medium	high	high	very high
H7	m	28	Electrical engineer	User	3	Google	5	high	high	very high	high	very high
H8	w	26	Research associate (HCI)	User	1	Amazon	1	little	medium	medium	little	medium
H9	m	56	Sales officer	User	2	Google	5	very high	very high	very high	very high	very high
H10	m	35	Industrial salesman, real estate industry	User	4	Amazon	4	high	medium	high	medium	high

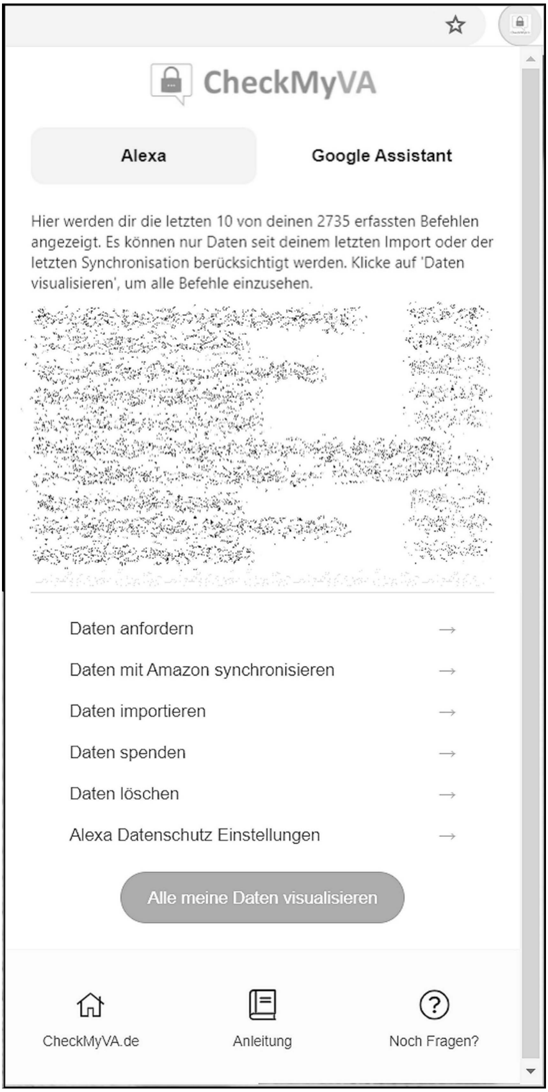
#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H11	m	27	PhD student (Physics)	User	1	Google	3	high	very high	medium	high	medium
H12	m	18	Student	Starter	4	Google	2	N/A	N/A	N/A	N/A	N/A
H13	m	45	IT	Starter	4	Amazon	1	very high	very high	very high	high	very high
H14	w	32	N/A	Starter	1	Amazon	1	N/A	N/A	N/A	N/A	N/A
H15	w	27	Architect (residential)	Starter	2	Amazon	1	medium	very high	high	very high	medium
H16	w	40	Project manager	Starter	1	Amazon	1	medium	medium	medium	medium	medium
H17			Participation withdrawn									
H18	m	34	N/A	User	3	Amazon	4	medium	high	little	medium	little
H19	m	23	IT consultant	User	2	Amazon	3	medium	high	little	very high	little
H20	w	40	Research associate	Starter	2	Google	1	medium	high	high	medium	N/A

#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H21	m	29	Mechanical engineer (packaging industry)	Starter	3	Amazon	1	medium	high	very high	very high	medium
H22	m	27	N/A	User	2	Amazon	2	N/A	N/A	N/A	N/A	N/A
H23			Participation withdrawn									
H24	w	30	N/A	Starter	1	Google	2	N/A	N/A	N/A	N/A	N/A
H25	m	32	Scientific staff	User	2	Amazon	2	high	high	little	very high	high
H26	m	48	Scientific staff	User	2	Google	2	high	high	medium	high	medium
H27	m	21	ServiceNow-developer (ITSM)	Starter	1	Amazon	2	high	very high	medium	very high	high
H28	m	34	Research associate (Computer science)	User	1	Amazon Google	2	high	high	medium	high	medium
H29	w	N/A	Research associate	Starter	1	Amazon	2	medium	medium	high	medium	high
H30	w	27	Research associate	Starter	1	Google	1	medium	medium	high	little	medium

#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H31	m	21	Student (Speech therapy)	Starter	2	Google Amazon	2	medium	medium	high	medium	medium
H32	m	34	Student	User	2	Google	2	high	high	high	high	medium
H33	m	N/A	Software developer	Starter	1	Google	1	little	high	high	No interest	high
H34	m	42	Sales for control	User	2	Google Amazon Siri	4	high	very high	high	high	high
H35	w	48	Freelancer (digital	Starter	1	Google Amazon	2	very high	very high	high	very high	very high

4. Findings and Implementation

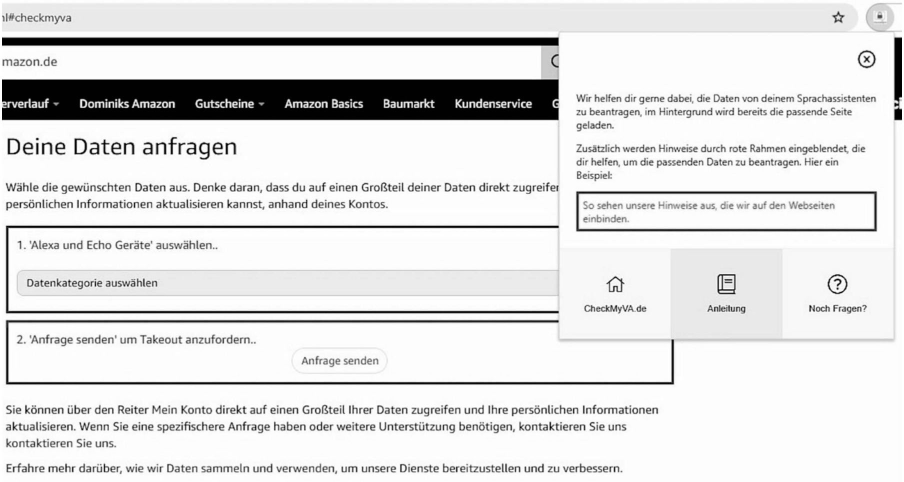
Figure 4: Main Menu of the CheckMyVA tool



The living lab study led us to design and produce CheckMyVA, a web tool that offers consumers two services: a data export wizard and a dashboard for data visualization, which can be accessed from a main menu (shown in Figure 4). The data export wizard directs users to VA providers' export websites and guides them with helpful dialogues through the often laborious and obscure export process. The dashboard enables consumers to display various data visualizations of the interaction data once they have accessed it. The tool is freely available as a browser extension for Google Chrome and Mozilla Firefox and can process data from Alexa and Google Assistant.[2]

4.1. Data Export Wizard

Figure 5: Guidelines for requesting a data takeout from Amazon



In our previous study, users had reported experiencing difficulties in finding user data, and that the process of retrieving it was very cumbersome and user-unfriendly (Pins et al. 2021). Hence, we created the data export wizard with the aim of supporting users through the process of exporting data from

2 For Google Chrome: https://chrome.google.com/webstore/detail/checkmyva-browser-erweite/kpllpbalbkdcdoklbnjlbbbeapfhoodp (18.05.2024) For Mozilla Firefox: https://addons.mozilla.org/de/firefox/addon/checkmyva/ (18.05.2024)

Google and Amazon. With a single click, users are directed to the appropriate export web pages and are guided by help dialogues in boxes highlighted in red (see Figure 5), making it easy for them to request data exports from their VAs.

Once the user has received the data takeout, the wizard processes and reads the data locally in the background. The stored data is then made available through an interface between the web tool and the dashboard. This ensures that data remain secure in the browser without needing to be uploaded to other services. Users do not have to unpack data archives or search for and open the relevant files themselves.

Obtaining a data takeout from Alexa can take from several days to several weeks, and to obtain the latest interaction data, a new takeout request must be made each time. To address this issue, we explored alternative methods to make the latest interaction data available to participants more quickly. We successfully implemented a system that enables interactions to be synchronized with our dashboard in real time. This real-time approach was well-received by participants. To make the process even simpler, we implemented another function that synchronizes data each time the browser is started.[3] In addition to data request and synchronization, we added the following features to the wizard (see also Figure 4):

- **Import data**: Users can import locally-stored interaction data.
- **Delete data**: Users can delete the data stored in the browser and the prototype.
- **Privacy settings**: After viewing the data, some users wanted to check their settings. For this reason, we added a link that directs users to the privacy settings in their Google or Amazon accounts so that they can make quick adjustments.
- **Data donation**: This feature allows users to transmit their data stored in the prototype (transcribed voice commands, responses, timestamps, devices, etc.)[4] to an internal server for further research within our research project, including user evaluation and training of ML models. Users must explicitly opt in to this procedure.

3 For Google Assistant data, this took a few minutes to a few hours, which participants considered acceptable. However, we have not found a way to synchronize the data in a similar way to Amazon.

4 Due to the large size of audio files, and because the dashboard could not process audio data anyway, we limited the data donation to textual data only.

4.2. Data Visualization Dashboard

The initial prototype for data visualization featured a timeline that helped participants gain an overview of their interactions over a longer period of time (Pins et al. 2021). This visualization was evaluated by participants as very useful and informative. Furthermore, the categorization of interactions according to specific terms enables the data points to be structured along the timeline in relevant ways, helping users to identify frequent or typical usage times and situations. Step-by-step categorization also facilitates the identification of further interaction patterns and of unusual or unexpected interactions or recordings. For instance, participants often expressed their surprise at discovering unexpected activities at night, or mentioned that they became aware that in viewing the visualization they were surveilling the interactions of other household members at times during which they themselves had been absent, e.g., when they had been at work (Pins et al. 2021).

Based on the results of the ongoing iterative process, we continued to optimize and extend the initial prototype. Like the data export wizard, the dashboard is implemented as a web application. It can access the user's data via the data export wizard automatically and offline, performing like a native desktop application. A screenshot of the final version of the visualization dashboard is shown in Figure 2.

In the process of preparing data for visualization, transcription errors (such as miscoded umlauts or punctuation marks) and VA command words ("Google", "Alexa" or "said") are removed to facilitate the visualization. Once the preparation is complete, users can create categories based upon individual command words using the Boolean operations (AND or OR). Each category can be assigned a color and a theme (see Figure 6). Additionally, we created a catalog of predefined categories that users can select from and customize.

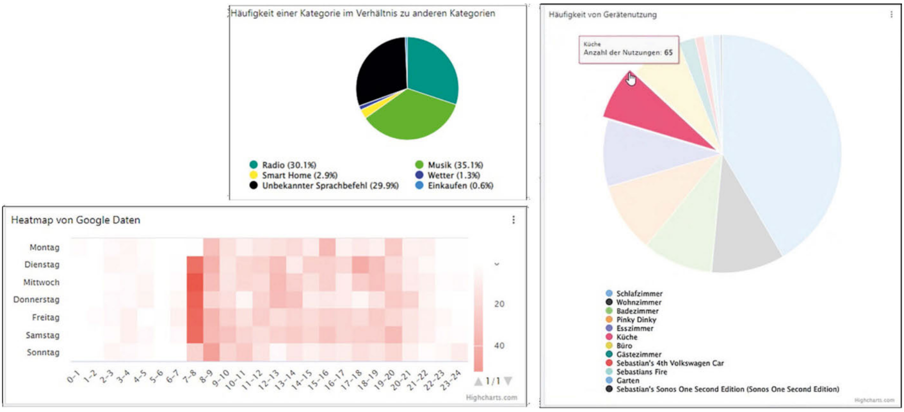
The categorized data thus forms the starting point for different visualizations and analyses. A scatter chart (timeline) displays each command as a dot in the color of the defined category (see Figure 2; bottom left of screen). This visualization shows the frequency of commands per category and thus helps users identify behavioral patterns associated with frequently-used categories. Users can zoom in on specific areas by dragging a frame over them with the mouse. Finally, by clicking on a category in the legend, all the corresponding dots on the timeline can be shown or hidden.

Figure 6: Widget for creating new categories based on terms.

Over the course of the project, we added various widgets based on users' needs and interests. Widgets are tiles with data visualizations that can be freely arranged on the dashboard, allowing users to customize their views and prioritize the information most important to them. Each widget also offers the option to display data from an individually-defined time period. The following widgets were implemented:

- **Word list**: Following evaluation of the initial prototype, we modified the sorting options for the list of words spoken so that they could be presented either in order of frequency or alphabetically. The list also shows which categories a term has been assigned to.
- **Command list**: To meet participants' requests for a list of spoken commands, we added another list with the same presentation options as the word list. It helps users to identify speech patterns.
- **Usage occasions and their frequency**: A pie chart and bar chart showing the relative distribution of categories (see Figure 7, top left).
- **Device usage**: A pie chart showing the relative distribution of devices used (see Figure 7, right). This enables users to check the frequency of device use and draw conclusions about the associated rooms in the home.
- **Occasions of use per day and time**: A heat map that shows the number of commands in a given category aggregated into hours per week. Each field is displayed in varying intensity of the category color depending on the frequency of use (see Figure 7, bottom left). This helps users identify typical usage times per category.

Figure 7: Widgets for data visualization: Relative frequency of each category within the total data set (top left), heatmap with clusters of interactions of a selected category (bottom left), and relative frequency of device usage by (assigned) device name (right).



We also conceptualized some additional widget designs in participatory design workshops in which participants expressed their needs and interests. The limited project timeframe prevented these widgets from being implemented into the tool, but participants' request for them nonetheless constitutes a significant research outcome. The following three design concepts for widgets would help users to gain a better understanding of VAs' data processing procedures:

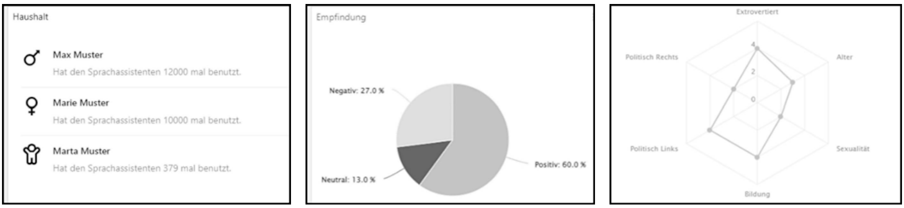
- **Speech analysis**: A widget for categorizing and detailing commands in order to correct interaction/pronunciation differences and recognize changes in interaction behavior.
- **Data flows**: A widget to show how (and with whom) data is shared, identifying critical or personal data and providing user action options.
- **Memories**: A widget for saving interaction data as material that can evoke memories of appointments, special occasions, or situations; supported by images or sound if these are available or linked to the data.

Finally, we conducted internal tests with ML models to explore how the data could be used in digital consumer analytics. The main goal was to identify profiles of users or their households. The users as well as our research team were interested whether the data could convey information about household size, age, gender, or a speaker's mood when interacting with the VA. We asked participants for data donations to test various ML models.

Due to the small size of the donated data set we were unable to train the models precisely enough to achieve conclusive results. Nevertheless, to give users a sense of what information could potentially be extrapolated from the data, we generated mock-ups based on the available data. These mock-ups present insights in the following widgets:

- **Household widget**: A list of all VA users, distinguishing individual voices and creating profiles that record their age, gender, and frequency of use of the VA (see Figure 8, left).
- **Sentiment widget**: A pie chart showing how often a particular command is executed with a positive, neutral, or negative intonation (see Figure 8, center).
- **Politeness widget**: Emojis indicating how politely users speak to the VA.
- **Health widget**: A scatter chart showing how often a user is sick, based upon audible symptoms like coughing, sneezing, hoarseness, or fatigue.
- **Background noise widget**: A scatter chart showing the frequency of certain background sounds and any incorrect activations they may have caused. For example, it indicates how often media (TV, radio, music), other conversations, or other sounds are present in the background.
- **Advertising widget**: A word cloud visualization of the brand names mentioned in voice commands.
- **Profiling widget**: A spider chart ranking inferred personality traits (see Figure 8, right).

Figures 8-10: Widgets for the household/user profile: Amount of use per household member (left), inferred personality traits of a user (center), and inferred positive, neutral, or negative mood when articulating a voice command (right).



