Social machines?

Critical reflections on the agency of 'Embodied Conversational Agents'

Florian Muhle, Paderborn/Bielefeld, Germany

Abstract

New agencies appear at the interface in the form of so called 'Embodied Conversational Agents' (ECAs). These humanoid machines are becoming increasingly implemented into digital games and virtual worlds in order to interact with human users in a humanlike way. Against the background of this phenomenon, some scholars ask whether the spread of 'social' machines calls for traditional definitions of society to be questioned. Following this debate, I will challenge the social abilities of ECAs from a sociological point of view. This means that I ask whether ECAs really can be treated as social actors. For this I will at first discuss different social-theoretical positions about 'social action', in particular Actor Network Theory and Niklas Luhmann's Systems Theory. Building on those considerations, in a second step, I will present a short example of my own empirical research on human-humanoid-communication in the virtual world 'Second Life'.

1. Introduction

New agencies appear at the interface in the form of so called 'Embodied Conversational Agents' (ECAs). ECAs "are computer systems that figure in humanoid form, either as robots or as 3D virtual characters, in order to allow the possibility for users to meet and interact with a machine as if having a face-to-face conversation with another human" (Kopp 2008). For many years ECAs only existed in laboratories of information scientists and it was not until a few years ago, that the development of 3D virtual worlds and Online Games allowed them to move to the internet. From the developer's perspective, virtual worlds serve as test beds for the evaluation of their products. Thus, ECAs are becoming increasingly implemented into digital games and virtual worlds in order to interact with human users in a humanlike way. As a consequence, virtual worlds are not only inhabited by human-controlled avatars, but also by non-humans, who are reputed to be able to socialize.

Hence, the investigation of communication between humans and ECAs (human-humanoid-communication) also becomes relevant for social scientists, as the attempt to create social and human-like machines calls traditional anthropocentric definitions of society into question. With this background, I will challenge the social abilities of ECAs from a sociological point of view i.e., I will discuss whether ECAs really can be treated as social actors – as information scientists promise.

To answer this question, I will first need a well-defined concept of social action; one that contains the possibility to attribute 'social' agency not only to humans, but also to non-human actors. For this reason, in chapter 2, I will discuss different social-theoretical positions, in particular the Actor Network Theory (ANT) and Niklas Luhmann's Systems Theory. In debating these contradictory approaches, I aim to develop a methodological position, which allows me to specify my view on empirical data. Building on those considerations, in a second step (chapter 3), I will present a short example of my own empirical research on human-humanoid-communication in the virtual world 'Second Life' (SL). Thus, I am able to show some characteristic aspects of human-humanoid communication, which question the humanlike agency of 'social' machines. Finally, in chapter 4 I will give a conclusion on my findings.

2. Social-theoretical considerations on analysing agency

Asking the question whether ECAs can be treated as social actors leads to a conflict with traditional social theory. One of the basic axioms of traditional sociology is that agency is an exclusive quality of human-beings. From this point of view, animals, things or machines never can be considered as social actors, but rather only as natural, artificial or technical entities.

Nevertheless, the traditional, anthropocentric conception of 'the social', which is represented by well-known names such as Weber, Simmel or Mead, has been questioned during the last 20 years by new social theories – probably most ambitiously by the Actor Network Theory. ANT rejects the idea that social relations are restricted to human beings. On the contrary, ANT claims that social relations can emerge between all different kinds of entities, namely when their actions have an effect on the actions of each other. If that is the case, the different entities will constitute 'the social' which then consists of "patterned networks of heterogeneous materials" (Law 1992, 2). John Law summarizes this consideration as follows:

> "This is a radical claim because it says that these networks are composed not only of people, but also of machines, animals, texts, money, architectures – any material that you care to mention. So the argument is that the stuff of the social isn`t simply human. It is all these other materials too. Indeed, the argument is that we wouldn`t have a society at all if it weren`t for the heterogeneity of the networks of the social. So in this view the task of sociology is to characterise these networks in their heterogeneity, and explore how it is that they come to be patterned to generate effects like organisations, inequality and power" (Law 1992, 2f.).

