Position paper presented at CONVERSATIONS 2020 – the 4th International Workshop on Chatbot Research, an online virtual event hosted by the University of Amsterdam, the Netherlands, November 23-24.

Impact of Control Features on Human Autonomy and Trustworthiness of Affective Chatbots

Ivo Benke[1], Ulrich Gnewuch[1], and Alexander Maedche[1]

[1] Karlsruhe Institute of Technology, Karlsruhe, Germany
[prename.surname]@kit.edu

Abstract. Virtual team communication at the workplace has changed through the leverage of innovative collaboration technologies like Slack, Zoom, and MS Teams. However, virtual teams face emotional obstacles during remote communication. To mitigate these challenges, affective chatbots can be applied by increasing the teams' emotional management capabilities. While, through advances in affective computing, these chatbots can understand human affective signals, chatbots that analyze and intervene in team emotions can increase feelings of surveillance and lack of autonomy. This ultimately reduces users' trust. In order to address this challenge, one strategy might be to design trustworthy affective chatbots through human agency-based control features. Since research is scarce about the effects of control features on users' trust of chatbots, in this paper, we present a research design for the evaluation of affective chatbot control features on users' trust and cognitive load. With our work, we plan to contribute theoretical knowledge about the effects of human agency-based control features in affective chatbots. Further, we pave the way for trustworthy affective chatbots through overcoming downsides of existing affective chatbots designs.

Keywords: Affective chatbots, Human autonomy, Control features, Trustworthiness, Virtual team

1 Introduction

Advances in affective computing have enabled chatbots to evolve from machine-like entities to partners for humans [1]. Research has started to identify such emotion-aware chatbots that can sense and understand human emotions enabled through artificial intelligence (AI) [2]. Since these chatbots can moderate emotion management in virtual team work, they are described as affective chatbots [3]. Simultaneously, innovative collaboration technologies, such as Slack and Microsoft Teams, have changed communication patterns at the workplace and empower virtual team communication [4]. Today, over 50% of the working population in the United States works remotely and thereby has become part of virtual teams using such tools [5]. However, in distributed communication, virtual teams face challenges, such as conflicts, breakdowns, or groupthink, which disrupt the flow of communication. These issues can be traced back to obstacles in the management of emotions in the team [6]. Therefore, research has proposed to introduce affective chatbots into virtual team communication (e.g. [7]) in order to support team members' emotional capabilities.

Simultaneously, human emotions are very intimate and sensitive [8]. AI-based disclosure of innermost emotions is associated with personal vulnerability [9]. Previous research on affective computing has shown that providing emotional feedback can increase feelings of surveillance, impaired transparency, and lack of autonomy [10]. These feelings lead to distrust against systems that can expose and use human emotions. However, trust is an important driver for the acceptance and use of information systems [11]. Therefore, affective chatbots need to regain trustworthiness which requires careful reflection on when and how to apply them and their unique abilities.

Although it is unclear how to enable human autonomy in the context of affective chatbots, human agency-based control has proven to help against this backdrop in other research contexts [12]. Human agency refers to the intentional influence on one's functioning and life circumstances [13], and determines control. Examples of instantiated control features include location-based service settings in navigation applications (e.g. through range sliders on mobile phones) or control bars of autopilot flight dashboards. In contrast, previous research on machine automation, aircraft, and military has also identified risks when applying additional control features to machine interfaces because these features increase users' cognitive load [14]. Taken together, it remains unclear what the effects of control features of affective chatbots on virtual teams' trusting beliefs are. Hence, we address the following research question:
What are the effects of human agency-based control features of affective chatbots on virtual team members' trust and cognitive load?

In order to answer this research question, we plan to conduct a between-subjects online experiment. We applied existing knowledge on privacy control and developed five control features for affective chatbots. We instantiated these features in an affective chatbot control center next to the group chat in order to increase human autonomy over affective chatbots. Each feature represents a specific control dimension (e.g. visual appearance, behavior, or timing) which effect will be evaluated in the planned experiment. Through this study, we aim to contribute to research in three ways. First, we derive theoretical knowledge about control feature effects in affective chatbots. Second, we aim to overcome downsides of affective chatbots and pave the way for trustworthy affective chatbots. Finally, we offer prescriptive knowledge for designing trustworthy affective chatbots building on human agency-based control features.