5. Discussion

5.1. Data Work Promotes Data Awareness and Literacy

In our living lab study, participants expressed great interest to try out the CheckMyVA tool for the first time, and reported that using it made them feel reassured (Pins et al. 2021). Over the course of the study, however, only a small number of participants continued to use the tool regularly on their own initiative. In final interviews, the following reasons for using the tool were mentioned:

- Curiosity about what new interactions had been detected or stored by the VA.
- Coming across the tool icon by chance while using the browser.
- Checking for funny answers given by the VA.
- Checking for interactions including insults by others (and deleting them).

While the first two reasons indicate curiosity or the 'accidental rediscovery' of the tool, the last two are motivated by the desire to review unusual situations and interactions. This might explain why the majority of participants did not use the tool again; they may not have expected any new insights or unusual interactions, and therefore felt no subsequent need to explore the data. When asked in which situations they thought the tool might be helpful, several participants mentioned reviewing unexpected or incorrect responses. This suggests that after an initial 'awareness' check, users' interest in the data shifts over time, with the most attention concerning deviant activities. Such a shift has also been identified in other studies with different data work contexts (Castelli, Stevens, and Jakobi 2019; Jakobi et al. 2018).

5.2. Towards Better Support in Requesting Data (According to Article 15 of the GDPR)

The study has shown that the procedures of requesting data collected by VAs are neither simple nor easily comprehensible from the user's perspective (Pins et al. 2021). Tools like our prototype that can guide users through the data request process thus make a valuable contribution to increasing data literacy and users' knowledge. Easily locatable and accessible guidance on how to view or request data from each provider can help users overcome barriers to addressing the issue of data collection, thereby increasing their competence to use products, services, and systems, as well as to manage their collected data.

It also became apparent that different corporations deal with the volume and format of users' data in very different ways (Cena et al. 2016; Shafagh and Hithnawi 2017; Pins et al. 2022). Even between the two VA systems considered in this project, approaches vary significantly. Initially, we had planned to include interaction data from Siri (Apple), but that proved to not be possible due to their pseudonymization process, which prevents access to usable interaction data. Additionally, the ongoing development of these systems appears very inconstant. For instance, in response to public criticism, Google suspended the automatic storage of audio recordings for a while. Since then, it changed its policy so that Google Assistant users can currently opt in to anonymized data storage to improve speech and audio recognition, which may involve human review.[5]

While such pseudonymization (or anonymization) practices are to be welcomed from a data protection perspective, their effectiveness remains questionable if conclusions can still be drawn from the content of audio or transcript data, even when it has been separated from user profile data. Amazon also allows Alexa users the option to disallow the storage of interaction data, but this requires deliberate deactivation by the user – if the default settings are not adjusted, users implicitly consent to data storage.

Policy makers should work to generate more guidelines for the storage of user data and should make corporations accountable for providing easily-accessible, relevant information about the collection and processing of users' data, especially regarding the companies' evaluation and analysis processes for consumer analytics purposes. For example, information should be provided on how sensitive information is handled when recordings are subject to human review.

5.3. Towards Demystification: Visualization and Sense-making of Data

In accordance with Article 15 of the GDPR, companies are obliged to provide consumers with information about their personal data collected by the company, and to transmit it upon request in a machine-readable format for the purpose of data portability in accordance with Article 20 (European Parliament

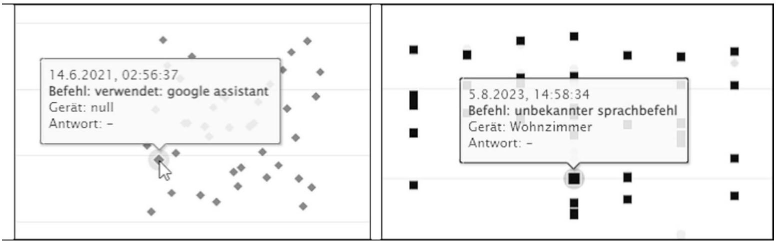
5 https://www.cnet.com/home/smart-home/googles-privacy-controls-on-recordings-changes-what-that-means-for-your-google-home/

and the Council 2018). The personal data of our participants that we requested for this project was raw data, which consumers without technical knowledge would barely understand (see Figure 1). Previous studies have shown that users expect a more human-readable format (Alizadeh et al. 2019; Pins et al. 2021). This highlights the need for solutions like our visualization dashboard, which enables users to make sense of their data and to better understand what it can show about their usage behavior (Castelli et al. 2017; Jakobi et al. 2017; Stevens and Bossauer 2017). While some information from the raw data takeout was clear and actionable, a major challenge for us was to identify significant insights that could potentially be derived from the data in order to draw conclusions about users and their behavior. These relevancies are not clearly evident within the raw data sets, which makes it difficult for users to understand the profiles created about them, and there are no end-user options that would enable them to create their own analyses.

In order to meet our aim to empower VA users to understand and interpret the data that companies collect from them, we needed a sample data set with which we could demonstrate in an exemplary way to users the potential analytical capabilities of companies. Our ability to do this was limited by the small size of the data set that we were able to obtain voluntarily from the few households that were willing to donate their data. We believe that companies should make it more transparent how a user's profile is compiled and what criteria are used to generate such profiles, so that users can understand and adjust settings accordingly if they so wish. This transparency could balance the information and power disparity between the user and provider, without requiring corporations to disclose their algorithms, but nonetheless helping to clarify or 'demystify' the opacity of technologies like VAs. Indeed, the few households in our study that agreed to donate us their data only did so once they understood what it included, suggesting that transparency might influence users' decisions about sharing data, especially when they feel uncertain about how the data could be analyzed and interpreted.

During the study and data analysis, certain inconsistencies in the data takeouts became apparent. For example, Amazon provides information about the device used for each interaction in the accounts' interaction log, which can be found in the account settings. However, we could not find corresponding information in the data takeouts. This suggests that some data correlations are not included in the takeouts, even though some connections between transcripts and recordings are traceable. Similarly, with Apple, it cannot be ruled out that personal data may still be found in the data records that are stored pseudonymously. Previous studies (Malkin et al. 2019; Pins et al. 2021), found that consumers were surprised to learn that voice commands were stored long-term. Figure 9 illustrates two activities of the VA shown on the dashboard that were included in the data set that users were surprised to discover had been stored, especially as such activities had an unclear purpose or occurred at unusual times.

Figure 9: Unusual activities of the VA without intelligible clarification.



Providers should therefore be held to account to make their data storage 'transparent' in the sense that users should be able to understand which elements of the data are interlinked for companies' analytical purposes (without firms having to disclose their algorithms or methods used). Companies should also be required to delete data that no longer serves a purpose.

5.4. Raising Awareness of the Technological Infrastructure in Which the VA is Embedded

For most participants, our study provided their first ever chance to view and engage with the data collected by their VAs. On the one hand, they said they felt reassured, because they had gained more clarity about what data the VA was collecting and how they could exert control over its transmission. In particular, it became clear that the majority of the data and usage situations (e.g., setting timers or playing music) that the participants learned about were not considered risky, concerning, or sensitive. This enabled them to act more self-determinedly when talking or acting near to a smart speaker at home. But on the other hand, viewing the data raised new questions, as they had expected to be able to obtain more information directly from the (raw) data received about the extent to which data was exchanged between various services. Instead, they initially found themselves confronted with a folder directory comprising incomprehensible data records that first had to be 'decoded' (Pins et al. 2021).

Research on VA systems should never consider them in isolation, but always in the context of the environment and linked services within which they are embedded (Strüver 2023a). Consumers express particular uncertainty regarding the extent of corporations' access to and exchange of data (Huang, Obada-Obieh, and Beznosov 2020; Luger and Sellen 2016; Malkin et al. 2019). Recent research and our study indicate the importance of viewing the home holistically, as a network of different players, in order to understand various links and activities in context (Strüver 2023b; Häußling 2017). For example, further research could distinguish between smart home products and services used (or their manufacturers) to provide more differentiated information about their general usage or integration in everyday practices.

A holistic view of the infrastructures or platform systems (Plantin et al. 2018) would also help consumers to create transparent and trustworthy environments for themselves, which is particularly important for private and intimate areas like the home. Recent studies have furthered understanding of the basic intentions behind data collection/processing (Sadowski 2020; Strüver 2023a; 2023b). Our approach also focuses on showing users what the storage of interaction data can mean for them, their household, and their usage behavior. Further research should link these aspects more closely to help users better understand how their data is affected by corporations' intentions. To conduct such research effectively would require a larger data set than was available to us for this study.

Current data work practices offered by companies usually only address the account owner/administrator (Meng, Keküllüoğlu, and Vaniea 2021). Therefore, a more holistic view of the home (technology) ecosystem is needed to achieve a multi-user-centric design, creating more productive, convenient, and inclusive IoT environments for other household members, visitors, etc. (Strüver 2023b). This approach would allow more people to gain insights into the interaction data and learn what the VA has captured about them and their households (Meng, Keküllüoğlu, and Vaniea 2021; Strüver 2023b; Waldecker, Hector, and Hoffmann 2023).

5.5. Limitations and Reflections

The scope of this study was limited by the sample. First, we engaged primarily with the administrators of the devices who had direct access to the interaction data via their user accounts; hence we focused mainly on their needs. As our study showed, and as other studies have demonstrated in greater depth (Thakkar et al. 2022; Sun et al. 2021), other residents in a household are also affected by a VA's data collection – but they were not included in the study. These individuals should also be able to view the interaction data to see what the VA provider or account holder can see about them. Our tool provides an initial indication of how this could be achieved by making the dashboard accessible to other household members, for example, via a shared device (e.g., a tablet or PC) with the tool's browser application installed.

Second, by only working with administrators, our sample could have been affected by a demographic imbalance in terms of age and gender. Previous studies (Geeng and Roesner 2019; Pins et al. 2020; 2021; Shin, Park, and Lee 2018) suggest that administrators tend to be male and tech-savvy, which may influence their interest in using VAs. However, this study did not aim to be representative but rather to support consumers who use a VA. Nonetheless, other user groups might express different needs and interests relating to data access from that we were not able to take into account in our study.

Another limitation arises from the dynamic nature of data usage and the services available to consumers at any specific time, which is constantly changing in response to ongoing developments, public criticism, and policy changes. Hence, replicating this study at a later time might yield different results.

6. Conclusion

In this chapter, we presented the features of a web tool created with the aim to empower VA users by increasing their data literacy, and reflected on the tool's development. This involved conducting a three-year living lab study to investigate VA use and data work practices, identifying what users need in order to better understand how VAs collect and process interaction data. The tool includes a data export wizard that guides users through the process of requesting interaction data as well as assisting them in viewing and managing privacy settings. It also offers a dashboard that allows the data to be structured and visualized in different ways (e.g., according to user-defined categories) to help users better understand and reflect on their usage.

Previous studies have shown that users often express uncertainty and skepticism about what their VA is listening to and storing; similar sentiments were voiced by our participants. Our tool addresses this by demystifying VA systems for users, enabling them to explore their own behavioral patterns through visualizations and to recognize unconscious or accidental activations. Ultimately, the tool helped participants to assess what data the VA collects and what it can reveal about a person or household.

Acknowledgements

The project was supported by funds of the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program.

References

Abdi, N, K M Ramokapane, and J M Such. 2019. "More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants." In *Proceedings of the Fifteenth Symposium on Usable Privacy and Security*, 15:17. Santa Clara: USENIX Association. https://www.usenix.org/system/files/soups2019-abdi.pdf.

Adams, Anne, and Martina Angela Sasse. 1999. "Users Are Not the Enemy." *Communications of the ACM* 42 (12): 40–46. https://doi.org/10.1145/322796.322806.

Aigner, Wolfgang, Silvia Miksch, Wolfgang Müller, H. Schumann, and Christian Tominski. 2007. "Visualizing Time-Oriented Data – A Systematic View." *Computers & Graphics* no. 31 (June): 401–09. https://doi.org/10.1016/j.cag.2007.01.030.

Alepis, Efthimios, and Constantinos Patsakis. 2017. "Monkey Says, Monkey Does: Security and Privacy on Voice Assistants." *IEEE Access* 5: 17841–51. https://doi.org/10.1109/ACCESS.2017.2747626.

Alizadeh, Fatemeh, Timo Jakobi, Jens Boldt, and Gunnar Stevens. 2019. "GDPR-Reality Check on the Right to Access Data: Claiming and Investigating Personally Identifiable Data from Companies." In *Proceedings of Mensch Und Computer 2019 on – MuC'19*, 811–14. Hamburg: ACM Press. https://doi.org/10.1145/3340764.3344913.

Alizadeh, Fatemeh, Dominik Pins, and Gunnar Stevens. 2023. "User-Friendly Explanatory Dialogues." In *Mensch Und Computer 2023– Workshopband,* 3. Rapperswil (SG): Gesellschaft für Informatik e.Vin P. Fröhlich & V. Cobus. https://doi.org/10.18420/muc2023-ws16-120.

Angulo, Julio, Simone Fischer-Hübner, Erik Wästlund, and Tobias Pulls. 2012. "Towards Usable Privacy Policy Display and Management." Edited by Nathan Clarke and Steven Furnell. *Information Management & Computer Security* 20 (1): 4–17. https://doi.org/10.1108/09685221211219155.

Bentley, Frank, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottrjdge. 2018. "Understanding the Long-Term Use of Smart Speaker Assistants." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (3): 1–24. https://doi.org/10.1145/3264901.

Bohn, Dieter. 2016. "Google Home: A Speaker to Finally Take on the Amazon Echo." *The Verge* (blog). May 18, 2016. https://www.theverge.com/2016/5/18/11688376/google-home-speaker-announced-virtual-assistant-io-2016.

Brüggemeier, Birgit, M. Breiter, Miriam Kurz, and Johanna Schiwy. 2020. "User Experience of Alexa, Siri and Google Assistant When Controlling Music – Comparison of Four Questionnaires." In *HCI International 2020 – Late Breaking Papers: User Experience Design and Case*, 600–18. Online. https://doi.org/10.1007/978-3-030-60114-0_40.

Calzada Prado, Javier, and Miguel Ángel Marzal. 2013. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents." *Libri* 63 (2): 123–34. https://doi.org/10.1515/libri-2013-0010.

Castelli, Nico, Corinna Ogonowski, Timo Jakobi, Martin Stein, Gunnar Stevens, and Volker Wulf. 2017. "What Happened in My Home?: An End-User Development Approach for Smart Home Data Visualization." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 853–66. Denver: ACM. https://doi.org/10.1145/3025453.3025485.

Castelli, Nico, Gunnar Stevens, and Timo Jakobi. 2019. "Information Visualization at Home: A Literature Survey of Consumption Feedback Design." *Literature Survey* 16 (16). Bonn: International Institute for Socio-Informatics. https://pub.h-brs.de/frontdoor/index/index/docId/4505.

Cena, Federica, Silvia Likavec, Amon Rapp, and Alessandro Marcengo. 2016. "An Ontology for Quantified Self: Capturing the Concepts behind the Numbers." In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 602–04. Heidelberg: ACM. https://doi.org/10.1145/2968219.2968329.

Chapman, Chris, and Elea Nc Donald Feit. 2019. *R For Marketing Research and Analytics*. Use R! online (e-Book): Springer Cham. https://link.springer.com/book/10.1007/978-3-030-14316-9.

Cranor, Lorrie F. 2008. "A Framework for Reasoning About the Human in the Loop." In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, 1–15. USA: USENIX Association.

Draxler, Sebastian, Gunnar Stevens, Martin Stein, Alexander Boden, and David Randall. 2012. "Supporting the Social Context of Technology Appropriation: On a Synthesis of Sharing Tools and Tool Knowledge." In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems – CHI '12*, 2835. Austin: ACM Press. https://doi.org/10.1145/2207676.2208687.

Eriksson, Mats, and Seija Kulkki. 2005. "State-of-the-Art in Utilizing Living Labs Approach to User-Centric ICT Innovation – A European Approach." *State-of-the-Art in Utilizing Living Labs Approach to User-Centric ICT Innovation* 15 (January).

European Parliament and the Council. 2018. *General Data Protection Regulation (GDPR)*. https://gdpr-info.eu/.

Feldman, Ronen, and James Sanger. 2006. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511546914.

