

Users, Bystanders and Agents: Participation Roles in Human-Agent Interaction

Antonia L. Krummheuer^(✉)

Aalborg University, Aalborg, Denmark
antonia@hum.aau.dk

Abstract. Human-agent interaction (HAI), especially in the field of embodied conversational agents (ECA), is mainly construed as dyadic communication between a human user and a virtual agent. This is despite the fact that many application scenarios for future ECAs involve the presence of others. This paper critiques the view of an ‘isolated user’ and proposes a micro-sociological perspective on the participation roles in HAI. Two examples of an HAI in a public setting point out (1) the ways a variety of participants take part in the interaction, (2) how the construction of the participation roles influences the construction of the agent’s identity, and (3) how HAI, as a mediated interaction, is framed by an asymmetric participation framework. The paper concludes by suggesting various participation roles, which may inform development of ECAs.

Keywords: Embodied conversational agent · Human-agent interaction · Participation role

1 Introduction

An embodied conversational agent is a computer interface in the form of a “virtual human” ([4], p. 39). ECAs are said to have social, embodied and linguistic abilities so that they can engage in ‘human-like’ face-to-face interaction with the aim of developing “computers that untrained users can interact with naturally” ([4], p. 60). Core studies have conceptualized the ECA interaction as a dyadic conversation between one agent and one user (e.g. [2, 5]) and do not address how to include several human participants simultaneously. This is a striking omission, as standard application scenarios tend to be settings in which other people are likely to be present and thus influence the interaction, as in a museum or clinical settings ([13, 18]). One exception is the anti-bullying system in [23], in which several agents with different participation roles were created. The system, however, is designed for individual use, even though it is tested in a school setting. The scenario in [20] is the only study in which an agent addresses two users at the same time.

This paper sheds light on situations in which many participants are present during the HAI and shows how the identities of the participants (users, bystanders and the

I am grateful to the Artificial Intelligence Group (Bielefeld University) and the participants of the recorded data for their cooperation. I also want to thank the reviewers for their constructive comments. This research has been supported by the German Research Foundation.

agent) shape and are shaped by the ways the parties take part in the ongoing production and interpretation of the HAI. This paper adds to the discussion of the term ‘user’ in HCI, which is often criticized for its simplistic perspective on the human relationship with the computational machine. While the term ‘user’ evokes the idea of a ‘typical user’ often conceptualised as interacting in isolation, [10, 12] point out that systems are used by several users with different background knowledge – such as experts and novices, developers and test-persons, or ‘marginal’ people, such as those maintaining the machine. Furthermore, the different practices of the individual’s usage, the routines and practices of a task and the social-material context of e.g. the company shape the understanding of the user [1].

This paper understands ‘user’ and ‘agent’ as categories that describe participation roles within the participation framework of an HAI. These concepts stem from an ethnomethodological point of view, taking the perspective of participants who mutually negotiate their own and each other’s action and identity in the ongoing production and interpretation of interaction, as in Suchman’s classical study on the user’s situated interpretation of an interactive copying machine [22]. A micro-sociological concept of participation ([7, 9]) is applied to two video recordings of a HAI in a public setting. While in both examples a communication problem is solved successfully, the encounter is framed by an asymmetric and mediated participation framework, in which different kinds of participants align in different ways to the production and interpretation of the event. In one, the user solves a communication problem together with the agent. In this case the agent is treated as an equal communication partner. In the other, the user solves the problem involving bystanders as helpers, while the agent is excluded from the interaction and treated as an unequal partner. On the basis of these examples, this paper emphasises that the construction of categories such as ‘user’ and ‘agent’ is situated, dynamic and reflexive, and suggests a variety of participation roles in HAI that may inform the development of ECA in social settings.

