Reeves, Stuart (2017) Some conversational challenges of talking with machines. In: Talking with Conversational Agents in Collaborative Action, Workshop at the 20th ACM conference on Computer-Supported Cooperative Work and Social Computing (CSCW '17), 25 February - 1 March 2017, Portland, Oregon, USA.

Some Conversational Challenges of Talking with Machines

Stuart Reeves

Mixed Reality Lab

School of Computer Science

University of Nottingham

Nottingham, UK

stuart@tropic.org.uk

Abstract

A surge of interest in the capabilities of so-called 'conversational' technologies—both from research and industrial contexts—furnishes CSCW and HCI with opportunities to enrich and leverage its historic connection to conversation analysis (and relatedly, ethnomethodology) in novel ways. This paper explores a number of preliminary interactional troubles one might encounter when 'talking to' conversational agents, and in doing so sketches out possible routes forward in the empirical study of agents as collaborative technologies, as well as touching on further conceptual challenges that face research in this area.

Author Keywords

Conversational agents; conversation analysis; ethnomethodology.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Recently, a broad range of what are often marketed as 'conversational' technologies have been productised and made commercially available. These span various forms and descriptors: personal assistants (Siri,

Amazon Echo) and their variants (e.g., LG's Hub Robot), chatbots (cf. in Slack, Facebook Messenger), and virtual humans (e.g., healthcare applications such as SimSensei [4]). Connected with this are various possible interaction modalities: voice, text, gesture, and facial recognition. Contributing drivers of this shift have been the growing tractability and commoditisation of machine learning (ML) techniques, developments in cloud computing infrastructures, and the establishment of and access to pertinent 'big data' to feed ML systems. It is often suggested conversational technologies are "transformative", and represent a shift towards a more "natural" computing paradigm [12]. This growing interest in what I'll class collectively as *conversational agents* presents new challenges to CSCW and HCI—from the pragmatics of the study and design of collaborative interactions between heterogeneous human-agent groupings [10], to deeper questions regarding the conceptual character of human-machine 'conversation' itself [2], as well as the possible implications of these for broader notions of human-machine 'interaction' in HCI research [5].

Firstly, I must note that conceptualising the design of interactions between human and machine as a conversation of sorts—or perhaps more commonly as a 'dialogue' in HCI—is very much *not* a novel matter (e.g., see [5]). Secondly, we also must remind ourselves that the dissonance embedded in the use of this metaphor (i.e., as dialogue, or conversation) is *not* new either; Suchman's work on the use of "artificially intelligent 'expert'" photocopiers [17, p. 4; 18] articulated the shape of this problem well. Thirdly and finally, the introduction of conversation analysis to this space of conversational agents is also not a novel move; rather, it was preliminarily addressed over 20

years ago in CSCW (see [8, 2]), although since then interest seemed to have dissipated.

The final element at play in this picture is texture of the relationship between HCI/CSCW and AI/ML's concerns. HCI—and by overlap, CSCW—has consistently been an advocate for the importance of 'the user' (via evolving notions of 'usability', 'user-experience', 'user-centredness', or 'human-centredness' etc. This infuses HCI and CSCW with a specific orientation to the design of human-machine dialogue/conversation, one that is strongly weighted in favour of developing deep understanding human practices. Technical development in this sense is subordinate to such understandings, which themselves offer the primary path to establishing a *fit* between human and machine. In contrast, AI/ML-informed dialogic approaches to human-machine interaction have tended to take an opposite perspective, seeking to *inhere* technologies with human-like capabilities and thus offer an alternative route to a better fit (e.g., the development of 'smart' machines) [1]. Yet without negating the value of such technical developments and their potential applications, some of the *conceptual foundations* of AI/ML perspectives have been challenged [2, 1]. Given the long-running nature of this relationship, it remains to be seen if more recent shifts in underlying technologies and dispersal of conversational agents into *everyday settings* might help refine, clarify or revise these critiques in ways that inform disciplinary relations between HCI/CSCW and AI/ML.

Revisiting and reviving past debates such as these in order to make sense of present excitement and hyperbole around conversational agents is critical to CSCW and HCI themselves (re)turning to address the

Figure 1: The Amazon Echo (with ring on top illuminated to indicate that the device is 'listening')

topic as a research matter. Informed by conversation analysis (CA) [13], this paper attempts to suggest some possible lines of investigation by contrasting human *conversation*—itself a practical, collaborative achievement of competent speakers—with the organisation of 'talking' to / with conversational agents. In order to remind ourselves of this conceptual distinction, from this point I will mark a differentiation between 'conversation' and 'talk' between people, and interactions between humans and machines (i.e., conversational agents). To do this I will bracket any conceptually muddy use with an alternate typography LIKE THIS to indicate its provisionality. Fundamentally this is because, like Button et al. [2], I am wary of the status of any descriptions of and ascriptions to the conversational agent as a CONVERSATIONAL PARTNER.