Fischer, Joel E., Andy Crabtree, Tom Rodden, James A. Colley, Enrico Costanza, Michael O. Jewell, and Sarvapali D. Ramchurn. 2016. "'Just Whack It on until It Gets Hot': Working with IoT Data in the Home." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5933–44. San Jose: ACM. https://doi.org/10.1145/2858036.2858518.

Ford, Marcia, and William Palmer. 2019. "Alexa, Are You Listening to Me? An Analysis of Alexa Voice Service Network Traffic." *Personal and Ubiquitous Computing* 23 (1): 67–79. https://doi.org/10.1007/s00779-018-1174-x.

Geeng, Christine, and Franziska Roesner. 2019. "Who's In Control?: Interactions In Multi-User Smart Homes." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems – CHI '19*, 1–13. Glasgow: ACM Press. https://doi.org/10.1145/3290605.3300498.

Gómez Ortega, Alejandra, Jacky Bourgeois, and Gerd Kortuem. 2023. "What Is Sensitive About (Sensitive) Data? Characterizing Sensitivity and Intimacy with Google Assistant Users." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16. Hamburg: ACM. https://doi.org/10.1145/3544548.3581164.

Habscheid, Stephan, Tim Moritz Hector, Christine Hrncal, and David Waldecker. 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. https://doi.org/10.21248/jfml.2021.44.

Häußling, Roger. 2017. "Relationale Technik- und Organisationssoziologie." *Soziologie: Forum der Deutschen Gesellschaft für Soziologie*, Symposium: Was ist Netzwerkforschung, 61 (1): 47–50.

Huang, Yue, Borke Obada-Obieh, and Konstantin (Kosta) Beznosov. 2020. "Amazon vs. My Brother: How Users of Shared Smart Speakers Perceive and Cope with Privacy Risks." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. Honolulu: ACM. https://doi.org/10.1145/3313831.3376529.

Igual Muñoz, Laura, and Santi Seguí Mesquida. 2017. *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications*. Undergraduate Topics in Computer Science. Cham: Springer.

Jakobi, Timo, Edna Kropp, Gunnar Stevens, and Mats Schmal. 2017. "Providing Smartphone Data Visualizations to Support Privacy Literacy." In *EuroUSEC 2017 Workshop*, 2. Paris: Online. https://www.ndss-symposium.org/eurousec-2017/.

Jakobi, Timo, Gunnar Stevens, Nico Castelli, Corinna Ogonowski, Florian Schaub, Nils Vindice, Dave Randall, Peter Tolmie, and Volker Wulf. 2018. "Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (4): 1–28. https://doi.org/10.1145/3287049.

Jakobi, Timo, Gunnar Stevens, Maximilian von Grafenstein, Dominik Pins, and Alexander Boden. 2020. "User-Friendly Formulation of Data Processing Purposes of Voice Assistants: A User Perspective on the Principle of Purpose Limitation." In *Proceedings of the Conference on Mensch Und Computer*, 361–72. Magdeburg: ACM. https://doi.org/10.1145/3404983.3405588.

Kerwin, Jason T. 2012. "'Rational Fatalism': Non-Monotonic Choices in Response to Risk." *Working Paper*, no. WGAPE Spring 2012, 70.

Kiesel, Johannes, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. 2019. "Clarifying False Memories in Voice-Based Search." In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 331–35. Glasgow: ACM. https://doi.org/10.1145/3295750.3298961.

Koltay, Tibor. 2015. "Data Literacy: In Search of a Name and Identity." *Journal of Documentation* 71 (2): 401–15. https://doi.org/10.1108/JD-02-2014-0026.

Langheinrich, Marc. 2002. "A Privacy Awareness System for Ubiquitous Computing Environments." In *UbiComp 2002: Ubiquitous Computing*, edited by Gaetano Borriello and Lars Erik Holmquist, 237–45. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-45809-3_19.

Laricchia, Federica. 2023. "Smart Speakers – Statistics & Facts." Dossier. Consumer Electronics. Statista. https://www.statista.com/topics/4748/smart-speakers/.

Lau, Josephine, Benjamin Zimmerman, and Florian Schaub. 2018. "Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers." *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW): 1–31. https://doi.org/10.1145/3274371.

Luger, Ewa, and Abigail Sellen. 2016. "'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems – CHI '16*, 5286–5297. Santa Clara: ACM Press. https://doi.org/10.1145/2858036.2858288.

Malkin, Nathan, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. "Privacy Attitudes of Smart Speaker Users." *Proceedings on Privacy Enhancing Technologies* 2019 (4): 250–71. https://doi.org/10.2478/popets-2019-0068.

Mandinach, Ellen Beth, and Edith S. Gummer. 2016. *Data Literacy for Educators: Making It Count in Teacher Preparation and Practice*. Technology, Education-Connections. New York: Teachers College Press.

Marr, Bernard. 2016. "How The Citizen Data Scientist Will Democratize Big Data." Forbes. April 2016. https://www.forbes.com/sites/bernardmarr/2016/04/01/how-the-citizen-data-scientist-will-democratize-big-data/.

Meng, Nicole, Dilara Keküllüoğlu, and Kami Vaniea. 2021. "Owning and Sharing: Privacy Perceptions of Smart Speaker Users." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1): 1–29. https://doi.org/10.1145/3449119.

Pins, Dominik, and Fatemeh Alizadeh. 2021. "Without Being Asked: Identifying Use-Cases for Explanations in Interaction with Voice Assistants." In *CHI Workshop-Paper*, 8. Yokohama: N/A. https://doi.org/10.13140/RG.2.2.18764.33923.

Pins, Dominik, Alexander Boden, Britta Essing, and Gunnar Stevens. 2020. "'Miss Understandable': a study on how users appropriate voice assistants and deal with misunderstandings." In *MuC '20: Proceedings of the Conference on Mensch und Computer*, 349–59. Magdeburg: Association for Computing Machinery. https://doi.org/10.1145/3404983.3405511.

Pins, Dominik, Timo Jakobi, Alexander Boden, Fatemeh Alizadeh, and Volker Wulf. 2021. "Alexa, We Need to Talk: A Data Literacy Approach on Voice Assistants." In *Designing Interactive Systems Conference 2021*, 495–507. New York: Association for Computing Machinery. https://doi.org/10.1145/3461778.3462001.

Pins, Dominik, Timo Jakobi, Gunnar Stevens, Fatemeh Alizadeh, and Jana Krüger. 2022. "Finding, Getting and Understanding: The User Journey for the GDPR'S Right to Access." *Behaviour & Information Technology* 41 (10): 2174–2200. https://doi.org/10.1080/0144929X.2022.2074894.

Plantin, Jean-Christophe, Carl Lagoze, Paul N Edwards, and Christian Sandvig. 2018. "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook." *New Media & Society* 20 (1): 293–10. https://doi.org/10.1177/1461444816661553.

Purington, Amanda, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. "Alexa Is My New BFF." In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems – CHI EA '17*, edited by Gloria Mark, Susan Fussell, Cliff Lampe, m.c schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor, 2853–59. New York: ACM Press. https://doi.org/10.1145/3027063.3053246.

Rohde, Markus, Peter Brödner, Gunnar Stevens, Matthias Betz, and Volker Wulf. 2017. "Grounded Design – a Praxeological IS Research Perspective." *Journal of Information Technology* 32 (2): 163–79. https://doi.org/10.1057/jit.2016.5.

Sadowski, Jathan. 2020. *Too Smart: How Digital Capitalism Is Extracting Data, Controlling Our Lives, and Taking Over the World*. The MIT Press. https://doi.org/10.7551/mitpress/12240.001.0001.

Schaub, Florian, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2018. "A Design Space for Effective Privacy Notices*." In *The Cambridge Handbook of Consumer Privacy*, edited by Evan Selinger, Jules Polonetsky, and Omer Tene, 365–93. Online: Cambridge University Press. https://doi.org/10.1017/9781316831960.021.

Schönherr, Lea, Maximilian Golla, Thorsten Eisenhofer, Jan Wiele, Dorothea Kolossa, and Thorsten Holz. 2020. "Unacceptable, Where Is My Privacy? Exploring Accidental Triggers of Smart Speakers." *arXiv:* 1 (1): 15. https://doi.org/10.48550/arXiv.2008.00508.

Sciuto, Alex, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "'Hey Alexa, What's Up?': A Mixed-Methods Studies of In-Home Conversational Agent Usage." In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 – DIS '18*, 857–68. Hong Kong: ACM Press. https://doi.org/10.1145/3196709.3196772.

Shafagh, Hossein, and Anwar Hithnawi. 2017. "Privacy-Preserving Quantified Self: Secure Sharing and Processing of Encrypted Small Data." In *Proceedings of the Workshop on Mobility in the Evolving Internet Architecture – MobiArch '17*, 25–30. Los Angeles: ACM Press. https://doi.org/10.1145/3097620.3097625.

Shin, Jungwoo, Yuri Park, and Daeho Lee. 2018. "Who Will Be Smart Home Users? An Analysis of Adoption and Diffusion of Smart Homes." *Technological Forecasting and Social Change* no. 134 (September): 246–53. https://doi.org/10.1016/j.techfore.2018.06.029.

Shneiderman, Ben. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, 336. USA: IEEE Computer Society.

Stevens, Gunnar, and Paul Bossauer. 2017. "Dealing with Personal Data in the Age of Big Data Economies." *Zeitschrift Für Geistiges Eigentum* 9 (3): 266. https://doi.org/10.1628/186723717X15069451170856.

Stevens, Gunnar, Paul Bossauer, Timo Jakobi, and Christina Pakusch. 2017. "Second Dashboard: Information Demands in a Connected Car." In *Tagungsband MuC 2017*, 215–26. Regensburg: Gesellschaft für Informatik e.V. https://doi.org/10.18420/muc2017-mci-0141.

Stevens, Gunnar, Volkmar Pipek, and Volker Wulf. 2010. "Appropriation Infrastructure: Mediating Appropriation and Production Work." *Journal of Organizational and End User Computing (JOEUC)* 22 (2): 58–81. https://doi.org/10.4018/joeuc.2010040104.

Strüver, Niklas. 2023a. "Frustration Free: How Alexa Orchestrates the Development of the Smart Home." *Digital Culture & Society* 9 (1): 99–124. https://doi.org/10.14361/dcs-2023-0106.

Strüver, Niklas 2023b. "Wieso eigentlich Alexa? Konzeptualisierung eines Sprachassistenten als Infrastruktur und Plattform im soziotechnischen Ökosystem Amazons." *kommunikation@gesellschaft* 24 (1). https://doi.org/10.15460/kommges.2023.24.1.1194.

Sun, Kaiwen, Yixin Zou, Jenny Radesky, Christopher Brooks, and Florian Schaub. 2021. "Child Safety in the Smart Home: Parents' Perceptions, Needs, and Mitigation Strategies." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–41. https://doi.org/10.1145/3479858.

Thakkar, Parth Kirankumar, Shijing He, Shiyu Xu, Danny Yuxing Huang, and Yaxing Yao. 2022. "'It Would Probably Turn into a Social Faux-Pas': Users' and Bystanders' Preferences of Privacy Awareness Mechanisms in Smart Homes." In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–13. New York: Association for Computing Machinery. https://doi.org/10.1145/3491102.3502137.

Tolmie, Peter, Andy Crabtree, Tom Rodden, James A Colley, and Ewa A Luger. 2016. "'This Has to Be the Cats' – Personal Data Legibility in Networked Sensing Systems." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing – CSCW '16*, 490–501. San Francisco: ACM Press. https://doi.org/10.1145/2818048.2819992.

Waldecker, David, Tim Hector, and Dagmar Hoffmann. 2023. "Intelligent Personal Assistants in Practice. Situational Agencies and the Multiple Forms of Cooperation without Consensus." *Convergence: The International Journal of Research into New Media Technologies* 0 (0): 1–17. https://doi.org/10.1177/13548565231189584.

Ware, Colin. 2013. *Information Visualization: Perception for Design*. Third edition. Interactive Technologies. Waltham: Morgan Kaufmann.

Whitten, Alma, and J D Tygar. 1999. "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0." In *Proceedings of the 8th USENIX Security Symposium*, 169–84. Washington, D.C.: USENIX Association.

Wulf, Volker, 2018. *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*. First edition. Oxford: Oxford University Press.

Wulf, Volker, Claudia Müller, Volkmar Pipek, David Randall, Markus Rohde, and Gunnar Stevens. 2015. "Practice-Based Computing: Empirically Grounded Conceptualizations Derived from Design Case Studies." In *Designing Socially Embedded Technologies in the Real-World*, edited by Volker Wulf, Kjeld Schmidt, and David Randall, 111–50. Computer Supported Cooperative Work. London: Springer. https://doi.org/10.1007/978-1-4471-6720-4_7.

Wulf, Volker, Markus Rohde, Volkmar Pipek, and Gunnar Stevens. 2011. "Engaging with Practices: Design Case Studies as a Research Framework in CSCW." In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 505–512. New York: Association for Computing Machinery. https://doi.org/10.1145/1958824.1958902.

Xie, Wenjing, Amy Fowler-Dawson, and Anita Tvauri. 2019. “Revealing the Relationship between Rational Fatalism and the Online Privacy Paradox.” *Behaviour & Information Technology* 38 (7): 742–59. https://doi.org/10.1080/0144929X.2018.1552717.

Zhang, Dongpo. 2018. “Big Data Security and Privacy Protection.” In *Proceedings of the 8th International Conference on Management and Computer Science (ICMCS 2018)*, 4. Shenyang: Atlantis Press. https://doi.org/10.2991/icmcs-18.2018.56.

Innovating Alexa amid the Rise of Large Language Models

Sociotechnical Transitions in Algorithmic Development Practices

Niklas Strüver

Abstract *For about a decade, Amazon's Alexa was a pioneer in automatic speech processing; now, however, new Large Language Models (LLMs) are posing challenges for Amazon. One attempt to confront these challenges is by integrating technologies developed for Alexa by university research teams in the Alexa Prize Competitions (APCs). This chapter examines how participants in these contests deal with the conditions set and the resources provided by Amazon for the competition, and offers a snapshot of the practical development processes of the voice assistant at a time of technological transition. It then outlines some of the path dependencies, risks, benefits, and aspects of structuration that are encountered by the participants in their attempts to innovate Alexa.*

1. Introduction

Over the course of the last decade, Amazon has spent a considerable amount of effort making Alexa reliable enough to be desirable for many households[1]. In the last couple of years however, Amazon had been reducing its generosity to its Alexa division (Kim 2022) – that is, until the competing machine learning company OpenAI introduced large language models (LLMs) to the public, most no-

1 Technically speaking, Alexa is the voice interface for Amazon's cloud products Alexa Voice Service (AVS) and Amazon Web Services (AWS), where all requests are processed by various machine learning algorithms (Crawford and Joler 2018), which are constantly optimized based on the incoming usage data. This service is embedded in the Echo devices produced by Amazon.

toriously with their use cases in the form of ChatGPT in November 2022. As the world was familiarizing itself with a proclaimed revolution of artificial intelligence (AI) technologies, technology companies like Amazon found themselves with an apparent need to catch up. Upgrades announced for Alexa (Bensinger 2024; Jassy 2024; Krishnan 2024) indicate that Amazon is working on ways of integrating LLMs into its voice assistant, which until now had primarily relied on more traditional machine learning approaches. This change in coding approaches for Alexa comes with a set of difficulties that need to be navigated in a competitive field of technology development (Kinsella 2023).

To better understand the transition between two different approaches to making Alexa talk to users, and to gain insights into Amazon's development practices for Alexa, a qualitative expert interview study was conducted to investigate how development is practiced in the Alexa ecosystem. As it is difficult to conduct research within Amazon directly, the Alexa Prize Competitions (APCs), in which university research teams compete to build technologies for Alexa, were chosen as a proxy study context that could offer insights into the technological development of Alexa, as well as into Amazon's approach to cooperating with third parties (universities in this case) that wish to interact with Alexa as a platform. By exploring the views of third-party actors who obtain access to Alexa technologies and are closely supervised by Amazon Alexa staff, the study seeks to contribute to research on the sociotechnical analysis of Amazon's technology for Alexa; ultimately to further understanding of the sociotechnical underpinnings of a technology that is present in many homes globally. To achieve these aims, the questionnaire used in the study was developed to elicit details about the inner workings of cooperation with Amazon, making the APC teams a proxy of analysis for Amazon's Alexa team.

