

ABSTRACT: When we talk about AI-driven systems there is a tendency by researchers to treat people encountering them as '*participants*' in human-machine interactions. This seems particularly true for so-called conversational AI, such as voice interfaces or chatbots. The pervasiveness of this position is encapsulated by the popular adoption of Nass et al.'s statement that "Computers are Social Actors" (CASA), which argues people are "mindlessly" applying human "social scripts" to AI systems; in other words, people act like participants as a kind of social reflex action. We think this is mistaken and find that a cursory look at actual interactions with (in our case) conversational AI systems reveals a different picture. Taking an ethnomethodological and conversation analytic perspective, we present a series of recorded fragments of people interacting with domestic voice interfaces. These show the organised ways in which conversational AI systems are embedded into everyday action. In doing this we reframe people's use of interactive AI technologies: far from being mindless or perfunctory, interactions with conversational AI are inextricably situated and interwoven with the sociality of a setting. Crucially, we show how AI systems are *regulated* within that sociality, via a wide range of practical (in our case conversational) methods. Understanding mundane regulatory work, then, is more pressing from a design perspective than working out how to design AI-driven systems to be better '*participants*'.

Stuart Reeves is Assistant Professor in Computer Science at the University of Nottingham, where he is member of the Mixed Reality Lab and Horizon Digital Economy Research Institute. Stuart's research interests span human-computer interaction, collaborative computing, and design research, while his approach is strongly informed by ethnomethodology and conversation analysis (EMCA). He recently held an EPSRC Early Career Fellowship (EP/K025848/1) investigating the connections between academic HCI research communities and the work of practitioners in UX/IxD/IA and other design professions, and is presently Co-I on the EPSRC grant "From Human Data to Personal Experience" (EP/M02315X/1). Stuart has authored publications in a wide variety of venues (ACM CHI, ACM CSCW, ACM ToCHI, Design Issues, Topics in Cognitive Science, etc.), and is author of the book *Designing Interfaces in Public Settings* (Springer, 2011).

Martin Porcheron is a Lecturer in Computer Science at the Swansea University Computational Foundry. His research is predominantly in the areas of human-computer interaction and computer-supported cooperative work, in which he draws upon human-centred design approaches to analyse, reflect upon, and inform the design of 'interactive AI' technologies. In his PhD, he studied interactions with technologies in settings including public places like cafés and pubs to understand how we use devices while we are socialising with friends and family. His work now spans multiple domains including everyday life, workplaces, and healthcare, focusing on the design, implementation and use of new Machine Learning-enhanced technologies.

<Chapter number>

Conversational AI: Respecifying Participation as Regulation

Stuart Reeves

University of Nottingham, UK

stuart.reeves@nottingham.ac.uk

Martin Porcheron

Swansea University, UK

m.a.w.porcheron@swansea.ac.uk

1. Introduction

When people interact with AI-driven systems, is it right to say that these systems are *participating* in an encounter? And is it right that we should design them with this aim in mind? Our purpose in this chapter is to examine the relationship between ‘conversational AI’¹ and conceptualisations of *participation* in social interaction. The need for this is because of the strong interest across many fields of research about how to design and engineer moments of human-AI interaction (HAI, see Amershi et al., 2019). This raises the question of what qualities such AI technologies should exhibit, and how are such notions of participation are construed. We think the version of participation adopted by researchers and designers of interaction with AI systems tends to be naïve, rarely dealing with the constructive organisation of the social situations into which AI systems—like all technologies—are actually embedded. We instead contend that we cannot understand AI without also appreciating the foundationally constitutive nature of the social situations that render such technologies meaningful. In doing this we need to rethink how we treat concepts like participation in human-AI interaction.

Few AI technologies are so readily examinable for the results of this naïve treatment of participation as conversational AI systems. These include voice interfaces and voice agent services such as the Amazon’s Alexa, Apple’s Siri, or Google’s Assistant. The conceit of these technologies is to create ‘conversational’ interactions with end-users through conversation design (Hall, 2018; Moore et al., 2018). The technologies variously aid access to information, shopping, music streaming, banking, gaming, calendaring, or even chatting with others (Alexa Prize, Khatri et al. (2018)). Conversational AI is often envisioned as a new modality of human-computer interaction and HAI, which strives to enable users to talk ‘naturally’ (Koefoed Hansen and Dalsgaard, 2015) and create systems that seamlessly embed themselves into social life.

Let us step back for a moment and look at the bigger picture. There are plenty of examples of claims that AI-driven systems are increasingly capable of providing us with something that tips towards ‘true’ participation in the disparate circumstances of social life. For instance: AI technologies are providing automated ‘decision making’ (ADM), or displaying ‘understanding’ and ‘knowledge’ about natural language, ‘recognising’ faces, actions, or in our case, acting as ‘personal assistants’, taking part in conversation by leveraging AI techniques and design processes that enable systems to enter into conversation or dialogue (Castle-Green et al., 2020). This is underpinned by a phenomenon we could call ‘agentification’², where AI systems’ supporting societal and

¹ We will refer to a class of AI systems as ‘conversational AI’ for the purposes of consistency throughout our chapter, although obviously we recommend caution for potential confusions that might result in face-value readings of such language: expressing conversational capabilities or agency they are really not in any *ordinary* sense. We will refrain from repetitive scare quotes to aid readability however their presence is universally intended.

² Agentification is a way of describing how AI systems have a tendency to become rarified at the expense of understanding ‘AI’ as constituted not only of specific technologies but *necessarily* the situations in which they are deployed, the sociality of which produces in totality such systems as ‘AI’ (Mair et al., 2020). A simple way to think about this is to consider how conversational AI becomes endowed with the

social infrastructures are routinely erased from the discussion. It is often admitted that present technological capabilities are limited, but a trajectory towards ‘true’ conversational AI (or even ‘true’ AI) typically follows this admission as a tacitly assumed position. Or more so, such a stance may even be explicitly embraced, as it tends to be with almost all AI technologies (Mitchell, 2019). We can point to relentless claims—emanating mostly from the technical Computer Science literature—that systems are not only meeting but even *exceeding* human-level performance for things like speech transcription (Xiong et al., 2016) or visual tasks like recognising the contents of images (He et al., 2015), or of faces (Toole et al., 2007). Another key step has been the results of large-scale language models such as GPT-3 (Brown et al., 2020) that produce supposedly uncannily human-like demonstrations of language production. These examples are, unfortunately for their proponents, categorically different technical feats to those claimed and confidence in teleology is a mistake that never seems to be committed to memory once revealed in prior waves of AI.

To some extent, this mistake around prospects for AI systems being participants in human life has been underpinned intellectually within fields like human-computer interaction (HCI), human-robot interaction (HRI) and other HAI-affiliated research communities by claims about computational systems’ role as “social actors” within interaction (see Nass et al. (1994) and Reeves and Nass (1996)). This “computers are social actors” (CASA) paradigm describes a *social orientation* towards computers displayed by users, who then treat said computers *as participants* in their activities. This is due to what Nass and colleagues argue is “over-learned social behavior, whereby users interact via socially mindless responses without meaning anything by them” (Nass and Moon, 2000). The effects of this phenomena identified by Reeves and Nass (1996) includes conversational AI; and most relevantly for us, many of their examples employ voice interaction. Social action towards machines in the CASA view is taken as “mindless”—i.e., the (mis)application by people of inappropriate “social scripts” to their interactions with computational systems that really should not be applied (Nass and Moon, 2000).

Nevertheless, Nass and colleagues advance the idea that “social scripts” play a role for people when they are found treating all manner of digital interactive technologies with this “social orientation”. It is not, Nass and Moon argue, that people are seeing computers anthropomorphically, but rather that they apply such social behaviours in spite of not anthropomorphising at all (*ibid.*, p. 93). This view—which seems to aid the design practices of interactive AI systems (and thereby conversational AI)—positions participation with computers as both leveraging and battling against a kind of fundamental confusion that is continually emergent in any and all instances of encounters between human and machine interlocutors. In this sense, CASA offers a stance on participation with AI systems that is built upon *mistaken* users—not strategic, nor methodical—who instinctively apply the scripts to many kinds of technology.

power of ‘conversation’—always shorn of the complex backend infrastructures, design process and constitutive totalities of their grounded circumstances of use. We will return to this point later.

