A more human side of a chatbot. Analysing anthropomorphism in conversations with a virtual agent depending on the level of elicited agent knowledge

Amelia La Torre[1]

[1] Jagiellonian University, Cracow, Poland
amelia.latorre@gmail.com

Abstract. The aim of the research is to analyse anthropomorphism in the context of short inter-action with a chatbot. Anthropomorphism is defined here as an individual tendency to perceive a nonhuman object as if it was a human (e.g. attributing human characteristics to it), which can be observed in individual reaction to an object. As chatbot is similar to a human due to its construction and can be anthropomorphised particularly easy, it is interesting to investigate: (1) how anthropomorphism manifests in the human-chatbot conversation, especially in the context of inconsistent chatbot's behaviour (2) if any individual differences appear in that level of contact (3) can the accessible knowledge about a chatbot influence anthropomorphism in a spontaneous conversation with it. Thirty-two chat conversations (16 male) witch *Eviebot* conversational agent will be analysed with qualitative methods. Half of it are conducted after experimental manipulation, including the information about *Eviebot* read by experimenter and short questions consolidating the provided knowledge, and second half with no additional manipulation. The analysis of conversations will put an accent to the specificity of anthropomorphism in the presented context, concentrating on the reactions to a chatbot's inconsistent behaviours. The possible differences between informed and non-informed group will be investigated. The study has an explorative character and has to help to characterise anthropomorphism of a chat-bot in a behavioural level. Conclusions and possible implications for the future research will be underlined.

Keywords: Anthropomorphism, Chatbot, HRI.

1   Introduction

Antropomorphism is a complex phenomenon, defined in terms of objects' observable similarity to a human being (e.g. Demeure, Niewiadomski & Pelachaud, 2011) or as an individual tendency to perceive an object as if it was a human (e.g. Epley, Waytz & Cacioppo, 2007). The second definition, common in psychological research, puts an accent to individual differences in attributing unobservable human characteristics to a nonhuman agent, which affects the person's behaviour towards it. The existing research shows that anthropomorphism influence the people emotional reactions towards an object, such as liking (Siino, Chung & Hinds 2008,: Waytz, Cacioppo & Epley, 2010), trust, or moral concern (Waytz Cacioppo & Epley 2010), but also antipathy (Shwartz 2003), which is manifested in behavioural level. Anthropomorphism is connected with including an object to the own social circle, and as so it can influence a social life of a person and the way he deals with other people. For example the stud-

ies suggest, that the contact with anthropomorphised objects can be adaptive for older people, who are lonely or lose their partners - in those situations also a social robots can be substitutes for people (e.g. Ring, Barry, Totzke & Bickmore, 2013). In this context it is interesting to investigate human interaction witch social robots and chatterbots, created especially to perform the social functions. Anthropomorphic reactions towards such objects can particularly exceed the level of automatic behaviours, and be connected with developing some level of attachment. Individual differences in anthropomorphic reactions towards robots and its determinants are yet to be investigated.

Three Factor Theory Of Anthropomorphism (Epley, Waytz & Cacioppo, 2007) indicates the determinants of that phenomenon, connected with the person's specific motives (*sociality* and *effectance* motivation) and to the accessibility of information about a human which can serve as an explanation of object's behaviour (*elicited agent knowledge*). The anthropocentric knowledge can be easily used especially if an object is very similar to a human or the information about an object is not available.

Human – object similarity is a factor especially significant considering anthropomorphism in interactions with robots and robot- like agents. The classic study by Reeves and Nass (1996) or Nass & Moon (2000), suggests for example that people automatically attribute human characteristics (e.g. sex and ethnicity) to computers and obtain the social rules in the contact with it. Those results are explained by the objective human-object similarities, including e.g. capability to use a natural language in communication. It can be assumed, that chatbots can provoke the same automatic reaction, not necessarily connected with anthropomorphism on a deeper level. Sustaining some social conventions in conversation can be an effect of the objective similarity of bot's and human behaviour.

On the other hand anthropomorphism can be seen as an effect of the lack of information about the object. In such situation the person can easily project invisible human characteristics into an object even if it is not objectively similar to a human (Epley, Waytz & Cacioppo, 2007). Such projection can be connected with actual belief that the object has the given characteristics, and be influenced by the individual characteristics and motives. In this case individual differences in anthropomorphism can be easily manifested. The accessibility of the information about an object can be a factor influencing those differences.

In interactions witch a chatbot it can be yet difficult to separate a motivated kind of reaction described in the last section from the automatic reaction to the objective similarity to a human. Although it can be possible to do that analysing the reactions to a chatbot's inconsistent behaviours. Illogical statements not related to the subject of conversation, which show the randomness of bot's behaviour can remain an occasion to test the interlocutor's reaction. The existing data suggests, that the machines manifesting malfunctions are easily anthropomorphised in the situation of the lack of information (Waytz, Morewedge, Epley, Monteleone et all, 2010). In that situation the

differences between a group who has the access to the information about an object and the group who has not could be possibly visible. The aim of the experiment described in this paper is to investigate this hypothesis, and analyse the way anthropomorphism manifest itself in a conversation with an anthropomorphic chatbot.

2   Methods and procedure

The thirty-two English Philology students took part in the study (age range = [20, 25], 16 men) in exchange for 5$ (equivalents in PLN). All participants were recruited by announcements in local universities.

Eviebot conversational agent used in the study is a chatbot construed by *Existors* and available on-line. It has a form of a female avatar, capable for verbal and non-verbal expression, who can make a conversation on any subject, by reacting to the words of interlocutor. In the experiment *Eviebot* is used for a chat conversation only (the voice turned off), and the avatar is displayed on the computer screen. The conversations' transcription is collected for analysis, with the assurance of confidentiality and anonymity of the data.