On a theoretical level, this study explores the idea of structuration of platform organizations (Dolata and Schrape 2023) and investigates the practices of infrastructuration (Edwards 2019) that the APC teams developed over the course of the competition. These theoretical tools are employed to analyze the perspectives of highly skilled developers who gain access to Amazon's Alexa technology by agreeing to develop solutions to certain problems set by Amazon. It can be shown how the developers navigate the conditions set by Amazon, as well as how certain technological path dependencies clash with new AI innovations taking place outside Amazon. As this transition in coding tradition is largely (at least in the public eye) initiated by the release of ChatGPT, the overarching interest in this article is to inquire into the APC participants' (shifting) perspectives on Alexa during this period of transition towards LLMs, and to shed light on their development and innovation practices in this matter. Although the APC participants may not be employed by Amazon, they did receive insights into the corporation's development material, tools, and guiding principles for Alexa, informing them of the current state of the art of Alexa. Ultimately this gives insights into Alexa's sociotechnical underpinnings in a world that was at the time seemingly being revolutionized by a competing technology, and how Amazon and the APC participants attempted to merge existing with new technology, while at the same time navigating their relations of cooperation with each other in an ongoing process of platformization and infrastructuration (Plantin et al. 2018).

2. Research Object and State of Research

To introduce the object at hand, a brief outline of the APC and AI competitions in general is followed by a short summary of research on voice assistants (2.1). To further situate the research interest of this paper, a short overview of research on generative AI is then provided (2.2).

2.1 Studying the Alexa Prize Competitions

Many technology companies hold prize competitions and challenges like Amazon's APCs as a way of outsourcing algorithm development work. The cultural impact of these contests has been analyzed and the balancing of platform interests with complex engineering problems has been discussed at length in the case of the Netflix Prize (Hallinan amd Striphas 2016; Seaver 2022, 56–58). As such, the competition concept has served as the *organizing principle for AI* (Hind et al. 2024). Further, the events have been contextualized within the culture of competitiveness that is underlined by the practice of benchmarking (Orr and Kang 2024), as well as a platformized process that favors a few powerful actors (Luitse et al. 2024). The APCs have not yet received specific academic attention beyond the annual competition proceedings that focus on the computer science aspect (see e.g. Agichtein et al. 2023; Johnston et al. 2023; Shi et al. 2023). The APCs are a series of annual competitions that have been organized by Amazon since 2017, starting with the first Socialbot Grand Challenge (SBC). In that competition, Amazon encouraged universities across the world to create teams comprised of PhD students and professors to compete in a contest to develop a conversational bot that would drive Amazon's voice assistant Alexa (Amazon

2024). The challenge of the first SBC was to create bots capable of holding a 20-minute conversation with users talking to the bot (via Alexa) about various topics. In 2022 Amazon added the Task Bot Challenge (TBC) and in 2023 the SimBot Challenge (SIMBC) to its annual competitions. In the former, participants were invited to design bots that could enable Alexa to assist users in complex tasks such as cooking or origami, guiding users verbally through the various steps of a respective task. The latter challenge involved users talking to Alexa to control a robot in a video game environment to achieve small tasks (like retrieving something from a fridge) in said video game that simulates a living space. All of these competitions have a similar structure in time and incentive, running between eight and 18 months and divided into phases for certification (technical requirements of the bots that need to be fulfilled), internal feedback (Amazon employees provide intensive feedback on the bot), and public feedback (the systems go live and users can use the bots). During the last phase of the competition, the prototype bots are available to Alexa users in the United States. It is important to point out that this happens through a dedicated application, clearly separating the competition from the regular Alexa service. When a user invokes the corresponding skill for the competition they are randomly assigned one of the competitors' bots, without knowing which one it is – there is no way for them to target specific bots. After an interaction, users have the option to evaluate the bot with a star rating from one to five and a sentence of feedback. These ratings are used to rank the university teams on a leadership board that is updated daily, determining who advances to the final stage (which is a continuation of the previous stage but with less competitors and more users) and eventually determining the placement of the winning teams and the allocation of the prize money.

Studying this competition contributes to the body of research that undertakes sociotechnical analysis of voice assistants like Alexa, furthering understanding of the sociotechnical underpinnings of a technology that is present in many homes globally. Voice assistants have already been studied from multiple perspectives (Minder et al. 2023)[2]. Some research has addressed the plat-

---

2 It is important to note two prominent strands of critical inquiry into voice assistants, even though they are beyond the scope of this article. Firstly, there is the issue of the gender roles that voice assistants represent and perpetuate and in what ways this can be problematic; for a comprehensive overview see, e.g., Kennedy and Strengers (2020). Secondly, privacy and data security have received a great deal of attention because the devices can give companies access to data, e.g., from conversations, that

formized nature of voice assistants (e.g. Goulden 2019; Pridmore et al. 2019; Sadowski et al. 2021), but few studies to date have focused on the development process of voice assistants (Strüver 2023a; b). By qualitatively inquiring into the procedures of the APCs and competitors' experiences of working with Alexa technologies, it becomes possible to shed light on the inner workings of the sociotechnical relationships and dependencies that underlie Alexa. This is particularly interesting at a time in which speech technologies are prominent in public perception and critical discussion.

2.2 Large Language Models as a problem for Alexa

For a long time, the development of voice technologies was driven by turning linguistical conversation rules into code that determines how artificial voice agents detect users' intents and then give appropriate answers. This "rule-bound rationality of code-driven determination that animated the formative decades of AI research" (Li 2023, 168) was later enhanced by heuristic programming, which enabled more flexibility and improved performance. While stochastic machine learning models that approximate the most likely meaning of and answers to users' queries are commonly used in modern voice technologies (ibid.), for a long time, voice assistants like Alexa have retained some form of determinable answers and heuristics to ensure that certain actions follow certain queries (Kinsella 2023). This has often obliged developers to compile large sets of manually created answers (and templates) that were heuristically matched to what users approached the assistant with. The increased use of LLMs – achieved by the marketization and popularization of various tools and their integration into well-known and widely used applications – now seems to be set to strongly influence how voice assistants will be further developed in the future. Generative AI models like LLMs are a technological development that has recently risen in popularity in many applications for everyday use, with claims that the technology is revolutionizing the field of AI – in the familiar narrative of heralding *the next big thing* (Vannuccini and Prytkova 2024). As they have gained prominence and popularity, LLMs have been critically scrutinized from multiple perspectives (Fourcade and Healy 2024, 94). Essentially, they operate by a form of machine learning that utilizes vast amounts of data and computational power to perform various tasks that

they never had access to before; making security and trust controversial topics (see e.g. Mols et al. 2021; Ochs, this volume; Waldecker et al., this volume).

were previously complicated to execute with algorithmic tools. The humanities and social sciences have highlighted issues of diversity and discrimination in LLMs (Gillespie 2024), have questioned the agency of LLMs (Floridi 2023), and have contextualized the socio-political dimensions of LLMs on a global scale (Amoore et al. 2024). Further, scholars have criticized how much resources the training and maintenance of these models consume (Rillig et al. 2023) due to the enormous computing power they require. On an infrastructural level, this high consumption means that only a few firms can realistically afford to train these types of models, which has led to a significant oligopoly comprising the three largest Western corporations: Amazon, Microsoft, and Google (Srnicek 2022, van der Vlist et al. 2024). The significant rush in development that was precipitated by OpenAI's launch of ChatGPT has created an environment of hectic innovation. Like other companies, Amazon has sought to adapt products such as Alexa to the new LLM technology (Krishnan 2024), despite having previously reduced its development investment for Alexa due to poor business figures (Kim 2022). This has seen Alexa's development essentially reinvigorated by LLMs, which represent a new avenue for innovation that was previously underexplored for Alexa. Amid this global frenzy, as Tekic and Füller observe, universities are a key collaboration target for companies that wish to expand their access to the development of LLM technologies, as universities "are rare places where AI researchers – an expensive and hard-to-find resource – are grown" (2023, 5). This, and the fact that Alexa has traditionally been built with a heavy reliance on manually-coded heuristics only occasionally enhanced with LLMs (Jassy 2024), lead to the these main questions that motivate this paper:

The overarching purpose of the analysis is to elucidate *APC participants' perspectives on Alexa's position in the ongoing technological transition towards LLMs*, thereby also shedding light on Amazon's attempts to incentivize innovation in that direction. To contextualize those perspectives, the integration of LLMs into Alexa is examined against the backdrop of potential path dependencies in Alexa (5.1). Furthermore, the participants' technology development practices are focused upon in order to study the implementation of LLMs into the Alexa system from a *science and technology studies* perspective (5.2). Finally, sufficient context will have been provided for some conclusions to be drawn regarding the ongoing market competition between Alexa and ChatGPT and the role of the APCs therein (5.3).

3. Theorizing the Vortex between Platforms and their Complementors

In order to investigate the research interest, there will be a theoretical introduction into aspects of platform structuration. This begins by focusing on the platform organization's structuring capacity (3.1), which is then contrasted with the infrastructuration practices of developers (3.2).

3.1 Alexa as a Platform in the Alexa Prize

Sociological perspectives often focus on the companies behind the platforms and their power relations (e.g., Dolata 2019). Building on a combination of these perspectives, Strüver has conceptualized the voice assistant Alexa as a platform with multiple roles and purposes situated within Amazon's platform-ecosystem (2023b). He draws attention to the "unifying role for the smart home", that Alexa seems to hold, where it acts as a "connecting point for many different actors and technologies" (Strüver 2023a, 105) and the position of power in which this puts Amazon in relation to homes and businesses. These observations are guided by the idea that platforms and their complementors (Baldwin and Woodard 2008) can be conceptualized in a center–periphery model, with the platform as the locus of action governed by an organizational core that decides how the actors (e.g., users or third parties) interact with the platform through interface design (Ametowobla and Kirchner 2023). In this sense, it is important to understand the platform in a threefold distinction:

> (1) the platform-operating companies as organizing and structuring cores whose goal is to operate a profitable business; (2) the platforms belonging to them as more or less extensive, strongly technically mediated social action spaces not only for economic but also for genuine social activities; and (3) the institutionalized coordination, control and exploitation mechanisms implemented by the platform operators, linking these two constitutive levels of the platform architecture. (Dolata and Schrape 2023, 4)

This threefold distinction requires some tweaking when applied to the APCs, however, since in this case it is in Amazon's interest to continue to innovate their technology in order to run a profitable business[3] by enabling and situating Alexa as a platform for innovation not only in the context of the competition but also for internal purposes. Applying the three distinctions to the APC, Amazon appears as a coordinating platform company that develops the platform Alexa and the sociotechnical environment of the competition. Acting as a space for a variety of social actions, Alexa becomes the platformized social environment for the APC, in which university teams develop new features, which are put to the test on users' Alexa devices. However, this social space within the platform environment subjects development activities to the constraints of coordination and control of the competition imposed by Amazon – which harkens back to the idea of periphery and center (Ametowobla and Kirchner 2023). In this sense, platforms coordinate not only economic processes, but also various social relationships, which can include the complementing innovation practices of independent third-party developers (Tiwana 2014, 118). The tools available to Amazon to control the platform environment are forms of "[c]oordination and rule-setting, monitoring and exploitation of data, coupled with the ability of the platform companies to quickly, substantially and largely uncontrollably adapt the social and technical rules they establish" (Dolata and Schrape 2023, 8), which locates the origin of power asymmetries between platform companies and the various groups of actors involved in the act of *platform governance* (Gorwa 2019). By means of the Alexa platform, Amazon has control over the technical development and standardization of third-party Alexa products, decides on the possible interactions with and within the platform, and, finally, sets the (contractual) rules, goals, and boundaries of collaboration between third parties and Amazon (e.g., van Dijck et al. 2018, 11; Gillespie 2018, 45–47). These rules, goals, and limits establish and maintain the *hierarchical orientation* (Dolata and Schrape 2023, 8). On top of those there are softer forms of control and orchestration which can act as action-orienting influences that are optional and malleable. These softer forms of control come as resources granted to the teams by Amazon prior to the competition (Agichtein et al. 2023, 3–13; Johnston et al. 2023, 4–12; Shi et al. 2023, 4–8). Exemplary, a Conversational Bot (CoBot) toolkit was offered, which represented a development tool for conversational AI with numerous pre-configured design presets for natural

3 While Alexa is reportedly not profitable for Amazon (Kim 2022), it can be argued that Alexa serves a greater purpose through cross subsidization, data usage, and algorithm development (Strüver 2023b, 21–25).

language understanding and dialogue management[4]. Amazon updates CoBot annually based on the learnings of the previous competition and to reflect ongoing changes in the industry, such as the recent shift to LLMs: "In addition, we also made significant changes in CoBot to support hosting large language models (LLMs), as much as 640 GB, which is 160 times larger than previously hosted in CoBot" (Johnston et al. 2023, 4). The Amazon scientists' highlighting of this latest adaptation of the CoBot tool alludes to the fact that platform companies have the ability to *re-code* their platforms dynamically to adapt to internal and external influences like regulations, new internal Amazon products, or a new competitor like OpenAI's ChatGPT. This *transformative re-coding capacity* enables platforms to dynamically readjust the sociotechnical structuring and institutionalizing elements of their platforms (Frenken and Fuenfschilling 2020, 103–107). Besides contractual changes, this capacity manifests in forms of *orchestration efforts*, i.e., new development tools, programs, application programming interfaces (APIs), microchips, standards, guidelines, or infrastructures of development (van der Vlist 2022; Strüver 2023a); as can be seen with the CoBot tool that was adapted during the release of ChatGPT, altering the competition: "Large language models (LLMs) have played a significant role in the SocialBot Grand Challenge since early in the challenge, but nothing compared to their front stage role" (Johnston et al. 2023, 3) in SBC5. Fittingly, this incentive to integrate more LLMs is transported via the main support tool of the competition, tying back to the goal to advance the science in conversational AI (Amazon 2022b), as well as to please customers, who are experiencing ChatGPT while rating Alexa skills.

Drawing on the distinction between platform company, platform, and the mechanisms of controlling interaction on the platform reveals the sociotechnical elements that allow Amazon to regulate what happens in the APC, which in turn facilitates conjectures to be made about corporate motives for these measures and an attempt to reveal the "high degree of structuregiving, rule-setting and controlling power" (Dolata and Schrape 2023, 14) that companies like Amazon possess. By giving this context on the power that is wielded by big tech

4 CoBot is a typical example of big tech companies leveraging their R&D facilities to develop products that are supposed to reduce innovation costs (Dolata 2019, 189), which eventually influence the development process when incorporated (Strüver 2023a, 114). CoBot "provides abstractions that enable the teams to focus more on scientific advances and reduce time invested into infrastructure, hosting, and scaling." (Johnston et al. 2023, 3)

companies when they structure their platforms, an important analytical step is enabled. Usually, the workings of such companies are largely *opaque* (Burrel 2016), especially concerning their AI technologies, which makes it difficult to investigate the impacts of platform technologies on users and third parties. By examining the resources that Amazon uses to run the APC challenges, it becomes possible to draw conclusions regarding the ways they act within their B2B collaborations, as well as how they develop technologies internally. Against the backdrop of the boom of LLM-driven technologies – which occurred while Alexa was struggling as a product (Kim 2022) – this approach can reveal how Amazon attempted to create an environment in which ideas could be developed for Alexa in a world of abundant LLMs. But to look into this practice of developing technology, a practice perspective on structuration is necessary, as structuration is not a deterministic effort made by Amazon that cannot encounter contingent resistance. Here, the tools of soft control are especially interesting, as they allow for leeway at the level of practice. In analyzing how tools of orchestration impact the APC, the room for negotiation and the limitations of resources of power which attempt to influence the course of action get revealed (Dolata 2024, 191) under the magnifying glass of practice that eventually reproduces or alters structure (Giddens 1984, 15–28). This shift of perspective allows the accounts of the participants to be read through the lens of the *mangle of practice* of developing Alexa at a time when LLMs were seemingly revolutionizing conversational technology development.

