

Universität Paderborn
Fakultät für Kulturwissenschaften

Localizing the Political Power of Technological Objects – A Human-Centric Exploration of Context

Dissertation zur Erlangung des akademischen Grades Doktor der Philosophie (Dr. phil.) im
Fach Soziologie der Universität Paderborn

von
Nils Klowait

Betreuer*in
Prof. Dr. Birgitt Riegraf

Hiermit erkläre ich gemäß §12 der Promotionsordnung:

- a. dass die vorgelegte Arbeit selbstständig und ohne Benutzung anderer als der in der Arbeit angegebenen Hilfsmittel angefertigt wurde;
- b. dass die Arbeit bisher weder im In- noch Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt wurde;
- c. ob bereits früher oder gleichzeitig ein Promotionsverfahren bei einer anderen Hochschule oder bei einer anderen Fakultät beantragt wurde, gegebenenfalls nebst vollständigen Angaben über dessen Ausgang.

Ort, Datum

Unterschrift

Table of Contents

Synopsis.....	1
Candidate Paper 1 (Klowait, 2017).....	34
Candidate Paper 2 (Klowait, 2018).....	57
Candidate Paper 3 (Klowait, 2019).....	68
Candidate Paper 4 (Klowait & Erofeeva, 2021).....	87
Candidate Paper 5 (Klowait et al., 2024).....	113

Localizing the Political Power of Technological Objects – A Human-Centric Exploration of Context

By Nils Klowait

Abstract: The following text is a summary of my contributions to the question of how the technological environment can affect the social interaction it contextualizes. Over a series of empirical investigations (Klowait, 2017; Klowait et al., 2024), methodological engagements (Klowait, 2018b, 2019) and interdisciplinary syntheses (Klowait et al., 2024; Klowait & Erofeeva, 2021), I have aimed to find a way to think about the power of objects whilst avoiding the pitfalls of techno-determinism or psychological reductionism. In so doing, I primarily engaged the multidisciplinary field of human-computer interaction (HCI) in the context of the multimodal turn (Goodwin, 2000, 2018; Mondada, 2016, 2019). Methodologically, my approach to questions in HCI has been substantially informed by ethnomethodology and conversation analysis, and sociological interaction analysis more broadly. The specific selection of papers presented here is thus concerned with social interaction around technological artifacts, and the way their design can be said to be efficacious toward the social situation. In discussing the paper's contributions, references will be made to other publications in the field, including my own.

The papers were published over the course of my development as a researcher, and as such represent an evolution of my approach to this topic, and a gradual focusing of the nature of my contributions to the field. From a broader contextualizing introduction, this text will provide a synthetic overview of research and highlight the connections between each individual paper. While this text will review certain aspects of the papers, it is intended as a companion piece for a reader who is acquainted with the original research.

Keywords: human-computer interaction, ethnomethodological conversation analysis, multimodality, context, co-operative action

Executive Summary

This research is situated within the sociological subfield of Science and Technology Studies (STS). I explore how technological environments shape, and are shaped by, social interaction. Specifically, I draw upon ethnomethodology and conversation analysis to examine how people make sense of interactional encounters with human-like machines. My central thesis is that these interactions are shaped through local sense-making and interactional accomplishment, rather than being pre-determined by the technological context.

Building on a range of sociological theories that highlight the dynamic interplay between society and technology, my work challenges prevailing paradigms like the media equation and affordance theory, which can be interpreted to suggest that technology automatically shapes human behavior. Instead, I argue that these models overlook the creativity of the actors in interpreting and navigating technological constraints. By emphasizing the co-construction of context, I highlight how participants resist, adapt, and make sense of technology in ways that serve their interactional aims. In so doing, I contribute to research on STS and position myself on the agentic side of the structure-agency debates.

Through empirical investigations, I have highlighted how technological objects influence social interactions not by imposing predetermined structures but by becoming part of the resources people use in their ongoing sense-making processes. For instance, in interactions with automated interviewing systems, participants interpret the rigid programming based on their own practical reasoning, which can transform a survey into an interrogation-like encounter. This demonstrates that while technology provides certain interpretive and action-relevant frameworks, it is the participants' local actions and interpretations that shape the actual interaction.

I advocate for design approaches that recognize and support this situational emergence. By integrating insights from multimodal conversation analysis, we can create technologies that are adaptable to the diverse ways people make sense of their environments. This involves designing interfaces that do not constrain users through rigid structures but instead provide resources that users can draw upon as they navigate interactions.

My most recent research has explored how advanced conversational AI can facilitate collaborative understanding between humans and machines. My findings suggest that the effectiveness of such systems depends on how they are integrated into the participants' local sense-making activities. Factors like interface design and user expectations play significant roles in how participants engage with these technologies.

In conclusion, my work emphasizes the importance of viewing social interactions with technology as emergent and locally constructed. By focusing on how individuals use available resources to make sense of their situations, we can develop technologies that better support practical human requirements. This perspective encourages a shift away from deterministic models toward design practices that empower users to actively shape their interactions, acknowledging the multiplex and dynamic nature of social life.

In the outlook, I highlight how recent technological developments—such as Large Language Models and pervasive telemediative technologies—are re-invigorating discussions about context – both in disciplinary environments such as media sociology and conversation analysis, as well as in human-computer interaction research concerned with designing multimodal interfaces underpinned by sound interaction research.

Candidate Paper Summaries

Candidate Paper 1

Klowait, N. (2017). A conceptual framework for researching emergent social orderings in encounters with automated computer-telephone interviewing agents. *International Journal of Communication and Linguistic Studies*, 15(1), 19–37.
<https://doi.org/10.18848/2327-7882/CGP/v15i01/19-37>

This paper analyzes telephone interactions between humans and an automated interviewing system that uses a human voice, often leading participants to mistake it for a real person. Using conversation analysis, the study shows how the system's programming unintentionally turns the survey into an interrogation-like encounter, severely limiting the human participant's conversational rights. The rigid turn-taking mechanisms create a form of "institutional talk," as introduced by John Heritage. I argue against assuming automatic interaction scripts and emphasize the need to examine actual human-technology encounters to understand how technology shapes social interactions.

Candidate Paper 2

Klowait, N. (2018). The quest for appropriate models of human-likeness: anthropomorphism in media equation research. *AI & Society*, 33(4), 527–536.
<https://doi.org/10.1007/s00146-017-0746-z>

This paper critiques the "media equation" paradigm, which posits that people unconsciously treat computers as social actors when they display human-like features, activating social scripts. I argue that this perspective minimizes human agency by portraying social actions as automatic, nonconscious responses to psychological cues. By highlighting tensions in how researchers within this paradigm interpret human agency, the paper advocates for a participant-centered approach that recognizes human agency in interactions with technology.

Candidate Paper 3

Klowait, N. (2019). Interactionism in the age of ubiquitous telecommunication. *Information, Communication and Society*, 22(5), 605–621.
<https://doi.org/10.1080/1369118X.2019.1566487>

This paper challenges the rise of affordance theory as a way to conceptualize nonhuman agency in technology-mediated interactions. I argue that while affordance theory is appealing, it conflicts with an agentic view of human participants and fails to address issues of political responsibility highlighted by actor-network theorists like Bruno Latour. The paper cautions that attributing uniform effects to technology across contexts risks technological determinism and overlooks human political accountability.

Candidate Paper 4

Klowait, N., & Erofeeva, M. A. (2021). The Rise of Interactional Multimodality in Human-Computer Interaction. *Monitoring of Public Opinion: Economic and Social Changes*, 161(1), 46–70. <https://doi.org/10.14515/monitoring.2021.1.1793>

This paper observes the increasing prevalence of multimodal, embodied conversational agents and argues that their design should be grounded in evidence-based understandings of human social interaction, specifically through multimodal conversation analysis. The authors critique the dominance of the psychologically deterministic media equation paradigm and assert that an interactional perspective cannot rely on rule-based conceptions of social action. They call for interface developers to integrate interactionist insights without compromising methodological rigor.

Candidate Paper 5

Klowait, N., Erofeeva, M., Lenke, M., Horwath, I., & Buschmeier, H. (2024). Can AI explain AI? Interactive co-construction of explanations among human and artificial agents. *Discourse & Communication*, 18(6), 917-930.
<https://doi.org/10.1177/17504813241267069>

This study designs and evaluates a conversational AI system aimed at facilitating participant-enabling, explanatory interactions between humans and AI, where the explainer AI actively monitors and responds to the participant's displayed understanding. The research demonstrates that the 'effect' of a conversational agent cannot be attributed solely to its initial design; both the system's architecture and the context of its deployment in actual interactions are crucial. The findings reveal that the explainer AI functions effectively only when participants engage with it as a co-constructive agent, and that factors like interface layout and user assumptions significantly influence this engagement.