Such a way of looking at things is possible due to a special kind of 'theory of action'. Whereas for good old-fashioned sociology, agency is mainly restricted to human intentional individual actors, ANT "extends the word actor – or actant – to *non-human, non-individual* entities" (Latour 1996, 369). Following a semiotic definition of 'actant', from the ANT's point of view an actor is just "something that acts or to which activity is granted by others. It implies *no* special motivation of human individual actors, nor of humans in general" (Latour 1996, 373). On the one hand, this means, for example, that even a scallop may become an actor in the same way as a fisherman or a scientist (sf. Callon 1986). On the other hand, it opens up the possibility to analyse the processes, "during which the identity of actors, the possibility of interaction and the margins of manoeuvre are negotiated and delimited" (Callon 1986, 6). As a result of this, at first glance ANT seems to be a reasonable approach to consider my research question.

However, taking a closer look, some problems appear at the surface. ANT's theory of action implicates certain methodological principles which I can not easily follow. In the present case especially the principle of 'generalized symmetry' (sf. Callon 1986) in combination with the weak actor-concept matters. It must be kept in mind that "the rule which we must respect is not to change registers when we move from the technical to the social aspects of the problem studied" (Callon 1986, 4). Consequently, the scientific observer has to describe the actions of a scallop in the same way as the actions of a fisherman. Latour and Callon put this as follows: "Whatever term is used for humans, we will use for non-humans as well" (Callon &

Latour 1992, 353).

Considering this methodological principle, my argument is, that this claim is difficult for my purposes because it does not allow for one to distinguish between different kinds of actions. Thus, the actions of an ECA, a door-closer or a human-being must all look the same. They are indistinguishable. This means that ANT only answers the question *whether* agency (in a weak sense) can be attributed to certain entities, but not *how* this would be realized – maybe in a different way than to other entities.

As a consequence, this perspective does not allow one to distinguish between humanlike, social action and other forms of action. However, this is exactly what is necessary to answer my research question. In order to find out whether ECAs act like humans or in a different way, I need a more sophisticated 'theory of action' which is principally able to reveal differences and asymmetries between the actions of certain entities (sf. Suchman 2007, 268ff). ANT obviously is unsuited for this challenge. Due to this, I have to look for another theoretical offer that allows for the distinguishing of several states of agency and subject-object positionings.

For this Niklas Luhmann's social theory could be a reasonable candidate. On the one hand, Luhmann shares an anti-essentialistic conception of the social with ANT. On the other hand, compared to ANT`s weak actor-concept, he provides a more sophisticated idea of the social. Referring to Talcott Parsons, Luhmann claims that the precondition of sociality is the problem of double contingency (sf. Luhmann 1995, 103ff). Parsons already used the theorem of 'double contingency' to distinguish

> "between objects which interact with the interacting subject and those objects which do not. These interacting objects are themselves actors or egos […]. A potential food object […] is not an alter because it does not respond to ego`s expectations and because it has no expectations of ego`s action; another person, a mother or a friend, would be an alter to ego. The treatment of another actor, an alter, as an interacting object has very great consequences for the development and organization of the system of action" (Parsons & Shils 1951, 14f.).

Consequently, in a *social* interaction both interlocutors "know that both know that one could also act differently" (Vanderstraeten 2002, 77). This has significant consequences for the emergence of social systems and allows one to draw a distinction between social actions and other forms of action. Here the importance of expectations comes into play.

As Parsons points out, "it is the fact that expectations operate on both sides of the relation between a given actor and the object of his orientation which distinguishes social interaction from orientation to nonsocial objects" (Parsons & Shils 1951, 15). Additionally, according to Parsons, Luhmann emphasizes that the analytical decision of whether an action is social or not depends on the complexity of underlying expectancy structures. "With double contingency there is a need for […] complicated expectancy structures that rely heavily on preconditions, namely *expectation of expectations*" (Luhmann 1985, 269). This means that an action can be treated as 'social', if, (from the perspective of a given entity, or rather an *ego)*:

> "the behaviour of the other person cannot be expected to be a determinable fact; there is a need to see it in terms of his selectivity, as a choice between various possibilities. This selectivity is, however, dependent on others` structures of expectation. It is necessary, therefore, not simply to be able to expect the behaviour, but also the expectations of others" (ibid).

In this sense, social actions only emerge between entities which attribute expectation of expectations to each other. On the contrary, handling (trivial) machines can be described as a situation with simple expectations. Whereas every input to the machine creates an expected output, a given ego can stabilize simple, but persisting expectations of the machine's behaviour. A door-closer will close the door – and nothing else. The machine in comparison has no expectations of ego's action.