2 Conceptual Foundation

2.1 Affective Chatbots in Virtual Teams

Virtual teams are comprised of individuals who work interdependently using computer-mediated communication to accomplish a shared organizational objective [15]. In contrast to face-to-face teams, virtual teams face unique obstacles to establish effective communication due to the lack of verbal and non-verbal cues supported through technology, and further ensuing problems like difficulties in conflict management or groupthink [6]. These obstacles and the impaired communication, however, complicate the management of emotions of team members [16]. Simultaneously, the ability to manage emotions has strong effects on the individual and the team [17] since this ability plays an important role and is strong predictor for job performance [6].

A promising approach for supporting emotion management in teams is the application of AI-based chatbots. Chatbots have been shown to successfully act as team facilitators [18]. Based on the concept of affective computing, affective chatbots are introduced. They extend and apply emotion-aware capabilities in order to support virtual teams by enabling emotion management through behavioral and persuasive cues [3]. First attempts have been promising and show the positive effect of these systems on emotion perception and communication efficiency [7]. However, previous research on affective systems has also identified negative effects. Humans tend to perceive the system's ability to recognize emotions as threatening and displeasing which leads to a significant decrease in trustworthiness and, thereby, aversion to use the affective system [19]. In the case of affective chatbots, this was emphasized through the loss of usefulness and loss of autonomy [3].

2.2 Human Agency, Control, and Autonomy

Human agency is the ability to exercise self-regulation, intentionality, and experience self-efficacy. It is a necessary condition for autonomy [20]. *"Unless people believe they can produce desired effects by their actions, they have little incentive to act"* (p.7) or accept affective systems [21]. Among mechanisms of agency, none is more central than people's beliefs in their control over their own functioning and over environmental events (p.1) [22]. Simultaneously, sense of control is a robust predictor of physical and mental well-being [23]. In order to establish control, several control features have been developed and defined as instantiations that "[..] *affect the behavior of another person as a means to achieve goals related to designing, developing, operating, using, and managing information systems."* (p.3) [24]. Exemplary control features are location-based service settings for navigation applications (e.g. through range sliders [12]). Design instantiations have been successfully applied for example in aircraft systems [25] and prominently for establishing of privacy control [26]. Lack of human autonomy was one central downside during the interaction with affective chatbots [3]. Human agency-based control features might serve as potential solution in order to tackle this downside.

2.3 Trust and Trustworthiness

To counteract these risks, the reestablishment of trust into a system is essential. Trust describes the trusting beliefs about system's competence, benevolence and integrity [27]. Trustworthiness of a system, in turn, is a characteristic of the system, which is informed by a set of values and previous behaviors [28, 29]. Trust is an important factor for a system's acceptance and usage of information systems [30] and has been proven to hold in the context of AI-enabled, intelligent agents as well (c.f. [20]). Therefore, we derive the necessity to instantiate trustworthiness for affective chatbots through system characteristics in order to eliminate the negative experiences with affective chatbots.

3 Hypothesis Development

Figure 1 presents the research model and the proposed hypotheses. The preservation of user autonomy should be explicitly defined as a characteristic of agents and addressed through the design [31]. Human agency and control may be a potential strategy to provide user autonomy [21]. To establish human agency, research suggests implementing control features. These features lead to control over an agent, which in turn allows for decisions by the user independent on the agent's behavior [32]. In contrary, user autonomy is reduced if agents do not provide the user with necessary abilities to achieve the intended goals [31]. Against this backdrop, we therefore hypothesize that the provision of control features has a positive effect on user autonomy.