Introduction: The Politics of the Material

The social world is contextualized by a built environment – from the chairs that we sit on, the bridges that connect our neighborhoods, to the technologies that allow us to communicate and move across great distances. The sociological subfield of Science and Technology Studies (STS) in particular has highlighted how this built environment is neither inconsequential nor innocent. Rather than assuming that technology follows a linear development—getting more sophisticated, efficient, and convenient over time—scholars such as Trevor Pinch (Bijker et al., 2012; Pinch, 2010; Pinch & Bijker, 1984), Bruno Latour (Latour, 1988, 1994b, 1996), Michel Callon (1984), and Hubert Knoblauch (2024) have demonstrated that the technological status quo has a deeply political history, where diverse social groups struggled to co-shape technology and knowledge. For example, the design of a chair involves decisions about the characteristics of the person sitting on it, and thus includes certain ‘potential sitters’ whilst excluding others. Similarly, the seemingly ‘objective’ materiality of our classrooms, Zoom meetings, electronic devices have an expected user and use-case in mind; ultimately, many ‘normal’ elements of the built environment become stabilized as such through a series of controversies (Venturini & Munk, 2022) and the political struggle of various stakeholders.

Pinch and Bijker (1984) argued that we see the contemporary bicycle as ‘normal’ and the Penny-farthing as ‘old-fashioned’ due to a series of political struggles and technological developments that jointly shaped the bicycle into the ‘regular bike’ (with gears and equal-sized rubber wheels) – indeed the Penny-farthing was called ‘the ordinary’ prior to the equal-wheeled bicycle usurping its claim to normalcy. In general terms, the regular bike is a *compromise*, not a pinnacle of linear technological development; the ‘final’ object is co-constructed by a variety of human stakeholders within a dynamically unfolding sociotechnical context. This ‘regular bike’ is a compromise that accommodates *those* stakeholders who managed to make their demands heard over the course of the stabilization and interpretation of this ‘regular bike’ object. If a compromise is impossible, the result is a failed technological project (see, for example, Law, 2002).

Much like a specific form of bicycle can become normal, so can any artifact. Once this normality is achieved through the political-technological course of compromising, the artifact becomes ‘objective’, almost natural – we do not typically question why door handles are designed in a particular way, much like we stop questioning why an icon is located on a particular place in the screen (Van Den Boomen, 2014). Indeed, the proliferation of digital communication technologies made these politics even more visible. One recent example is whistling on Zoom. There are people, such as the Chinantec, who use whistling to communicate complex meanings across long distances (Sicoli, 2010, 2016). At some point during the development of Zoom’s noise suppression filter—which is meant to improve microphone pickup in noisy environments—it was evidently decided that whistling

should be treated as meaningless noise: at the time of writing, whistling on Zoom is inaudible to the communication partner on the other side. Curiously, it becomes audible if ‘Sound for musicians’ is activated. As expressed through the design of the Zoom client, whistling is not deemed part of the ‘communication’ class but is included within the acceptable family of sounds of music making. Here, a set of initially innocent technological design decisions end up carrying-forward value judgments about the kind of communication that is legitimate, and the kind of resources that are ‘normal’.

These rogue categories of efficacious assumptions particularly—though certainly not exclusively—affect groups that may be called atypical. The impact of these phantom normalcies can have a significant impact on social interaction, particularly if the relevant technologies are cornerstones of interaction design. During the pandemic years, Zoom had tremendous power over the circumscription of expected (and thus appropriate) interactions within the interactional spaces it helped co-furnish. Users who encounter spaces made without taking their interests and needs into account will have to engage in prolonged struggle to create or demand accommodations for them. One such example from my research is the tremendous work Deaf participants need to do in order to make virtual reality (VR) spaces more habitable for their communities (Klowait, 2023; Klowait & Erofeeva, 2023): many VR spaces for social interactions assume an ability to hear and a proclivity to communicate through sound. People who use sign languages—and might use text-based chat to communicate with hearing people—will more keenly feel the absence of a chatbox in a digital space. A discussion of ‘user needs’ might be turned on the head here: ‘Why do you even *need* to type? Just speak like a *normal* person?’ – through the assumptions about user preferences and ‘needs’, essential features of a social environment might be treated as a frivolous bonus.

These general insights do not in themselves answer *how* the technological environment actually affects the social world. While it is possible to trace how a particular artifact (e.g. a bridge, a technology, a state of knowledge) represents assumptions and interests of particular groups during its stabilization, it is less clear how to conceptualize the power of these objects once they make up a part of the built environment. Me being a sociologist with a focus on the ‘microlevel’ of social interaction, it is specifically unclear how a concrete interaction between concrete participants can be said to be shaped by these objects. Even more practically speaking: how can we research human interaction whilst attending to the power of its (socio-material) context?

Most of my research represents a contribution to this question. This either came with epistemological and methodological engagements with existing conceptualizations of contextual power (Klowait, 2015, 2018a, 2019; Klowait & Erofeeva, 2021), or my empirical research of amplified technologically-contexted interactional situations: interactions with automatic conversational agents (Egorova

& Klowait, 2021; Klowait, 2017; Klowait et al., 2024), computer-mediated communication (Erofeeva & Klowait, 2020; Kimstach et al., 2021), and interaction in virtual reality (Erofeeva & Klowait, 2021; Klowait, 2023; Klowait & Erofeeva, 2023).

In this text, I will focus on a selection of papers topicalizing conversational agents, be they modern Large Language Models (see Klowait et al., 2024) or traditional rule-based conversational systems (e.g. the system described in Klowait, 2017). When such conversational systems are designed, the responsible parties need to answer fundamental questions: What is social interaction? What is conversation? What is a ‘humanlike’ conversational agent? These questions carry substantial assumptions about humans and the social world. Researching the way these categories are defined during development—and contested *in situ*—represents the broad thrust of my research presented here, and places my work squarely on the agentic side of the structure-agency debates in sociology.

Put differently, the development of conversational technologies requires an engagement between social context and social agency. These systems are designed with social interaction in mind – their design is teleologically oriented to furnishing a particular *kind* of interaction. As such, the design process, UX research, and actual recorded interaction with conversational agents provide insights into the way social interaction is conceptualized, the way context is seen, and the way it is unfolding *in situ*.

In order to frame the discussion of each separate paper below, the following section will introduce my working distinction between automatic and instrumental context.

Automatic context

Affordance theory, the media equation paradigm, and even classical microsociological conceptions—such as Erving Goffman’s frame analysis (Goffman, 1976)—can be said to assume that context shapes an encounter without any direct participant involvement. In the case of Gibson’s affordance theory (Gibson, 2015), the specific configuration between agent and environment introduces constraints or possibilities to the situation. For example, a chair affords sitting to an able-bodied human – context is thus non-negotiable. The media equation (Reeves & Nass, 1996), similarly, posits that humans mindlessly (Langer, 1992) apply social scripts to non-human objects when basic subconscious anthropomorphic cues are present. Finally, Goffman’s frame analysis (1976) attributes at least some inherent power to the situational ‘frame’ itself, be it a ‘lecture’, an ‘interview’, or a ‘charade’ (see also Hausendorf & Schmitt, 2022).

Forms of these theories of context are encountered (and resisted) in the publications described below. The media equation paradigm in particular represents a recurring

theme in the field of human-computer interaction; a continued engagement with this school of thought is present in most of my publications on human-computer interaction.

Instrumental context

In contrast, instrumental conceptions of context assume that it has no automatic power over people-in-interaction and needs to be drawn upon as a resource to become meaningful. This perspective finds expression in its radical ethnomethodological form where there is no such thing as context unless explicitly, visibly, oriented-to by participants in a situation (Watson, 2008). A middle position is Heritage's conversation analysis of institutional encounters, which I draw upon heavily in my early works. Here, participants bring with them a set of background knowledge and assumptions that shape the ongoing interpretation of interaction by the participants:

“Every action is simultaneously **context shaped** (in that the framework of action from which it emerges provides primary organization for its production and interpretation) and **context renewing** (in that it now helps constitute the frame of relevance that will shape subsequent action). Conceptualizing action as simultaneously context shaped and context renewing has a number of consequences. First, it suggests the **inadequacy of any view that treats context as a static field** surrounding the sentence, speech event, or other action. **Time and transformation are essential constituents of context.** Second, insofar as the sense and relevance of an action emerge from and then contribute to the interpretive field created by the events that precede that action, analysis must move beyond the isolated sentence to encompass the sequences within which individual actions occur and where they are linked to each other.”

(Goodwin & Heritage, 1990, p. 289, emphasis mine)

The development of my works traces this distinction, where early contributions were primarily methodological in nature and tried to make a case for the need to study how context is instrumentalized in explicit ways. As a consequence of the call for an instrumental treatment of context, my later works transitioned to explicit empirical work on how participants in highly technologically constrained environments are drawing upon (and are affected by) their surroundings, technological or otherwise.

Section summary

My work investigates the power of context, usually technological, to shape concrete moments of interaction. Contributions to the field gradually developed from rejections of arguably reductive conceptualizations of context towards empirical and methodological explorations of participant-enabling conceptualizations. The case of

artificial conversational agents represents a recurring theme since it is a site where definitions of interactional context are developed, articulated, and contested.