2 Participation Framework and Participation Status

Goffman stresses that interactions are *socially situated* [8], as many interactions take place in the presence of others, who are taken into account. In his paper on *footing*, Goffman deconstructs traditional linguistic concepts of hearer and speaker. He differentiates between the *participation status*, or *participation role* ([17], p. 162), that concerns the relation between the participants, what is said, and their understanding of their self, and the *participation framework* defined as the relations of all those present at a given moment of the encounter.

Goffman’s model not only suggests the perspective of interaction as a *social* encounter, but also defines a nonexclusive list of participation roles as dynamic concepts that affect each other and can change during the interaction. Participants can, for example, be involved in an interaction as *ratified* or *unratified* participants, *addressed* or *unaddressed* hearers, and *eavesdroppers* or *bystanders*. This also has consequences for the status of communication. Goffman identifies different kinds of *subordinate communication* in relation to a dominating communication, such as *byplay*, a conversation of a subset of ratified participants (e.g. two people whispering to each other

during multiparty conversation), or a *crossplay*, a communication between a ratified participant and a bystander across the boundaries of the dominant communication (e.g. calling the waitress during a conversation with a friend at a restaurant) ([7], p. 134). Furthermore, Goffman offers a complex understanding of a speaker, as he distinguishes for example between the *animator* producing the talk, the *author* who is responsible for the production of the words, and the *principal* who is socially responsible for what is said ([7], p. 144).

Goffman's work is often criticized for not being empirical [17] and offering only a "static set of categories" ([9], p. 225) that cannot account for the situated dynamics of interaction. Goodwin and Goodwin adjust Goffman's ideas for interactional analysis and define participation as "actions demonstrating forms of involvement performed by parties within evolving structures of talk" ([9], p. 222). In a detailed analysis of *talk-in-interaction*, they show the situated and interactive "practices through which different kinds of parties build actions together by *participating* in structured ways in the events" ([9], p. 225). Participation status and the ongoing interaction are thereby reflexively intertwined as they shape and reshape each other. From this theoretical basis, this paper will demonstrate how participants in a HAI in a public setting take part in different ways and how this influences their participation status, the participation framework and the actions themselves.

3 Data Collection

The data derives from an ethnographic study of recorded video, naturally occurring (not experimental) interactions between users and the ECA, Max, during a public presentation of the agent in a shopping centre where people could volunteer to communicate with Max ([15, 16]). The data were transcribed and analyzed according to the principles of conversation analysis, including embodied and material aspects [6]. Max is a human-sized agent who can be seen on a large screen. To communicate with Max, users send messages to the agent by typing text on a keyboard in front of the screen. The actual production of the text can be seen and corrected in the white space at the bottom of the screen. After the users press Enter, the text is sent to the system and can be seen in the grey space above the white one. The dialogue system searches the user's text for key symbols and grammatical phrases, assigns a single functional purpose to the message and selects a pre-programmed utterance combined with a bodily reaction performed by the agent (movement of lips, facial expression, gestures, etc.) ([13, 14]). During the presentation at the shopping centre, Max was constructed as a presenter. He could inform the user about certain topics (e.g. AI or the event), take part in small-talk or play a game, for example.

4 The Participation Framework of a Human-Agent Interaction

The recorded HAIs can be described as mediated interactions between two different entities located in two different situations and with different abilities to access each other's situation. Max's presence was mediated by screen and loudspeakers to the user,

and the user's presence was mediated to Max by written text. Furthermore, the programmed structures that form the agent's understanding of actions differ fundamentally from the user's situated interpretations of the technology during the course of interaction (see [15, 16, 22]).

As the user accesses the virtual world from the keyboard and text-production, he is present in two situations: his physical situation in front of the screen and his engagement (as text-message) in the virtual world. Thus, the HAI is situated in a participation framework that is asymmetrically assessed by user and agent. While Max is programmed to engage in communication with a single user, the user is a participant of a larger social encounter in which various parties are involved in constructing meaning from the HAI. The following analysis demonstrates how this asymmetric participation framework affects the construction of participation roles and activities.