Instead, aided by looking at actual encounters with AI systems in practice, we are going to show that conversational AI, and potentially other forms of interactive AI systems, can be more productively understood as placed within organised *interactional 'regulation'*, in contrast with the idea of them being thought of as 'social actors'. What we mean is that interactions with and around conversational AI systems are continuously, ongoingly *regulated* by people: people using them, subject to them, co-proximate to, around, or bystanding them and the relevant moments of interaction that unfold. Our aim is to 'let out some air' from the quite pervasive argument that computational systems can be considered participants or have 'conversations' (Reeves, 2019; Porcheron, 2021)—*at least not in any typical sense*. In this regard CASA is potentially misleading as a conceptual apparatus with which to reason about conversational AI. This matters because a confused basis for design can lead to confused designs and confused critical assessments of design.

Regulatory work is not always successful, rather it is something that we find to be typically strived-for by dint of the interactional environment. More so, it doesn't preclude or elide the significant, well-documented, profound problems (e.g., unfairness, racism, neo-phrenology, accessibility etc.; for example, see Trewin et al., 2019; Gebru, 2019, Buolamwini and Gebru, 2018; Bowyer et al., 2020) created and exacerbated by the persistent tactic of exempting AI systems from determinations and attributions of responsibility in society, along with the organisations, economic systems, and people that produce them. This chapter does not intend to suggest that such matters are somehow readily mitigated by such regulatory action. A conversational AI system that exploits workers (Wong, 2019), compromises protected characteristics (Hannon, 2016; Sutton, 2020), appears to readily promote regressive behaviour (e.g., Microsoft Tay; see Wolf et al. 2017) or conjure repressive politics (Phan, 2019) is still a problematic system, although they cannot (and should not) be disentangled from their umbilicals to the cloud and organisations' infrastructures and work practices. Rather, our point is that *mundane regulatory work* is—additively to the foregoing—simply part of this core texture of AI systems in everyday encounters, albeit often overlooked, and should form part of the growing understanding of the social implications of AI systems in society and practical use.

In this chapter, to respecify the concept of participation, we employ a set of 'reminders from mundanity'—recordings of actual conversational AI use in everyday life—that suggest alternate ways of considering the sheer force of regulation by users. First, though, before looking at these reminders, we want to briefly talk about our approach: ethnomethodology (Garfinkel, 1967), conversation analysis (Sacks, 1992) and the situated mundanity of conversational AI in-the-world.

2. Approaching conversational AI

Research into and with conversational AIs has the propensity to disavow the situated and mundane context in which the technology is encountered, and which is a fundamentally constitutive part of AI itself. For one reason or another, discourse routinely engages in the rarification of AI, underpinning agentification and the ultimate

shaping of participation. The rarified figurative technology that gets put on show in demos, or deployed into mass-market products, is built within corporate and university research labs, working towards an imagined figment of human-AI participation. This is achieved by shearing off *constitutive* features of AI in-the-world and engineering technology for this abstraction. Researchers interested in understanding (conversational) AI in-the-world then retroactively recover and attempt to ‘resolve’ the situation. Such a stance emphasises “reflecting how we find AI in the world (c.f. Neyland, 2019), rather than drawing boundaries around the technology in narrow ways [which] means treating all the work that goes into and is done with AI, including descriptions of what a given system might be said to be or be doing, as being as much part of the ‘assemblage’ as the hardware and software” (Mair et al., 2020).

On the other hand, however, the CASA paradigm does push towards *social* responses to computational technology—most obviously AI and conversational AI. However, the paradigm locates those social responses as a property of human psychology and misapplied reactions to such technologies. In doing so, CASA ignores the socially constituted production of those ‘responses’, and that embedded within local, reflexive orders of social action. For conversational AI this is often located in the organisation of domestic life at a given moment (Porcheron et al., 2018).

We look to everyday encounters with conversational AI for what they are, as constitutive phenomena in-the-world, rather than how we might imagine or desire them to be. We adopt the orientation of ethnomethodology and conversation analysis to explore the *mundanity* of action—to identify the things that we ‘know’ are the case but nevertheless tend to be abandoned in favour of agentified accounts. Through examining the situations that conversational AI systems are embedded amidst, we can identify the accountable and continual analysis of action by others, as well as the crafting and designing of action for recipients (Sacks et al., 1974), including ‘talk to’ such systems.

In one sense, none of the facets of interaction described here will be news to those familiar with daily encounters with conversational AIs, specifically of the voice kind (Porcheron et al. 2018; Beneteau et al. 2019), although the capacity for self-deception in spite of this seems present even for some AI and even HCI researchers. Thus, much of this may be new to technologists and some designers, who may tend to assume conversation or talk is a fundamentally ‘messy’ activity, and so adopt strictures of guidelines or frameworks, or indeed rely upon findings from linguistics about what ‘conversation’ is and how this variegated phenomenon may be formally understood and then *encoded* into infrastructures underpinning conversational AI. For those in the social sciences with a perhaps more critical perspective, there is a deflationary and maybe even ‘realist’ power to examining interaction with AI systems in detail—particularly for those not as familiar in the technical substrates of such systems.

Through this orientation to the mundane, we examine the orderly practices involved in using conversational AI, to aid our re-specification of participation. A strong idea of participation is fundamental to ethnomethodology. Speaking ethnomethodologically, everyone is a participant in some order of sociality, all the time. There is—to use

Garfinkel's aphorism—"no time out" from society during which participation is suspended (Garfinkel, 1988; McGrane, 1994). Our job then is to understand how that continuous work of participation is accomplished and, in turn conceptualised. Through this, we can identify what this all means for the design of the technologies, services and infrastructures that increasingly demand our involvement.

Empirical examples will help us in trying to get to grips with thorny conceptual issues and exactly what kinds of orientations to computational systems that members of mundane social settings actually have (as opposed to, say, those orientations found in experimental settings (Nass and Moon, 2000), or equally accounts of experiences with conversational AIs (Cowan et al., 2017). Our guiding insight is to say that language gets its meaning from its use—as argued by ethnomethodological conversation analysis via Wittgenstein. Correspondingly, we look at actual instances of human (social) action as it is—rather than as we might wish it to be. Applying this to a discussion of conversational AI means immediately asking how interactions with and around these technologies gain meaning from talk-in-interaction, and to what extent empirical instances might act as aids to us in tracing what life with them actually looks like. As we present data in this chapter, we will encounter problems with arguments that conversational AIs a) might count as "social actors" and b) that such roles turn on the mindless (mis)application of social rules by people interacting with conversational AIs.

3. Living with *social actors*

The transcripts we draw upon next are derived from a month-long study of conversational AI use in domestic environments. Specific details about the approach do not concern us here and instead we would direct readers to Porcheron (2019). In short, we have audio recordings captured from a number of households interacting with one such conversational AI—in these instances, an Amazon Echo 'smart speaker'. We take the Echo device—and the Alexa voice service that it acts as an interface to—as a reasonable representation of a whole range of similar voice-based conversational technologies (e.g., Google Home, Siri, Cortana, etc.). We remain sceptical of claims that changes to these systems (e.g., 'improvements' etc. since the data was collected or in the immediate future) invalidate what we address here. The basic necessity for people living with and around all manner of digital systems to develop sensitivity towards, and ultimately continuously tailor and design those interactions to accomplish particular ends doesn't change (a point we will return to briefly in our discussion). This is, of course, notwithstanding the utopian bulwarks of claimed 'naturalness' (Allen et al., 2001), 'intelligence' or 'exponential improvement' (see Mitchell, 2019). The thoroughly mundane and practical circumstances (often glossed as 'messy') we find ourselves in daily seem to resist this on any unwavering close inspection.