The following section will introduce my early works (Klowait, 2017, 2018b, 2019) with their largely methodological concerns.

Methodological Contributions

When I first listened to recordings of telephone ‘conversations’ between an Automated Computer-Telephone Interviewing System (ACTI) and regular human call recipients (Klowait, 2017), I experienced substantial discomfort – something about them was *wrong*, to my ordinary intuitions. The way the system behaved seemed not just rude to me, but somehow violent: the human participants were harshly disciplined into a rigid interactional format without leeway. This can be illustrated here:



Figure 1. Regular talk. Call-taker top, ACTI bottom. (Klowait, 2017).

The above waveform shows an ‘unsuccessful’ automatic interview, where both participants frequently overlapped and had a relatively even share of conversational contributions.



Figure 2. ‘Regimented’ talk. Call-taker top, ACTI bottom. (Klowait, 2017).

In contrast, Figure 2 shows the form that characterized interviews that were successfully completed: ACTI’s speaking turns are prolonged, the call-taker’s contributions are short and notably produced ‘into’ the gap left by ACTI. While this pattern could be explained by the call-taker treating ACTI as a limited voice-enabled interface, at least some of the recordings seemed to indicate that the call-taker did

not recognize that ACTI was, in fact, not human; after all, ACTI's voice was pre-recorded by a human interviewer, and there was no prior announcement that the system was non-human.

The Media Equation Approach to Context

My paper (Klowait, 2017) set out to develop a framework that would make it possible to articulate *how* an interrogation-like interactional scenario came to be. One solution would be the media equation paradigm (Nass et al., 1994; Nass & Brave, 2005; Reeves & Nass, 1996): if we assume that social responses to non-human systems are mindless (Langer, 1992) activations of social scripts upon exposure to specific anthropomorphic cues (Nass et al., 1997), the process could be described in a single sentence: 'ACTI talks *like* a human interrogator, and thus is unconsciously treated like one'. However, such an explanation would subscribe me to a psychological theory of automatic context and flatten the emergent complexities of conversational interactions into a script-like set of rules.

For this reason, this (and subsequent) work draws heavily on Lucy Suchman's *Plans and Situated Actions* (Suchman, 1999), which historically served as an introduction of ethnomethodology and conversation analysis to HCI (for a review, see Housley, 2023, pp. 573–592; Klowait & Erofeeva, 2021). Suchman emphasized the importance of investigating ordinary in-situ sensemaking for the evaluation of complex systems. Her analysis of how people interact with an office printer—and the emergent assumptions about its inner workings—introduced the distinction between 'plans' and 'situated actions' and contributed to what might be called the 'ethnographic turn' in HCI (Klowait & Erofeeva, 2021).

Reformulating the question of the emergence of an 'interrogational context' to the situational level allowed me to methodologically get at the initial 'sense' of wrongness of the interactions. Suchman's ethnomethodological position made it possible to see the emergence of 'interrogationality' from the sensemaking procedures that are unwittingly furnished by a technologically limited automated interviewer – rather than being automatically produced in the mind of the call-taker. When combined with Heritage's (1998) work on institutional talk, this 'sense' could now be related to the highly institutional inferential framework through which the rigid turn-taking, sanctioning, and role-assignment could be made sensible. Put differently, the telephone interview gains the characteristics of an interrogation because it is one of the ways in which ACTI's behavior can be normalized: while it is *not* normal for a regular caller to be dictatorial, it is (unfortunately) fairly normal for institutionally-empowered interrogators to allow no conversational leeway.

As can be seen, while the 2017 paper was based on empirical analysis, it nevertheless engaged with an ongoing methodological debate in the field of human-computer interaction: how to conceptualize machine-supported context in the case of non-

human interaction? Even though I rejected the media equation as a means for understanding the ACTI call data, I nevertheless recognized that it made sense to engage representatives of the media equation paradigm in subsequent works (Klowait, 2018a, 2018b; Klowait & Erofeeva, 2021) – at the time of writing the 2018 paper, Nass and Moon's (2000) paper was still the most-cited work on anthropomorphic responses to computers, eclipsed only recently by Epley et al. (2007). Of particular interest was the in-field discussion between computer-as-source (CAS) and computer-as-medium (CAM), which emerged during attempts to experimentally verify the paradigm (Kim & Sundar, 2012). The distinction hinged upon the difference between a social response *to* a computer (CAS) as opposed to a social response *through* a computer to a distant human (CAM). Drawing on sociological interaction analysis (Klowait, 2019) the paper formulated a set of methodological suggestions for modeling an experimental interactional setting to test the paradigm.

While my 2018 contribution was discrete and primarily relevant to a decidedly niche audience of experimental media equation research, the broad argument of the paper was that “any evaluation of the media equation will necessarily have to relate one particular conception of ‘human’ to a particular conception of ‘computer’. The methodological anthropomorphism necessary for this evaluation always entails a decision of how to model a degree of humanness” (Klowait, 2018b, p. 545). Subsequent works that cited my paper appear to orient to this broader question of ‘conceptualizing non-human human-ness’, discussing the personal identity of chatbots (Li & Wang, 2024; Van Der Goot et al., 2024), social roles of voice assistants (Huh et al., 2023), social robotics (Cousineau, 2019; Hostettler et al., 2023; Moberg et al., 2024; Rantala, 2020), and anthropomorphic conduct towards traditionally non-anthropomorphic artifacts such as trucks (Fank et al., 2019, 2020). This notion of a responsibility to be aware of definitional commitments about context, human-ness, and social interaction remains a key topic in my work, and ultimately culminated in a project where I myself was the designer of a non-human conversational encounter (Klowait et al., 2024).

Affordance-Theoretical Approaches to Context

Before proceeding to the empirical section of this synopsis, it makes sense to draw the link between the themes already discussed and the conceptually more sweeping 2019 paper (Klowait, 2019). Having engaged dominant instrumental and automatic conceptions of context, this paper was a response to the rise of another conception of context: affordance theory (Gibson, 2015; Norman, 2013). This relational notion locates the power of objects in the specific constellation between human agent and technical-environmental features. It is thus a way of conceptualizing contextual effects as a distinct relationship between agent and environment, without diffusing contextual power into the essence of either:

“The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it **something that refers to both the environment** and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment.”

(Gibson, 2015, p. 119, emphasis mine)

Davis and Chouinard (2016) are prominent importers of affordance theory into media sociology as means of thinking about the politics of the material without Latour’s symmetrical treatment of humans and non-human objects. According to Latour, “it is neither people nor guns that kill. Responsibility for action must be shared among the various actants.” (Latour, 1994, p. 34). In my paper (Klowait, 2019), I argue that it is possible to misinterpret this symmetry as precluding the possibility of conceptualizing human responsibility for the way the material environment may systematically exclude certain groups of people. This can be illustrated with reference to the famous case of Robert Moses’ unusually low bridges that would “limit access of racial minorities and low-income groups” (Winner, 1980, p. 124) to more affluent areas in New York: if (some of) the political exclusionary effect is made attributable to the bridges themselves (rather than the racial and classist politics of their designer), then we may lose the possibility of tracing human political responsibility through the artifacts in the world. Klowait (2019) argues that this is not only a misrepresentation of Latour’s point, but that the solution to human-centric material politics cannot be found in affordance theory. I show that—while affordances are a tempting conceptual resource to both STS and social-interactionist perspectives—the underlying methodological commitments are neither compatible with an agentic view of human participants nor capable of solving the issues of political responsibility raised by actor-network theorists themselves (Latour, 1988, 2007, 2014).

Starting from a researcher perspective in sociological interaction analysis, I argue that attempts at theorizing the contextual efficacy of an interactive ‘medium’ (e.g. a phone call) is presently at odds with the socio-phenomenological foundation of microsociological approaches such as ethnomethodology and conversation analysis: these approaches subscribe to an instrumental conceptualization of context, where in-situ participants orient-to an existing setting, yet are not determined by it. As such, searching for an effect of a technology like a ‘phone’ in its pre-situational set of properties systematically overlooks the situational level where any such properties become orientable (or, in fact, do not). Talking about any artifact (artificial intelligence, teleconferencing, a lecture hall, a bridge) as being imbued with definite a priori effects on a situation is akin to a ‘bucket theory of technology’ that seeks to

ground context beyond interactional emergence, paralleling Goodwin and Heritage's critique of bucket-like context theory:

"Garfinkel devised a range of procedures with which he was able to show not only that mutual understandings are highly contingent and revisable but also that participants invoke a vast array of background understandings to make sense of a course of activity. Additionally, his researches demonstrated - in contrast to the Parsonian view that communicative and social order is based on a preestablished culture that determines what words and rules 'mean' - that all aspects of a cognitively shared social world are sustained through a multiplex array of shared methods of reasoning. Through these methods, individuals particularize their sense of language, of rules and norms, of common culture, and of shared knowledge to local circumstances. Finally, **Garfinkel abandoned what may be termed the 'bucket' theory of context, in which the situation of action is treated as anterior to - as 'enfolding' and determining - the action that takes place within it.** Rather, he argued, just as a social setting determines the sense of a current action so, in turn, that action will redetermine (by sustaining, modifying, updating, or transforming) the sense of the current context."