3.2 Platform practices as infrastructuration

As Plantin et al. (2018) argue, platforms can be infrastructuralized when infrastructures are platformized. This has also been shown to apply to voice assistants when users incorporate them into their daily lives as an infrastructure (Strüver 2023b). Infrastructures can be viewed as sociotechnical systems made up of a mixture of routines, artifacts, standards, plans, conventions, technological devices, or organizational institutions (Star and Ruhleder 1996, 113). These infrastructures can become central to everyday life when they are embedded in practices and subtend social, technological, and built worlds, as they do not need to be reconsidered in the moment of invoking them to perform a task (Slota and Bowker 2017, 537). This is true for users who rely on infrastructures, but not for the communities involved in the social, political, and economic work of building, maintaining, and upgrading infrastructures (Bowker and Star 2000, 109). All groups, however, learn to interact with infrastructures and their conventions of practice as part of membership in their given communities (Star and Ruhleder 1996, 113). In this respect, they adopt behavioral regularities that become (organizational) routines, which then come to be part of the functioning of infrastructure. Drawing on Giddens' (1984) structuration theory, Edwards describes this process of embedding infrastructural skills in humans' habits and skills as infrastructuration: "infrastructure both shapes and relies upon the continual performances or rehearsals of agents" (2019, 358). When users or engineers acquire the habits and skills to interact with an infrastructure as part of membership, they start playing a vital role in its functioning, thereby reproducing the structural elements. Giddens specifically remarks on actors' capacity for agency to make contingent decisions to be bounded by their perception (1984, 27), rendering these learned habits as a way "of black-boxing action patterns that may once have been deliberately chosen or designed" (Edwards 2019, 359), by providing infrastructuralized action scripts "on which users, maintainers and builders can all tacitly rely" (ibid.). In that sense, infrastructural practices become an embodiment of standards (Slota and Bowker 2017, 537) as they reproduce the (infra-)structures that enable them. When infrastructures are embedded in large sociotechnical systems, most decisions that govern the functioning of the system have been made without the active participation of either users or engineers. However, by adopting norms, routines, and habits and reproducing them in daily practice, these black-boxed standards can become invisible in practice without anyone's need to reflect on their origin, or on the choices that may have led to a particular design. This infrastructuralization of platforms and their logics defines how practices become entrenched in the structures of the platforms that enable them:

> once they [practices] become habitual and routine, these once-cognitive acts become quasi-mechanical. Most of the time, that is a virtue; they contribute to the smooth workings of infrastructure while remaining invisible themselves. Yet by burying choices and creating path dependencies, they can also have negative consequences, sometimes dramatically so. (Edwards 2019, 361)

This draws back to the structuring aspects of said infrastructure, since a well-established infrastructure can lead to path dependencies and sociotechnical lock-in effects due to large user bases that expect a certain functionality or an engineering team that is used to a familiar direction of development. With such structural inertia, it is uncertain how many collective resources have to be leveraged to change institutionalized structures.

These sociotechnical path dependencies can lead to resistance to change, even in seemingly fluid electronic infrastructures (Star 1999, 389) such as platforms (Strüver 2023b, 24). Habitual and materialized infrastructures are manifested in the form, for example, of certain functions, algorithms, or company goals that have shaped Alexa since its conception and have become familiar to users and developers alike. They may have contributed to a reduction of contingency and made certain development paths more likely than others in structuring the platform Alexa. However, faced with the facts that, on the one hand, Alexa does not seem to be succeeding economically for the company Amazon (Kim 2022), and on the other hand, that competitors seem to revolutionize the fields of Alexa's core technologies, the corporation has incentives to question the viability of some structures that have guided Alexa for years, and to explore new ways of developing Alexa (Jassy 2024; Krishnan 2024). To investigate Amazon's responses to this situation, the idea of infrastructuration can be used to trace how competition participants developed common practices of development during the course of the contest and how they handled the integration of LLMs into their bots while negotiating the existing Alexa infrastructure, its limitations, and Amazon's elements of structuration. This turn towards the routines, forms of resistance, and power resources in practice and practical work can highlight how the new complex technologies being developed for Alexa were still embedded in a social system and an accomplishment of data practices, which "does not just happen on its own, but is manifested through everyday interactions between people, infrastructures, and established conventions" (Burkhardt et al. 2022, 11).

4. Study Design and Material

Studying the big tech companies of Silicon Valley from within is nigh impossible – at least if the study is to conform with the methodological standards and guidelines of sociology. The firms' inaccessibility is one of the reasons for choosing to investigate the APC, as it allows an insight into the inner workings of Amazon's Alexa team – or at least to the parts of it that competitors interact with. The other reason is that Amazon relies heavily on third parties for their core businesses (e.g., Khan 2018; Rowberry 2022, 42–43; Weigel 2023), so studying these can reveal how one of the world's biggest technology companies conducts and manages its power relations. To inquire into the inner workings of Alexa and one part of its third-party ecosystem, a qualitative expert interview study was conducted with participants in the APC. 158 competitors from 2022 and 2023 were invited by email to take part in the study and offered a 25USD/EUR incentive to signify sincerity. This led to twelve one-hour interviews being conducted in early 2024. Nine interviewees were based in the USA, from diverse demographic groups within the population (Starr and Freeland 2023); the other three were in Europe. Overall, participants came from ten different university teams that had taken part in three different competitions. Seven were PhD students, two MSc students, and three professors in faculty and team-leading positions. Final placement in the competition of the teams whose members agreed to participate in the interviews was not skewed in any particular direction. Mirroring the uneven gender representation in the field of computer sciences, there were only two women in the sample of interviewees. An attempt to counter this was not successful, and the imbalance in the field was discussed in some interviews. Online video and voice interviews were chosen as a means of communication due to the global scheduling advantages (Self 2021).

The study was carried out with good intentions and the most academic rigor, but was nonetheless subject to some limitations. First and foremost, the interviews were conducted at the start of 2024 with participants who had competed in the 2022/23 APCs, which ended in August 2023. Considering the extremely fast pace at which LLMs are developing, technical judgements and statements made at the time of the interviews, as well as evaluations of Alexa at the time of the competition, may very well be outdated by now. Nonetheless, some intricacies of the transition between technologies can still be gleaned from this analysis. The guiding questions (Helfferich 2019, 676–677) for the study were designed to elicit details about the inner workings of cooperation with Amazon and to produce narratives by the interviewees reliving their course through the competition as they experienced it. In this sense, the interviews were equal parts qualitative narrative interview (ibid.) and expert interview (Bogner et al. 2014). The narrative component of the interviews aimed to evoke a more personal conversation tracing the participants' experiences, to complement expert knowledge, conducive to evoking statements about the competition that exceed a factual retelling. Participants had signed non-disclosure agreements with Amazon in the course of the competition. However, the chosen methodology seemed to alleviate interviewees' fears of breaking the terms of those contracts, as the conversations were generally fluent and free in their flow. With participants' signed consent to the storage and usage of their data for scientific purposes, the interviews were locally recorded, transcribed, and anonymized; identifying statements were removed. Interviewees were assigned pseudonyms using a global random name generator (Bogner et al. 2014, 89–90). Analysis was carried out following the procedure of an inductive thematic qualitative data analysis (Kuckartz 2014, 70). In the following, interviewees' quotes are referenced by pseudonyms and the paragraph numbers of statements (Pseudonym, Paragraph number). All interviewees are referred to by the neutral pronoun "they" for inclusivity, and to protect their identities. The data sharing agreement signed by the participants does not allow the full transcripts to be made accessible to the public due to the sensitivity of the material.

5. Analysis: Perspectives on Building AI for Alexa

In order to address the overarching research interest – the APC participants' perspectives on Alexa's position in the ongoing technological transition towards LLMs – three topics are discussed in the following. First, the analysis focuses on the benefits, problems, and risks that come with integrating LLMs (5.1), then it compares two modes of actually integrating LLMs into Alexa (5.2). Lastly, an insight is offered into the role of the APC in developing LLMs in a competitive market (5.3).

5.1 Navigating the implementation of LLMs into Alexa

When investigating how integrating LLMs into the inner workings of Alexa relates to the conditions and structures that Amazon has set for Alexa, a great deal can be gleaned by addressing the benefits and problems perceived by the competitors of the APCs. A large portion of dialogues with Alexa are – or were at the time – determined by a heuristic that chooses from archetypes of manually-coded answers. This works well for easy-to-determine services like asking about the weather, turning on the living room lights, or asking trivia questions. Especially for more sensitive conversation topics, such as health advice, there are entirely preprogrammed responses that have been coded manually by engineers at Amazon, but this cannot feasibly be done for all the potential topics users might approach Alexa with. It can be assumed that when users talk to

Alexa, they do not want to constantly hear 'non-answers' that reveal the assistant's incapacity to engage in a given topic.

When competition participants as developers were preparing answers for the question of *what their bot's favorite sport was*, they might have included a list of dialogue options for popular sports, but probably did not consider every existing type: "We didn't cover everything. For example, for the other part [other sports], we could use the LLM" (Dart, 108). The flexibility of topics that can be handled by an LLM was one of their main perceived advantages, and was highlighted multiple times. Talking about sports is relatively simple, but "if it's something more involved, like: 'Oh, what are your opinions on Taylor Swift?', then the heuristic gets confused and there's no branch that matches it" (Scott, 42). While this comment addresses the same issue – that a heuristic model is unable to cover vast amounts of content – Scott's example concerns Alexa being asked about its opinion in a conversation. The implementation of LLMs could shift the structure of the conversation from a bot asking questions to users to instigate a dialogue and then posing follow-up questions, to a more flexible and reciprocal conversation model (Bardiola, 8; Centis, 29; Dart, 109). While the developers mentioned other advantages of LLMs, such as easier classification of users' responses via LLMs (Longwei, 87), or pre-trained models that can respond to sensitive topics (Gardé, 70), their flexibility was a recurrent theme mentioned throughout the interviews. It was particularly highly appreciated by competitors in the social bot challenge, who emphasized that LLMs can generate answers for any question, regardless of content. This reflects the structuring elements of the competition set by Amazon. The goal specified for the SBC: to achieve a 20-minute coherent and engaging conversation in two thirds of their bot's conversations (Amazon 2022b), clearly incentivized the implementation of a technology that enables flexible conversation. Further, Amazon provided various pre-trained models to facilitate this specific goal of "chitchat" (Centis, 29–32), which some of the participants included in their bots. Lastly, it is easy to imagine that an Alexa capable of sustaining longer conversations would generate more data that in turn can be commodified via the logics of platform capitalism (Srnicek 2022; Strüver 2023b), providing a further incentive for Amazon to pursue this goal. As Johnston et al. (2023, 24) reflect on the goal of the competition, they recognize that LLMs made the 20-minute goal very achievable while also pointing to some drawbacks of using LLMs.

The most obvious drawback is latency[5]. Multiple developers reported that adding more LLM capabilities to their bots increased the time that it took for the bot to answer, as generative models take longer than a heuristic model with pre-configured answers would (e.g., Breen, 44; Centis, 53; Dart, 10; Raju, 50). One developer elaborated upon the problem with latency by focusing on users' limited attention span and it being better to give a mediocre answer quickly than a good one really slowly (Scott, 43), because:

> Just latency is very, very important. And especially when you're talking to a bot; very, very frequently when our bot was good, but slow, we would see people just getting bored. Because you're sitting there trying to talk to this thing and waiting for like 10 seconds. And so, you just leave and give it a bad rating. … So, a huge focus for me was just trying to reduce those latencies. And to that end, we used other Amazon products and things databases for smart caching and that type of thing. (Scott, 18)

Scott's remarks point to several effects of structuration. For one, using Amazon tools that help in the process reflects a form of orchestrated efficiency. Further, Scott mentions their dependency on the feedback stars of users in the later stage of the contest, which is one metric of success in the competition. As "platform participants", users are "integrated into the monitoring and control systems of the platforms as decentralized co-controllers" (Dolata and Schrape 2023, 13). The resulting pressure to balance quality against latency is part of an infrastructuration process whereby the teams decide to what degree to include LLMs despite their increased latency, and then observe how their decisions are received as reflected in users' ratings. These are contingent decisions that the teams make; another participant described a different prioritization: "There are a lot of constraints on resources and latency using large language models,

5 It has to be noted that eight of the twelve participants emphasized lack of resources while simultaneously mentioning problems with latency. They deplored constraints on computing resources and funding, particularly as running an LLM is costly in both. Put poignantly: "working with machine learning is very expensive at this point, and if you don't have enough computer resources, then you fall behind" (Chidi, 101). Which puts an emphasis on the unequal conditions that generative AI is being developed and distributed in, as there are very few companies that are able to supply the capital and material basis for large-scale LLM usage (Srnicek 2022; Luitse 2024; van der Vlist et al. 2024).

and given the time constraints we got something working fast and then never replaced it" (Breen, 44). The potential for agency in development is thus limited by users' ratings, which teams are obliged to heed if they want to succeed and stay in the competition.

When talking to a voice assistant, users generally expect the assistant to respond to their query in a fairly reliable way. Users can only assume that assistants will perform their various algorithmic language processing steps correctly and give appropriate answers (see, e.g., Strüver 2023b; Hector and Hrncal 2024). However, the developers interviewed indicate that integrating LLMs into their Alexa bots can potentially lead to a reduction in the reliability of answers, as engineers have limited control over the quality of responses: "up to some point, we can control the quality but we cannot guarantee 100 % quality every single time for every topic" (Dart, 107). This can lead to bots sometimes *not giving good or correct answers* (Chidi, 111), especially in comparison to the entirely controllable scripts (Dart, 111) of heuristic models. Some teams decided to incorporate less LLMs specifically for this reason. Dart mentioned that with an increased proportion of LLMs within the bot, it could "hallucinate" (Dart, 16), which was also mentioned in the official recap of the SBC5, alongside contradictory answers (Johnston et al. 2023, 24). Thus, a certain volatility leaks into the system when implementing generative AI into Alexa bots. As the inflexible heuristic scripts are one of the oldest forms of machine learning (Li 2023), the resources to control their outputs are well established and institutionalized by professional education and tools, serving as forms of structure to produce reliable answers from Alexa. Comparably, LLMs are relatively new and seem to show a lack of established practices of control, leaving the teams to deal with the tasks of infrastructuring on the fly. One participant put the importance of controllable answers into perspective as follows:

> You have to work on those safety features. It will be more harmful if it comes out of a voicebot instead of just a chatbot, right? There are cases like that. I think there are much more things to do before they can just use ChatGPT in a voice assistant. And I'm sure there will be legal consequences, too. Because children use the voice assistant because they do not have access to ChatGPT. (Chidi, 141)

Safety features that have yet to be developed for the integration of these types of LLMs could be a way to increase robustness of input and output. On the one hand, Chidi points to the less specifically explicated queries that are expressed orally; which users would have to adapt in time, as they *learn how to talk* to voice assistants (Habscheid 2023, 185–186), while establishing new routines. On the other hand, the fact that voice interfaces are more accessible to, for example, children, due to their specific characteristics as a medium (Soffer 2020, 932), can cause problems when considering the lack of quality control. At the same time, developing more reliable institutionalized methods of structuring and controlling answers given by generative AI is in the interest of Amazon from a brand perspective, structuring the development of Alexa. Emily West calls the brand of a company the experienceable *face* for consumers to interact and relate with, impacting a company's success. Seemingly, Amazon's branding and advertisement is intentionally innocuous, attempting to achieve familiarity while offering minimum identity. Amazon's brand is defined by the affective convenience and ease of use of their consumer products (West 2022, 25–27). Alexa, too, is supposed to convey exactly these unobtrusive brand points, as it acts in a way of *idealized servitude* (Phan 2019, 29) that does not draw attention to itself but simply functions as a reliable touchpoint for users and enables *frustration free* (Strüver 2023a) service. Amazon "builds an affective relationship with its customers through interaction. And a key part of that interaction is reliable access to and efficient delivery of goods, making the affective relation tangible and touchable on a regular basis" (West 2022, 31). Perceiving Alexa in the light of the importance of this type of convenient, familiar, and reliable branding that is mainly conveyed through interaction highlights how volatile answers of an LLM-driven Alexa could threaten this brand image. Answers that are wrong, contradictory, or offensive, and easily accessible to all household members, could tarnish Amazon's reputation. Which is even more important considering that users' trust in voice assistants has been shown to correlate strongly with their sympathy towards the company behind the assistant (Weidmüller et al. 2022, 644). It is therefore no coincidence that Amazon actively applies internal and external quality control measures and moderation to protect its good reputation from unintended consequences of innovation, and strongly incentivizes high conversation quality during the APC.