Approach

When we attempt to design machines that TALK to their users, a broad range of interactional troubles arise for designers to tackle. In this paper I focus on use of the Amazon Echo (Figure 1) as merely one possible starting place for such investigations—to wit I provide three transcripts in order to assist the points made. The Echo is a 'smart' speaker intended for use in the home, assisting with activities such as answering queries (e.g., the weather), playing music, ordering products, and so on. The device provides voice interaction and minimal other modalities, specifically a status light that indicates when the Echo has HEARD something (which is perhaps evident in the 0.7s and 0.5s pauses at the start of Transcripts 1 and 2 (line 01)). Here I am not presenting a systematic study of methods of talk (and, speaking of the Echo itself, METHODS OF TALK) that emerge from dyadic or multiparty interaction with a conversational agent. Instead I want to raise a number

of preliminary considerations that arose from my own use of the Echo; these touch on well-known topics from conversation analysis. Accordingly I raise **four** initial conversational troubles that I encountered while TALKING TO the Echo, and reflect on three transcripts[1] of my own recorded interaction with the Echo (Transcripts 1-3, referred to as T1, T2 and T3; Transcript 3 includes another person, J).

Trouble 1: Action-formation

In everyday interaction conversationalists produce or 'formulate' particular actions in talk (e.g., inviting, questioning, or commanding), such that they are routinely recognisable to co-conversationalists as *those particular actions* and thus distinguishable from an extremely broad range of possible other classes of actions [16]. For instance, this might be how a question is heard as-a-question, hence how it comes to be sequentially organised as a first-pair-part of a question-answer adjacency pairing. The problem of action-formation is an ongoing collaborative endeavour in talk where speakers project and formulate future actions, such as their upcoming turn-at-talk, while the current speaker produces their utterance [7]. Further, action-formation connects with recipient design, i.e., the ways in which such actions are tailored to the recipient and relevant situational specificities.

The question is, how does a user formulate actions for a conversational agent? Action-formation is one of the troubles that I encountered immediately upon attempting to use the Echo. It begins in the problem of the user developing an understanding for what

---

[1] It is also intended that these transcripts can provide resources for further discussion the workshop this paper is submitted to.

```
01 S: Alexa? ↑what's: (0.5) er:::
02    what's the best way (0.3)
03    to get (.) to: (.)
04    Heathrow::=airport
05    (1.9)[ from Nottingham
06 A:      [sorry↓          (.)
07    [I'm not sure
08 S: [Alexa?      >Alexa stop<
```

Transcript 1: Asking a question (S=the author, A=Amazon Echo)

```
01 S: Alexa (0.7) er:: (0.5) add
02    to my calendar.
03    (2.0)
04 A: £got it£ (0.8) for when?
05 S: er:: (0.3) <twenty fifth
06    of> February twenty
07    seventeen
08    (1.6)
09 A: and what's the name of the
10    event?
11    (1.0)
12 S: eh (.) talking with
13    conversational agents in
14    collaborative action
15    workshop
16    (2.0)
17 A: that's talking with
18    conversation agents in
19    collaborative action
20    workshop on ↑Saturday (.)
21    the ↑twenty fifth of
22    February °at° five pm (.)
23    right?
24    (0.3)
25 S: .hh (0.4) a:t nine to five
26    pm.
27    (3.2)
28 A: that's talking with
29    conversation agents in
30    collaborative action
31    workshop on ↑Saturday (.)
32    the ↑twenty fifth of
33    February °at° four fifty one
34    pm (.) right?
35    (0.4)
37 S: no: (0.6) nine am=
38 A: =[okay    cancelled
39 S:  [(.) to five-    uh
```

Transcript 2: Adding to the calendar using the Echo

constitutes the adequate production of an action such as a 'queryable' that it will be RECOGNISABLE as that thing (e.g., a question about travel, see T1) by the Echo. Or, in more vernacular terms the trouble for any user of an agent is how to solve the problem of "what I can say?" and "how can I say it?", and in such a way that the conversational agent will successfully parse (or MAKE SENSE OF or UNDERSTAND) the utterance(s)? This is particularly sticky when the agent CO-CONVERSATIONALIST in question is seemingly unable to produce, in an ethnomethodological sense, any *reflexive action*, i.e., actions that do things whilst simultaneously and inextricably *showing* that they are doing those particular things, e.g., turn-taking in conversation and the production of transition relevance places in-and-through utterances—all of which the Echo dismally fails to do.

Trouble 2: Speaker selection

Speakers have sophisticated sets of methods for managing who is 'in play' in a conversation via methods of selection. These stem from two main methods: "current speaker selects next" and "self-selection" [14]. These are of course sensitive to pauses and other temporal features in talk. Within these broad forms, there are a further methods such as pre- or post-positioning names for managing turn-allocation [6].