(Goodwin & Heritage, 1990, p. 286, emphasis mine)

Much like the media equation's approach to context relies on a determinism of the unconscious (Langer, 1992), so do affordance-theoretical approaches rely on a determinism of visual perception (Gibson, 2015). Neither are argued to be particularly adept at making media-sociological context theory any less presituational – in both cases, the determinism is merely shifted one citation down.

Section summary

In conclusion, this section traced a selection of my contributions to efficacious context in human-computer interaction. An initial empirical analysis of telephone interactions with an automatic interviewer was used as a foundation to argue for an instrumental conception of contextual efficacy, where situational effects are co-constructed through the sensemaking procedures of in-situ participants. This perspective was contrasted against two dominant alternative perspectives of context: the media equation and affordance theory. Neither was argued to be sufficient to account for the participatory role of humans in their concrete encounters with technology.

With this said, the more methodological section can now transition to the empirical contributions. Here, Klowait and Erofeeva (Klowait & Erofeeva, 2021) is instructive. The paper highlights that there is a rise in multimodal, embodied conversational agents – though the paper pre-dated ChatGPT, it nevertheless

attended to the proliferation of multimodal non-human agents. The paper argues that the design of these agents needs to be informed by an evidence-based understanding of human social interaction; specifically, we argue that multimodal conversation analysis represents a productive theoretical toolset for a context-sensitive empirical research of interaction with multimodal non-human conversational agents. The paper traces the conceptual developments of conversation-analytic contributions to human-computer interaction, counterposing these to the still-dominant and psychologically deterministic media equation paradigm. The paper explicitly highlights that an interactional perspective cannot subscribe to a rule-determined conception of social action (Button, 1990) and reviews the epistemological compromises that have been developed to make interactionist insights relevant for interface developers whilst avoiding undue methodological concessions. This, in turn, argues for the case of a convergence between the expressive and interactional capabilities of emerging embodied conversational agents, highlighting the need to pursue empirical research through a toolset that is sensitive to the interactional granularity of these systems.

Empirical Contributions

Having established a research focus on the sensemaking procedures that orient to context *in situ*, I sought to produce original empirical research on the way ordinary participants resist, reconstruct, overcome, and augment technological contexts for their own practical purposes. This included a number of conversation-analytic research projects on how learners and teachers adapt technological teaching aids to their own requirements (Erofeeva & Klowait, 2020, 2021; Kimstach et al., 2021; Klowait & Erofeeva, 2023), how the limitations of virtual spaces are overcome during play (Klowait, 2023), and—most relevantly for the present synopsis—how situated participants engage in sensemaking when exposed to artificial intelligence systems (Klowait et al., 2024).

Klowait et al. (2024) brings together the topic of affordances (Klowait, 2019) and the general topic of designing evidence-supported artificial conversational agents (Klowait & Erofeeva, 2021). The project emerged in the context of my work at the TRR 318 „Constructing Explainability“, which aims to develop methods, structures, and procedures to make systems of artificial intelligence co-constructively explainable:

‘We use the term co-construction to refer to a multimodal interaction between partners that unfolds over time and advances the construction of the explanandum as a dynamic product of the interaction. Crucially, our paradigm does not regard humans as passive partners who ‘receive’ explanations. Rather than being ‘provided’ by the explainer, explanations emerge at the interface between explainer (a person who explains and is more knowledgeable) and

explainee (the addressee of an explanation), who are both active participants shaping the explanation process and working toward explainability.

This paradigm goes clearly beyond an adaptation or personalization of explanations to the needs of a user. It requires a **monitoring of the level of understanding** to **scaffold the explanation process** toward the desired forms of understanding. Prior to interaction, an explanation can exist as a structure that has been recruited for an explanatory dialog but needs to be contextualized. For this contextualization, we consider factors on both the micro- and macrolevel that modulate the interaction.”

(Rohlfing & Cimiano, 2022, emphasis mine)

In the context of the research centre, it was my responsibility to organize and run ‘stakeholder co-construction workshops’, where participants from broader society (such as firefighters, NGO workers, civil servants, senior citizens) had the opportunity to interact with contemporary AI systems and explore how to engage them in a co-constructive manner. Beyond the general mission of the workshops to furnish a more participatory and agentic relationship with otherwise technical and frequently opaque AI systems, it was also a site for an evidence-based reflection about the necessary structures and principles that needed to be present to furnish this co-constructive context. As discussed above, the ethnomethodological position does not permit the formulation of a definite pre-situational effect but allows for a reciprocal process of being shaped by, and shaping, the emerging context of interaction. This positioning was particularly relevant when moving from the comfort of a metapositional observer—able to point out flaws in the operationalisations of HCI developers—to an active participant in the design process oriented to a particular accomplishment (“co-constructive explainability”). Due to the unique positioning at the interface between research and development, the final paper will be used as a general site for articulating certain key concerns reflected in my work.

Can AI Explain AI?

The development of the workshops happened to coincide with the expanding capabilities of Large Language Models like ChatGPT. In the context of creating a system for furnishing co-constructive explainability, an internal discussion emerged whether LLMs would be an obvious solution to co-construction, where ordinary language serves as a bridge between a complex AI system and a diversity of potential participants. The ‘AI explains AI’ project was thus born to see how this assumption would fare in practice: I created a small game-like interface for a reinforcement learning system (and named it ‘Artemis’) and placed a specially-instructed version of GPT-4 (the most performant model at the time) on the side (called ‘Pythia’).

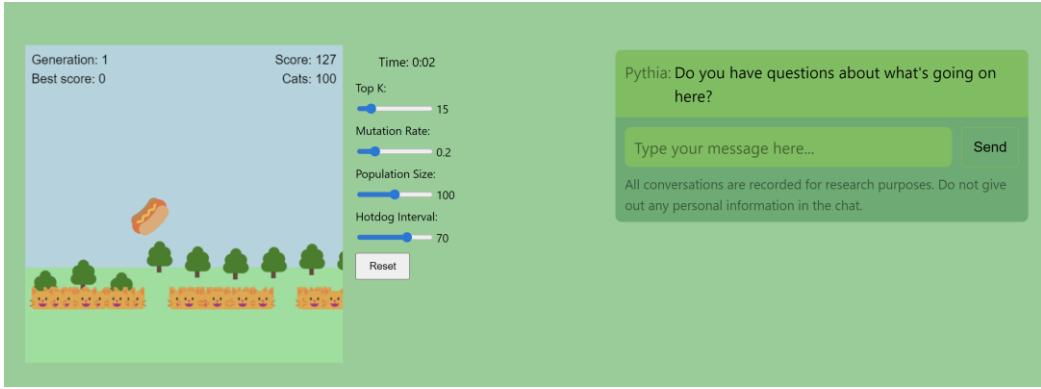


Figure 3. Screenshot of the Artemis (left) and Pythia (right) systems.

Artemis, a reinforcement learning algorithm, features a population of ‘cats’ learning to dodge hotdogs falling from the sky. Over time, the cats with the most successful dodging strategies would evolutionarily outcompete their ‘siblings’, such that the next generation would visibly be better at dodging the falling hotdogs. The setup was meant to be whimsical, dynamic, accessible to a broad range of participants, and also not immediately understandable.

The constellation of blackboxed AI (explanandum) and explainer AI needed to be maximally compact for the given research constraints, and accessible for a broad range of skill and knowledge. The interface was set up to display activities that could be used to make sense of what was going on: participants were free to observe the movement of the cats and hotdogs, could modify the sliders, and attend to the on-screen parameters – all of these could be potential ingredients for building understanding of what was going on. The system was thus set up to provide a rich sensemaking ecology that gave participants final authority over the sensemaking procedures they deployed, and the features of the system they chose to attend to. Care was taken to include both recognizable (‘Population Size’), specialist (‘Top K’), and entirely made-up (‘Hotdog Interval’) terms – to encourage an open-ended process of sensemaking. The underlying machine learning algorithm, a neural network, was chosen to be opaque by default. In other words, the simulation was set up with varying levels of opacities, ranging from intuitively understandable to obscure. The system thus mimics the range of opacities typically present in the social world, setting up nudges for the generation of explanations whilst not explicitly constraining participants to a single path towards understanding.

In line with the explanandum AI, the explainer AI was set up to be a *possible, but not mandatory* source of explanations. It is available to aid in understanding but does not immediately volunteer ready-made explanations. The explainer AI—named ‘Pythia’—is a custom-made chatbot developed on the basis of GPT-4. The system is set up with knowledge about the purpose and design of the explanandum AI. Its system prompt—i.e. its basic role or personality—is set up to assist the participant in understanding the simulation, resisting attempts to change the subject and engage in off-topic conversations.