The study is conducted in university laboratory. During each session only one participant and experimenter are present in the room. The participants are randomly divided into two samples, 16 persons (8 men) in each. Sample A is asked to make a 10-minute random chat conversation with *Eviebot* conversational agent. Sample B is given the same instruction but before the conversation is additionally introduced with the information about *Eviebot* and is given three paper tasks related to it. The information (read loud by experimenter) contains the operating principles of a given bot, description of its visual characteristics and skills (and also some information of other bots), as well as its functions and price of an application. After that information, Sample B is given the paper tasks to perform (to underline the most important information about the chatbot, to indicate which information was new for them, and to write which kind of people could be interested in using *Eviebot*). The function of the tasks and the information is to make the specific knowledge about a chatbot potentially available to the members of Sample B during the conversation.

3   Expected results and conclusions

The collected data are still processed, therefore in this section a brief description of the process of analysis and the tested hypothesis will be presented. The collected data are analysed with the use of qualitative data analysis software (QDA Miner). The investigation will include content analysis, the number of reactions qualified to distinguished categories will be also considered.

First, chatbot's behaviours during the conversation will be analysed, as it can directly influence the interlocutor reactions. This investigation will include specifically bot's inconsistent behaviors during the conversation. A chatbot's behaviors used to be similar to human's, and provokes the reactions typical to human-human conversation automatically. The chatbot inconsistent behavior: e.g. unlogical, random statements, can break that reactions schema. It is then important to control this factor as possibly moderating the results. On the other hand, the participants' reactions on a chatbot's inconsistent behaviours will be investigated. The reactions will be divided into
s) reactions suggesting that a participant expect from a bot the same behavior as from a human, b)reactions connected with understanding of a machine malfunctions. Foregoing reactions are interpreted as related to anthropomorphism, which is more than just an act of sustaining some convention in conversation (sustaining a schema of human-human conversation can be for participants just a funny and comfortable way to speak with a chatbot, the chatbot malfunctions can provoke more spontaneous reactions). Then, other participants reactions connected with anthropomorphism will be analysed: attributing human characteristics to a chatbot (emotions, personality traits, mind attribution) and expressing emotions towards a chatbot. The differences between Sample A and Sample B, as well as between men and women will be investigated in all of foregoing aspects.

To sum up, the following predictions will be examined:
  1) The group informed about an *Eviebot* characteristics and given the tasks consolidating that information will react differently to chatterbot's inconsistent behaviours in comparison to non-informed group
  2) Attributing human characteristics to *Eviebot* will appear in a comparable level in informed and non-informed group
  3) The differences between men and women will appear in the level and quality of anthropomorphic reactions (*Eviebot* is a female avatar thus it can be more easily anthropomorphised by women, see e.g. Eyssel, Kuchenbrandt, Hegel & de Ruiter 2012)

The content analysis will also concentrate on investigating, which reactions are appearing after confronting chatbot's inconsistent behavior in all of the sample.

The conclusions of the research could be used to characterize the phenomenon of anthropomorphism of a chatbot in the level of behavioral reactions. Reacting to an object which is very similar to a human can be analysed in two levels: the level of automatic reactions to the its similarity and the level connected with individual expectations towards an object possibly connected with individual factors. That level could be visible when analysing the reactions to the chatbot's inconsistent behaviors, and then the individual differences in anthropomorphism could be then manifested. The accessibility of information about a chatbot can additionally influence those differences. The results of the study could then suggest some implications for designing the context of interaction with a chatbot, signalizing the possible individual differences and specificity of anthropomorphic reactions to it.

References

1. Demeure, Virginie, Radosław Niewiadomski, and Catherine Pelachaud, 'How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment?', *Presence: Teleoperators and Virtual Environments*, 20 (2011), 431–48 <https://doi.org/10.1162/PRES_a_00065>

2. Epley, Nicholas, Adam Waytz, and John T. Cacioppo, 'On Seeing Human: A Three-Factor Theory of Anthropomorphism.', Psychological Review, 114 (2007), 864–86 <https://doi.org/10.1037/0033-295X.114.4.864>

3. Eyssel, Friederike, Dieta Kuchenbrandt, Frank Hegel, and Laura de Ruiter, 'Activating Elicited Agent Knowledge: How Robot and User Features Shape the Perception of Social Robots' (IEEE, 2012), pp. 851–57 <https://doi.org/10.1109/ROMAN.2012.6343858>

4. Nass, Clifford, and Youngme Moon, 'Machines and Mindlessness: Social Responses to Computers', Journal of Social Issues, 56 (2000), 81–103 https://doi.org/10.1111/0022-4537.00153

5. Reeves, Byron and Clifford Nass, 'The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.' (1996), New York: Cambridge University Press.

6. Ring, Lazlo, Barbara Barry, Kathleen Totzke, and Timothy Bickmore, 'Addressing Loneliness and Isolation in Older Adults: Proactive Affective Agents Provide Better Support' (IEEE, 2013), pp. 61–66 <https://doi.org/10.1109/ACII.2013.17>

7. Swartz, Luke. 'Why people hate the paperclip: labels, appearance, behaviour and social responses to user interface agents.' (2003), Master's thesis, Stanford University.

8. Waytz, Adam, John Cacioppo, and Nicholas Epley, 'Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism', Perspectives on Psychological Science, 5 (2010), 219–32 <https://doi.org/10.1177/1745691610369336>

9. Waytz, Adam, Carey K. Morewedge, Nicholas Epley, George Monteleone, Jia-Hong Gao, and John T. Cacioppo, 'Making Sense by Making Sentient: Effectance Motivation Increases Anthropomorphism.', Journal of Personality and Social Psychology, 99 (2010), 410–35 <https://doi.org/10.1037/a0020240>