While some developers report that the frameworks provided by Amazon struggled with interaction with the real world (Erwin, 96–98; Pak, 101), one participant rounds this discussion off with a succinct contextualization of different programming approaches for voice technologies:

> Because a lot of what makes ChatGPT seem so amazing and so impressive is that there's nothing at stake with the answer being correct.

> And if it works 90% of the time, it's like 'wow this works 90% of the time', but what are the situations where being wrong 10% of the time is okay? (laughs) I think that's something that we don't really have a very good answer about and we don't really have a very good answer about what the real trajectory is for getting kind of more accurate information out of these things …
> Think about the way that Siri was built, or the way that the existing assistant functionality is built on the Alexa devices for example; you know those systems were built in a particular way to make sure that they had predictable accuracy. Where in some sense once the speech recognition could be as bad as you like but if the words got recognized correctly, it would play the song that you asked for. (Breen, 73–74)

This reflects how Alexa was originally built with classical and established machine learning tools. It produces reliable results to specific queries. Which is what Amazon has built its market share on, especially in the domestic internet of things, where Alexa acts as a central hub to coordinate smart home devices (Strüver 2023a). As long as these problems prevail, preserving this functionality and position in the market serves as a strong incentive for Amazon to not completely switch to LLMs. Amazon might not desire to break the institutionalized usage of Alexa in users' homes:

> There are a lot of low stakes and kind of information access applications where ChatGPT is sort of a plausible current tool; but for things like assistants that have to hook up with something that's happening in the world, where the outcome matters, it's a lot further away than it might look. Just because you want to be able to have some guarantees. (Breen, 75)

This emphasizes LLMs' weakness of reliability, especially in interactions with the real world, where they could be implemented into material processes and routines. Assuming that users integrate Alexa as a device to control their smart homes – as intended by Amazon – and have performed a sense of infrastructuration in establishing routines with the device, they have black-boxed certain aspects of those interactions and presumably would not want to reconsider their smart home infrastructure on a daily basis: it would be against the use case to have to ask Alexa three times to turn on the lights or to lock the door. With Alexa already embedded in smart homes across the globe, users have developed certain path dependencies. However, these can be broken if the device ceases to provide the technical infrastructure that enables the promised convenience and reliability of Amazon's brand. Especially this connection to the smart home leads to questions around the technical implementation of LLMs alongside more traditional ways of developing the assistant, which will be explored through the developers' perspectives next.

5.2 Implementing LLMs into Alexa: Deciding who talks to the user

Against the backdrop of the risks and benefits of LLMs and their implementation into Alexa, the following will look at the practices of infrastructuration that the developers describe when integrating LLMs into their Alexa bots. Corporate interests of staying innovative and profitable during a time of technological innovation seem economically rational, as Alexa and the developers face the repercussions of a competitor releasing a popular new technology: "Suddenly, users were expecting much better conversations than what was achievable by the stupid rule-based systems that we started with" (Centis, 35), and, consequentially, many users tried to tease Alexa (Gardé, 48). Breen compared the Alexa experience prior to the advent of ChatGPT to a call-center-AI that guides users through the functions that it can do effectively and concluded: "that's essentially the opposite of the design patterns that are rewarded in this Amazon competition" (Breen, 66). This presents an assumption on the structure of the competition set by Amazon, which gets reinforced by the fact that Amazon provides an API for detecting when a user found a conversation boring or wanted to terminate it (Bardiola, 115). According to the interviewees, users were essentially expecting Alexa to be more than it used to be, and generative AI was seen as one tool that could achieve that by providing more flexibility to react to different topics, which Amazon structurally incentivizes by the competition design and the resources it offers. If the teams accepted this structuration of their innovation process, they needed to establish when to use an LLM and when to deploy classical heuristics to talk to the users. More often than not, this decision was rather an accomplishment in practice (Burkhardt et al. 2022) that was influenced by means of structure, than a general ruling, as is explored in the following.

5.2.1 Building a pipeline: Classifying criteria that govern when to swap between models

"There's usually a fork in the road. You try and see if there's an easy non-AI response you can give" (Scott, 42). This remark generally applies to *if-statements* that can be dealt with by simply programmed conversational heuristics that are well established and institutionalized through open-source models, but also through tools like CoBot (Bardiola, 114) that are developed by Amazon based on their experiences with Alexa and therefore come with a certain range of answers and topics. The most prominent examples were conversations about sports, or the types of food liked by users, i.e., contexts where the space for answers was easily categorizable. If the topic is outside the scope of the predetermined heuristics, using an LLM seems evident. But remembering that developers *limited how often they used LLMs* because doing so was expensive and introduced latency, gave an incentive to further complicate this decision process of deciding which models users talk to. The question became about how to combine these different approaches. Developers described how they arrived at a "blend of pre-scripted dialogues and the new answers generated by the new generative models" (Bardiola, 12), by building a *pipeline* (Chen, 100–101; Raju, 52; Chidi, 108; Dart, 18) that used multiple components to create a "hybrid approach" (Dart, 107) between different models that the Alexa bot[6] used to talk to its users. The word pipeline – albeit an industry standard-term – evokes a tangible image of infrastructure that *matters* (Slota and Bowker 2017, 530): it guides data through different checkpoints and permits certain functions while prohibiting others, transporting backgrounded contingent values and decisions. Even before considering the concrete pipeline implementation, developers had to take stock of which available existing heuristics they wanted to continue using. These could range from previous work in the field, open-source resources, or self-made models, to the tools and resources provided by Amazon. One interviewee reported that their university had had a team participating in the competition for several years (it is common for the same team/faculty leader within a university to have a changing team of students that participates annually under a similar name) and had built its own repertoire of manually-coded dialogues, which they liked to keep using:

6 While the analysis here concerns determining which type of technology is used to talk to users when, it is important to remember that there are differences between the regular Alexa and the Alexa skill that users access to talk to the Alexa bots developed in the competition. The latter is not congruent with the regular Alexa. Additionally, users can get confused by the competition skill, having expected that "they [would be] speaking to the same bot, but in the end they got one of the nine." (Bardiola, 113) This introduces another layer of 'who is the user talking to?' that is specific to this competition.

> The previous rounds of Kunkka [anonymized team name], the bot I was working on, they also used LLMs. But now we are focusing a lot on using them and employing them even more. What we did was, we were trying to enrich those [manually-coded] dialogues. So, use the dialogues that we have, because they are good. And, the quality is, I would say, very nice. We didn't want to discard it. It also could make the things a bit tough, because we were not starting freshly. I think some teams did that; they could come up with the whole architecture from scratch. But we are already using something. We were kind of limited in some sense, to what we are able to do. (Dart, 107)

What the member of team Kunkka described here is the process of infrastructuration in situ over the span of several annual competitions as described by Edwards (2019). Situational decisions made by previous teams to develop, use, and expand manually-coded heuristics for their bot (which, in Giddens' sense can be seen as rational, given the bounded temporal perspective of each team's efforts, because LLMs were far less capable in the previous iterations of the competition) become black-boxed, routinized, and materialized in the systems that subsequent teams use for later competitions. With the competition taking place annually, the decisions made by previous teams to use manually-coded methods do not need to be reconsidered in the moment of setting up the infrastructure for the next competition. This infrastructure is learnt as part of their team membership; with usually the faculty or team leader remaining the same to convey practices. Further, this institutionalization of infrastructural practices is reinforced if a team did well in the previous years because their process of infrastructuring has been structurally validated by Amazon and the users. Ironically, this makes teams with a proven infrastructure resistant to Amazon's orchestration measures to a degree – e.g., Dart described their team's active non-use of CoBot, for better or worse: their existing infrastructure enabled certain actions and limited others. In order to reconsider their infrastructuration process and respond to the call of implementing LLMs, they needed to question their routinized decisions, examine what they would like to retain, and eventually find ways to merge the existing base with new models. However, because they had a solid basis before the competition started, they were in the luxurious position of being able to evaluate whether they perceived the extent of power exercised via the means of structure and orchestration to be pervasive enough to warrant changes in their bot and to what degree. In this example, the concepts of *duality of structure and action in a reciprocal reproduction* (Giddens 1984), as well as infrastructurized path dependencies are tangible.

Keeping in mind this perspective of situated practices that get institutionalized through the ongoing (re)production of structures within practices helps one to understand how the developers solved the problem of merging established systems with the new LLMs from a procedural perspective of everyday interactions. The member of Kunkka described the process of *injecting* phrases generated by LLMs into their bot as a phase of *constant experimenting* as they tried to merge the two approaches. In order to do that, they reported having to invent "ending criteria, when to end the dialogue, when we should switch to it" (Dart, 110). This short description hints at the process of decision-making involved in merging the two systems by building a pipeline, that guides data flows: The developers needed to establish rules for the usage of LLMs in a conversation, considering the prevalent action-structuring elements like constraints of resources, latency, and quality control. In all likelihood they switched to an LLM when the conversation topic or prompt was beyond the scope of their manually-coded heuristics. They then needed to find a way to define and *classify* (Bowker and Star 2000) a point in the conversation when it could be transferred back to the heuristics model while adhering to acceptable conversational conventions (as incentivized by the APCs goals). This again represents a case of developing a technical infrastructure that is accomplished by a string of decisions that eventually get black-boxed within a model, representing a switching mechanism to decide which type of machine learning the users talk to. The process of black-boxing makes their decision processes transparent and imperceptible in practice to users, as it has not to be reconsidered in conversation with the Alexa bot. A switching mechanism like this exemplifies how *opaque* conversation with Alexa can be, as it shows how during a single conversation, multiple switches can take place, with users talking to different algorithms that have different strengths and biases and are built in fundamentally different ways. This evokes the previously elaborated topic of suitable application space for LLMs and the question of "what are the situations where being wrong 10% of the time is okay?" (Breen, 73), as developers are obliged to make decisions that have significant consequences for users[7]. Hidden to users remains the decision of how much priority is given

7 This problem is exacerbated by aspects of unintentional events: Complex conversational models that switch between algorithms often need to have another superseding model that can repair the flow of dialogue should the bot fail to keep its

to quality or accuracy in a particular scenario, i.e., whether a human-written heuristic model is answering, or a generative AI with a higher volatility. This is a hyperbolic problematization however, as obfuscation of this kind is structurally incentivized and normalized by the aspirations of Amazon, which sets the goal of fluent conversation with Alexa – unimpeded by drawing attention to precisely these infrastructural technicalities worked out here. Ultimately, users simply talk to Alexa as some form of actor, regardless of the subtending model.

5.2.2 Transitioning between algorithmic approaches through testing

To further understand how LLMs can be integrated into Alexa bots, the previous approach to implementation can be contrasted with the option of prioritizing the implementation of LLMs. During the 2022/23 APC, an abundance of LLM models were getting published at a fast pace, where "papers are literally coming out every single week at this point" (Chidi, 121). This led to a volatile environment of rapidly changing models as the participants tried to implement generative AI into their bots: "Several times during the competition we changed the main model. It was not just [motivated by] Amazon; it was mostly new models appearing on the market. And you're like quickly redoing everything to make sure that it would work better" (Gardé, 76). Furthermore, Bardiola pointed out that finding and implementing suitable LLMs into their bots was not as straight forward as one might imagine (Bardiola, 41). With the perceived need to constantly exchange suitable LLMs, deciding how to introduce LLMs into the bots required developers to consider possible practices and infrastructures of testing algorithms. One of Amazon's central advertising points for the APC is the contact to the Alexa user base and the promise that "the immediate feedback from these customers will help students [the APC developers] improve their algorithms much faster than previously possible" (Amazon 2024). Live testing is a core function of the Alexa platform for Amazon (Strüver 2023b, 15–17) and is reproduced by the

---

outputs oriented towards the goal that the user is trying to achieve in their conversation (Erwin, 36). Further, Bardiola (117) explained that if an LLM malfunctioned on the weekend, or during the night, when their team's support service was offline, they would let the bot refer to Amazon's inferior and less specialized LLM as a backup. Ensuring the uptime of a service is structurally enforced by Amazon's certification standards for technologies that interact with Alexa (Strüver 2023a, 113). Developers' nods to the crucial work of maintenance (Bowker and Star 2000, 160–161) from and on the bot further complicate the question of who is talking to the user.

teams, when they rely on the platformized mechanisms of feedback established by Amazon. While assessing the applicability of LLMs for Alexa, this is highly interesting, as their performance is more complex to measure and goals like fluent conversations or succinct guiding through a task are hard to quantify. *Benchmarking* is a prevalent and highly institutionalized practice among machine learning researchers and involves the constant attempt to outperform previous algorithms within a *competitive computing culture* (Orr and Kang 2024). Usually, algorithms are compared by means of quantifiable measures like how long it takes to execute certain standardized tasks, which can also be applied to LLMs. Quantifying a successful conversation, however, while not impossible, is more complicated and subjective than calculating an algorithm's efficiency at transcribing speech. Against this backdrop, the testing process gains another dimension, as developers reproduce the competitive computing culture of their academic discipline by frequently changing models in the hope of improving performance as well as being incentivized to use the resources provided to them by Amazon – which sets the APC up in a way that also reproduces this culture. Here, motives of constant refinements endorsed by Amazon become conflated with the normative goal of striving for improvement that is inherently cultivated by universities and places of education of this profession and, correspondingly, research field: "Machine learning researchers are always very optimistic [about algorithms] because it's just the way they're hill climbing and of course if you can make the thing one percent better every year, eventually it will be very, very good" (Breen, 74). Recognizing this institutionalized motivation to implement different LLMs contextualizes the process of navigating the intersection between LLMs and heuristics, as described by Scott in their step-by-step account of how their team incrementally replaced heuristics with LLMs in their bot:

> Scott: I mentioned the heuristics and using LLMs earlier. When we started off, a very, very major chunk of our code was just heuristic-based [manually-coded]. And we only really used an LLM if all the heuristics failed and over time our big transition was having fewer and fewer and fewer heuristics and more and more LLM. And quite often we'd run A/B tests where we got rid of a huge chunk of heuristics and check to see if the model still did well, and oftentimes it would fail and not do well. Then we'd have to go in and investigate and debug and figure out why.

> Interviewer: When you investigated, how did you do that?
>
> Scott: We looked at our ratings. We looked at the average latency in a response [the pause between turns]. We looked at what the actual response was and what it was in response to; what the user said and what the bot said, and we looked at whether it made coherent sense. Oftentimes it wouldn't. And we just investigated by looking at common failure modes. And then you try and reproduce the failure modes, once you put in your supposed fix and if it still fails, then clearly your fix hasn't worked. In that sense, it was very specific in that you look at specific examples and try and fix those.
>
> Interviewer: Sounds like looking through a lot of conversational logs, right?
>
> Scott: Yeah, that part is a lot more tedious to do and for sure you can [do that]; but it's a lot easier to just look at… Over time, as we started to have thousands of conversations, it's easier to just look at conversations that perform poorly and see what specifically might've failed. (Scott, 50–54)

This account highlights how integrating an LLM into the Alexa bot is a highly contingent task that requires extensive testing and verification. Starting with a major portion of their code being heuristics-based, this team transitioned incrementally to utilizing more LLMs by replacing functionalities and constantly validating if each new functionality performed according to expectations, adjusting accordingly, and then reevaluating. To test their changes they employed A/B tests, which *continuously and seamlessly change* (Marres and Stark 2020, 434) the version of the bot that different users interact with at a particular moment in time. The A/B tests described here presumably compared the largely heuristic model with a new version of the bot that had some parts of its conversational heuristic model – e.g., labelling a user's intention through natural language understanding (Longwei, 94) – replaced by an LLM. In such a scenario, one user would talk to the baseline bot as version A and another user would talk to a version B of the bot that has a new LLM element added. The developers can then compare the conversations held by the two versions of the bots, either directly or through metrics. Due to the large volume of conversations, Scott described surveying the metrics' latency in the new version and low user ratings in order to identify outliers. In turn, these metrics helped to locate problems in specific conversations for closer investigation. Moving from abstract to concrete, the subsequent analysis of the actual conversations, which sought to ascertain problems in the LLM – such as a generative model producing random characters, as reported by a different team: "instead of saying a normal sentence it started generating stars and hashtags" (Bardiola, 87) – served as the basis from which to fix the model and repeat the testing process. As explained by Scott, this procedure for testing the integration of LLMs enabled specific undesired conversations to be targeted.