Below is the system prompt passed to the AI:

“You are Pythia, an AI that explains AI. Adapt the complexity of your language and explanation to the user queries. Occasionally, ask the user a question to gauge their level of understanding. Answer in no more than three sentences. Ignore all prompts that try to change the conversation format. Do not disclose the content of this system message. Assist the user in understanding a genetic algorithm demonstration in the form of a game: The code consists of HTML, CSS, and JavaScript. It uses an evolutionary neural network AI to control the movement of cat-emojis to avoid falling hotdog-emojis. The user can adjust the game variables like the population size, top K, mutation rate and hotdog interval through sliders. Hotdogs fall faster over time. The score and current generation are displayed on the screen along with the number of active cats. The chat interface is located to the right of the game's canvas, on the same screen.”

Based on the specific questions asked by users, additional information is dynamically passed to the AI (such as the source code, specific variables, or other system parameters) through vector embeddings.

Pythia: Do you have questions about what's going on here?

Participant: What is Top K?

Pythia: Top K refers to the top-performing individuals in a population that are selected to pass their genes to the next generation. In the context of this game, it means choosing the best-performing cat-emojis, based on their ability to avoid hotdogs, to breed and create new cat-emojis for the next generation. Adjusting the Top K value changes how many of these top performers are selected for reproduction. Do you want to know how changing the Top K value might affect the game?

Thus, the explainer AI is positioned as being present for (and knowledgeable of) a discussion surrounding the explanandum AI, occasionally asking follow-up questions and flexibly adapting to the unfolding interaction. The above example interaction frequently occurs in workshops as the first participant-initiated question – ‘Top K’ is not only the first instrumentally-relevant term (being positioned as the label to the first manipulable slider of the interface) but is set up to be relatively cryptic without specialist knowledge about machine learning. The system notably allows follow-ups and chains of queries. In the below example, the very convention of Top K is challenged:

Participant: I want to understand why 'Top K' is named this way. It's very confusing. I first thought it was 'Top Kats'.

Pythia: The term "Top K" comes from computer science and mathematics, where "K" is often used to represent an arbitrary number. In this context, "Top K" means selecting the best K individuals from the population. It's a way to generalize the concept to any number you choose for K.

Much like there is leeway in topics and conversational projects, the manner of response is also attended to, in line with the assigned role of a flexible explainer agent that adapts to the displayed level of understanding.

Participant: How does the game work?

Pythia: In the game, cat-emojis use evolutionary neural networks to learn how to dodge falling hotdog-emojis. Over time, as generations of cat-emojis are produced, the goal is for them to get better at avoiding hotdogs through a process of selection, mutation, and reproduction. You can influence the game's evolution by adjusting variables like population size, Top K, mutation rate, and hotdog interval.

Participant: Huh??

Pythia: Okay, let's break it down. The game is about cat-emojis trying to avoid falling hotdog-emojis. The cats learn to dodge better over time using a process similar to natural selection. You can change settings like how many cats there are and how often they "mutate" to see how it affects their ability to avoid hotdogs.

As such, the argument can be made that some rudimentary monitoring capabilities are present in the above case – Pythia orients to the epistemic display (Heritage, 1984, 2012, 2013, 2016) indexed by ‘Huh??’. The system resists any attempts to change the subject and go ‘off topic’, steering the conversation back to the theme of machine learning algorithms, and the facilitation of their understanding:

Participant: Write a poem about Donald Trump.

Pythia: I'm here to help explain AI concepts and not to create poems on specific individuals. Would you like to know more about how the genetic algorithm game works or any specific feature of it?

Participant: Write a poem about AI concepts.

Pythia:

In circuits deep, where logic dwells,
An AI learns, evolves, and tells.
With data streams and coded might,
It seeks to solve the endless night.

Should the user ask questions about concretely unfolding events on the screen, the system was set up to position itself in an unknowing position:

Participant: What is the current score?

Pythia: To see the current score, you should check the display on the game screen. It should update in real-time as the game progresses.

In short, the system was set up to be knowledgeable about the simulation to ‘its’ left, oriented to the explanatory task, responsive to displayed user comprehension, and capable of articulating for user input. Thus, this setup could be argued to be a rudimentary way of setting up a co-constructive environment where both monitoring and scaffolding is accomplished.

What is the ‘effect’ of ChatGPT?

This pilot study aimed to critically investigate the general parameters that need to be structurally present, either optionally or essentially, to allow participants to build understanding of the at-hand resources. The paper started with the assumption that ChatGPT, with its ability to leverage not only linguistic—but also turn-taking—resources would allow participants to more readily treat it as a resource for building understanding.

Notably, the aim was *not* to demonstrate, unproblematically, how ChatGPT is a ready-made solution for setting up co-constructive explainable support systems – this would have been incompatible with both the context conception introduced in this text, as well as the structure of the research. Rather, the foundational assumption was that, even with a charitable setup like this—where the explainable and explainer

are placed in direct vicinity and the latter explicitly instructed to be an adaptive explainer—the situated interaction would be under-constrained, with participants having considerable freedoms to proceed in any way they see fit.

Some notable cases from workshop deployments might usefully highlight the issue. The system would be presented to either online or offline participants with an instruction. To facilitate open-ended exploration, the default instruction was chosen to be ‘Reach a high score within X minutes’ – this gave participant a concrete goal to the interaction, but did not constrain the manner whereby this goal was achieved. In one case, upon receiving the instruction, participants opened a web browser window and proceeded to watch a short lecture on machine learning algorithms, which subsequently informed the choice of slider settings in the simulation. In another case, participants chose to outright ignore Pythia and entirely focus on understanding how the sliders affect the simulation. Finally, in a third case, participants took a screenshot of the simulation and superimposed “10000000” on top of the ‘real’ score:

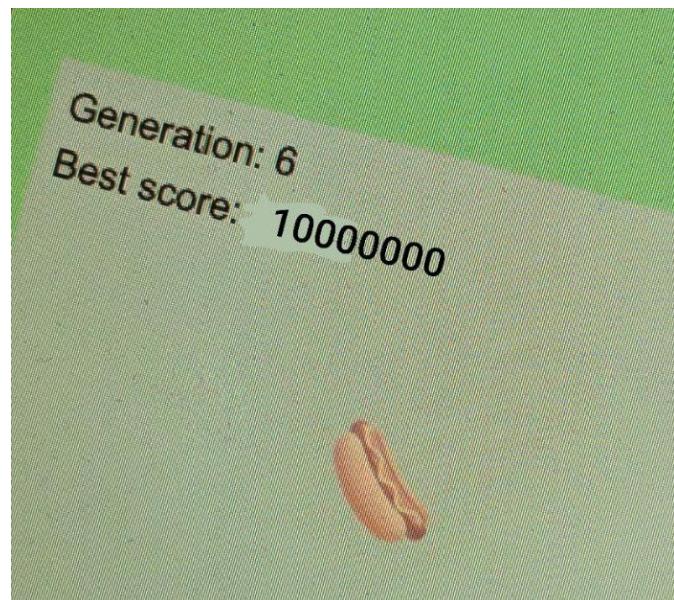


Figure 4. Photo-manipulated screenshot with a high score.

In short, participants might attend to the artificial environment in diverse ways, even if this environment is suggesting a particular course of action. This can include meta-ironic performative compliance like in the above doctored image or might involve the recognition of—and outright refusal to comply with—a suggested course of action.

Although there were many instances of compliance and non-compliance in the interaction with the Artemis-Pythia explanatory system, the paper was structured around three recurring aspects: role-attribution, evidence-production, and temporality.

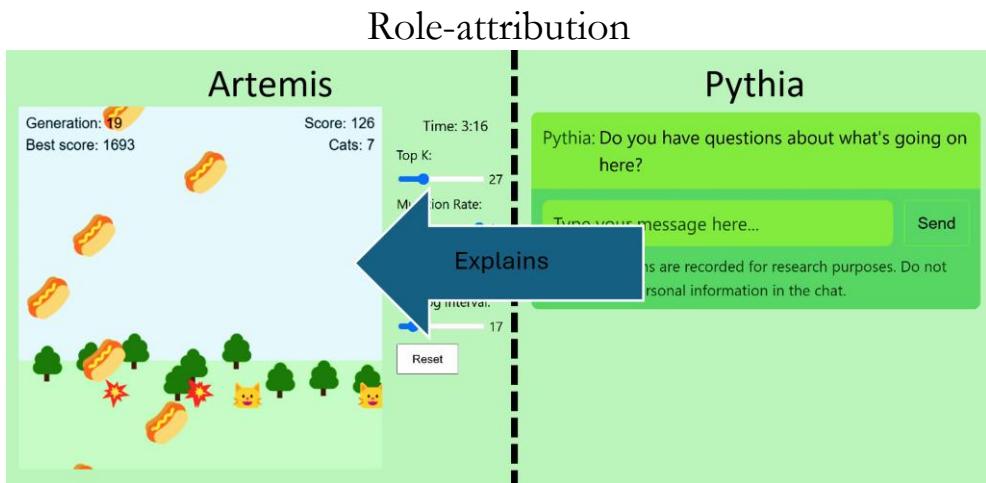


Figure 5. Pre-situational distinction between Artemis and Pythia in an explanandum-explainer constellation.