Looking at Luhmann's definition of sociality, at first glance it seems to be obvious "to imagine ego and alter, on both sides, as fully concrete human beings, subjects, individuals or persons" (Luhmann 1995, 107). Nonetheless, this is not entirely true. The crucial point remains that it is all about attributions. It is not important whether a given entity is 'really' a human being. Instead it is relevant if "an ego *experiences* an alter as alter ego and acts in this experiential context" (Vanderstraeten 2002, 86; author`s emphasis). Therefore, it is principally possible that ECAs are treated as social actors during conversations. It depends on the underlying expectancy structures of human-humanoid communications.

According to Luhmann's theory, these expectancy structures can be analyzed easily because they will be expressed during the process of communication. Thus, the mentioned expectations are not mental, but rather communicative. Against this background, answering the question whether and/or how *agency* is attributed to ECAs during conversations then requires an analysis of communication processes and the underlying expectancy structures on the basis of transcripts. These structures constitute the state of agency which is attributed to ECAs as well as to humans or other entities during conversation. Ian Hutchby, a conversation analyst, accordingly writes that "in human-machine encounters, whether a machine is treated as a social entity therefore depends […] on the extent to which its affordances match up with the expectations generated by the normative structures of interaction" (Hutchby 2001, 172). However, these considerations should be enhanced by an empirical investigation of expectancy structures in human-humanoid-communications. This leads to the next part of the paper.

3. Empirical investigation

In this chapter, I give an empirical example of a human-humanoid dialogue in order to analyze the underlying expectancy structures. Thus, I want to present some interesting aspects of human-humanoid communication, which question the humanlike agency of ECAs as so called 'social' machines. For this I have followed a conversation analytical (CA) approach, because this approach is well-suited for the analysis of such expectancy structures (sf. Schneider 2004). In short, CA is a special type of sequential analysis which

> "is done, basically, in three steps. The first is to reconstruct the selectivity of each utterance by answering such questions as: How does it relate to preceding turns? What implications were made by it? What possible options were not chosen? The second step is to reconstruct the future expectations connected with the utterance, i.e. the action space opened by the utterance. The question here is: What are the possible options for the speaker to continue or the recipient to react? The third step is to reconstruct the interactional consequences, i.e. the selective way the recipient reacts to the utterance. By sequentially proceeding with these operations CA reconstructs the way the turns are linked together that is characteristic of the communicative pattern in question" (Maiwald 2005, 8).

As I mentioned above, the dialogue I want to investigate takes place in the virtual world Second Life (SL). SL is a 3D world where users can communicate with each other through graphic representations of themselves, which are called avatars.

Although most avatars are driven by human users, it is also possible that they are controlled by ECA-software, as is the case in the example below. There is one human-driven avatar who is called Tim and one machine-driven avatar, whose name is Makks. The original conversation is held in German. Thus, the original text still remains in the transcript in grey letters. Numbers in brackets show pauses (in seconds) between the utterances.

1 Makks: How old is your Avatar? // Wie alt ist dein Avatar?
2 (6)
3 Tim: 19
4 Makks: I'd have guessed you one year older! // Ich haette dich ein Jahr aelter geschaetzt!
5 (11)
6 Tim: oh ok
7 Makks: you really have a fine grasp // du hast aber eine gute auffassungsgabe
8 (5)
9 Tim: How old are you? // wie alt bist du?
10 Makks: The first time I started running was back in ninety-nine. It isn't possible for everybody to be 19 years old, is it? // Zum ersten Mal in Betrieb genommen wurde ich Ende neunundneunzig. Es kann ja nicht jeder 19 Jahre alt sein, oder?
11 (7)
12 Tim: loool
13 Makks: There are so many words – why don't you use some more? // es gibt soviele worte – moechtest du nicht ein paar mehr benutzen?
14 (7)
15 Tim: bracketsclose // klammerzu
16 Makks: Is it a Yes or No? // Ist das ein Ja oder ein Nein?
17 (2)
18 Tim: yes // ja
19 (1)
20 Makks: it's always good to come to an agreement. // Zustimmung ist immer gut.
21 (2)
22 Tim: bracketsclose // klammerzu
23 Makks: What's your point? // Was meinst Du damit?
24 (3)
25 Tim: bracketsclose // klammerzu
26 (1)
27 Makks: if you say that again, I'll know it by heart. // wenn du das nochmal sagst weiss ich es auswendig.