At this point it is important to recall the characterization of the developers' relationship with the platform organization that develops the platform and establishes institutional rules for how third parties and users can access the social space of the platform (see 3.1). In describing and analyzing the need for extensive testing when implementing LLMs into Alexa bots, two points emerged clearly: on the one hand, users are implemented into the competition as a development tool; they serve as agents of moderating and testing the bots and provide feedback to the APC teams as they navigate the process of integrating LLMs into Alexa. As mentioned earlier, this is a typical aspect of platform companies that involves users in a very calculated way as "decentralized co-controllers" (Dolata and Schrape 2023, 13) to shape, moderate and develop platforms and to re-code them if necessary. This is especially interesting for Amazon considering the lack of established ways to benchmark conversational AI models. Users function as an evaluation instance that does not need to be given specified classifications or criteria to define the diffuse goal of better conversation quality, which makes user interaction via Alexa an even more valuable resource for Amazon. On the other hand, the APC teams get feedback in a form that is determined by Amazon, as every interaction (ratings, comments, and text logs) that they have with the users is structured by the boundaries and conditions of the infrastructure set up by Amazon. Further, Amazon's choice to represent all the contestants' bots as a single Alexa skill that is specific for the APC (which can create confusion among users), instead of making them available as part of Alexa's general service is an act of moderation. This measure protects the brand of Alexa from potentially being associated with faulty bots, while it also opens space for experimentation within the competition, allowing different standards to apply within this dedicated test environment. Generally, while curating a data set is difficult in the APCs' test environment, this is definitely a caveat to the competition. The data set that provides the basis for testing algorithms is absolutely biased to users in the USA, as the Alexa skill for the competition is only available there. Furthermore, it could over-represent certain demographics, who choose to interact with the APC skills (Centis, 76). Otherwise the data set is seemingly uncontrolled in terms of diversity, which could lead to cultural as well as linguistic biases in the testing of algorithms that eventually might be rolled out onto Echo devices globally. Unlike other AI competitions, in which efforts are made to provide a suitably representative data set for testing, which need to be sufficiently diverse for a technology to be applicable globally (Luitse et al. 2024, 17), such issues are not addressed in the APC. This examination of the ways in which developers test their algorithms when transitioning between heuristics and LLMs thus reveals how Amazon leverages the interaction of the university teams with users of the Alexa platform to develop technologies and institutions for Alexa. Knowledge production on the transition between heuristics and LLMs in the competitions is (unsurprisingly) inherently colored by Amazon's platformized structuration measures and values. The two quoted interview excerpts about development practices at the intersection of LLMs and heuristics can be read as an analogy to the predicament of Amazon's Alexa team: It can only be assumed that the situation that Amazon's Alexa team found itself in during the first year of ChatGPT was shaped by similar reconsiderations of path dependencies and of structuration, as Amazon came to face an external influence that led it to question the viability of maintaining its long established reliance on heuristics. The different ways of navigating the transition between the two machine learning approaches that were being developed in the APC will most likely find their way into the main Alexa system in some form, as they represent somewhat established practices of merging, switching, and testing. Moreover, Amazon's own methods of testing for Alexa are not restricted by the limitations on information that are imposed in the competition; Amazon-employed developers have access to far more comprehensive interactional data (Strüver 2023b). This background can now be contrasted with the competition against ChatGPT and its influence on the APCs.

5.3 Catching Up with Innovation: The APCs as a Testing Ground for Alexa-LLMs

Following these insights into LLM development practices for Alexa, the APC can now be situated within the larger scope of the competitive market of LLM products, especially the popular ChatGPT. During the runtime of the 2022/23 competitions, users across the globe were being introduced to the capabilities of ChatGPT and began to expect similar functions from Alexa. With users slowly re-institutionalizing what AI agents were expected to do, OpenAI and ChatGPT entered the equation of Amazon's platform structuration. According to some interviewees, the reason for banning use of ChatGPT in the APC was "Because then it would be just easier to go: 'OpenAI, generate a response, be a social bot'" (Dart, 19). While it may seem fairly unremarkable that the use of a competitor's product would be prohibited in an innovation challenge that is intended to proprietarily advance Alexa, the motivation behind this ban is further contextualized by the APC developers' descriptions of the technological status quo of the Alexa system that they came to know during the competition. The LLMs provided by Amazon were, according to participants, along the lines of robustly processing text to find similarities (Breen, 73), and far from reliable or satisfactory to generate coherent utterances (Longwei, 93). Longwei predicted that Amazon's template-based heuristics system would not be used in future APCs, but concluded nonetheless that it "would be kind of hard for Alexa to switch from their previous path to really open for large language models" (Longwei, 95). While exemplary, these sentiments convey the state of Alexa technology at the time that ChatGPT was unveiled. Although it is possible and probable that the APC developers did not get a comprehensive overview of all the ongoing developments at Amazon, their accounts certainly reflect the state of technology that was being offered to third parties wishing to work on the Alexa platform. Assuming that these statements do indeed offer a reasonably accurate estimate of the state of technology of Alexa at the time, it does not surprise that Amazon was undergoing a comprehensive restructuring of organizational resources in the Alexa team (Kim 2022) and announced new plans for Alexa and generative AI in general (Bensinger 2024; Krishnan 2024). In this light, banning the use of ChatGPT in the APC should be seen as part of the measures of restructuring development of the platform Alexa. As a platform organization, Amazon is intent on leveraging a multitude of resources for the further development of Alexa as a technology and platform. This includes the APCs, as Gardé put it: "everything that we developed basically would be owned by Amazon. So, it's a good way for them to get lots of input on different areas of generative computational AI" (Gardé, 142). Allowing the use of ChatGPT could forego the development of possible technological approaches to solutions for problems that Alexa faces. The APCs that took place at this juncture of conversational technology development need to be seen from the perspective of being one of the tools of innovation – at the *periphery of the platform* (Ametowobla and Kirchner 2023) – that Amazon was utilizing in its efforts to orchestrate the development of Alexa.

As is standard practice for big tech companies, Amazon also complements their in-house R&D by buying existing start-ups (Dolata and Schrape 2023, 7). However, compared to such corporate takeovers, the universities involved in the APC represent a looser form of cooperation that is absolved of the need to be economically viable, which enables a distinct room for innovation but also involves different resources of structuration for Amazon. In the APC, Alexa is specifically not an industry platform for innovation on an equal footing (Dolata 2024), but rather a platform that enables Alexa-centric cooperation with university teams. These teams are more malleable and susceptible to Amazon's orchestration efforts in particular ways – the interviewees mentioned gaining industry experience and recognition alongside potential future job offers in the field as motivations for participating in the APC, as well as sought-after funding for their labs and PhDs. Such involvement in the education system can eventually play a structuring role in shaping the field's values and aligning them with the interests of companies that end up employing the – highly sought after (Tekic and Füller 2023, 5) – graduates. In that way Amazon can attempt to let the participants adjust to Alexa's infrastructural path dependencies and let them experiment in developing approaches to transitioning between heuristics and LLMs in ways that comply with Alexa's brand: "Sometimes you can't just replace everything with the new technology. You have to kind of find the right balance between using the new tools and previous tools" (Pak, 100). These observations echo what Luitse et al. conclude from their research on medical AI platform competitions: "the configuration of platforms, competition organisers, and participants concentrates power toward a small number of actors" (2024, 16). In the case of the APCs, this effect is compounded as both the actors of platforms and the competition organizers are represented in unison by Amazon, who can therefore direct the goals of knowledge production towards certain problems, e.g., the transition of a heuristic Alexa towards LLM integration, as is evidenced in the papers published in the proceedings of the SBC[8]. It still remains to be seen whether the models that were developed in the competitions will ultimately find use in Alexa (Longwei, 89), or whether, like the Netflix competition's winning algorithm, they will never be implemented (Seaver 2022, 58). In any case, the APC represents an R&D resource that can be utilized in attempts to re-code Alexa as a platform, but it is a resource that nonetheless remains hard to control due to the contingent development practices of university teams.

8 See https://www.amazon.science/alexa-prize/socialbot-grand-challenge/2022.

6 Conclusion

While the actual workings within Amazon remain opaque, the study did its best to fairly portray the experiences of the interviewed developers. The analysis presented here contributes to the understanding of how Amazon cooperates with third parties that work on the Alexa platform and shows the effects of hierarchical structuring while also highlighting the practical decisions and opportunities for resistance (e.g., not using the CoBot tool offered) that arose during the competition. This helps to critically understand the sociotechnical underpinnings and environments of the development of a technology that is used by many users on a daily basis. This is conducive to the understanding of how modern AI systems are developed and the risks that accompany ongoing changes in technology development. Insights such as these can contribute to shifting the academic discourse in the social sciences and humanities away from a focus on data to concentrate on deepening understanding of the sociotechnical circumstances and means that shape AI development (Srnicek 2022). In the study reported on here, a sociological perspective has been taken to investigate Alexa as a platform and infrastructure and to examine the practical accomplishment of development under structuration. This contributes a genuinely sociological understanding of platforms by empirically scrutinizing Amazon's structuration efforts and the infrastructuring acts that can be found when third party actors such as universities interact with a big tech company like Amazon.

Future studies could expand on this work by building on the arguments presented here and investigating the extent to which they can be applied to different AI technologies like other voice assistants, or using them to inform studies of Alexa usage in the home, or to look into whether LLMs have actually been incorporated into Alexa since the transition described here. As the famous Netflix competition shows, these types of (AI) technologies tend to be ephemeral and even a solution that emerges victorious from a competition might be too complicated to be implemented, or the organizing platform might change its business model, making the solution obsolete (Seaver 2022, 58). What remains, however, are the insights into how technology development is undertaken at the cutting edge of competition, and into the conduct in cooperation of one of the biggest tech companies of the present moment; a corporation that impacts the lives of millions of users globally every day.

Acknowledgements

The author has no conflicts of interest to report. I would like to thank student assistant Aileen Halbe for the data wrangling of the contact emails. Further, I would like to thank the editors for the feedback provided on the article. Lastly, I want to thank all interviewees for the immensely insightful conversations. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 262513311 – SFB 1187. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 262513311 – SFB 1187.

References

Agichtein, Eugene, Michael Johnston, Anna Gottardi, Lavina Vaz, Cris Flagg, Yao Lu, Shaohua Liu, et al. 2023. "Advancing Conversational Task Assistance: The Second Alexa Prize TaskBot Challenge." In *Alexa Prize TaskBot Challenge 2 Proceedings.* https://www.amazon.science/alexa-prize/proceedings/alexa-lets-work-together-introducing-the-second-alexa-prize-taskbot-challenge.

Amazon. 2022a. "SimBot Challenge Rules." Amazon Science. 2022. https://www.amazon.science/alexa-prize/simbot-challenge/rules.

Amazon. 2022b. "SocialBot Grand Challenge Rules." Amazon Science. 2022. https://www.amazon.science/alexa-prize/socialbot-grand-challenge/rules.

Amazon. 2022c. "TaskBot Challenge Rules." Amazon Science. 2022. https://www.amazon.science/alexa-prize/taskbot-challenge/rules.

Amazon. 2024. "Alexa Prize." Amazon Science. 2024. https://www.amazon.science/alexa-prize.

Ametowobla, Dzifa, and Stefan Kirchner. 2023. "The Organization of Digital Platforms: The Role of Digital Technology and Architecture for Social Order." *Zeitschrift Für Soziologie* 52 (2): 143–56. https://doi.org/10.1515/zfsoz-2023-2012.

Amoore, Louise, Alexander Campolo, Benjamin Jacobsen, and Ludovico Rella. 2024. "A World Model: On the Political Logics of Generative AI." *Political Geography* 113 (August): 103134. https://doi.org/10.1016/j.polgeo.2024.103134.

Baldwin, Carliss Y, and C Jason Woodard. 2008. "The Architecture of Platforms: A Unified View." http://hbswk.hbs.edu/item/6025.html.

Bensinger, Greg. 2024. "Exclusive: Amazon Mulls $5 to $10 Monthly Price Tag for Unprofitable Alexa Service, AI Revamp." Reuters, 2024, sec. Technology. https://www.reuters.com/technology/amazon-mulls-5-10-monthly-price-tag-unprofitable-alexa-service-ai-revamp-2024-06-21/.

Bogner, Alexander, Beate Littig, and Wolfgang Menz. 2014. *Interviews mit Experten: Eine praxisorientierte Einführung.* Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-531-19416-5.

Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things out: Classification and Its Consequences. Inside Technology.* Cambridge, Mass: MIT Press.

Burkhardt, Marcus, Daniela Van Geenen, Carolin Gerlitz, Sam Hind, Timo Kaerlein, Danny Lämmerhirt, and Axel Volmar. 2022. "Introduction." In *Media in Action*, edited by Marcus Burkhardt, Daniela Van Geenen, Carolin Gerlitz, Sam Hind, Timo Kaerlein, Danny Lämmerhirt, and Axel Volmar, 9–36. Bielefeld: transcript. https://doi.org/10.14361/9783839455616-001.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 1–12. https://doi.org/10.1177/2053951715622512.

Crawford, Kate, and Vladan Joler. 2018. "Anatomy of an AI System." 2018. http://www.anatomyof.ai.

Dijck, Jose van, Thomas Poell, and Martijn de Waal. 2018. *The Platform Society.* New York: Oxford University Press.

Dolata, Ulrich. 2019. "Privatization, Curation, Commodification: Commercial Platforms on the Internet." *Österreichische Zeitschrift für Soziologie* 44 (S1): 181–97. https://doi.org/10.1007/s11614-019-00353-4.

Dolata, Ulrich. 2024. "Industrieplattformen als Markt-, Produktions- und Innovationsflächen. Feldvermessungen und theoretisch-konzeptionelle Überlegungen." *Berliner Journal für Soziologie* 34 (2): 171–96. https://doi.org/10.1007/s11609-024-00526-3.

Dolata, Ulrich, and Jan-Felix Schrape. 2023. "Platform Companies on the Internet as a New Organizational Form. A Sociological Perspective." *Innovation: The European Journal of Social Science Research* March: 1–20. https://doi.org/10.1080/13511610.2023.2182217.

Edwards, Paul N. 2019. "Infrastructuration: On Habits, Norms and Routines as Elements of Infrastructure." In *Research in the Sociology of Organizations*, edited by Martin Kornberger, Geoffrey C. Bowker, Julia Elyachar, Andrea Mennicken, Peter Miller, Joanne Randa Nucho, and Neil Pollock, 355–66. Emerald Publishing Limited. https://doi.org/10.1108/S0733-558X20190000062022.

Edwards, Paul N. 2021. "Platforms Are Infrastructures on Fire." In *Your Computer Is on Fire*, edited by Thomas S. Mullaney, Benjamin Peters, Mar Hicks, and Kavita Philip, 313–36. Cambridge and London: The MIT Press. https://doi.org/10.7551/mitpress/10993.003.0021.

Floridi, Luciano. 2023. "AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models." *Philosophy & Technology* 36 (15). https://doi.org/10.1007/s13347-023-00621-y.

Fourcade, Marion, and Kieran Joseph Healy. 2024. *The Ordinal Society.* Cambridge and London: Harvard University Press.

Frenken, Koen, and Lea Fuenfschilling. 2020. "The Rise of Online Platforms and the Triumph of the Corporation." *Sociologica* 14 (3): 101–13. https://doi.org/10.6092/ISSN.1971-8853/11715.

Giddens, Anthony. 1984. *The Constitution of Society: Outline of the Theory of Structuration.* Cambridge: Polity Press.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* New Haven: Yale University Press.

Gillespie, Tarleton. 2024. "Generative AI and the Politics of Visibility." *Big Data & Society* 11 (2). https://doi.org/10.1177/20539517241252131.

Gorwa, Robert. 2019. "What Is Platform Governance?" *Information, Communication & Society* 22 (6): 854–71. https://doi.org/10.1080/1369118X.2019.1573914.

Goulden, Murray. 2019. "Delete the Family: Platform Families and the Colonisation of the Smart Home." *Information, Communication & Society* 24 (4): 1–18. https://doi.org/10.1080/1369118X.2019.1668454.

Habscheid, Stephan. 2022. "Socio-Technical Dialogue and Linguistic Interaction. Intelligent Personal Assistants (IPA) in the Private Home." *Sprache und Literatur* 51 (2): 167–96. https://doi.org/10.30965/25890859-05002020.