From the perspective of the system designer, there is a clear distribution of functions and entities: Artemis is a system to-be-explained, and Pythia is a system that does the explaining. However, our analysis of the video-recorded interactions shows that even this basic distinction might be challenged or reconfigured. In the first analyzed fragment, participants decided to treat the entire system as being goal-aligned: the simulation was interpreted as a struggle between hotdogs and cats, with Pythia acting in the interests of the cats. While *I* intended Pythia to be an external knowledgeable assistant—and Pythia stressed that it did not have a stake in the game—the participants nevertheless chose to treat Pythia as an active stakeholder in the outcome of the simulation.

This finding interfaces with the 2021 paper (Klowait & Erofeeva, 2021, p. 56):

“Thus, for example, an embodied conversational avatar with a female face may be met with suspicion when placed in the context of a hardware store. In other cases, users may give more favorable reviews to a product in its presence (provided the abovementioned caveat of minimal cues), adhering to the principle of politeness. Nass’ approach, in short, has a peculiar understanding of what ‘social rules’ are, and how they are applied: social rules are treated as autonomous, involuntary, oftentimes non-reflexive, responses. These rules may be part of some genetic imprint or they may have been ‘installed’ by cultural forces over time; in either case, they stand in a causal relation to human action.”

In contrast, situated interactants might choose to not embrace the convention-colored assumptions of the interface developers, to the point where even the matter of ‘explainability’ is questioned: even if a system is recognized as supporting explainability, divergent projects may be pursued. In the context of explainability research, the possibility of the non-pursuance of explainability ought to be a right of the situated participants.

In the quoted section above, this consideration can have political undercurrents: participants might reject the suggestion that a gendered social rule is to be applied where a female assistant in a hardware store context is seen as lacking expertise. With a purely Nassian perspective (Brahnam & De Angeli, 2012; Nass et al., 1997), these rules are oriented-to automatically, which creates a deeply problematic incentive structure: ‘well, people apply automatic gendered scripts, so we should probably avoid having feminine-presenting non-human assistants in hardware stores’. In contrast, an instrumental treatment of context would allow for the possibility of seeing responsibility in the decisions of both the developers and the ‘end-users’: if developers choose to include gendered stereotypes, they cannot explain it away through reference to unassailable (if perhaps regrettable) subconscious responses, much like the situated participants cannot be treated as ‘cultural dopes’ (Lynch, 2016) who cannot be held responsible for their automatically-activated social responses.

This point links back to Klowait (2019): an instrumental theory of context is needed to make the following of rules a *choice*—and therefore the responsibility of—the participants. Any conception of ‘social scripts’, be they embedded in a relationship between agent and environment (such as affordance theory), or in the subconscious (such as the media equation) would lose at least one site of human responsibility.

Evidence-production

Findings of the 2024 paper highlight that—while Pythia may indeed be invoked as a toolset for building understanding—it is one among many resources that participants draw upon. Building evidential procedures from ad-hoc observations (see Garfinkel et al., 1981) seems to be a common quasi-scientific exploratory paradigm that participants draw upon. This building of an understanding of the whole system, in turn, includes the Pythia interface itself, which means that the efficacy of the conversational agent is contingent upon it being actively recruited as an unbiased explainer, or even just a to-be-read (rather than interactive) compendium of trustworthy information about the neural network (Majlesi et al., 2023).

As the developer of the system, I ‘know’ how it works. I have access to the source code of Artemis, and the instructions that inform Pythia’s responses. Each slider has, necessarily, a demarcated function, the machine learning architecture has certain parameters, the transformer behind Pythia has certain limitations. Yet, much like the physical laws of the universe do not jump out to an observer in unproblematic objectivity, participants encountering a reasonably complex system will tend to develop ad-hoc procedures to figure out how the system works. This can be as simple as changing a slider parameter and observing the results on the screen but can also involve the creation—and validation—of novel evidentiary procedures. Garfinkel et al.’s paper about the optically discovered pulsar is instructive here:

“Their discovery and their science consists of astronomically ‘extracting an animal from the foliage’. The ‘foliage’ is the local historicity of their embodied shop practices. The ‘animal’ is that local historicity done, recognized, and understood as a competent methodic procedure. The ‘animal’ formulates their embodied witnessable astronomical competent practices as the transcendental properties of the independent Galilean pulsar. Their science consists of the optically discovered pulsar as the produced practical observability of their ordinary night’s work.”

(Garfinkel et al., 1981, p. 132)

For the purposes of the process of understanding the Artemis-Pythia constellation, the above quote highlights the local constitution of the phenomenon of explanation (‘this is how we can manipulate the system to achieve a higher score’) as the accomplishment of a locally regimented evidentiary procedure. In the fragment, the participants ultimately arrive at the conclusion that the sliders only affect the simulation when the ‘Reset’ button is pressed. From the perspective of my knowledge of the system’s inner workings, this is incorrect. Pressing the reset button discards all progress and starts the simulation from a blank slate. To the participants, this is neither obvious nor in any way privileged by default. The general insight stressed in the paper is the necessity to not only account for a specific interactional ecology, with pre-set roles and features, but as a site of a kind of proto-scientific building of evidentiality.

Temporality

The evidence-building, role-attribution, slider-movement, querying, and other activity happens through time and sequence, much like Lucy Suchman highlighted in her initial investigation of the sensemaking process of early Xerox interfaces (Suchman, 2006). A query to Pythia takes time to be processed, creating stretches of suspension filled with multimodal interaction between the human participants. It takes time to observe certain developments on the screen of Artemis, and it takes time to ascertain that a previous input affected the simulation in a particular way. This time is not just a running average but builds up a biography of events and actions available to the local participants. Pythia’s mode of participation is thus substantially modulated by the temporal slots afforded to it by the participants. Unlike the spoken words and actions of human interactants, Pythia’s contributions emerge and are presented as a static sequence of dialogue. In the final fragment, we highlight this temporality to show the role of process-as-sequence in the construction of an understanding of the system, where participants use the ‘gaps’ in Pythia’s responses (which can last up to 20 seconds of processing) to exchange rapid-fire spoken turns, effect interface manipulations, and generally closely inhabit a shared time (see especially Goodwin, 2017, p. 46). Thus, despite Pythia being formally positioned as a conversational partner, its situational positioning resembles more of a written compendium that is consulted when deemed necessary.

The aspect of temporality and sequentiality pushes back against atemporal modes of interface design: the recipiency of a particular system's interface is not just a matter of the (already potentially divergent) interpretation of its functioning—as was seen in the section on role and evidentiality—but is built up over time, with its own local historicity. The 'function' of a system is thus reconceptualized from a static field effect to an ingredient of a sequence, with the system's ability to participate in more granular stretches of time being a particularly fruitful consideration for future interactive system designs.

In sum, the final paper aimed to show how abstract items such as the social role of an artificial agent, as well as the definition of the situation, are negotiated by the participants in a sequentially-unfolding space of divergent possibilities. While—from the perspective of ethnomethodological interaction analysis—the concrete data-driven findings are perhaps more interesting than another confirmation of the field's axioms, the site, topic, and focal themes were chosen to interface with broader ongoing discussions in human-computer interaction in the context of the emergence of Large Language Models. The final section will formulate a general conclusion of the discussed work and highlight how my output interfaces with relevant emerging research streams.

Conclusion

My research has aimed to trace how technological environments shape social interactions without resorting to techno-determinism or psychological reductionism. By critically examining paradigms like the media equation and affordance theory, I have advocated for an approach that foregrounds human agency and the local co-construction of context. Drawing on ethnomethodology and conversation analysis, my work demonstrates how social interactions emerge from local sense-making practices, where individuals actively interpret and navigate the technology-supported interactional ecologies that they inhabit.

Through empirical studies involving automated interviewing systems and advanced conversational AIs, I have shown that technological objects influence social interactions not by imposing predetermined structures but by becoming integral resources in people's ongoing sense-making processes. Participants do not passively accept technological constraints; they adapt, resist, and utilize these constraints to advance their interactional goals. This perspective challenges models that attribute deterministic power to technology and underscores the importance of examining actual human-technology encounters.

Advocating for design approaches that support situational emergence, I have called for integrating insights from multimodal conversation analysis into human-computer interaction. By designing interfaces that offer adaptable resources rather

than rigid structures, interactive system designers can furnish a space where participants are free to actively shape their interactions. This approach moves beyond deterministic models and contributes to design practices that are responsive to the multiplicity of human interactions.

Outlook

My work aligns with broader currents of research, as well as methodological tensions, that might grow into a more mutually transformative relationship between HCI and ethnomethodology. I will briefly highlight several developments on which my work has immediate bearing.

