Show *Me* Yours if I Show You Mine

Self-disclosure in Conversational Agents

Loes van Uffelen
Human Media Interaction,
University of Twente
Enschede, The Netherlands
j.m.l.vanuffelen@student.utwente.nl

Jelte van Waterschoot
Human Media Interaction,
University of Twente
Enschede, The Netherlands
j.b.vanwaterschoot@utwente.nl

Mariët Theune
Human Media Interaction,
University of Twente
Enschede, The Netherlands
m.theune@utwente.nl

ABSTRACT

Self-disclosure describes the act of revealing personal information to others; in other words, revealing a part of one's own identity. Implementing a form of self-disclosure in conversational agents, including social robots, might contribute to the user's feeling of engagement and stimulate them to reveal more information about themselves. This paper addresses whether self-disclosure in dialogue of conversational agents could benefit interaction with the user and how dialogue scripts should be designed. We also briefly discuss our own use case, an embodied conversational agent able to talk with older adults about their well-being.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design theory, concepts and paradigms**; User models; *HCI theory, concepts and models*; Natural language interfaces.

KEYWORDS

self-disclosure, older adults, conversational agents

1 INTRODUCTION

People can learn much about somebody's identity in casual conversation, often more even than by asking questions directly [9]. Sharing stories and finding commonalities in each other's identities creates a bond, rapport [6]. In this paper, we focus on self-disclosure as a way of doing this between conversational agents and their human users.

Conversational agents, including social robots, have now reached a point where smooth, natural communication with a person is possible. This is eminently beneficial for older adults, who often lack proficiency in interaction with archetypal computer systems

or are unfamiliar with modern technology [7]. This has encouraged researchers to design healthcare applications for older adults with conversational interaction. Spoken dialogue with virtual characters or companion robots can support services that aim to reduce loneliness [1, 5, 6, 17, 24, 25, 31, 33], bring entertainment, track health care needs, or provide assistance in daily life [13, 27–29]. We argue that part of an agent's identity design should be designing self-disclosure behaviour to build rapport with users.

Conversational agents often have more added value to users if they successfully obtain information about them, mainly through conversation. This information can be used to improve the interaction through personalization, or to gather user data for research. Self-disclosure in agents can be a relevant factor in encouraging users to share information about themselves, by triggering interlocutors' disclosure reciprocity. In human communication, disclosure reciprocity is an interpersonal effect of self-disclosure: recipients disclose information about themselves in a comparable level of intimacy as the self-disclosure input they receive [21]. Disclosures typically gradually escalate from casual exchanges to increasingly intimate ones over time. The four stages that describe the self-disclosure process are categorized as orientation, exploratory, affect-exchange, and stable-exchange [2]. According to the social penetration theory (see below), disclosure reciprocity can only be expected if the pattern of escalation is not violated.

The principle of disclosure reciprocity, which is valid in human-human communication, may also be found in human-computer interaction. This is based on two communication theories. Firstly, the social exchange theory explains that people attempt to maintain equality in self-disclosure: imbalance makes them uncomfortable. This reinforces a so-called "norm of reciprocity": reciprocating disclosure is a social norm [8, p. 111]. Secondly, the theory of social response explains that humans generally tend to treat computers as social actors, disregarding the realization that machines do not have emotions, expectations, "selves", or human intentions [14, 18, 19]. This is often referred to as the CASA (Computers Are Social Actors) paradigm [15, 16, 20].

Not only could self-disclosure by conversational agents elicit disclosure reciprocity in users, it might also boost the users' engagement. Establishing common ground between human interlocutors is often indispensable for the development of interpersonal relationships that are built on trust and engagement. This could also be true for conversational agents, with a positive effect on the duration or frequency of use of such agents. However, design and implementation of self-disclosure in conversational agents is not self-evident, as conversational practice between humans is a sophisticated, complex process. Machines must learn to conform to socially appropriate

norms of interaction. For instance, gradual escalation of intimacy of self-disclosures by an agent might be easily accidentally violated. Furthermore, statements about human behaviour by an agent may as well be perceived as disingenuous or preposterous, depending on the kind of self-disclosure by the agent.

In this paper we discuss current developments of self-disclosure in conversational agents and dialogue design and briefly discuss our use-case of self-disclosure with older adults.

2 EFFECTS OF SELF-DISCLOSURE

We discuss the related work on self-disclosure in agents with two questions in mind: first, how a self-disclosing agent can influence the user's behaviour, including communication behaviour such as disclosure reciprocity, and second, what self-disclosure in agents does with the perception of the agent.

2.1 Communication Behaviour

Ravichander and Black found that a conversational agent with self-disclosing behaviour correlates with instances of self-disclosure by the user [21]. Kang et al. compared virtual agents that have either human or computer backstories (e.g. "I was born and raised in ..." versus "I was built and designed in ..."). They showed that participants express more non-verbal behaviour related to anxiety in the computer backstory condition [11]. van der Drift et al. studied the use of a self-disclosing robot for motivating diabetic children to keep a diary, and concluded that "The robot's self-disclosure appeared to be a strong motivator for the child to disclose information in return" [30, p. 468]. Lee et al. studied the effect of self-disclosure in a chatbot over a longer period of time, and concluded that self-disclosure in agent is more beneficial for deeper conversations with sensitive questions than for journaling and small-talk.