4.1 Solving a Communication Problem with the Agent

Figure 1 shows a sequence in which Max and the user, Dave, are playing a game. The agent tries to guess an animal the user has in mind. The agent asks questions and the user answers with yes or no.¹ In line 1, Max asks the user if the animal he has in mind has a mane. Dave does not answer the question but pauses briefly before he types and sends the text "pardon?" (line 3), which marks a problem of understanding. Dave leans forward (lines 4–5) as if he wants to listen closely to Max's next utterance. Max announces that he will repeat his former utterance and does so word by word (line 6). Dave slowly straightens up, puts his hands on the keyboard and eventually says "no" to Max's former question (line 12). Max treats the problem as solved. He asks his next question and continues the game (line 13). In this sequence, Max and Dave mark and repair [21] a communicative problem in understanding so the game can continue.²



```

01 MAX: Has it a mane
02 DAVE: ((looks at Max for 1 sec., turns to keyboard, types))
03 TEXT: *pardon?*
04 DAVE: ((rests his right hand on the table, bends slightly
05 forwards and looks down))
06 MAX: I have said [has it a mane.
07 DAVE: [((turns his head a bit to the side))
09 DAVE: [((straightens up, gaze and hands on the keyboard))
10 [(2.0)
11 DAVE: ((types))
12 TEXT: *no*
13 MAX: Has is claws?

```

Fig. 1. Repair: solving a problem with Max

¹ The transcripts are simplified and translated from German [16]. Pauses are shown in brackets. Double brackets indicate nonverbal communication. Square brackets mark overlapping actions. Capital letters indicate stress and text within ** is the message sent to the agent.

² The pause in line 9-10 indicates that Dave still might not have understood Max. However, Dave produces the relevant answer to Max's question (line 12), which gives the communicative basis for continuing the game. On an interactional level, the problem is solved.



```

01 MAX: Has it a mane?
02 SONJA: [((looking at the keyboard))
03 [(0.75)
04 SONJA: ((turns to the audience)) Has it WHAT?
05 ?: [Ma:ne ]
06 CARL: [((turns to Sonja))] MAne
07 CS: Ma:ne
08 SONJA: [((turns to the keyboard))
09 SONJA: No [it does not ((typing))
10 TEXT: *no*
11 MAX: Has it claws?

```

Fig. 2. Troubleshooting: solving a problem without Max

Dave and Max treat each other as ratified and directly-addressed communication partners and are engaged in a focused interaction.

4.2 Solving a Communication Problem Without the Agent

In contrast to Dave, who solved the problem *with* Max, Fig. 2 shows how the user, Sonja, excludes Max from the problem-solving activity. Sonja came to the stand with her friend, Carl. He is standing in the background and observes the interaction with the computer scientist who programmed Max (indicated by the initials “CS” in the transcript), as well as other people observing the interaction. In common with Dave, Sonja experiences the same problem of understanding during the guessing game.

After Max asks if the animal has a mane (line 1), Sonja pauses and looks down at the keyboard, before she turns to the audience and asks, “has it what?” (line 2–4). Emphasising the word “what”, she identifies the communication problem as specific – she could not understand the word “mane”. Sonja’s question is taken up by several people in the audience who repeat the problematic term “mane” for her (lines 5–7). Turning back to Max, Sonja announces “no it does not” (line 9), demonstrating to the audience that the problem is solved, and turns back to the communication with Max, writing the missing answer “no” (line 10). Max formulates his next question and the game continues.

By turning to the audience Sonja engages in a crossplay, a side-sequence with the bystanders to inform the dominant communication (playing a game with Max). The participation status of former observing-only bystanders changes to that of directly addressed participants. In this side sequence, Sonja and the bystanders engage in troubleshooting [11] and achieve a mutual understanding of the communication problem and its remedy. While the observers become addressed helpers, the agent is no longer treated as an addressed participant. Max is excluded from parts of the communication and part of the HAI becomes an object of a subordinate communication. Sonja seems to assume that Max is not able to repair her problem (at least not in the way she might hope) and thus treats Max as an unequal or inexperienced communication partner. At this point, the consequences of the asymmetric participation are evident – from Max’s perspective, the crossplay did not happen. He is engaged in conversation with one person only.