We begin with the simplest work of *regulating* conversational AI—addressing an agent, i.e., being 'heard' and avoiding being 'heard', by the technology. We will then move onto a more complex embroilment: the depths of *doing politeness*, a signature consideration of computers being treated as social actors.

3.1 Addressing conversational AIs and managing ‘hearing’

Methods of address and the regulation of turns-at-talk have been extensively explored in conversation analysis (Sacks et al., 1974; Schegloff, 2007). Address is also naturally a key part of the technical construction of conversational AIs—epitomised through ‘wake words’, which preface subsequent talk directed at an AI. However, this most basic of features often leads to accidental, unexpected, or unintended ‘use’. Such issues of address figure in wider need to manage trouble, something quite forcefully obvious on initial use. This has a strong bearing on how talk to and around conversational AIs proceeds because people quickly become oriented to this as a relevant feature of their embedding in situ. The design of voice services tends to be ‘greedy’ in that it is often addressed unwittingly by users (Lau et al., 2018). Thus, people need to regulate input—to manage being ‘heard’ and ‘not heard’ (see Bellotti et al. 2002).

Fragment 1. Let’s look at a family dinner scene—with parents Susan and Carl, and young siblings Emma and Liam. The Amazon Echo sits on a sideboard near the dining table and is being used while members of the family share a meal. Susan is wondering out loud how she might change her shopping list (a feature that Alexa provides via its connection with Amazon’s core shopping service):

```
SUS  so how do I tell Alex:a (1.1) that I bought some stuff
      off my: shopping list?
      (2.0)
ALE  you have [three] items on your shopping [list ]
SUS  [°oh° ] [↑oh: ]
?    hu hu huh
      (0.4)
ALE  rosemary (1.1) marshmallow (1.1) beef
      (0.5)
?    huhh
      (0.8)
SUS  [Alexa (2.0) ] Alexa I’ve bought everything off my shopping
      list.
?    [((laughter))]
      (2.1)
ALE  you have three items on your [shopping list (1.4)]
?    [((laughter)) ]
ALE  rosemary [ (1.2) marsh mallow (1.1) beef
CAR  [say:: (.) please delete shopping list=]
SUS  =oh
```

We want to draw attention to two key moments here: 1. Susan’s opening request for assistance in developing an appropriate formulation of input to their Echo regarding the shopping list (“so how do I tell Alexa that I bought some stuff off my shopping list?”); and 2. Carl’s delayed provision of that assistance in which he utters a fragment of a command that will do what Susan asks (“say ‘please delete shopping list’”).

Susan’s initial request leads immediately to Alexa ‘interjecting’ with an unexpected turn-at-talk (“you have three items on your shopping list”). Alexa has ‘heard’ Susan’s call for assistance as appropriate input (due to Susan’s use of the wake word), located relevant

components that are parseable (probably just “shopping list”), and produced a response, which in this case offers a report of Susan’s current shopping list. But it is clear that Susan’s opening turn is a request for help, not an address to the device. We know this because of Susan’s overlapped, surprised “oh”s and the subsequent laughter by another member of the family at the out-of-turn response from Alexa. Carl’s later assistance also analyses Susan’s initial turn as a request for help and *not* as input designed for Alexa.

It is tempting to think about this moment as analogous to being careful what one says around children; we could imagine an extension to CASA studies which demonstrates how people apply similar social rules to computers in such situations, not wishing a computer to ‘overhear’. Instead, however, we see formulations and ways of talking that are designed in quite specific ways to gear into witnessed, gradual surfacings of the technical capabilities that emerge from particular instances of interaction, and thus enable their regulation and that of input. Here Susan either fails to design her request for help to the assembled family in a way that avoids breaking the threshold of what can be heard by the Echo (e.g., her “Alexa” is simply too loud), or she momentarily forgets to exclude it in this. Either way, her subsequent overlapped “oh”s offer a surprised accounting for this lapse—Susan clearly was not expecting to trigger a response from Alexa. At this moment, the adequacy of her methods to this end—her competence in living alongside the conversational AI and demonstrating a sensitivity towards matters of input, ‘address’ and ways to do ‘not being heard’ by the conversational AI—seem to be lacking. Susan does not treat Alexa’s interjection as ‘out of turn’ in a way we might treat something similar in conversation. Susan does not castigate Alexa for this or seek to smooth over the infraction. She does not mitigate these problems or compensate for them in some way according to some violation of ‘social rules’. In fact, the occurrence is treated by her and the assembled family as mistaken device input. This is evident in the surprise, in the extended laughter and also in Susan’s subsequent direct attempt to clear her shopping list (“Alexa I’ve bought everything off my shopping list”).

Now, the tempting analogy of a ‘talking in front of child’ that one could try to characterise Susan’s response as akin to would also necessitate a projection that the assembled family treat the conversational AI as such, which we can see does not happen. Rather than addressing the Echo device as some form of interjecting social actor, Carl’s later utterance “say ‘please delete shopping list’” is *for* Susan. Carl also regulates his talk in such a way as to avoid further triggering of the Echo. Through projecting a gap within the device output, he inserts a response to Susan’s earlier request for assistance. He does so in a way that provides the second part of the necessary input to the Echo that Susan will need to then use, and crucially *without* the “Alexa” wake word. The absence of this wake word is striking—a clear demonstration of conversationalists’ competencies in managing accidental input: one such way of avoiding being heard is simply to omit any wake word. In this case, Carl’s “say” in particular takes its sense from Susan’s prior request (i.e., to “say” the following to Alexa).

This example shows us two contrasting moments of designing input and its regulation, underscoring the notion that regulation is a developed competency of people living with conversational AI.

3.2 Talking to conversational AIs as reflexive social phenomena

We have just seen how users of conversational AIs must make sense of things like ‘hearability’ when designing input and regulating their place within social organisation. We now want to begin to foreground how such common moments of interaction with conversational AIs take place *around others*, and the critical implications of this. In other words, we want to examine conversational AIs’ entanglements in the circle of reflexivity of social actions. Ethnomethodology puts central importance upon the reflexivity³ of social actions and argues reflexivity fundamentally permeates social organisation. With an ethnomethodological orientation, reflexivity is the property that social actions—verbal, bodily, and so on—*simultaneously and inextricably* both do things *and* offer accounts of themselves. It is this which enables their recognisability and analysability to others as *this* or *that ‘thing’* being done (such as: recruiting assistance, instructing someone, or listening to music). This is somewhat akin to but goes well beyond Austin’s notion of “performative utterances” (Austin, 1962).

Fragment 2. We need to make this more concrete, so turn to a new example. Here we join Joanne and Rob who are trying to get their Echo to play some music. Just prior to this Joanne and Rob have been asking to “play music by Kate Bush” to which the device responded “I can’t find songs by Kate Bush”. So now Joanne and Rob try a different strategy.

JOA	who- <u>Alexa</u> ? (1.0) who sin:gs: (0.3)
ROB	°wuh- (.) Wuthering Heights°
JOA	Wuthering Heights (1.8)
ALE	Wuthering Heights is by Kate Bush
JOA	((laughter))

Here we must pay attention to how Rob provides assistance to Joanne at just the right moment and in just the right way. As Joanne begins producing her Alexa-addressed utterance, she starts extending and drawing out “sings” and then stops (“Alexa, who sings-”). There is then a momentary, but short, pause (0.3s). Rapidly, Rob then steps with a completion to Joanne’s question fragment of seemingly half-finished input addressed to Alexa (ultimately which becomes “who sings Wuthering Heights”). Joanne then repeats what Rob has offered her in her next turn, and we see a response to the completed question from Alexa after a short pause, followed by Joanne’s laughter⁴.

³ Ethnomethodology’s definition and treatment of ‘reflexivity’ is different to many other senses of the term; see Lynch (2000).