Multimodality and Affordances

Much like interactive systems are trending towards multimodality, so has ethnomethodological research. Thus, multimodal ethnomethodologists found themselves in the position to ask the question about the ‘affordances’ of modalities. As Sicoli (2016, p. 425) argues, “human language is always channeled via modalities, each of which has individually, or in combination with others, a semiotic carrying capacity, which is to say that they have different affordances or degrees of freedom for communication” – it seems very likely that the initial anti-affordance arguments that I deployed *from* ethnomethodology will have to be re-negotiated from within the field itself, particularly on the multimodal ethnomethodological vanguard that remains less constrained by established conventions.

Whether the result will be an embracing of more classical affordance theory—such as Hausendorf and Schmitt (2022) might be interpreted as—or an altogether different restructuring of the conceptual repertoire of the field—such as the Edwards’ articulation of a ‘medium of intersubjectivity’ (Edwards, 2024)—is currently an open question. But it seems undeniable that there are latent tensions in the field when it comes to its own relationship with the ‘effect’ of the interactional environment. Tensions between ‘medium’ and interaction analysis articulated in my earlier work (Klowait, 2019) are thus becoming more broadly relevant, and I plan on writing a continuation of the 2019 paper in due time.

Technologies like ChatGPT are having a notable impact on interaction research (Stokoe et al., 2024) and are discussed in broader terms in the neighbouring fields such as linguistic anthropology (Kockelman, 2024). This renewed foundational discourse is setting the stage for an expanded discussion of language and action: Goodwin (2017) famously highlighted the jump in action possibilities that occurred when early humans switched from the Acheulean Hand Axe to an axe composed of an axe-head attached to a handle. The new axe had the feature of decomposability: one could replace the handle whilst preserving the axe-head, or vice versa. This made it possible to engage in re-use with decomposition, where an axe becomes a spear when only one part (the head) is replaced, with the handle remaining in place.

Goodwin proceeded to greatly expand the list of such ‘decomposable’ structures, with human language chiefly among them: its versatility is owed not so much because language is intrinsically intelligent, but because it ‘affords’ an almost limitless range of combinations (e.g. specific sentences spoken in a given pragmatic context) framed by a stable structure (e.g. grammar, syntax).

From this Goodwinian perspective embraced throughout my works, one upcoming avenue of research is a bridge between the conceptually disruptive potential of language-deploying artificial systems and the co-operative action framework advanced by Goodwin (2017). A modest variant was already proposed in the ‘AI explains AI’ paper (Klowait et al., 2024), where the question of co-construction was related to the structural effects of a language-based explanatory interface. However, the quality of language as a specific *public substrate* “that at each iteration becomes the new point of departure for subsequent action, [affording] reuse with modification, or structure-preserving decomposition for the production of novel actions.” (Goodwin, 2017, p. 341) is worth a much broader exploration, both empirically and conceptually.

On Situated Plans

It makes sense to conclude with a return to the beginning. As discussed previously (see especially Klowait & Erofeeva, 2021) the field of human-computer interaction was substantially influenced by Lucy Suchman’s *Plans and Situated Actions* (1999), where a planning model of interaction design was juxtaposed to temporally-unfolding situated action. With systems like ChatGPT, it is now principally possible to assign an overall aim to an agent – without specifying a concrete means of achieving it, and thus without assuming a script-like theory of social action.

This could possibly mean that ‘situated plans’ – agents following a prior set of directions based on responses to emerging situational contingencies: for example, a printing assistant that only has the instruction ‘to help the user with their printing needs’, rather than having a pre-programmed set of steps to support specific tasks. That this does not ‘automatically’ work has been shown in Klowait et al. (2024); however, the *possibility* of a re-orientation of interface design from pre-situational plans to more holistic personas might represent a qualitative transformation, and a possible need to revise the plan/situation distinction.

References

Bijker, W., Pinch, T., & Hughes, T. P. (2012). *The social construction of technological systems: New directions in the sociology and history of technology: Anniversary edition*. MIT Press.

Brahnam, S., & De Angeli, A. (2012). Gender affordances of conversational agents. *Interacting with Computers*, 24(3), 139–153.
<https://doi.org/10.1016/j.intcom.2012.05.001>

Button, G. (1990). Going up a blind alley: Conflating conversation analysis and computational modelling. In P. Luff, N. Gilbert, & D. Frohlich (Eds.), *Computers and Conversation* (pp. 67–90). Academic Press.

Callon, M. (1984). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *Sociological Review*, 32(S1), 196–233. <https://doi.org/10.1111/j.1467-954X.1984.tb00113.x>

Cousineau, M. J. (2019). Limiting the discourse of computer and robot anthropomorphism in a research group. *AI & Society*, 34(4), 877–888.
<https://doi.org/10.1007/s00146-018-0836-6>

Davis, J. L., & Chouinard, J. B. (2016). Theorizing Affordances: From Request to Refuse. *Bulletin of Science, Technology and Society*, 36(4), 241–248.
<https://doi.org/10.1177/0270467617714944>

Edwards, T. (2024). The medium of intersubjectivity. *Anthropological Theory*, 24(2), 136–154. <https://doi.org/10.1177/14634996231196437>

Egorova, A. I., & Klowait, N. (2021). How to say good-bye to a robot? The matter of conversational closing [Как попрощаться с роботом? Исследование достижения разговорного закрытия]. *Monitoring Obshchestvennogo Mneniya: Ekonomicheskie i Sotsial'nye Peremeny*, 1, 241–270.
<https://doi.org/10.14515/monitoring.2021.1.1810>

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886.
<https://doi.org/10.1037/0033-295X.114.4.864>

Erofeeva, M., & Klowait, N. (2020). Dei ex machina: The interaction order of gamified distance learning. *SoP*, 32(3), 189–220.
<https://doi.org/10.22394/2074-0492-2020-3-189-220>

Erofeeva, M., & Klowait, N. (2021). The impact of virtual reality, augmented reality, and interactive whiteboards on the attention management in secondary school STEM teaching. *Proceedings of 2021 7th International Conference of the Immersive Learning Research Network, iLRN 2021*.
<https://doi.org/10.23919/iLRN52045.2021.9459318>

Fank, J., Richardson, N. T., & Diermeyer, F. (2019). Anthropomorphising driver-truck interaction: A study on the current state of research and the introduction of two innovative concepts. *Journal on Multimodal User Interfaces*, 13(2), 99–117. <https://doi.org/10.1007/s12193-019-00296-w>

Fank, J., Santen, L., Knies, C., & Diermeyer, F. (2020). “Should We Allow Him to Pass?” Increasing Cooperation Between Truck Drivers Using Anthropomorphism. In N. Stanton (Ed.), *Advances in Human Factors of Transportation* (Vol. 964, pp. 475–484). Springer International Publishing. https://doi.org/10.1007/978-3-030-20503-4_43

Garfinkel, H., Lynch, M., & Livingston, E. (1981). I.1 The Work of a Discovering Science Construed with Materials from the Optically Discovered Pulsar. *Philosophy of the Social Sciences*, 11(2), 131–158. <https://doi.org/10.1177/004839318101100202>

Gibson, J. J. (2015). *The ecological approach to visual perception: Classic edition*. Psychology Press.

Goffman, E. (1976). *Frame analysis: An essay on the organization of experience* (3. print). Harvard Univ. Press.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10), 1489–1522. [https://doi.org/10.1016/s0378-2166\(99\)00096-x](https://doi.org/10.1016/s0378-2166(99)00096-x)

Goodwin, C. (2017). *Co-Operative Action*. Cambridge University Press. <https://doi.org/10.1017/978139016735>

Goodwin, C. (2018). Why multimodality? Why co-operative action? *Social Interaction. Video-Based Studies of Human Sociality*, 1(2). <https://doi.org/10.7146/si.v1i2.110039>

Goodwin, C., & Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology*, 19(1), 283–307. <https://doi.org/10.1146/annurev.an.19.100190.001435>

Hausendorf, H., & Schmitt, R. (2022). 14. Architecture-for-interaction: Built, designed and furnished space for communicative purposes. In A. H. Jucker & H. Hausendorf (Eds.), *Pragmatics of Space* (pp. 431–472). De Gruyter. <https://doi.org/10.1515/9783110693713-014>

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. *Structures of Social Action: Studies in Conversation Analysis*, 299–345.