H1: *Providing control features for the affective chatbot increases users' perceived autonomy.*

Supporting human autonomy has generally been associated with higher trust between humans [33] since it creates an environment conducive to trust development through greater interest, less pressure and tension, and higher self-esteem [34]. Moreover, trust is one of the most important affective elements which influence human-computer interaction [35]. Affective chatbots, on the one hand, are AI-based interactive technology. They use anthropomorphic design elements, such as anthropomorphic appearance, interactive features like breaks, and a moderative style [36], to appear human-like and are perceived as social actors [37]. Therefore, we transfer this prospering relationship from human-human to human-computer interaction and propose:

H2a: *Perceived autonomy increases system-like trust in the affective chatbot.*

H2b: *Perceived autonomy increases human-like trust in the affective chatbot.*

A high degree of control gives users freedom over all potential actions. The concept of progressive disclosure states that too many available options are distracting and lead to an increased cognitive load when interacting with user interfaces of interactive systems [38]. Translating this argumentation to the context of affective chatbots, we propose:

H3: *Providing control features for the affective chatbot increases users' cognitive load.*

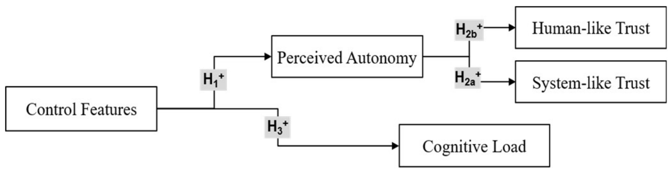

Fig. 1. Research Model and Hypotheses.

4 Method

The hypotheses will be tested in an online experiment using five treatment groups with varying control features as stimuli (between-subjects design). In the experiment, participants are confronted with a text-based chat of three team members. One of these members is an affective chatbot that supports the team's emotional intelligence. The affective chatbot senses the team emotions and interacts with the team through previously designed behavior which adapts content, social roles, dynamics and interaction format like chat breaks.

Participants. The online experiment will be conducted with 305 participants equally distributed on two experimental panels, Prolific and Amazon Mechanical Turk (MTurk). The sample size is calculated through an a priori power analysis. The participants are randomly assigned to one of the four experimental conditions or the control condition.

Affective Chatbot and Experimental Stimuli. The affective chatbot serves as a moderator in the group chat and intervenes when negative emotions arise. It implements design features, such as social cues [39], behavioral interventions (e.g. chat breaks and feedback), as well as comparing images and motivating speech. Through this design, it stimulates emotional capabilities of virtual team members. Based on privacy control research [26], we added an affective chatbot control bar to the chat window. The chatbot is displayed in Fig. 2. It contains control features that represent different control levels: an (1) overall "power" feature to turn the chatbot on and off, a (2) representative level on which users can adjust the appearance of the affective chatbot (i.e., visual/textual presentation, individual/group values, proactive/reactive timing), and a (3) behavioral level which allows to adapt blocking/unblocking behavior of the affective chatbot.

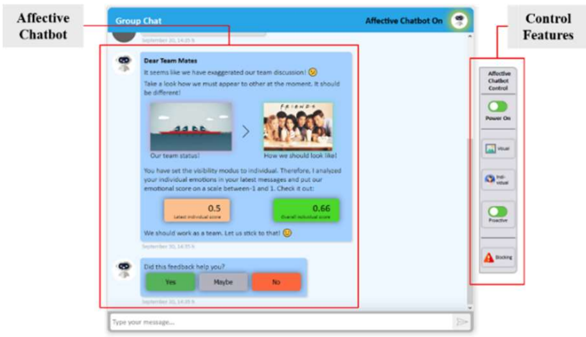

Fig. 2. Scenario group chat with affective chatbot (center) and control features (right). The figure shows the full treatment with all available control features.

Procedure and Task. The online experiment is planned for 15 minutes via an online questionnaire. Participants will be welcomed and watch an introductory video that explains the functionalities of the control features and describes the scenario. Afterwards they experience the treatment where they run through a pre-scripted group chat conversation where the affective chatbot intervenes. Finally, they will answer a questionnaire and finish with a debriefing.

Measurement. We use established measures for all constructs. Items will be adapted to the study context and measured using 7-pt. Likert scales. Perceived autonomy will be measured using three items taken from [40]. System-like trust will be measured following the conceptualization of [41]. It is composed of three subdimensions reliability, functionality, and helpfulness. Human-like trust will also be measured based on [41]. It is composed of three subdimensions integrity, competence, and benevolence. We will assess cognitive load via the Cognitive Load Component Survey, a ten item questionnaire [42]. We include several control variables (e.g., demographics, prior experience). As attention check, we report the time the participants take for conducting the scenario conversation with the team. As a manipulation check, we will ask participants to rate the extent to which they had control over the affective chatbot.