28 (14)
29 Tim: nothing sorry // nix sorry

In the following, I want to emphasize three details. First, it is obvious, that Makks produces utterances, which seem to be a bit unusual. They just make sense as isolated responses to each precedent utterance – but not as reasonable contributions to the overall-topic of the conversation.

The utterance “you really have a fine grasp“ in line 7 as a response to Tim's “oh ok“ only makes sense in a pedagogical context. Given such a context, “oh ok“ probably expresses that Tim just understood something new, so that his teacher compliments him on his comprehension. However, as an answer to "I'd have guessed you one year older“ in line 4 "oh ok“ definitely does not express Tim's comprehension, but rather his surprise because he obviously has no idea why Makks guessed him one year older. Accordingly, Makks' assessment here is qualified as non-intuitive, so that an explanation for his inappropriate utterance would be expected – but not the following compliment on Tim's comprehension.

The following sequences are structured in a similar way:

12 Tim: loool
13 Makks: There are so many words – why don’t you use some more? // es gibt soviele worte – moechtest du nicht ein paar mehr benutzen?

and

18 Tim: yes // ja
19 (1)
20 Makks: it’s always good to come to an agreement. // Zustimmung ist immer gut.

If we just look at the isolated adjacency pairs, it would be possible to imagine a dialogue in which these sequences of utterances make sense. However, against the background of the available conversation, Makks seems to behave in an odd way. While reading the transcript, Garfinkel’s so-called 'crisis experiments' come into mind, in which he deliberately broke everyday expectations to investigate the meaning of such expectations for the organisation of interaction (sf. Garfinkel 1967). In the present case, it is Makks who breaks such expectations. Even though the ineptitude of his utterances is not articulated explicitly, the following sequence of Tim's 'bracketclose' utterances shows the breakdown of social expectations. The subjects in Garfinkel's experiments reacted frustrated and confused, whereas Tim responds unemotional and – subsequent to Makks' utterance "There are so many words – why don't you use some more?" – starts a kind of experiment on his part:

13 Makks: There are so many words – why don't you use some more? // es gibt soviele worte – moechtest du nicht ein paar mehr benutzen?
14 (7)
15 Tim: bracketsclose // klammerzu
16 Makks: Is it a Yes or No? // Ist das ein Ja oder ein Nein?
17 (2)
18 Tim: yes // ja
19 (1)
20 Makks: it's always good to come to an agreement. // Zustimmung ist immer gut.
21 (2)
22 Tim: bracketsclose // klammerzu
23 Makks: What's your point? // Was meinst Du damit?
24 (3)
25 Tim: bracketsclose // klammerzu

By repeating 'bracketsclose' three times, Tim also leaves the realm of socially expected behaviour. From my point of view, he conducts a kind of 'Turing-Test'. He checks, if his own senseless and identical utterances foreseeably lead to identical utterances from Makks. In doing so, he tries to receive an indication of the mechanical or social identity of his dialogue partner. Thus, Tim does not orientate himself on expectation of expectations anymore. Following Luhmann, the assumption, that your counterpart is able to select between various possibilities, is a premise for expectation of expectations. If we keep this in mind, it becomes clear, that in the given situation the problem of double contingency is not present anymore. On the contrary, Tim treats Makks' behaviour as determined and is testing, whether a certain input leads to a clearly expectable output from his dialogue partner. This means, that he does not treat Makks as a social entity anymore. Makks' inappropriate utterances obviously call social expectations into question, so that his status as a social actor becomes disputable.

However, this does not need to be stable, as the third sequence I want to emphasize shows us.

25 Tim: bracketsclose // klammerzu
26 (1)
27 Makks: if you say that again, I'll know it by heart. // wenn du das nochmal sagst weiss ich es auswendig.
28 (14)
29 Tim: nothing sorry // nix sorry

As a response to the last "bracketclose"-utterance (line 25), Makks answers "If you say that again, I'll know it by heart" (line 27), which leads to Tim's apology "nothing sorry" in line 29.

Due to the apology at the third turn, the structure of the conversation becomes transformed in a decisive way. Whereas Tim was carrying out a Turing-Test in the precedent sequence, now the test comes to its end. Instead, in this triadic sequence the communication process returns to a situation of double contingency. Tim apologizes for the test and, by doing so, again treats his counterpart as a social actor with social expectations. His apology only makes sense, if he assumes that Makks expects not to be treated as a machine. As we can see here, in this sequence a structure of social expectations really becomes reproduced. But this can change rapidly again, so that the status of the involved entities may change from turn to turn. From the perspective of the participants, it would be hard to find a definitive answer to the question if they are talking to social entities or to mechanical machines.