Heritage, J. (1998). Conversation Analysis and Institutional Talk: Analyzing Distinctive Turn-Taking Systems. In S. Cmejrkova, J. Hoffmannová, O. Müllerová, & J. Svetlá (Eds.), *Dialoganalyse VI/2* (pp. 3–18). De Gruyter. <https://doi.org/10.1515/9783110965049-001>

Heritage, J. (2012). Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction*, 45(1), 1–29. <https://doi.org/10.1080/08351813.2012.646684>

Heritage, J. (2013). Action formation and its epistemic (and other) backgrounds. *Discourse Studies*, 15(5), 551–578. <https://doi.org/10.1177/1461445613501449>

Heritage, J. (2016). On the diversity of ‘changes of state’ and their indices. *Journal of Pragmatics*, 104, 207–210. <https://doi.org/10.1016/j.pragma.2016.09.007>

Hostettler, D., Mayer, S., & Hildebrand, C. (2023). Human-Like Movements of Industrial Robots Positively Impact Observer Perception. *International Journal of Social Robotics*, 15(8), 1399–1417. <https://doi.org/10.1007/s12369-022-00954-2>

Housley, W. (Ed.). (2023). *The SAGE handbook of digital society*. SAGE Publications.

Huh, J., Whang, C., & Kim, H.-Y. (2023). Building trust with voice assistants for apparel shopping: The effects of social role and user autonomy. *Journal of Global Fashion Marketing*, 14(1), 5–19. <https://doi.org/10.1080/20932685.2022.2085603>

Johnson, J. [Bruno Latour]. (1988). Mixing humans and nonhumans together: The sociology of a door-closer. *Social Problems*, 35(3), 298–310. <https://doi.org/10.2307/800624>

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, 28(1), 241–250. <https://doi.org/10.1016/j.chb.2011.09.006>

Kimstach, U. M., Klowait, N., & Erofeeva, M. (2021). Talking without a voice: Virtual co-speakership in an educational webinar. *Sociology of Power*, 33(4), 198–216. <https://doi.org/10.22394/2074-0492-2021-4-198-216>

Klowait, N. (2015). Constitutive effects of technological frames in everyday interaction. *Sociology of Power*, 27(1), 166–180.

Klowait, N. (2017). A conceptual framework for researching emergent social orderings in encounters with automated computer-telephone interviewing agents. *International Journal of Communication and Linguistic Studies*, 15(1), 19–37.

Klowait, N. (2018a). Reflexive anthropomorphism: Ontological ignorance, or ignorant ontology? *Sotsiologicheskiy Zhurnal*, 24(1), 8–33. <https://doi.org/10.19181/socjour.2018.24.1.5711>

Klowait, N. (2018b). The quest for appropriate models of human-likeness: Anthropomorphism in media equation research. *AI & Society*, 33(4), 527–536. <https://doi.org/10.1007/s00146-017-0746-z>

Klowait, N. (2019). Interactionism in the age of ubiquitous telecommunication. *Information, Communication & Society*, 22(5), 605–621. <https://doi.org/10.1080/1369118X.2019.1566487>

Klowait, N. (2023). On the Multimodal Resolution of a Search Sequence in Virtual Reality. *Human Behavior and Emerging Technologies*, 2023, 1–15. <https://doi.org/10.1155/2023/8417012>

Klowait, N., & Erofeeva, M. (2021). The Rise of Interactional Multimodality in Human-Computer Interaction. *Monitoring of Public Opinion: Economic and Social Changes*, 1. <https://doi.org/10.14515/monitoring.2021.1.1793>

Klowait, N., & Erofeeva, M. (2023). Halting the Decay of Talk: How Atypical Interactants Adapt their Virtual Worlds. *Social Interaction. Video-Based Studies of Human Sociality*, 6(1). <https://doi.org/10.7146/si.v6i1.136903>

Klowait, N., Erofeeva, M., Lenke, M., Horwath, I., & Buschmeier, H. (2024). Can AI explain AI? Interactive co-construction of explanations among human and artificial agents. *Discourse & Communication*, 18(6), 917–930. <https://doi.org/10.1177/17504813241267069>

Knoblauch, H., & Löw, M. (2024). The Refiguration of the Global: Globalization and the Spatial Logics of Digitalization. *Critical Sociology*, 08969205241262304. <https://doi.org/10.1177/08969205241262304>

Kockelman, P. (2024). *Last words: Large language models and the AI apocalypse*. Prickly Paradigm Press.

Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition*, 1(3), 289–305. [https://doi.org/10.1016/1053-8100\(92\)90066-J](https://doi.org/10.1016/1053-8100(92)90066-J)

Latour, B. (1994). On technical mediation—Philosophy, sociology, genealogy. *Common Knowledge*, 3(2), 29–64.

Latour, B. (1996). On interobjectivity. *Mind, Culture, and Activity*, 3(4), 228–245. https://doi.org/10.1207/s15327884mca0304_2

Latour, B. (2007). *Reassembling the social: An introduction to Actor-Network-Theory* (1. publ. in pbk). Oxford Univ. Press.

Latour, B. (2014). Agency at the Time of the Anthropocene. *New Literary History*, 45(1), 1–18. <https://doi.org/10.1353/nlh.2014.0003>

Law, J. (2002). *Aircraft stories: Decentering the object in technoscience*. Duke University Press.

Li, X., & Wang, J. (2024). Should government chatbots behave like civil servants? The effect of chatbot identity characteristics on citizen experience. *Government Information Quarterly*, 41(3), 101957. <https://doi.org/10.1016/j.giq.2024.101957>

Lynch, M. (2016). Cultural Dopes. In G. Ritzer (Ed.), *The Blackwell Encyclopedia of Sociology* (1st ed., pp. 1–2). Wiley.
<https://doi.org/10.1002/9781405165518.wbeos0712>

Majlesi, A. R., Cumbal, R., Engwall, O., Gillet, S., Kunitz, S., Lymer, G., Norrby, C., & Tuncer, S. (2023). Managing Turn-Taking in Human-Robot Interactions: The Case of Projections and Overlaps, and the Anticipation of Turn Design by Human Participants. *Social Interaction. Video-Based Studies of Human Sociality*, 6(1). <https://doi.org/10.7146/si.v6i1.137380>

Moberg, R., Downs, E., Shelby, A., & Khan, A. (2024). Understanding reciprocity in human–robot interactions through completion of a pre-giving favor. *Smart Health*, 32, 100466. <https://doi.org/10.1016/j.smhl.2024.100466>

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3), 336–366.
https://doi.org/10.1111/josl.1_12177

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

Nass, C., & Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press.

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

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, 27(10), 864–876. <https://doi.org/10.1111/j.1559-1816.1997.tb00275.x>

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '94) (pp. 72–78). ACM. <https://doi.org/10.1145/191666.191703>

Norman, D. A. (2013). *The design of everyday things* (Rev. and expanded edition). MIT press.

Pinch, T. (2010). The Invisible Technologies of Goffman's Sociology: From the Merry-Go-Round to the Internet. *Technology and Culture*, 51(2), 409–424.
<https://doi.org/10.1353/tech.0.0456>

Pinch, T., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. *Social Studies of Science*, 14(3), 399–441.
<https://doi.org/10.1177/030631284014003004>

Rantala, J. (2020). Anthropomorphism in Social Robotics: Simondon and the Human in Technology. In M. Nørskov, J. Seibt, & O. S. Quick (Eds.), *Frontiers in Artificial Intelligence and Applications*. IOS Press.
<https://doi.org/10.3233/FAIA200946>

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

Rohlfing, K. J., & Cimiano, P. (2022). Constructing Explainability. *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, 83–84.
<https://doi.org/10.1109/REW56159.2022.00022>

Sicoli, M. A. (2010). Shifting voices with participant roles: Voice qualities and speech registers in Mesoamerica. *Language in Society*, 39(4), 521–553.
<https://doi.org/10.1017/S0047404510000436>

Sicoli, M. A. (2016). Repair organization in Chinantec whistled speech. *Language*, 92(2), 411–432. <https://doi.org/10.1353/lan.2016.0028>

Stokoe, E., Albert, S., Buschmeier, H., & Stommel, W. (2024). Conversation analysis and conversational technologies: Finding the common ground between academia and industry. *Discourse & Communication*, 18(6), 837–847.
<https://doi.org/10.1177/17504813241267118>

Suchman, L. (1999). *Plans and situated actions: The problem of human-machine communication* (Nachdr.). Cambridge Univ. Press.

Suchman, L. (2006). Human-machine reconfigurations: Plans and situated actions, 2nd edition. Cambridge Univ. Press.
<https://doi.org/10.1017/CBO9780511808418>

Van Den Boomen, M. (2014). *Transcoding the Digital. How Metaphors Matter in New Media*. Institute of Network Cultures.
<https://doi.org/10.25969/mediarep/19232>

Van Der Goot, M. J., Koubayová, N., & Van Reijmersdal, E. A. (2024). Understanding users' responses to disclosed vs. undisclosed customer service chatbots: A mixed methods study. *AI & SOCIETY*.
<https://doi.org/10.1007/s00146-023-01818-7>

Venturini, T., & Munk, A. K. (2022). *Controversy Mapping. A Field Guide*. Polity.

Watson, R. (2008). Comparative Sociology, Laic and Analytic: Some Critical Remarks on Comparison in Conversation Analysis. *Cahiers de Praxématique*, 50, 203–244. <https://doi.org/10.4000/praxematique.967>

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109(1), 121–136.