2.2 Perception

Schulman and Bickmore showed that embodied agents that simulate human conversation through social dialogue are positively perceived and could persuade participants to change their attitudes towards exercise [26]. Moon showed that self-disclosure in text-based human-machine interaction increased trust, closeness and liking [18]. Razavi et al. designed the LISSA dialogue manager for long-term casual conversation with older adults. Compared to an earlier Wizard-of-Oz version, their fully automated LISSA system received higher ratings on criteria such as being encouraging and polite, which the authors suspect was due to its use of backstories and self-disclosure [22]. This is in line with the findings of Bickmore et al. that users rate interactions with agents that have backstories in dialogue as more enjoyable [4].

3 DIALOGUE DESIGN

Agents that aim for casual or long-term conversation with humans depend on the user's willingness to converse with them. This willingness could be increased by considering the implementation of self-disclosure in dialogue scripts. Self-disclosure can manifest itself in various ways: Altman and Taylor distinguish between personal history, thoughts, opinions, beliefs, feelings, preferences, attitudes, aspirations, likes, dislikes and favourites [2]. Jourard developed the 25-item self-disclosure questionnaire [10], comprising a list of

Figure 1: An older adult interacting with our agent.

conversational topics rated on their degree of intimacy. Examples of high-intimate topics are love, personal deficiencies, income and the user's appearance. Low to moderately intimate topics include spare time, music, sports, food and travelling.

3.1 Believability

For some types of self-disclosure it is debatable where we should draw the line between believability and deceit. This ethical issue has been raised by Bickmore et al. [4]. Arguably, believability of the agent depends among other things on the intimacy of topics, where higher intimacy topics could be seen as less believable and more "deceitful". An agent "missing his mother" seems less believable than an agent mentioning nice weather, suggesting that high-intimacy topics are less suitable for self-disclosure in agents than low-intimacy topics. However, discussing "safe" low-intimacy topics does not guarantee believability of agents' self-disclosure. Agents conversing about doing human activities seem less believable than having an opinion on that activity. For example, on the topic of food an agent could say: "Yesterday I had spaghetti, I used a recipe from a cookbook I own", or "I really like Italian cuisine, lately I saw a nice recipe." It seems harder to believe an agent could experience the human activity of eating than that an agent could give an opinion on this activity.

Paradoxically, self-disclosure by agents on high-intimacy topics may be less believable, but it may also be more important. High-intimacy topics are generally considered high-revealing, and revealing information from users could be the objective of a conversational agent. Pickard et al. [20] discussed several reasons why users would prefer an embodied conversational agent (ECA) or a human interviewing them on more or less sensitive topics, and found that an ECA interviewer was preferred for more sensitive topics, while a human interviewer was preferred for less sensitive topics. Possible reasons for this include the agent's lack of judgement and more personal comfort when conversing with an ECA due to reduced negative feelings, such as anxiety, embarrassment, or guilt.

The deceit versus engagement issue is hard to overcome. The deceit that might result from implementing self-disclosure in agents could create misleading expectations about the system's capabilities. To avoid this, dialogue could be scripted that suggests the agent

recognizes his own environment and refers to itself as a virtual character rather than human [4, 18]. Examples of casual dialogue for agents that recognize their environment could be: "Do you have plans for the holidays? [...] I will only have a vacation if you unplug this computer from the power!"

3.2 Emotion

Displays of emotion (feelings) are also a form of self-disclosure. As demonstrated in a study about a self-compassionate chatbot, participants can feel conflicted by a chatbot's display of emotion [14, p. 9]. There are differences between people in how they experience this: the study showed that some appreciated the display of emotions by the chatbot, while others felt discomfort. Accordingly, any emotional display by an agent must be adjusted to the target group and the user must have a sense of control of how an agent "feels" towards the user. However, for agents that are not necessarily involved in establishing a close relationship with the user, it would be prudent to minimize emotional displays explicitly directed towards the user such as "I miss you" or "I really love talking to you".

4 USE CASE: OLDER ADULTS

In our research project *BLISS* we focus on the design of conversational agents that are capable of having casual conversations with older adults. The agent's purpose is to gather information on the factors influencing users' personal well-being [32], as a step towards improving healthcare. In the context of this research project, we designed an experiment with two versions of an embodied conversational agent (ECA), to investigate whether self-disclosure by an ECA would increase positive attitude towards the ECA and disclosure reciprocity of older adults.