⁴ For context, Joanne’s laughter at the end of this fragment seems to be calling to attention the strangeness of Alexa not successfully ‘understanding’ who Kate Bush is whilst at the same time ‘knowing’ that the song they are interested in is indeed by Kate Bush. Such fractures in sequential order and the organisation of joint understanding are common and are characteristic of both symbolic and sub-symbolic

The reflexive property of actions like turns-at-talk means that things done with conversational AIs (such as asking a question) are simultaneously indexed within, and accountable to the local social order in which they unfold (i.e., the conversation, its situation, and so on). Note how Joanne's drawn out "sings" projects a possible trouble: that her success in producing appropriately formulated input is in question. The same can be said for the pause, which reinforces the sense of mounting trouble by virtue of its position after the extended "sings". Rob then comes to the rescue with the necessary missing component. From this we can add that Rob's assistance demonstrates his analysis of this series of occurrences as troubled—so his subsequent offer of "Wuthering Heights" points to him hearing that trouble emerging. The design of Rob's assistance to Joanne shows his orientation to, and competence for, the design of appropriate conversational AI input—his analysis is oriented towards noticing various verbal 'disfluencies' as they are produced including a sensitivity towards any elongations, pauses etc. which may disrupt the successful design of input to Alexa. This is something that must be learned by users of conversational AIs.

Further, Rob's assistance and how it is designed points to a necessity to produce that assistance in ways that are *not* readily hearable by the conversational AI, like Carl's sotto voce help in the previous example. To this end, Rob's offer of "Wuthering Heights" is spoken much more quietly than Joanne's surrounding turns and delivered pretty promptly, given that there is a need to produce fluent input without longer pauses. We might even want to call these backstage⁵ remarks (Stokoe and Sikveland, 2020): Rob's quiet interjection of "Wuthering Heights"—clearly designed for Joanne and not the Echo—as well as Carl's wake-word-less insertion "say please delete shopping list". In summary, Rob's methods here are directed towards regulating interactions with the conversational AI in just the right way.

Again, as we saw with the previous example, these methods come in handy for the direct user of the Echo too. So, the sum effect is seamlessly co-produced 'Alexa talk' where each member of the local setting is monitoring one another's production practices and their relevance and appropriateness as 'Alexa talk'. Use of the conversational AI is thus regulated further, not only in terms of how people regulate 'hearability' of utterances to it and around it, but also how that particular form of talk is itself recognisably and analysably 'conversational AI talk', and therefore open to the kind of co-produced resolutions of trouble we see here.

Conversational AIs are thus easily pulled into ethnomethodology's explication of the reflexivity of social actions: utterances addressed to systems are accountable to the local setting of talk while also doing the work of input. Conversely, talk *not* designed as

forms of AI (e.g., displaying that one is unaware of X at point T1 yet at point T2 displaying awareness that Y is related to X). At a basic level this is a failure of tracking 'state', however provides for a shared moment of amusement (or maybe frustration) for Rob and Joanne. It is little different from the phenomena reported by Suchman decades ago in her studies of (what we might now call) 'smart' photocopiers (Suchman, 1987).

⁵ Although this hints at a Goffmanian dramaturgical metaphor (Goffman, 1959), we want to use it as a mere aid to illuminate agent-oriented practical competencies and not suggest anything more 'formal'.

input must be crafted in ways that avoid being scooped up by wake-word sensitive conversational AI. In other words, in this second form, there is the embedding of features that seek to ensure that non-input is analysable and accountable for other co-conversationalists. On this point, consider how Joanne demonstrates an understanding of what Rob is doing: not missing a beat she immediately picks up “Wuthering Heights” and reproduces it, more loudly this time, and in doing so also presents an analysis of her recognition of Rob’s prior turn as demonstrably *not* ‘conversational AI talk’. It is on this point about how reflexivity of social actions is at play in.

We start to see—at a fundamental level—the expressive paucity of conceptualising a kind of participation with conversational AIs as formulated by misapplied social rules. There is *a lot* more going on here even for quite ‘basic’ interactions with conversational AIs as soon as we think of human-AI interactions as inextricably drawn into reflexive social order. Rather than the conversational AI being treated as a participant in interaction, it is more conceptually coherent to think of it as being ‘stage managed’ through people regulating their actions as part of the on-going social order.

Next, we look at two seemingly more ‘complex’ kinds of interactions with (and around) conversational AIs, both of which directly speak to questions about conceptualisations of conversational AIs as social actors and the ways in which their regulation within social interactions actually unfolds. To this end, we firstly examine how scenes of ‘politeness’ play out, and then secondly how conversational AIs become entangled with routine domestic politics such as the exertion of control between family members.

3.3 Treating conversational AIs with politeness

Ideas about treating conversational AIs with a kind of politeness or lack thereof is frequently brought up in the popular media (Searles, 2019) and by research (Nass et al., 1999). It tends to be used as a cultural touchstone, sometimes to berate technologists, end-users, or society. For instance, discussions about politeness might raise concerns about children’s use of language, technology, and conversational AIs’ impact on domestic life in general (Garg and Sengupta, 2020). Politeness is also something argued by Nass and Moon (2000) is potentially the product of a “mindless” social orientation to computational systems and misapplication of social rules. Building on this, Lopatovska and Williams suggest that for conversational AI, “mindless politeness is more prevalent in a group than a lonely setting” (Lopatovska and Williams, 2018). Instead, we want to point out that for such group settings, conversational AIs are often drawn into the social organisation of manners simply because they are potentially in play as resources for those moments.

Fragment 3. Let’s now look at a perspicuous example of moral work as it relates to politeness and manners (Lopatovska and Williams, 2018). We join an example with two people, Nikos and Isabel at a party⁶. Their Amazon Echo is proximate to them. Nikos and Isabel (jointly) ask Alexa to play something for the party.

⁶ We have previously discussed this fragment in a prior publication (Reeves et al., 2018).

NIK	Alexa (2.6)
ISA	play some New Year's music (1.7)
ALE	here's a station for jazz music (.) instrumental jazz. (1.4) ((music starts playing)) (4.4)
ISA	Al(h)exa this is not what we w(h)anted ((laughter))
NIK	Alexa: (0.8) <u>shut</u> up. (0.8)
ISA	H↑E:Yuh (0.5) Alex(h)a (.) Nikos apologises for being <u>so</u> rude
ALE	hi there ((music continues playing)) (2.4s)
NIK	Alexa stop stop

The most immediately interesting thing we see here is Nikos telling Alexa to “shut up”, only to be subsequently brought to account for his directive by Isabel in her address to Alexa: “heyuh” and “Alexa, Nikos apologises for being so rude”. On the face of it, this looks like a clear incidence of being polite *to* a conversational AI and therefore might lead us to think that Alexa is being afforded politeness, maybe even that Alexa is becoming a participant in the ongoing unfolding of conversation. In any case, a clear “mindless” application of “social rules” of politeness.

First, we observe that Alexa is designed to respond to commands like “stop” and also “shut up” by subsequently stopping text-to-speech generation. In that sense Nikos is using valid input (valid from a computational parsing sense). However, a point we have continually turned to is that this interaction, as a sequentially organised matter, is grounded in the local context of talk. Nikos’s “shut up” is a well-formatted input or command to halt device output (music). So as input it does nothing wrong and seems like a category mistake to treat it as problematic in this sense. But—and this is a critical point—“shut up” *simultaneously* stands in the local order of talk, investing and delivering its meaning-in-interaction from proximate actions. Nikos’s “shut up” can be heard by members of the setting as simultaneously a device-relevant command *and* as building upon (and gaining further relevance from) Isabel’s prior assessment, i.e., a laughter-inflected negative assessment of the music being played by Alexa as “not what we wanted”. Nikos’s “shut up” then sets up Isabel’s next turn: a chiding third-person “apology” to Alexa, “Nikos apologises for being so rude”. In this we see how talk-in-action retrospectively and prospectively shapes the conversation as a result of co-conversationalists working to anticipate the ways that talk will unfold (Goodwin, 1979).