5 Expected Contribution and Next Steps

In this position paper, we describe our ongoing research on the need for autonomy provision for the system class of chatbots with enhanced capability of emotion awareness, called affective chatbots. Subsequently, we presented instantiations of control features and derived hypothesis in order to measure their positive effect on trusting beliefs, and their negative effect on cognitive load, potentially identifying a balanced sweet spot for their implementation. With this work, we aim to contribute to chatbot research by providing theoretical knowledge about the effects of control features and about the reestablishment of trustworthiness in affective chatbots. Simultaneously, we plan to derive prescriptive knowledge for autonomy provision of affective chatbots in the form of control features. Our work also comes with limitations. Autonomy is only one determinant of trustworthiness. Therefore, trustworthiness might also be addressed from different perspectives (e.g. transparency). Moreover, conducting the experiment via MTurk and Prolific might affect results due to the scenario-based experience of participants without a real virtual team environment. As next steps, we will execute the presented online experiment and analyze the results to test our hypotheses. Through our results, we hope to further facilitate the trustworthiness and acceptance of advanced, emotion-aware affective chatbots for real-world virtual teams and enable the advancement of chatbots from basic functional entities into capable team moderators and acknowledged team members.

References

1. Poria, S., Cambria, E., Bajpai, R., Hussain, A.: A review of affective computing: From unimodal analysis to multimodal fusion. Inf. Fusion. 37, 98–125 (2017).
2. Mensio, M., Rizzo, G., Morisio, M.: The Rise of Emotion-aware Conversational Agents. In: Companion Proceedings of the The Web Conference 2018. International World Wide Web Conferences Steering Committee. pp. 1541–1544 (2018).
3. Benke, I., Knierim, M.T., Maedche, A.: Chatbot-based Emotion Management for Distributed Teams: A Participatory Design Study. Proc. ACM Human-Computer Interact. 4, 1–30 (2020).
4. Stoeckli, E., Dremel, C., Uebernickel, F., Brenner, W.: How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electron. Mark. 1–35 (2019).
5. Forbes.com: 10 Remote Work Trends That Will Dominate 2019, https://www.forbes.com/sites/abdullahimuhammed/2018/12/21/10-remote-work-trends-that-will-dominate-2019/#6aea2b847c72, (2019).
6. Pitts, V.E., Wright, N.A., Harkabus, L.C.: Communication in Virtual Teams : The Role of Emotional Intelligence. J. Organ. Psychol. 28, 2046–2054 (2012).

7. Peng, Z., Kim, T., Ma, X.: GremoBot: Exploring emotion regulation in group chat. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing. pp. 335–340 (2019).
8. Brave, S., Nass, C.: Emotion in Human–Computer Interaction. 53–68 (2009).
9. Moon, Y.: Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. J. Consum. Res. 26, 323–339 (2000).
10. IEEE: ETHICALLY ALIGNED DESIGN: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. (2019).
11. Wang, W., Qiu, L., Kim, D., Benbasat, I.: Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. Decis. Support Syst. 86, 48–60 (2016).
12. Ataei, M., Degbelo, A., Kray, C.: Privacy theory in practice: designing a user interface for managing location privacy on mobile devices. J. Locat. Based Serv. 12, 141–178 (2018).
13. Moore, J.W.: What is the sense of agency and why does it matter? Front. Psychol. 7, 1–9 (2016).
14. Wickens, C.D.: Designing for Situation Awareness and Trust in Automation. IFAC Integr. Syst. Eng. 28, 365–370 (1995).
15. Martins, L.L., Gilson, L.L., Maynard, M.T.: Virtual teams: What do we know and where do we go from here? J. Manage. 30, 805–835 (2004).
16. Hassib, M., Buschek, D., Wozniak, P.W., Alt, F.: HeartChat: Heart Rate Augmented Mobile Messaging to Support Empathy and Awareness. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. pp. 2239–2251 (2017).
17. Kelly, J.R., Barsade, S.G.: Mood and emotions in small groups and work teams. Organ. Behav. Hum. Decis. Process. 86, 99–130 (2001).
18. Følstad, A., Brandtzæg, P.B.: Chatbots and the new world of HCI. Interactions. 24, 38–42 (2017).
19. McDuff, D., Czerwinski, M.: Designing emotionally sentient agents. Commun. ACM. 61, 74–83 (2018).
20. Banks, J.: A perceived moral agency scale: Development and validation of a metric for humans and social machines. Comput. Human Behav. 90, 363–371 (2019).
21. Bandura, A.: Toward a Psychology of Human Agency. Perspect. Psychol. Sci. 1, 164–180 (2006).
22. Bandura, A.: Human Agency in Social Cognitive Theory. Am. Psychol. 44, 1175–1184 (1989).
23. Skinner, E.A.: A Guide to Constructs of Control. J. Pers. Soc. Psychol. 71, 549–570 (1996).
24. Cram, W.A., Brohman, M.K., Gallupe, R.B.: Information systems control: A review and framework for emerging information systems processes. J. Assoc. Inf. Syst. 17, 216–266 (2016).
25. Norman, D.A.: The "Problem" with Automation: Inappropriate Feedback and Interaction , not "Over- Automation." Philos. Trans. R. Soc. London. 327, (1990).
26. Schaub, F., Balebako, R., Cranor, L.F.: Designing Effective Privacy Notices and Controls. IEEE Internet Comput. 21, 70–77 (2017).
27. McKnight, D.H., Lankton, N.K., Nicolaou, A., Price, J.: Distinguishing the effects of B2B information quality, system quality, and service outcome quality on trust and distrust. J. Strateg. Inf. Syst. 26, 118–141 (2017).

