Clélie Amiot (2020): Trustworthy chatbots assisting large-scale collaboration. In: Proceedings of the 18th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.18420/ecscw2020_dc03

Trustworthy chatbots assisting large-scale collaboration

Clélie Amiot
Inria, Université de Lorraine, CNRS
Contact Author: clelie.amiot@loria.fr

Abstract. Cognitive assistants are a promising solution to the increasing complexity of large-scale collaboration. By providing support in data processing and decision making, they could lighten the cognitive load put on human collaborators. More precisely, conversational agents are especially suited to large-scale collaboration as they are user-friendly and could be integrated into existing collaboration tools. However, to be successfully integrated into the collaboration process, the assistant needs to be trusted. My doctoral research aims to identify which factors determine trust in a chatbot's advice during collaboration.

1 Introduction

Large-scale collaboration describes a joint effort between dozens of people working towards the same goal, who often come from multiple organizations with different work processes and have varying levels and areas of expertise. The release of a new product by a multinational brand, the organization of the Olympics, and the coordination of relief efforts after a flood are all examples of large-scale collaboration. Successful collaboration relies on trust between the agents, in their intentions and competences, or at the very least trust in their organizations (Dodgson, 1993). Without trust, communication lines deteriorate as

they get disregarded more often, and poor decisions might be taken (Daim et al., 2012). In large-scale, this is made even more difficult as most agents do not know each other, and their work processes might clash.

While modern communication technology can ease collaboration, it also is the source of growing complexity. More data sources, such as sensors, crowd-sourcing, and social media, are leveraged (McAfee et al., 2012), and it has become harder to make suitable decisions, as there is too much available information to process (Vieweg et al., 2014). Our goal is to introduce cognitive assistants to large-scale collaboration who could lighten the cognitive load put on collaborators, by summarizing data and keeping better track of relevant contextual information. We are mainly interested in conversational agents, also referred to as chatbots, which are text or vocal interfaces that simulate human speech (Dale, 2016). We hope that chatbots could be integrated into already existing tools for collaboration, who often provide direct messaging functionalities, and provide assistance to users without prior training, as conversational interfaces are relatively user-friendly (Zadrozny et al., 2000).

2 Related work

Chatbots are a booming technology that sees more use in recent years but is yet still relatively overlooked in large-scale collaboration or crisis management (Misiura and Verity, 2019). Its primary use case is commercial prospecting and customer service. Concerning trust, Følstad et al. (2018) identified multiple trust factors for commercial chatbots and sorted them as either internal or external. The internal factors (relevance of the answers, personality, transparency on its limitations, and general appearance) are determined by the chatbot itself. In contrast, the external factors depend on the environment of the chatbot (company, security, privacy, and importance of the task). Müller et al. (2019) have described three different personality profiles when it comes to trusting voice assistants using the HEXACO model of personality. Those profiles are "introverted careless distrusting user", "conscientious curious trusting user", and "careless dishonest trusting user". However, those studies are not interested in measuring how this trust affects future actions and the dissemination of the information given. One study by Ramchurn et al. (2016) about a crisis management cognitive assistant did show that the ability to debate the assistant's decisions was an important factor when following its advice.

As chatbots get more reliable, we see an increase in interest for health chatbots who are closer to our application domain, as the conversations tend to be more critical and time-sensitive. The trust factors for medical chatbots identified by Wang and Siau (2018) are similar to the ones for commercial use with an emphasis given to data security and the explainability of the chatbot's advice.

Aside from the application domain of the chatbot, most of the current research have in common that they are interested in dyadic conversations and do not explore how group dynamics are affected by the introduction of a conversational agent.

3 Proposed experimental design

We are working on a minimal protocol to study the impact of a chatbot assistant on problem-solving. A subject has to answer a sequence of questions with an assistant helping her over a real-time chat. We propose a between-subject design with a first group interacting with an assistant presented as a human researcher and a second group interacting with an assistant presented as a conversational agent. However, we do not want technical difficulties to impair the conversation and deteriorate the participant's trust in the chatbot. For this reason, we will use a wizard of Oz approach for the group interacting with the "chatbot" (Dahlbäck et al., 1993). This means that while the participants think that they are conversing with a chatbot, they will actually be conversing with an experimenter following a script. The assistant for the two groups will be working off the same script to avoid introducing any disparity in the experimental setups. The subject will have to answer a list of multiple-choice questions requiring her reasoning and logical skills. The questions are designed to be non-trivial enough that the assistant can be helpful and provide advice. For example, some of them are based on NASA's survival on the moon scenario, where the participant has to rank a list of various objects by their usefulness on the moon. Some non-obvious answers are that a compass is useless without Earth's magnetic field, and a parachute can provide protection against the sun's rays. When questioning the subject, the assistant will have different interactivity levels, assigned randomly: simply asking the question, providing a tip beforehand, or challenging the answer provided by the subject. Whatever the level of interactivity, the assistant will also provide additional advice to the subject if prompted.