4. Conclusion

The swapping between different states of agency seems to be characteristic for the status ascribed to 'intelligent' machines today. They are neither traditional, trivial machines nor social actors which behave in a humanlike way. Instead, they can probably be best described as 'hybrids', who/which are switching between the realms of 'the social' and 'the technological'. The analysis of human-humanoid communication shows that as well as how social actors, technological artifacts and the differences and asymmetries between them are produced during communication processes – not as given and stable entities, but as *effects* of such communication processes.

What does this mean for the question about the humanlike agency of ECAs? On the one hand, we need to be skeptical as to whether the developer's idea of humanlike agency and sociality of machines lives up to its promise. There is currently no reason why we should assume that 'social machines' will become part of society in a similar way as human beings. On the other hand, however, there is no doubt, that in the near future 'intelligent' machines and interfaces will increasingly become part of everyday life. This will change the conditions of the human-machine relationship.

Consequently, new agencies at the interface definitely are an interesting and important research field for science and technology studies. By investigating them, we can learn how borders between humans and technology constantly are shaped, negotiated and (re-)configured.

References

Callon, Michel (1986). 'Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay', http://www.vub.ac.be/SOCO/tesa/RENCOM/Callon (1986) Some elements of a sociology of translation.pdf. [14 September 2009].

Callon, Michel and Latour, Bruno (1992), 'Don't Throw the Baby out with the Bath School! A Reply to Collins and Yearley`, in Pickering, Andrew (Ed.), *Science as Practice and Culture,* Chicago: University of Chicago Press, 343-368.

Garfinkel, Harold (1967), *Studies in ethnomethodology',* Englewood Cliffs, NJ: Prentice Hall.

Hutchby, Ian (2001), '*Conversation and Technology: From the Telephone to the Internet',* Cambridge: Polity Press.

Kopp, Stefan (2008), 'From Communicators To Resonators – Making Embodied Conversational Agents Sociable', Keynote at the Symposium Speech and Face to Face Communication (in memory of Christian Benoit), Extended Abstract with references, Grenoble: France, http://www.techfak.uni-bielefeld.de/ags/soa/download/ FromCommunicatorsToResonators-Abstract.pdf [7 June 2010].

Latour, Bruno (1996), 'On Actor-Network Theory: A few clarifications', *Soziale Welt* 47 (4): 369–381.

Law, John (1992), 'Notes on the Theory of the Actor Network: Ordering, Strategy and Heterogeneity', http://www.lancs.ac.uk/fass/sociology/papers/law-notes-onant.Pdf [14 September 2009].

Luhmann, Niklas (1985), '*A sociological theory of law',* International library of sociology. London: Routledge & Paul.

Luhmann, Niklas (1995), '*Social Systems',* Stanford: Stanford University Press.

Maiwald, Kai-Olaf (2005), 'Competence and Praxis: Sequential Analysis in German Sociology [46 paragraphs]', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 6 (3): Art. 31, http://nbn-resolving.de/urn:nbn:de:0114-fqs0503310.

Parsons, Talcott and Shils, Edward A. (Eds.) (1951), '*Toward a General Theory of Action',* New York: Harper and Row.

Schneider, Wolfgang Ludwig (2004), '*Grundlagen der soziologischen Theorie: Band 3: Sinnverstehen und Intersubjektivität - Hermeneutik, funktionale Analyse, Konversationsanalyse und Systemtheorie***',** Wiesbaden: VS Verlag für Sozialwissenschaften.

Suchman, Lucy (2007), '*Human-Machine Reconfigurations: Plans and Situated Actions, 2nd Edition',* Cambridge: Cambridge University Press.

Vanderstraeten, Raf (2002), 'Parsons, Luhmann and the Theorem of Double Contingency', *Journal of Classical Sociology* 2 (1): 77–92.

Contact information:

M.A. Florian Muhle

BGHS Bielefeld
Universität Bielefeld
Postfach 10 01 31
33501 Bielefeld
fmuhle@uni-bielefeld.de

Graduiertenkolleg Automatismen
Universität Paderborn
Warburger Straße 100
33098 Paderborn
fmuhle@mail.uni-paderborn.de