5 Conclusion and Discussion

This paper described a HAI as an asymmetric, mediated and social encounter in which different kinds of participants take up different participation roles in producing and interpreting a HAI. User and agent have different abilities in accessing the situation of the other and are, as such, involved in different participation frameworks. The agent's actions are designed on the basis of a dialog with one single user, while the user is situated in a larger social encounter with other participants. While example 1 showed how Max could be constructed as a quasi-equal conversation partner in an interaction, example 2 highlighted the impact of the asymmetric participation framework – while the user was engaged in a crossplay with the bystanders, Max was not aware of it. The agent's participation status changed to that of an unequal or inexperienced communication partner.

Considering the whole event at the public presentation of the agent, it is possible to differentiate various kinds of participation roles that could inform other studies of HAIs in public or multiparty settings. The *user* was a person who left the audience and went to the table to write a message to the agent. Becoming a user included a change of footing as they 'took the stage' and performed the interaction with Max in front of an observing audience. Sometimes another person (e.g. a friend or the computer scientist helping out) joined the user at the keyboard, construing himself or herself as *co-user*. Other participation roles included those of *by-passers* who glimpsed or more or less ignored the HAI on their way through the shopping centre. *Observers* or *bystanders* formed the audience that surrounded the table and the screen. Most of the bystanders were observing silently, sometimes laughing, commenting on the performance or engaging in subordinated conversation. Some parts of the audience became *helpers* during the course of the interaction, as in example 2, suggesting how to solve a problem or what to write to the agent. People helping out the user often became *authors* or *co-authors* of the text that was written to Max, which also affected the role of the user who became the *animator* of another person's text. Some helpers demonstrated themselves as *experts* in this case, showing a familiar knowledge of the agent-system and how it worked. These were often, but not always, the computer scientist who developed the agent. They often became *principals* of the agent, as they spoke in favour of the agent or explained why it was doing what it did.

This is not a closed list of participation roles in HAI, but it does indicate that the view of an isolated and standardised user is a reductionist way of conceptualizing HAI. Users engage in HAI in various ways and often together with others. Different participation roles have also been found by other authors. Blomberg already has mentioned the special role of advanced users [3] and Woolgar observes *commentators* and *observers* during usability trials. He and argues that people take different roles towards the technology as "they can speak as insiders who know the machine and who can dispense advice to outsiders" ([24], p. 88).

Studies of ECA aiming for applications in social encounters where several parties are present therefore need a theoretical framework that accounts for the different participation roles. This is already done in other areas of interactive technologies. In [19], for example, Goffman's concepts are taken for developing a robot that shapes participation

roles by gazing differently at addressees, bystanders and eavesdroppers of the human-robot interaction. However, more work needs to be done to understand the complex and situated ways in which people engage with technologies, so that interaction can be made more effective.