I found the Echo to provide very limited resources for speaker selection. Apart from specific cases, the Echo largely does not SELF-SELECT and as such the user of it must *initially* always use the 'wake up' word, e.g., "Alexa (.) what's the time"—see line 01 in T1 and T2. In other words, this is analogous to current speaker selects next with a pre-positioning of the 'name'. It is not possible to post-position the name with the Echo,

e.g. "what's the time (.) Alexa?". This means there is a lack of resources available to the user of the Echo. For example, in talk a conversationalist can use post-positioning to do interactional work so as to manage speaker allocation multiparty circumstances [10]; for instance, post-positioning names *in the course of* the production of an utterance may be used to, say, rapidly shift recipiency [6] of said utterance to a particular speaker (or in our case, agent).

Trouble 3: Turn-taking, adjacency pairs, and repair

The study of the organisation of turn-taking in talk (as well as other actions) has been a key focus conversation analysis. Seminal work by Sacks, Schegloff and Jefferson examined how conversationalists manage and regulate turns-at-talk between themselves [14] (turn-taking and their accomplishment between speakers is also strongly tied to methods of speaker selection). The adjacency pair—such as question-answer pairs—are one such sequence-organisational method.

As I came to use the Echo, TURN-TAKING with it felt like a somewhat impoverished affair. This is perhaps because, as with other conversational agents, the majority of the turn-taking the Echo PARTICIPATES IN is based in (highly fragile) question-answer adjacency pairings [10] that struggle with utterance repair at a basic level. T1 lets us refine these observations a little; in T1, S uses the wake up word itself with a questioning prosody in line 01 before producing and repairing a question. Some of the normative orientations to turn-taking identified by Sacks et al. [14] and many others since seem suspended for the Echo. Further, as T2 shows, there are (interactionally-speaking) very long

```
01      ((responding to a question
        about the weather in
        Denver, USA; S and J are
        testing A's capabilities))
02 A: today's forecast has snowy
03    weather (.) with a high of
04    mi[nus ten degrees
05 S:    [>hang on< it jus–=
06 A: =and a low of minus
07    [(twenty) degrees
08 S: [it just said
09    ((coughing))
10 J: Alexa (.) what about next
11    week
12    (1.8)
13 A: ↑ here's the weather in
14    Nottingham for the next
15    [seven days
16 J: [ah:↓
```

Transcript 3: Enquiry about the weather (involving another person, J)

pauses (round-trip latency etc.) at turn transition relevance places (lines 03, 08, 16 etc.). This is in contrast with Sacks et al. [14] who point out that the "vast majority of transitions" in talk involve either no or minimal gap or overlap between turns. T1 then reveals how this becomes sequentially-relevant: at line 05, S attempts to take advantage of one such usually long pause to self-select as speaker and retain his turn, producing a repair to his question ("from Nottingham"). However this belated repair does not then shape the Echo's SECOND PAIR-PART APOLOGY on line 06, and neither does the emergence of overlaps (lines 05-06) do so.

Trouble 4: Sequence-organisational problem

Conversation analysis has extensively documented the ways in which conversationalists collaboratively work to manage and maintain the coherence of their conversations [16]. For example, various methods are employed by speakers to project (or 'foreshadow') certain future sequential turns-at-talk. A simple case is where conversationalists collaboratively work to shut down a conversation using a pre-closing "oh-kay" or "we::ll" to project the possible end of a phone call [15].

The problem of sequence-organisation arose with the Echo during moments where I assumed that subsequent utterances addressed to the Echo are HEARD as sequentially tied to prior ones. T2 and T3 provide two contrasting examples of this. In T2, the Echo ASKS a sequence of further questions (lines 04, 09-10) after the initiation by S in order to fill out the calendar event details. Her, the question-answer adjacency pairs between S and the Echo in lines 04-15 *appear* to maintain sequential coherence, albeit one that is clearly fragile (see the breakdown in lines 37-39). Compare this with T3 where J asks a question, lines 10-11, that

is sequentially tied to a prior question-answer pair (the second-pair-part being lines 02-04, and 06). As a competent conversationalist, J's response cry in line 16 marks heard trouble with the (mundane) unaccountability of the Echo's switch from Denver, USA to Nottingham, UK.

Conceptual and methodological challenges

In the prior sections I provided a brief, terse sketches of four initial conversational troubles encountered with the Amazon Echo. Beyond this there are two broader challenges that I think are pertinent; these include: 1. language and concomitant conceptual troubles when we apply concepts from understandings of human conversation to agent-based CONVERSATIONALISTS; and 2. methodological troubles regarding the relevance of the conversation analysis literature to human-agent interactions.