When we see this exchange between Isabel and Nikos take place in order, we see that the apparent politeness displayed by Isabel to the conversational AI can only really be understood by correspondingly taking into account its production within this local, sequential unfolding. Isabel uses the form of a 3rd person apology as a method for holding Nikos accountable for his turn. Isabel is not *just* apologising to the conversational AI here. Instead Isabel employs the simultaneity of input-at-talk to provide an analysis of Nikos’s behaviour, an assessment if you will, of how this

behaviour does or does not align with this particular, local normative moral order established in their interactions with and around Alexa. In that sense, ‘politeness’ to a conversational AI cannot readily be extricated from the present shared, agreed-upon ways-of-talking-at-a-party against which Nikos is being reprimanded (albeit with a hint of humorous insincerity). In this sense, the conversational AI can be deployed by members of a social setting as a resource to e.g., chide or joke. Thus ‘politeness’ is simply doing other things for the social situation. It is on this basis that we should have a discussion about what work concepts like politeness are doing for us, and not acting as if the successful accomplishment of appropriate language (whether polite, non-discriminatory, and so on) can be achieved by only technically fixing valid input.

The previous three fragments have shown in different ways how the conversational AI acts as a resource that is brought into the folds of talk by our co-conversationalists⁷. Participation in the form of politeness seems ever more distant from CASA’s solution to this, that people are simply applying social rules of politeness to machines. The situated unfolding of such actions argues for much more than this.

3.4 Conversational AIs’ embroilments in domestic politics

As we saw how regulating input and proxemic talk—i.e., the ‘hearing’ and ‘hearability’ of input, so to speak—shaped how address to conversational AIs as well as talk around them unfolded. Here we want to look at how people regulate *one another’s input* in more explicit ways than the previous example, where they deploy methods of control or ‘gatekeeping’ such input to agents.

Fragment 4. As devices that are designed for domestic environments, conversational AIs like Alexa are destined to become just another potential resource for participants in the organisation of domestic politics—as subject to or exerting ‘control’ over others, designing or demonstrating compliance. To exemplify this point, we return to the family seen earlier. Here we have another family dinner scene. The Amazon Echo sits on a sideboard near the dining table and is being used while members of the family share a meal. There has been an ongoing minor ‘struggle’ between parents Susan and Carl, and their children Emma and Liam, about who gets Alexa to play exactly what music and when. We join a moment where Emma addresses Alexa in order to resume the music they had on previously—which has been expressly dispreferred by the parents previously—at which point Susan intervenes.

```
EMM  Alexa
SUS  no hold on a minute=
EMM                                     =resume [RESUME music=]
SUS                                     [Alexa Alexa  ]=oh:
ALE  ((music starts playing))
```

⁷ While there is some sense of potential pro-activity being a desirable design possibility for conversational AIs (Cha et al., 2020), currently such designed elements are *not* part of the standard set of functionalities supported.

We see the ‘competitive’ production of a summons to the conversational AI as well as subsequent directive by Emma to play (or not play) the music. Firstly, we pick apart how this is done. Emma launches—or perhaps more correctly resumes—her ongoing project to play the music by producing the wake word “Alexa”. Swiftly after this, Susan locates and takes advantage of the characteristic pause between wake word and subsequent utterance to express opposition to Emma’s nascent project: “no hold on a minute”. But Emma continues, producing the second part of the directive—“resume music”—neatly latched onto the end of Susan’s prior turn. In this way it seems as if Emma is taking advantage of the propensity of the underlying parsing of transcribed speech to search for specific keywords that may map to a space of possible well-formatted input. In this case “no hold on a minute” has a possibility of being ignored by the conversational AI’s natural language parsing and dialogue management components in favour of “resume music” which more clearly pertains to specific commands around the control of music⁸. Analytically setting to one side the relevance of our knowing Emma’s ‘true’ comprehension of the particularities of underlying technologies, she nevertheless upgrades her attempts towards control subsequent, louder repeat of “resume”. Simultaneously Susan, with overlapping talk, tries to wrestle control from Emma of the Echo by using the wake word herself, twice. Doing so can sometimes reset the way Alexa awaits input, effectively letting a user start again from the beginning (see prior footnote). This is clearly unsuccessful as subsequently the music does indeed start playing, leading to Susan’s somewhat surprised “oh”.

Both Emma and Susan employ prosodic methods to regulate one another’s access to the conversational AI. They design their input to Alexa in ways that are sensitive to the technical capabilities of the Echo’s microphones and underlying software. This sensitivity is not gathered through reading manuals or learning about conversational AI system architecture but through the sum of prior interactional experiences coupled with a range of common-sense practical experience of computational devices in general. In a sense, they display some semblance of Sacks’s concept of recipient design (Sacks et al. 1974). However, it is perhaps more correct to talk about a sensitivity towards the *parameters of input* to the technology as we have done throughout this chapter. In either case, there does not seem to be a need for arguing any substantive difference between talk and, say, the how any technologies of recognition (or input devices in general) necessarily constrain the parameters of that input (cf. Burak and Reeves, 2017). Ultimately, most technologies require human users to provide appropriately ‘disciplined’ actions so as to get something done with them, be it an opportune mouse click or sequence of verbal utterances.

There are other subtleties apparent in this domestic scene to focus on too. The conversational AI becomes embroiled in the everyday workings of family ‘small-p’

⁸ Note that this is not a technical point about the inner workings of natural language processing at play, but rather a point about how users of conversation agents learn about what kinds of technical foibles they have and how such foibles may be co-opted to achieve certain kinds of effects. In this case, one can rely upon an overzealous approach towards keyword matching to achieve this kind of effect. We cannot be certain that this is what Emma is doing here, but an ethnographic perspective on the data would suggest as much.

politics and in this case what we could call the “politics of control” (Porcheron et al. 2018) or perhaps something here we might recast as the politics of regulating conversational AI. Encompassed in this are matters of entitlement to address Alexa—who gets to say what and when. In this case parents’ interactional projects with the conversational AI are in tension with the children’s ongoing projects, i.e., to get to play music, to play music they want to listen to, and equally to some extent use it to leverage some enjoyment in engaging in a level of antagonism with the parents. There is no model for rights to conversational AI input going on here, so members of the family instead use a raft of methods to manage those rights, entitlements, and so on, including (as we have seen) prosodic methods, overlap avoidance / latching, and eventually sheer volume to get desired results.

In sum, the conversational AI looks more like a resource that—as with any other technology—turns upon social organisation to manage, distribute, and share access—i.e., to regulate it. Like politeness, describing the methodical work of participants in social interaction unfolding as the (mis)application of social rules to non-human ‘participants’ seems to be an unnecessary stretch beyond a simpler apparent truth. Our data shows that those interactions with and around conversational AIs take advantage of and adapt to fit mundane methods of social interaction—talk in these cases—to the emerging textures, and foibles of the technology in question. It is unclear why this needs to be positioned as a ‘confused’ state for such participants, that they are “mindlessly” applying (and adapting to fit) methods of social interaction to the practical tasks they are faced with.