References

1. Agre, P.E.: Conceptions of the user in computer system design. In: Thomas, P.J. (ed.) *The Social and Interactional Dimensions of Human-Computer Interfaces*, pp. 67–106. Cambridge University Press, Cambridge (1995)
2. Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.): *Intelligent Virtual Agents*. Springer, Berlin (2013)
3. Blomberg, J.L.: Social interaction and office communication: effects on user's evaluation of new technologies. In: Kraut, R.E. (ed.) *Technology and the Transformation of White-Collar Work*, pp. 195–210. Lawrence Erlbaum, Hillsdale (1987)
4. Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsón, H., Yan, H.: Human conversation as a system framework: designing embodied conversational agents. In: Cassell, J., et al. (eds.) *Embodied Conversational Agents*, pp. 29–63. MIT Press, Cambridge, Mass. (2000)
5. Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.): *Embodied Conversational Agents*. MIT Press, Cambridge (2000)
6. Heath, C., Hindmarsh, J., Luff, P.: *Video in Qualitative Research: Analysing Social Interaction in Everyday Life*. Sage, London (2010)
7. Goffman, E.: Footing. In: Goffman, E. (ed.) *Forms of Talk*, pp. 124–159. Blackwell, Oxford (1981)
8. Goffman, E.: The interaction order. *Am. Sociol. Rev.* **48**, 1–17 (1983)
9. Goodwin, C., Goodwin, M.H.: Participation. In: Duranti, A. (ed.) *A Companion to Linguistic Anthropology*, pp. 222–244. Blackwell, Malden (2004)
10. Grudin, J.: Interface: an evolving concept. *Commun. ACM* **36**(4), 110–119 (1993)
11. Haase, J.: *Computer aus Nutzerperspektive von der Nutzeranalyse zum Interface-Design*. DUV, Wiesbaden (2005)
12. Hyysalo, S., Johnson, M.: The user as relational entity. Options that deeper insight into user representations opens for human-centered design. *Inf. Technol. People* **28**(1), 72–89 (2015)
13. Kopp, S., Gesellensetter, L., Krämer, N.C., Wachsmuth, I.: A conversational agent as museum guide – design and evaluation of a real-world application. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) *IVA 2005. LNCS (LNAI)*, vol. 3661, pp. 329–343. Springer, Heidelberg (2005)
14. Kopp, S., Jung, B., Lemann, N., Wachsmuth, I.: Max – a multimodal assistant in virtual reality construction. *Künstliche Intelligenz* **4**, 11–17 (2003)
15. Krummheuer, A.: Conversation analysis, video recordings, and human-computer interchanges. In: Kissmann, U.T. (ed.) *Video Interaction Analysis. Methods and Methodology*, pp. 59–83. Peter Lang, Frankfurt am Main (2009)
16. Krummheuer, A.: *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*. Lucius & Lucius, Stuttgart (2010)
17. Levinson, S.C.: Putting linguistics on a proper footing: explorations in Goffman's concepts of participation. In: Drew, P., Wootton, A. (eds.) *Erving Goffman. Exploring the Interaction Order*, pp. 161–227. Polity Press, Cambridge (1988)

18. Moosaei, M., Gonzales, M.J., Riek, L.D.: Naturalistic pain synthesis for virtual patients. In: Bickmore, T., Marsella, S., Sidner, C. (eds.) IVA 2014. LNCS, vol. 8637, pp. 295–309. Springer, Heidelberg (2014)
19. Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., Hagita, N.: Footing in Human-Robot conversation: how robots might shape participant role using gaze cues. In: HRI 2009, March 11–13, La Jolla California, USA (2009)
20. Rehm, M.: ‘She is just stupid’ – analyzing user-agent interactions in emotional game situations. *Interact. Comput.* **20**, 311–325 (2008)
21. Schegloff, E.A., Jefferson, G., Sacks, H.: The preference for self-correction in the organization of repair in conversation. *Language* **53**(2), 361–382 (1977)
22. Suchman, L.: *Human-Machine Reconfigurations Plans and Situated Actions*. Cambridge University Press, Cambridge (2007)
23. Vannini, N., Enz, S., Sapouna, M., Wolke, D., Watson, S., Woods, S., Dautenhahn, K., Hall, L., Paiva, A., André, E.: “Fearnot!”: a computer-based anti-bullying-programme designed to foster peer intervention. *Eur. J. Psychol. Educ.* **26**(1), 21–44 (2011)
24. Woolgar, S.: Configuring the user: the case of usability trials. In: Law, J. (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination*, pp. 58–99. Routledge, London (1991)