On point (1), I ask the following: can we say we are 'talking to' an agent *in the same sense* that we say we talk to one another? Might we call the conversational agent a speaker? Is a dyadic interaction between a user and an agent something that might reasonably be called a 'conversation'? There are considerable conceptual challenges to address here, many of which, as noted earlier, have perhaps been encountered before in a prior confluence between conversational agents and CA over 20 years ago [2]. Moving on from this perhaps involves a deliberative plotting out of the ways in which we deploy understandings of metaphor from language to the design and study of conversational agents. I would also argue that looking to our research participants' *own* lay sociological reasoning, analysis, and orientation to agents could be a productive way to address these questions.

Turning to point (2), I ask what the relationship might be between the methods deployed in conversation, as extensively documented by the conversation analytic literature, and those (new? adapted?) methods emerging from interactions as conversational agents come to feature in dyadic and multiparty settings. While it seems tempting to directly lift understandings of the organisational structures of, say, conversational repair and apply them to new contexts (e.g., repair in search found in Moore, Churchill and Kantamneni [9]), there are possible concomitant dangers in doing so. The application of CA to conceptually vexed circumstances (i.e., the above question of agents as ratified 'speakers') might lead to a familiar criticism levelled at the methods of the social sciences, specifically that of Cicourel's "deterministic grid" [3, p. 109] where social phenomena are rendered in terms of a predetermined framework (conversation analytic findings) that can occlude the phenomena of multiparty and individual human-agent interaction for what it is.

Conclusion

In this paper I have attempted to provide some initial thoughts on the relevance of conversation analysis to the study of conversational agents. Yet there are many central aspects of conversation analysis which have not been mentioned (such as membership categorisation) yet could find some relevance in studies of human-agent interaction. Further, my account is lacking in how such an endeavour—i.e., the conceptual and empirical study of conversational agents—ties to design. While I do not wish to return to the idea of developing computational models of conversation based in conversation analysis (see Luff et al. [8]), at the same I think shifts in the technology landscape and everyday dispersal of agents does recommend a second look.

References

1. Alan F. Blackwell. 2015. Interacting with an inferred world: the challenge of machine learning for *humane* computer interaction. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives* (AA '15). Aarhus University Press 169-180. http://dx.doi.org/10.7146/aahcc.v1i1.21197

2. Graham Button, Jeff Coulter, John R. E. Lee and Wes Sharrock. 1995. *Computers, Minds and Conduct*. Cambridge: Polity.

3. Aaron Cicourel. 1964. *Method and Measurement in Sociology*. New York: The Free Press of Glencoe.

4. David DeVault, et al. 2014. SimSensei kiosk: a virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (AAMAS '14). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1061-1068.

5. Kasper Hornbaek and Antti Oulasvirta. 2017. What is interaction? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (May 2017), ACM Press.

6. Gene H. Lerner. 2003. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society*, 32, 177–201.

7. Stephen C. Levinson. 2012. Action Formation and Ascription. In Jack Sidnell and Tanya Stivers (eds.), *The Handbook of Conversation Analysis*, Blackwell Publishing.

8. Paul Luff, David Frohlich and Nigel Gilbert (Eds.). 1990. *Computers and conversation*. Academic Press.

9. Robert J. Moore, Elizabeth F. Churchill, and Raj Gopal Prasad Kantamneni. 2011. Three sequential positions of query repair in interactions with internet search engines. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (CSCW '11). ACM, New York, NY,

USA, 415-424.
http://dx.doi.org/10.1145/1958824.1958889

10. Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '17). ACM, New York, NY, USA. https://doi.org/10.1145/2998181.2998298

11. Stuart Reeves and Barry Brown. 2016. Embeddedness and sequentiality in social media. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 1052-1064. http://dx.doi.org/10.1145/2818048.2820008

12. The Economist, "Now we're talking", Jan 7th 2017. http://www.economist.com/news/leaders/2171383 6-casting-magic-spell-it-lets-people-control-world-through-words-alone-how-voice

13. Harvey Sacks, Gail Jefferson, and Emanuel Schegloff. (Eds.) 1992. *Lectures in Conversation*. Oxford: Blackwell.

14. Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, Vol. 50, No. 4, Part 1 (Dec., 1974) , pp. 696-735.

15. Emanuel A. Schegloff and Harvey Sacks. 1973. Opening up Closings. *Semiotica* 8 (4):289-327, De Gruyter Mouton.

16. Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis, Volume 1*, Cambridge: Cambridge University Press.

17. Lucy Suchman. 1985. *Plans and situated actions: The problem of human-machine communication*. Tech report ISL-6 (PhD thesis), Palo Alto Research Center, February 1985.

18. Lucy Suchman. 2007. *Human-Machine Reconfigurations.* Cambridge University Press, New York.