4. Rethinking ‘conversational’ technologies

The discourses on AI that cross research, industry, and the media, share a tendency to concoct some semblance of deeper ‘participation’ in everyday life as the ultimate goal of AI systems design, moving it well beyond HCI’s more ‘realist’ conceptions of interaction (see review by Hornbæk and Oulasvirta (2017)). The promise of participation is embedded deeply within AI’s early figuration via its strongly Cartesian framing of ‘intelligence’. This has resulted in substantial branches of AI categorising intelligence as a cognitive phenomenon via a representationalist theory of mind, which then lands us into a discursive space organised around the apparent similarity between this and computation⁹. The stage was thus set for establishing a debate about participation as a potential property of human-machine interactions and has tended to be one of—we argue unwarranted and confused—connection and comparison between human and AI capabilities. While this has been subject to critique by many others (Dreyfus, or Collins

⁹ We should temper this presentation a bit. Significant work in AI such as that by Rodney Brooks has explored non-representational AI (in distinction to cognitive-inspired ‘Good Old Fashioned AI’ (GOFAI)) whereby intelligence is an emergent property of agents *in* environments (Brooks, 1991). Equally, we must point out that cognitive science has too pushed towards the importance of considering embodiment via ecological psychology, enactivism, and so on (“4E cognition”, see Hutchinson (2019)). Nevertheless, while one could argue that contemporary surges in sub-symbolic AI also may be seen as rejecting representation, the bases for much lauded branches like neural network based architectures take their inspiration quite directly from cognitivist thinking (even though the ‘neural’ in neural network is only loosely connected at best).

more recently, (Dreyfus, 1992; Collins, 2018)), such arguments are mostly grounded on the basis of impossibility or insufficiency of Cartesian AI visions to achieve what has been premised, rather than a full dismissal of this approach entirely as categorically mistaken (Button et al., 1995). This established ground level of the discourse is the first hurdle in concretely understanding conversational AI in-the-world.

Often within more technical communities AI participation is portrayed as the *present* state of affairs, albeit subject to various technical issues which need ironing out (assessments of the scale of such problems varies, see Mitchell, 2019). More rosy visions of participation are frequently provided by technology corporations; for example, presentations of conversational AI skirt between framing as an unobtrusive ‘assistant’ that blends into the domestic setting (e.g., Amazon Echo), through to more extreme cases that build the semblance of participation through outright deception of people involved in interactions¹⁰. Underlying this comfort with participation is technical communities’ apparent confidence in advances in machine ‘understanding’. For instance, Google CEO Sundar Pichai stated in his Google I/O 2018 conference keynote that their conversational AI “can actually understand the nuances of conversation”¹¹. More critical AI researchers, however, point to meaning being a serious and well-known obstacle within contemporary AI (Mitchell, 2019), which suggests unreflective use of such concepts can hinder proper understanding of a non-agentified view of AI-as-situated, inextricably, in-the-world. Whilst there are obvious strategic business reasons to maintain such a position, it forms a big chunk of mainstream hype around the promise of conversational AI.

In our view, these various versions of participation bear little chances of survival when measured against concrete social practices into which conversational AI becomes intertwined. In our fragments we have shown how AI in-the-world is embedded into continual regulation work by people interacting with and around them. Foundational work on AI in-the-world by Suchman, studying, as Puig de la Bellacasa (2011) neatly describes, “‘smart’ interfaces in software ‘assistant technology’” (Suchman, 2007; Suchman, 1987) demonstrated the nature of this embedding well before the advent of sub-symbolic, machine learning based AI, but the lesson remains the same:

“[Suchman] shows how the search for ‘autonomous machine agency’ and for the artefact that ‘speaks for itself’ contributes to an erasure of ‘artificiality’. In general, what disappears is ‘the human labour’ involved ‘in technological production, implementation [and] maintenance’.” Puig de la Bellacasa (2011)

So, we turned to the CASA paradigm to furnish us with something that tries to tackle participation in ways that seem to move beyond the limitations we describe. CASA is by

¹⁰ For instance, consider Google’s well-publicised ‘Duplex’ demo (see <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>, also Chen and Metz (2019)) in which a voice agent is used to automate tasks like booking a restaurant or a hairdresser appointment via phone. The agent uses disfluencies and hesitations to sound more ‘natural’ and thus ‘dupe’ the (human) recipient of the call. Also consider ELIZA (Weizenbaum, 1966). In all cases it is worth considering Garfinkel’s comments on the documentary method of interpretation (Garfinkel, 1967, ch.3) as

¹¹ See <https://www.youtube.com/watch?v=ogfYd705cRs>

our reckoning is probably one of the most well-adopted theories of human-machine participation particularly for fields of HCI, HRI and the wider gamut of HAI research. For starters CASA takes notice of the apparent sociality of situations with machines and pulls it out for investigation. In doing so CASA provides an explanation: that people employ (misplaced) *social* orientations to computational technologies like conversational AI, (mis)applying their internal stock of social rules in order to build their interactions with AI. At the same time CASA argues for a subtle shift in the idea of anthropomorphisation as phenomenon—thus rejecting the idea that people apply social behaviour to computational systems because they actually believe in a machines' status as equivalent to human participants¹². Instead, the solution is that anthropomorphism is a *result* of the aforementioned mindlessness mode of operation. As Reeves and Nass argue, people are “more simple than we often imagine” (Reeves and Nass, 1996, p. 254). This move is advantageous to AI-related research because it fits with a representationalist, cognitive conception of human actors in human-AI interactions: retaining humans as subject to AI-fuelled modelling of them. Often CASA seems to be taken at face value by many researchers or at most criticised for its implications (e.g., that it may not actually be desirable to leverage CASA and its claimed effects for the purposes of design (Shneiderman, 2020)).

In some aspects, CASA has a trivial resonance with the ethnomethodological argument we have pursued here in our fragments. We saw repeatedly how people interacting with conversational AI systems must bring to bear the variety of interactional *methods* they have at their disposal to accomplish anything with a conversational AI system. For the CASA perspective these methods might be thought of as its ‘social rules’. Across our fragments exhibiting moments of voice-based conversational AI interaction, people employ something *similar* to the methods normally relied upon to produce and therefore participate in everyday talk—whether that is in regulating turn-taking, managing address and self-selection of speakers, or repairing utterances. But the difference is that those methods must be *interactionally, situatedly tilted and tailored* by people in such a way as to fit with the noticeable strictures of the underlying technologies—automated speech recognition, natural language processing, conversation or dialogue management, and so on—and the ways in which those strictures manifest as the ‘interactional surface’ of conversational AI. At the same time those methods are designed by people as to be recognisable as regulating the machine—i.e., conversational AI. So, utterances are designedly input or designedly not input (‘hearing’, fragment 1). Such input reflexively shapes and is shaped by the local sequential contexture (fragment 2), or in other words it is accountable to the ongoing conversation and situation people find themselves in. Methods of managing politeness or domestic politics may fold in and adopt conversational AI as a *resource* to accomplish social interaction (fragments 3 and 4). And so on.

¹² We note that some research adopting a CASA perspective treats matters differently to CASA's original description (e.g., Nass et al. 1994), often emphasising a transformation in the object itself rather than seeing all the action as being on the human side in the CASA account. This is perhaps a result of a declarative paper title (“Computers are Social Actors”).

One of the glaring issues, though, is that the situational, contingent, always unfolding, context-shaped and context-renewing nature of interaction (Heritage, 1984, p. 280) does not feature in CASA (see Muller's argument about differentiation (2004)), much in the same way that such erasures are a signature in wider AI research and discourse (Mair et al. 2020; Puig de la Bellacasa, 2011; Suchman, 2007). Instead, CASA presents a view of participation which is based around *rules*. We can see this, for instance, in the suggestion that in "overlearning" of social rules which "automatically came into play" meant that "people were polite to a computer" (Nass and Moon, 2000). Yet we saw how the local sequential context of our example from fragment 3 and how the manifestation of such 'politeness' had a significant range of social relevancies (e.g., talk to devices is done around others and that talk is accountable). These had little to do with the applying a set of rules-from-without to the computational device in question (Alexa), and instead had everything to do with the situation of co-producing conversation and accounting for here-and-now conventions of what conduct is acceptable and what isn't (see Dourish (1996) for a related critique).

Relatedly, CASA's conception of participation as straightforwardly ("mindlessly") about social rules governing interactions with e.g., conversational AI, seems difficult to square with the socially organised nature of rules and rule-following. To take one example, it is argued that socially-oriented participants apply rules of politeness like "other-praise is more valid than self-praise" (Nass et al., 1994) to the computational systems they encounter in certain situations. In contrast, an ethnomethodological treatment of rules and rule-following via Wittgenstein points out how 'rules' are inseparable from their production. Looking for rule-based explanations of why something was said ignores the retrospective-prospective index in which utterances sit. In this example, we can imagine producing "other-praise" which is e.g., sarcastic or ironic, and instead uses this example 'rule' in order to produce such irony or sarcasm. This is what happens in our example between Isabel and Nik (fragment 3). Rules are not a 'script' that is applied by people but rather acts as an ongoing resource in the production of action, reflexively.