28. Ben-Ner, A., Halldorsson, F.: Trusting and trustworthiness: What are they, how to measure them, and what affects them. J. Econ. Psychol. 31, 64–79 (2010).
29. Chee-Wee, T., Benbasat, I., Cenfetelli, R.T.: Building citizen trust towards e-government services: Do high quality websites matter? Proc. Annu. Hawaii Int. Conf. Syst. Sci. 1–10 (2008).
30. Lee, S.Y., Choi, J.: Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. Int. J. Hum. Comput. Stud. 103, 95–105 (2017).
31. Friedman, B., Nissenbaum, H.: Software Agents and User Autonomy. In: Autonomous Agents (1997).
32. Piccoli, G., Ives, B.: Trust and the Unintended Effects of Behavior Control in Virtual Teams. MIS Q. 27, 695–395 (2003).
33. Deci, E.L., Ryan, R.M.: The Support of Autonomy and the Control of Behavior. J. Pers. Soc. Psychol. 53, 1024–1037 (1987).
34. Wiener, M., Mähring, M., Remus, U., Saunders, C.: CONTROL CONFIGURATION AND CONTROL ENACTMENT IN INFORMATION SYSTEMS PROJECTS: REVIEW AND EXPANDED THEORETICAL FRAMEWORK. MIS Q. 40, 741–774 (2016).
35. Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Hum. Factors. 46, 50–80 (2004).
36. Wang, W., Benbasat, I.: Empirical Assessment of Alternative Designs for Enhancing Different Types of Trusting Beliefs in Online Recommendation Agents. J. Manag. Inf. Syst. 33, 744–775 (2016).
37. Feine, J., Gnewuch, U., Morana, S., Maedche, A.: A Taxonomy of Social Cues for Conversational Agents. Int. J. Hum. Comput. Stud. 132, 138–161 (2019).
38. Springer, A., Whittaker, S.: Progressive Disclosure: Designing for Effective Transparency. (2018).
39. Araujo, T.: Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput. Human Behav. 85, 183–189 (2018).
40. Pirkkalainen, H., Salo, M., Makkonen, M., Tarafdar, M.: Coping with Technostress: When Emotional Responses Fail. In: Proceedings the 38th International Conference on Information Systems. pp. 1–18 (2017).
41. Lankton, N.K., Mcknight, D.H., Tripp, J.: Technology, Humanness, and Trust: Rethinking Trust in Technology. J. Assoc. Inf. Syst. 16, 880–918 (2015).
42. Morrison, B.B., Dorn, B., Guzdial, M.: Measuring cognitive load in introductory CS: Adaptation of an instrument. ICER 2014 - Proc. 10th Annu. Int. Conf. Int. Comput. Educ. Res. 131–138 (2014).