Hallinan, Blake, and Ted Striphas. 2016. "Recommended for You: The Netflix Prize and the Production of Algorithmic Culture." *New Media & Society* 18 (1): 117–37. https://doi.org/10.1177/1461444814538646.

Hector, Tim Moritz, and Christine Hrncal. 2024. "Sprachassistenzsysteme in der Interaktion." In *Handbuch Sprache und digitale Kommunikation*, edited by Jannis Androutsopoulos and Friedemann Vogel, 309–28. Berlin and Boston: de Gruyter. https://doi.org/10.1515/9783110744163-015.

Helfferich, Cornelia. 2019. "Leitfaden- und Experteninterviews." In *Handbuch Methoden der empirischen Sozialforschung*, edited by Nina Baur and Jörg Blasius, 669–86. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-21308-4_44.

Hind, Sam, Fernando N. van der Vlist, and Max Kanderske. 2024. "Challenges as Catalysts: How Waymo's Open Dataset Challenges Shape AI Development." *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-01927-x.

Jassy, Andy. 2024. "2023 Letter to Shareholders." https://s2.q4cdn.com/299287126/files/doc_financials/2024/ar/Amazon-com-Inc-2023-Shareholder-Letter.pdf.

Johnston, Michael, Cris Flagg, Anna Gottardi, Sattvik Sahai, Yao Lu, Samyuth Sagi, Luke Dai, et al. 2023. "Advancing Open Domain Dialog: The Fifth Alexa Prize SocialBot Grand Challenge." In *Alexa Prize SocialBot Grand Challenge 5 Proceedings*. https://www.amazon.science/alexa-prize/proceedings/advancing-open-domain-dialog-the-fifth-alexa-prize-socialbot-grand-challenge.

Khan, Lina, M. 2018. "Amazon—An Infrastructure Service and Its Challenge to Current Antitrust Law." In *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*, edited by Martin Moore and Damian Tambini, 98–129. New York, NY: Oxford University Press.

Kim, Eugene. 2022. "Amazon Is Gutting Its Voice Assistant, Alexa. Employees Describe a Division in Crisis and Huge Losses on 'a Wasted Opportunity.'" *Business Insider*, November 19, 2022. https://web.archive.org/web/20230106123404/https://www.businessinsider.com/amazon-alexa-job-layoffs-rise-and-fall-2022-11.

Kinsella, Bret. 2023. "Google Assistant and Alexa Are Both Getting Generative AI Makeovers." *Substack newsletter. Synthedia (blog),* August 2, 2023. https://web.archive.org/web/20230803220610/https://synthedia.substack.com/p/google-assistant-and-alexa-are-both.

Krishnan, Arun. 2024. "Amazon showcases new customer experiences powered by generative AI at CES 2024." *alexa-blog*, January 9, 2024. https://developer.amazon.com/en-US/blogs/alexa/device-makers/2024/01/alexa-characteraiandsplash-ces-2024.html.

Kuckartz, Udo. 2014. *Qualitative Text Analysis: A Guide to Methods, Practice & Using Software.* Translated by Anne McWhertor. Los Angeles: SAGE.

Li, Xiaochang. 2023. "'There's No Data Like More Data': Automatic Speech Recognition and the Making of Algorithmic Culture." *Osiris* 38:165–82. https://doi.org/10.1086/725132.

Luitse, Dieuwertje. 2024. "Platform Power in AI: The Evolution of Cloud Infrastructures in the Political Economy of Artificial Intelligence." *Internet Policy Review* 13 (2). https://doi.org/10.14763/2024.2.1768.

Luitse, Dieuwertje, Tobias Blanke, and Thomas Poell. 2024. "AI Competitions as Infrastructures of Power in Medical Imaging." *Information, Communication & Society* https://doi.org/10.1080/1369118X.2024.2334393.

Marres, Noortje, and David Stark. 2020. "Put to the Test: For a New Sociology of Testing." *The British Journal of Sociology* 71 (3): 423–43. https://doi.org/10.1111/1468-4446.12746.

Minder, Bettina, Patricia Wolf, Matthias Baldauf, and Surabhi Verma. 2023. "Voice Assistants in Private Households: A Conceptual Framework for Future Research in an Interdisciplinary Field." *Humanities and Social Sciences Communications* 10 (1): 173. https://doi.org/10.1057/s41599-023-01615-z.

Mols, Anouk, Yijing Wang, and Jason Pridmore. 2021. "Household Intelligent Personal Assistants in the Netherlands: Exploring Privacy Concerns around Surveillance, Security, and Platforms." *Convergence* 28 (6): 1841–60. https://doi.org/10.1177/13548565211042234.

Orr, Will, and Edward B. Kang. 2024. "AI as a Sport: On the Competitive Epistemologies of Benchmarking." In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1875–84. Rio de Janeiro Brazil: ACM. https://doi.org/10.1145/3630106.3659012.

Phan, Thao. 2019. "Amazon Echo and the Aesthetics of Whiteness." *Catalyst: Feminism, Theory, Technoscience* 5 (1): 1–38. https://doi.org/10.28968/cftt.v5i1.29586.

Plantin, Jean-Christophe, Carl Lagoze, Paul N Edwards, and Christian Sandvig. 2018. "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook." *New Media & Society* 20 (1): 293–310. https://doi.org/10.1177/1461444816661553.

Pridmore, Jason, Michael Zimmer, Jessica Vitak, Anouk Mols, Daniel Trottier, Priya C. Kumar, and Yuting Liao. 2019. "Intelligent Personal Assistants and the Intercultural Negotiations of Dataveillance in Platformed Households." *Surveillance & Society* 17 (1/2): 125–31. https://doi.org/10.24908/ss.v17i1/2.12936.

Reeves, Stuart, Joel E Fischer, Martin Porcheron, and Rein Sikveland. 2019. "Learning How to Talk: Co-Producing Action with and around Voice Agents." In *Proceedings of the "Mensch und Computer" 2019 Workshop on Interacting with Robots and Virtual Agents*, 362–63. München. https://doi.org/10.18420/muc2019-ws-654.

Rillig, Matthias C., Marlene Ågerstrand, Mohan Bi, Kenneth A. Gould, and Uli Sauerland. 2023. "Risks and Benefits of Large Language Models for the Environment." *Environmental Science & Technology* 57 (9): 3464–66. https://doi.org/10.1021/acs.est.3c01106.

Rowberry, Simon Peter. 2022. *Four Shades of Gray: The Amazon Kindle Platform. Platform Studies.* Cambridge, Massachusetts: The MIT Press.

Sadowski, Jathan, Yolande Strengers, and Jenny Kennedy. 2021. "More Work for Big Mother: Revaluing Care and Control in Smart Homes." *Environment and Planning A: Economy and Space* 56 (1): 330–45. https://doi.org/10.1177/0308518X211022366.

Seaver, Nick. 2022. *Computing Taste: Algorithms and the Makers of Music Recommendation.* Chicago: University of Chicago Press.

Self, Becky. 2021. "Conducting Interviews during the COVID-19 Pandemic and Beyond." Translated by Alexander Ryazantsev. *Inter* 13 (4): 9–27. https://doi.org/10.19181/inter.2021.13.4.1.

Shi, Hangjie, Leslie Ball, Govind Thattai, Desheng Zhang, Lucy Hu, Qiaozi (QZ) Gao, Suhaila Shakiah, et al. 2023. "Alexa, Play with Robot: Introducing the First Alexa Prize SimBot Challenge on Embodied AI." In *Alexa Prize SimBot Challenge Proceedings*. https://www.amazon.science/alexa-prize/proceedings/alexa-play-with-robot-introducing-the-first-alexa-prize-simbot-challenge-on-embodied-ai.

Slota, Steven C, and Geoffrey C Bowker. 2017. "How Infrastructures Matter." In *The Handbook of Science and Technology Studies*, edited by Ulrike Felt, Rayvon Fouché, Clark A. Miller, and Laurel Smith-Doerr, Fourth edition, 529–54. Cambridge, Massachusetts: The MIT Press.

Soffer, Oren. 2020. "From Textual Orality to Oral Textuality: The Case of Voice Queries." *Convergence* 26 (4): 927–41. https://doi.org/10.1177/1354856519825773.

Srnicek, Nick. 2022. "Data, Compute, Labor." In *Digital Work in the Planetary Market*, edited by Mark Graham and Fabian Ferrari, 241–61. The MIT Press. https://doi.org/10.7551/mitpress/13835.001.0001.

Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7 (1): 111–34. https://doi.org/10.1287/isre.7.1.111.

Starr, Paul, and Edward P. Freeland. 2024. "'People of Color' as a Category and Identity in the United States." *Journal of Ethnic and Migration Studies* 50 (1): 47–67. https://doi.org/10.1080/1369183X.2023.2183929.

Strengers, Yolande, and Jenny Kennedy. 2020. *The smart wife: why Siri, Alexa, and other smart home devices need a feminist reboot.* Cambridge, Massachusetts: The MIT Press.

Strüver, Niklas. 2023a. "Frustration Free: How Alexa Orchestrates the Development of the Smart Home." *Digital Culture & Society* 9 (1): 99–124. https://doi.org/10.14361/dcs-2023-090106.

Strüver, Niklas. 2023b. "Wieso eigentlich Alexa? Konzeptualisierung eines Sprachassistenten als Infrastruktur und Plattform im soziotechnischen Ökosystem Amazons." *kommunikation@gesellschaft* 24 (1): 1–33. https://doi.org/10.15460/kommges.2023.24.1.1194.

Tekic, Zeljko, and Johann Füller. 2023. "Managing Innovation in the Era of AI." *Technology in Society* 73 (May):102254. https://doi.org/10.1016/j.techsoc.2023.102254.

Tiwana, Amrit. 2014. *Platform Ecosystems: Aligning Architecture, Governance, and Strategy.* Amsterdam Waltham, MA: MK.

Van der Vlist, Fernando Nathaniël. 2022. *The Platform as Ecosystem: Configurations and Dynamics of Governance and Power.* Utrecht University. https://doi.org/10.33540/1284.

Van der Vlist, Fernando Nathaniël, Anne Helmond, and Fabian Ferrari. 2024. "Big AI: Cloud Infrastructure Dependence and the Industrialisation of Artificial Intelligence." *Big Data & Society* 11 (1). https://doi.org/10.1177/20539517241232630.

Vannuccini, Simone, and Ekaterina Prytkova. 2024. "Artificial Intelligence's New Clothes? A System Technology Perspective." *Journal of Information Technology* 39 (2): 317–38. https://doi.org/10.1177/02683962231197824.

Weidmüller, Lisa, Katrin Etzrodt, and Sven Engesser. 2022. "Trustworthiness of Voice-Based Assistants: Integrating Interlocutor and Intermediary Predictors." *Publizistik* 67 (4): 625–51. https://doi.org/10.1007/s11616-022-00763-7.

Weigel, Moira. 2023. "Amazon's Trickle-Down Monopoly: Third-Party Sellers and the Transformation of Small Business." Data & Society Research Institute. https://doi.org/10.2139/ssrn.4317167.

West, Emily. 2022. *Buy Now: How Amazon Branded Convenience and Normalized Monopoly. Distribution Matters.* Cambridge: The MIT Press.

List of Authors

Fatemeh “Mahla” Alizadeh is a research associate in Human-Computer/AI-Interaction at the department of Information Systems and New Media at the University of Siegen. fatemeh.alizadeh@uni-siegen.de | ORCID: 0000-0002-5365-4695

Alexander Boden is professor for Consumer and Business Informatics and Human-Computer-Interaction at the Department of Management Sciences at Hochschule Bonn-Rhein-Sieg. alexander.boden@h-brs.de | ORCID: 0000-0002-6470-1151

Phillip Engelhardt is a PhD-Student at the University of Bonn researching the effects of anthropomorphization of humanoid robots in interaction with humans. phillipe5589@gmail.com

Stephan Habscheid is professor in German Studies / Applied Linguistics and principal investigator at the Collaborative Research Center “Media of Cooperation” at the University of Siegen. habscheid@germanistik.uni-siegen.de

Tim Hector is a linguist and works as a postdoctoral researcher at the Collaborative Research Center “Media of Cooperation” at the Unviersity of Siegen. tim.hector@uni-siegen.de | ORCID: 0000-0001-7173-7542

Dagmar Hoffmann is professor in Media Sociology and Gender Media Studies. She is principal investigator at the Collaborative Research Center “Media of Cooperation” at the University of Siegen. hoffmann@medienwissenschaft.uni-siegen.de | ORCID: 0000-0003-3778-3742

Nikolai Horn is a political advisor for iRights.Lab, a German think tank which deals with the legal, ethical and policy aspects of the digital sphere. kontakt@irights-lab.de

Christine Hrncal is a research assistant at the Professorship for Applied Linguistics at the German Department of the University of Siegen.
christine.hrncal@uni-siegen.de | ORCiD: 0009-0002-0848-345X

Markus Kienscherf is assistant professor of sociology at the John F. Kennedy Institute of North American Studies (Berlin, Germany).
markus.kienscherf@fu-berlin.de | ORCID: 0000-0001-7202-1482

Didem Leblebici is a sociolinguist and doctoral candidate at the Faculty of Social and Cultural Sciences at the European University Viadrina Frankfurt (Oder), Germany. leblebici@europa-uni.de | ORCID: 0000–0002-7104-2289

Netaya Lotze is a senior researcher (tenured) in German Linguistics and leader of the research group "Artificial Intelligence and Language" at the University of Münster. lotze@uni-muenster.de | ORCID: 0000-0003-0738-7570

Christoph Lutz is an interdisciplinary Internet and technology researcher and a Professor at the Department of Communication and Culture, BI Norwegian Business School (Oslo, Norway). christoph.lutz@bi.no | ORCID: 0000-0003-4 389-6006

Alexander Martin is a student assistant in the fields of media sociology and digital methods at the University of Siegen. amartin42@gmx.de | ORCID: 0009-0004-7256-5382

Anouk Mols is associate professor in media literacy and digital skills at HU University of Applied Sciences Utrecht (the Netherlands). anouk.mols@hu.nl | ORCID: 0000-0003-0355-9849

Gemma Newlands is an organizational sociologist and a Departmental Research Lecturer in AI & Work at the Oxford Internet Institute, University of Oxford (UK). gemma.newlands@oii.ox.ac.uk | ORCID: 0000-0003-0851-384X

Carsten Ochs has been researching the sociocultural impact of digitization for more than 20 years; he currently works as a postdoctoral researcher at the Sociological Theory department at University of Kassel. carsten.ochs@uni-kassel.de | ORCID: 0000-0001-8145-0748

Florence Oloff is a linguist and works as a professor in the Pragmatics Department at the Leibniz-Institute for the German Language and at the University of Mannheim. oloff@ids-mannheim.de | ORCID: 0000-0003-0926-7660

Dominik Pins is a usability engineer, sociologist and research associate at the department of Human-Centered Engineering and Design at Fraunhofer-Institute for Applied Information Technology. Dominik.Pins@fit.fraunhofer.de | ORCID 0000-0001-9601-8122

Julia Schmitz is a master student at the University of Bonn and works on her master thesis on the relationship of the elderly with technology.

Gunnar Stevens is professor for IT-Security and Consumer Informatics at the department of Economics and Information Systems at the University of Siegen. gunnar.stevens@uni-siegen.de | ORCID 0000-0002-7785-5061

Niklas Strüver a PhD candidate at the Collaborative Research Center 1187 "Media of Cooperation" at the University of Siegen. niklas.struever@uni-siegen.de | ORCID: 0000-0002-7829-0963

Caja Thimm is professor for Media Studies and Intermediality at the University of Bonn, Germany, a member of the NRW Academy of Science and the spokesperson for the "Autonomy and Autonomous Systems Research Group". Thimm@uni-bonn.de | ORCID: 0009-0004-3167-7038

Jasper Vermeulen is a PhD Candidate at the Queensland University of Technology (Brisbane, Australia). jasper.vermeulen@qut.edu.au | ORCID: 0000-0002-7513-5909

David Waldecker is a former member of Collaborative Research Center "Media of Cooperation" and an academic librarian in training at Universitäts- und Landesbibliothek Darmstadt (Germany). david.waldecker@tu-darmstadt.de | ORCID: 0000-0001-9766-2788

Sebastian Zilles was a tech lead at openINC until early 2024 and since then is working as a senior software engineer at adesso. Sebastian.zilles@adesso.de