It is high time to move on from misleading aphorisms like "computers are social actors" as a useful approach to conceptualising human-machine participation as we begin to see widespread resurgence of interest in human-AI interactions. Reification and agentification will not help us address the significant challenges posed by AI in-the-world. We need to take seriously AI systems' *constitutively* socially situated production. In line with this programme, our work here has shown how crucial the embeddedness of conversational AI systems in-the-world turns on *regulation work*, from designing inputs to managing device outputs in and as the embedding into local courses of action. This is the not inconsiderable interactional work of embedding a (domestic) object into the various situations of life that one finds oneself in—perhaps little different to 'any' household object in some senses, but ones in which the means of their manufacture (Crawford and Joler, 2018) and infrastructural entanglements (Taylor, 2015) become far more pertinent. We need a respecification of how and in what ways we talk about conversational AI just as we do for AI in general. The onus is on the human to make the vagaries of conversation AI frontend design and infrastructures, such as they present / reveal themselves in the course of interactions, to become an accountable part of

ongoing circumstances. This discussion applies to other technologies of recognition all the same—of voice, face, gesture, other biometrics—that are adopted to drive interaction in one way or another. Technologies of recognition do not ‘recognise’ in any conventional sense, and so people encountering them in whatever situation must work to *regulate* them, whether that is in designing appropriate input, managing and embedding what comes out, or locating ways to avoid such things. Regulation work for people is thus simultaneously regulation *of* these systems and *of themselves*—designing input, hearability, etc. Elsewhere bodily work employed to get technologies of recognition working has been referred to as “disciplining” (Tekin and Reeves (2017)). In many respects, with all the connotations that come along with such a term, this applies too to regulatory work.

5. Conclusion

We’ve taken a critical stance to examining interaction with conversational AI as ‘participation’. This has knock-on implications for understanding AI as constituted in-the-world and the mundane regulation work by people that is entailed in this. Often we see people conceptualised within fields designing and studying human-AI interactions as some kind of ‘participant’, acting ‘mindlessly’, numbly applying social rules to the technology. The critical issue with CASA is that the approach examines interactions with computational systems, but fail to consider how the actions unfold in and as social organisation. CASA approaches attempt to infer meaning of actions from the inner, the cognitive, but find none, so default to distinctions constructed by Langer’s mindfulness / mindlessness (Langer, 1992). The problem is that this approach treats context as separate from action and so the proper import of these actions is lost.

By re-orienting towards an ethnomethodologically-informed picture of conversational AI, aided by empirical reminders, we see how orderly, interwoven, embedded, and regulated conversational AI technologies are for people that live with them. This forms a response to CASA, mindlessness, and adherence to social rules by respecifying ‘participation’ as situated and regulatory, in which meaning is produced interactionally.

References

- James F Allen, Donna K Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent. 2001. Toward Conversational Human-Computer Interaction. *AI Magazine* 22, 4 (Dec.2001), 27. <https://doi.org/10.1609/aimag.v22i4.1590>
- Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Paper 3, 1–13. DOI:<https://doi.org/10.1145/3290605.3300233>

John L. Austin (1962). *How to do things with words*. Cambridge: Harvard University Press.

Maria Puig de la Bellacasa. 2011. "Matters of Care in Technoscience: Assembling Neglected Things." *Social Studies of Science* 41, no. 1 (February 2011): 85–106. <https://doi.org/10.1177/0306312710380301>.

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '02). Association for Computing Machinery, New York, NY, USA, 415–422. DOI:<https://doi.org/10.1145/503376.503450>

Erin Beneteau, 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* (Glasgow, Scotland Uk) (CHI '19). ACM, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300473>

K. W. Bowyer, M. C. King, W. J. Scheirer and K. Vangara. 2020. The "Criminality From Face" Illusion. In *IEEE Transactions on Technology and Society*, vol. 1, no. 4, pp. 175-183, Dec. 2020, doi: 10.1109/TTS.2020.3032321.

Rodney A. Brooks. 1991. Intelligence without representation. *Artificial Intelligence* 47(1-3), January 1991, pp. 139-159. Elsevier. DOI: [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Computation and Language* (Jul 2020), 75. arXiv:2005.14165

Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research* (January 2018), 77-91.

Graham Button, John R. Lee, Jeff Coulter, and Wes Sharrock. 1995. *Computers, Minds and Conduct*. Blackwell Publishers, Inc., USA.

Teresa Castle-Green, Stuart Reeves, Joel E. Fischer, and Boriana Koleva. 2020. Decision Trees as Sociotechnical Objects in Chatbot Design. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (CUI '20). Association for Computing Machinery, New York, NY, USA, Article 27, 1–3. DOI: <https://doi.org/10.1145/3405755.3406133>

Narae Cha, Auk Kim, Cheul Young Park, Soowon Kang, Mingyu Park, Jae-Gil Lee, Sangsu Lee, and Uichin Lee. 2020. Hello There! Is Now a Good Time to Talk? Opportune Moments for Proactive Interactions with Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 74 (September 2020), 28 pages. DOI: <https://doi.org/10.1145/3411810>

Brian X. Chen and Cade Metz. 2019. Google's Duplex Uses A.I. to Mimic Humans (Sometimes). *The New York Times*, May 22, 2019. <https://www.nytimes.com/2019/05/22/technology/personaltech/ai-google-duplex.html> (retrieved 27th March 2021)

Harry Collins. 2018. *Artificial Intelligence: Against Humanity's Surrender to Computers*. Polity Books, Cambridge, 2018.

Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (Vienna, Austria) (MobileHCI '17). Association for Computing Machinery, New York, NY, USA, Article 43, 12 pages. <https://doi.org/10.1145/3098279.3098539>

Kate Crawford and Vladan Joler. 2018. Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources. *AI Now Institute and Share Lab*, (September 7, 2018). <https://anatomyof.ai>

Paul Dourish. 1996. Review of "The Media Equation" (Cliff Nass and Byron Reeves, 1996). In Apple Labs Review.

Hubert L Dreyfus. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason* (rev ed.). MIT Press, Cambridge, MA, USA.

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

Harold Garfinkel. 1988. Evidence for Locally Produced, Naturally Accountable Phenomena of Order, Logic, Reason, Meaning, Method, etc. In and as of the Essential Quiddity of Immortal Ordinary Society, (I of IV): An Announcement of Studies. *Sociological Theory* 6, 1 (spring 1988), 103–109.

Radhika Garg and Subhasree Sengupta. 2020. He Is Just Like Me: A Study of the Long-Term Use of Smart Speakers by Parents and Children. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 11 (March 2020), 24 pages. DOI: <https://doi.org/10.1145/3381002>

Timnit Gebru. 2019. Oxford handbook on AI ethics book chapter on race and gender. *Computers and Society* (aug 2019). arXiv:1908.06165

Erving Goffman. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Charles Goodwin. 1979. The Interactive Construction of a Sentence in Natural Conversation. *Everyday Language: Studies in Ethnomethodology* 97 (1979), 121.

Erika Hall. 2018. *Conversational design. A Book Apart*, New York, NY, USA.