We will measure the number and length of the messages sent to the assistant, as well as their time distribution (evenly distributed, clustered towards the beginning or the end of the conversation). The messages will also be classified by type (for example, questions about the assistant or requests for help). We will evaluate the performance of the participant by the reflection time for each question and the number of right answers given. As we want to assess the participant's trust in the assistant, we will determine how many times the subject followed the assistant advice and changed her answer. The participant will also fill a post-experiment survey with her perceived fluency with instant messaging, her estimated score, and her trust in the assistant's advice.

4 Expected contributions and future works

With this first experiment, we should get our first results on how communication and trust with an agent differ depending on whether he or she is perceived as human or as a robot. We theorize that the participants will exchange more when they think they are discussing with a chatbot, as they should feel less self-conscious and more inclined to ask for advice (Brandtzaeg and Følstad, 2017). However, we are aware that the results might be affected by the observer's effect as the subjects know that

their conversation logs will be read. We also expect them to use shorter sentences, as people tend to speak less courteously to robots and more directly (Hill et al., 2015). Next, we expect the human assistants to be given a better score in the survey and to have their advice followed more often. Indeed, chatbots tend to be perceived less favorably than humans when their tasks are not purely technical (Seeger et al., 2017).

With those results, our next step will be to introduce more human participants in the experiment by having three humans communicating by chat and trying to efficiently resolve a task with the help of the assistant. This next experiment will allow us to see how group dynamics are affected by the help of a conversational agent and how the participants integrate the agent in their reflection. This will bring us one step closer to our research goal, which concerns large-scale collaboration.

Additionally, we aim to organize focus groups with some of the anticipated users of the conversational agent: firemen or the army, for example. With this, we will be able to have a more accurate picture of the actual needs and expectations for this chatbot and design the first prototype accordingly.

References

Brandtzaeg, P. B. and A. Følstad (2017): 'Why people use chatbots'. In: *International Conference on Internet Science*. pp. 377–392.

Dahlbäck, N., A. Jönsson, and L. Ahrenberg (1993): 'Wizard of Oz studies: why and how'. In: *Proceedings of the 1st international conference on Intelligent user interfaces*. pp. 193–200.

Daim, T. U., A. Ha, S. Reutiman, B. Hughes, U. Pathak, W. Bynum, and A. Bhatla (2012): 'Exploring the communication breakdown in global virtual teams'. *International Journal of Project Management*, vol. 30, no. 2, pp. 199–212.

Dale, R. (2016): 'The return of the chatbots'. *Natural Language Engineering*, vol. 22, no. 5, pp. 811–817.

Dodgson, M. (1993): 'Learning, trust, and technological collaboration'. *Human relations*, vol. 46, no. 1, pp. 77–95.

Følstad, A., C. B. Nordheim, and C. A. Bjørkli (2018): 'What makes users trust a chatbot for customer service? An exploratory interview study'. In: *International Conference on Internet Science*. pp. 194–208.

Hill, J., W. R. Ford, and I. G. Farreras (2015): 'Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations'. *Computers in human behavior*, vol. 49, pp. 245–250.

McAfee, A., E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton (2012): 'Big data: the management revolution'. *Harvard business review*, vol. 90, no. 10, pp. 60–68.

Misiura, J. and A. Verity (2019): 'Chatbots in the humanitarian field - Concepts, uses and shortfalls'. https://www.digitalhumanitarians.com/chatbots-in-the-humanitarian-field-concepts-uses-and-shortfalls.

Müller, L., J. Mattke, C. Maier, T. Weitzel, and H. Graser (2019): 'Chatbot Acceptance: A Latent Profile Analysis on Individuals' Trust in Conversational Agents'. In: *Proceedings of the 2019 on Computers and People Research Conference*. pp. 35–42.

Ramchurn, S. D., F. Wu, W. Jiang, J. E. Fischer, S. Reece, S. Roberts, T. Rodden, C. Greenhalgh, and N. R. Jennings (2016): 'Human–agent collaboration for disaster response'. *Autonomous Agents and Multi-Agent Systems*, vol. 30, no. 1, pp. 82–111.

Seeger, A.-M., J. Pfeiffer, and A. Heinzl (2017): 'When do we need a human? Anthropomorphic design and trustworthiness of conversational agents'. In: *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS, AISeL, Seoul, Korea*, Vol. 10.

Vieweg, S., C. Castillo, and M. Imran (2014): 'Integrating social media communications into the rapid assessment of sudden onset disasters'. In: *International Conference on Social Informatics*. pp. 444–461.

Wang, W. and K. Siau (2018): 'Trust in Health Chatbots'.

Zadrozny, W., M. Budzikowska, J. Chai, N. Kambhatla, S. Levesque, and N. Nicolov (2000): 'Natural language dialogue for personalized interaction'. *Communications of the ACM*, vol. 43, no. 8, pp. 116–120.