Charles Hannon. 2016. Gender and Status in Voice User Interfaces. *Interactions* 23, 3 (April 2016), 34–37. <https://doi.org/10.1145/2897939>

John Heritage. 1984. *Garfinkel and ethnomethodology*. Cambridge [Cambridgeshire: Polity Press.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Proceedings of the IEEE International Conference on Computer Vision (Santiago, Chile) (ICCV '15)*. IEEE, New York, NY, USA, 1026–1034. <https://doi.org/10.1109/ICCV.2015.123>

Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5040–5052. DOI: <https://doi.org/10.1145/3025453.3025765>

Phil Hutchinson. 2019. The Missing 'E': Radical Embodied Cognitive Science, Ecological Psychology and the Place of Ethics in Our Responsiveness to the Lifeworld. In *Moral Foundations of Philosophy of Mind*, Joel Backström, Hannes Nykänen, Toivakainen Niklas, and Thomas Wallgren (Eds.). Palgrave Macmillan, Cham, Switzerland, 103–127. https://doi.org/10.1007/978-3-030-18492-6_4

Chandra Khatri, Anu Venkatesh, Behnam Hedayatnia, Raefer Gabriel, Ashwin Ram, and Rohit Prasad. 2018. Alexa Prize — State of the Art in Conversational AI. *AI Magazine* 39, 3 (sep2018), 40–55. <https://doi.org/10.1609/aimag.v39i3.2810>

Lone Koefoed Hansen and Peter Dalsgaard. 2015. Note to self: stop calling interfaces "natural". In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives (CA '15)*. Aarhus University Press, Aarhus N, 65–68. DOI: <https://doi.org/10.7146/aahcc.v1i1.21316>

Ellen J. Langer. 1992. Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition: An International Journal*, 1(4), 289–305. [https://doi.org/10.1016/1053-8100\(92\)90066-J](https://doi.org/10.1016/1053-8100(92)90066-J)

Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart

Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (November 2018), 31 pages. DOI:<https://doi.org/10.1145/3274371>

Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (New Brunswick, NJ, USA) (CHIIR '18)*. ACM, New York, NY, USA, 265–268. <https://doi.org/10.1145/3176349.3176868>

Michael Lynch. 2000. Against Reflexivity as an Academic Virtue and Source of Privileged Knowledge. *Theory, Culture & Society*. 2000;17(3):26-54. doi: <https://doi.org/10.1177/02632760022051202>

Michael Mair, Phillip Brooker, William Dutton, and Philippe Sormani. 2020. “Just What Are We Doing When We’re Describing AI? Harvey Sacks, the Commentator Machine, and the Descriptive Politics of the New Artificial Intelligence.” *Qualitative Research*, (December 2020). <https://doi.org/10.1177/1468794120975988>.

Bernard McGrane. 1994. *The Un-TV and the 10 MPH Car: Experiments in Personal Freedom and Everyday Life*. Fort Bragg CA: The Small Press.

Melanie Mitchell. 2019. *Artificial Intelligence: A Guide for Thinking Humans*. Penguin, UK.

Robert J Moore, Margaret H Szymanski, and Raphael Arar (Eds.). 2018. *Studies in Conversational UX Design (1 ed.)*. Springer International Publishing. 204 pages. <https://doi.org/10.1007/978-3-319-95579-7>

Michael Muller. 2004. Multiple paradigms in affective computing, *Interacting with Computers*, Volume 16, Issue 4, August 2004, Pages 759–768, <https://doi.org/10.1016/j.intcom.2004.06.005>

Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. Association for Computing Machinery, New York, NY, USA, 72–78. DOI: <https://doi.org/10.1145/191666.191703>

Clifford Nass and 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. and Carney, P. 1999. Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1. *Journal of Applied Social Psychology*, 29: 1093-1109. <https://doi.org/10.1111/j.1559-1816.1999.tb00142.x>

Thao N. Phan. 2019. Amazon Echo and the aesthetics of whiteness. *Catalyst: Feminism, Theory, Technoscience*, 5 (1), 1-39 (April 1, 2019). DOI: <https://doi.org/10.28968/cftt.v5i1.29586>

Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 640, 1–12. DOI: <https://doi.org/10.1145/3173574.3174214>

Martin Porcheron. 2019. "Understanding conversation around technology use in casual-social settings", PhD thesis, University of Nottingham.

Martin Porcheron. 2021. What's in a name and does CUI matter? In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (CUI '21). Association for Computing Machinery, New York, NY, USA, Article 24, 1–3. <https://doi.org/10.1145/3469595.3469619>

Stuart Reeves, Martin Porcheron, and Joel Fischer. 2018. 'This is not what we wanted': designing for conversation with voice interfaces. *interactions* 26, 1 (January - February 2019), 46–51. DOI:<https://doi.org/10.1145/3296699>

Stuart Reeves. 2019. Conversation considered harmful? In *Proceedings of the 1st International Conference on Conversational User Interfaces* (CUI '19). Association for Computing Machinery, New York, NY, USA, Article 10, 1–3. DOI: <https://doi.org/10.1145/3342775.3342796>

Byron Reeves and Clifford Nass. 1996. The Media Equation: How People Treat Computers, Television, and New Media Like Real People. Number 2. Cambridge university press Cambridge, UK, Chapter Politeness, 19–36.

Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 4 (jan 1974), 696–735. <https://doi.org/10.2307/412243>

Harvey Sacks. 1992. *Lectures on conversation*. Oxford: Blackwell.
Sacks 1992b: 438

Emanuel A Schegloff. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis (Vol. 1)*. Cambridge University Press, Cambridge, UK. DOI: <https://doi.org/10.1017/CBO9780511791208>

Michael Searles. 2019. Shut it, Siri: Most of us are really rude to our smart speakers. *City AM*, May 10, 2019. <https://www.cityam.com/shut-siri-why-so-rude-our-smart-speakers/> (retrieved 27th-March 2021)

Ben Shneiderman. 2020. Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 12(3), pp. 109-124.
DOI: <https://doi.org/10.17705/1thci.00131>

Elizabeth Stokoe and Rein Ove Sikveland. 2020. The backstage work negotiators do when communicating with persons in crisis. *Journal of Sociolinguistics* 24, 2 (apr2020), 185–208. <https://doi.org/10.1111/josl.12347>

Lucy A. Suchman. 1987. *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press, USA.

Lucy A. Suchman. 2006. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, USA.

Selina Jeanne Sutton. 2020. Gender Ambiguous, Not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Bilbao, Spain) (CUI '20)*. ACM, New York, NY, USA, Article 11, 8 pages. <https://doi.org/10.1145/3405755.3406123>

Alex Taylor. 2015. After interaction. *interactions* 22, 5 (September-October 2015), 48–53. DOI: <https://doi.org/10.1145/2809888>

Burak S. Tekin and Stuart Reeves. 2017. Ways of Spectating: Unravelling Spectator Participation in Kinect Play. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1558–1570. DOI: <https://doi.org/10.1145/3025453.3025813>

A. J. O'Toole, P. Jonathon Phillips, F. Jiang, J. Ayyad, N. Penard and H. Abdi, "Face Recognition Algorithms Surpass Humans Matching Faces Over Changes in Illumination," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1642-1646, Sept. 2007, doi: 10.1109/TPAMI.2007.1107.

Shari Trewin, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. 2019. Considerations for AI Fairness for People with Disabilities. *AI Matters* 5,3 (Dec.2019), 40–63.
<https://doi.org/10.1145/3362077.3362086>

Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (Jan. 1966), 36–45. DOI: <https://doi.org/10.1145/365153.365168>

M J Wolf, K W Miller, and F S Grodzinsky. 2017. Why We Should Have Seen That Coming: Comments on Microsoft's Tay "Experiment," and Wider Implications. *The ORBIT Journal* 1, 2 (2017), 1–12. <https://doi.org/10.29297/orbit.v1i2.49>

Julia Carrie Wong. 2019. 'A white-collar sweatshop': Google Assistant contractors allege wage theft. *The Guardian*, June 25, 2019.

<https://www.theguardian.com/technology/2019/may/28/a-white-collar-sweatshop-google-assistant-contractors-allege-wage-theft> (retrieved 27th March 2021)

Wayne Xiong, Jasha Droppo, Xuedong Huang, Frank Seide, Mike Seltzer, Andreas Stolcke, Dong Yu, and Geoffrey Zweig. 2016. Achieving human parity in conversational speech recognition. arXiv preprint arXiv:1610.05256 (2016).