4.1 Dialogue design: Self-disclosure

Taking into account the considerations discussed in Section 3, we designed the dialogues with the ECA as follows. We built in a gradual escalation of self-disclosure by the agent, but not moving beyond the minimum level of intimacy required to meet the agent's goal (the topics loneliness and family representing the highest level of intimacy). We avoided emotional statements from the agent, as well as statements that could potentially mislead the user about the agent's capabilities.

4.2 Technical setup

We conducted a Wizard-of-Oz pilot study with a virtual avatar. This was done fully online due to COVID-19 restrictions. The setup consisted of three computing devices: one computer running the software of the avatar, another running a wizard interface and a tablet, phone or laptop for the participant. We created a male avatar in Unity, with the UMA package [1] (see Figure 1). ASAPRealizer and the UnityBridge served as behaviour realizers for the avatar [12, 23]. This software was run on one computer. The wizard interface on the other computer was used to send behaviours to the ECA. These consisted of self-disclosure questions and statements, communicative intents such as affirmation and deny, or custom messages typed by the wizard (the experimenter).The only non-verbal behaviours

---

[1]http://umawiki.secretanorak.com/Unity_Multipurpose_Avatar

implemented were idle breathing and blinking. Finally, participants used their own tablet or computer to connect to a video call in which both the ECA and experimenter were present.

4.3 Procedure

Prior to the experiments, participants received an information brochure and gave informed consent. Participants were videocalled via their computer or tablet and instructed to converse with the conversational agent. Each participant had two conversations, one with and one without self-disclosure by the ECA, and with a different list of topics for each conversation. Both the topic list and with/without self-disclosure variables were counterbalanced to circumvent any habituation effects.

After the participants finished the two conversations with the agent, the experimenter conducted a semi-structured interview with them, asking them about their attitude towards the two versions of the ECA. Interview questions were based on descriptors of the Godspeed scale [3].

4.4 Analysis Method

To answer our research question, besides analysing the outcomes of the interviews, we also manually annotate transcriptions of the audio recordings using a coding scheme for self-disclosure reciprocity inspired by [2]. Self-disclosure is measured by counting snippets of personal information. Additionally, we analyze the word count and duration of each conversation. We expect that number of snippets, word count and duration will be higher in the self-disclosure conversations.

4.5 Status

We recently carried out the pilot with eight participants, recruited through the *BLISS* project. All participants were Dutch citizens aged 65 or above and had some experience with video calls, but none with virtual agents. The results are currently being analysed. The results will be reported elsewhere.

5 CONCLUSION

We believe that designing self-disclosure behaviour of an agent (either a robot or a virtual character) is a vital part of their identity. It can lead to increased engagement and stimulate disclosure reciprocity by users. However, designing spoken dialogues for interaction with specifically older adults requires a careful deliberation of how to incorporate a self-disclosure strategy in the agent. Disclosure that is perceived as deceitful is rather hard to overcome for user acceptance. Additionally, self-disclosing too much or too little might have detrimental effects on reciprocal disclosure. Finally, disclosure by conversational agents might be easier for low-intimate topics, though high-intimate topics might encourage the user to reveal more valuable information. Our future work is finalizing the pilot study analysis and developing an autonomous ECA with self-disclosure that we can evaluate on a larger scale.

REFERENCES

[1] H. Abdollahi, A. Mollahosseini, J. T. Lane, and M. H. Mahoor. 2017. A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. IEEE, Birmingham, UK, 541–546. https://doi.org/10.1109/HUMANOIDS.2017.8246925

[2] Irwin Altman and Dalmas A Taylor. 1973. *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston, Oxford, England.

[3] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (1 1 2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[4] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2009. Engagement vs. Deceit: Virtual Humans with Human Autobiographies. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents* (Amsterdam, The Netherlands) *(IVA '09)*. Springer-Verlag, Berlin, Heidelberg, 6–19. https://doi.org/10.1007/978-3-642-04380-2_4

[5] Timothy W Bickmore, Lisa Caruso, Kerri Clough-Gorr, and Tim Heeren. 2005. 'It's just like you talk to a friend'relational agents for older adults. *Interacting with Computers* 17, 6 (2005), 711–735.

[6] Timothy W Bickmore and Rosalind W Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 293–327.

[7] Sara J Czaja, Neil Charness, Arthur D Fisk, Christopher Hertzog, Sankaran N Nair, Wendy A Rogers, and Joseph Sharit. 2006. Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (CREATE). *Psychology and aging* 21, 2 (2006), 333.

[8] Valerian J Derlaga and John H Berg. 1987. *Self-disclosure: Theory, research, and therapy*. Springer Science & Business Media, New York.

[9] Suzanne Eggins and Diana Slade. 2001. *Analysing Casual Conversation* (1st ed.). Continuum, London, UK.

[10] Sidney M Jourard. 1961. Self-disclosure patterns in British and American college females. *The Journal of Social Psychology* 54, 2 (1961), 315–320.

[11] Sin-Hwa Kang, Albert (Skip) Rizzo, and Jonathan Gratch. 2012. Understanding the Nonverbal Behavior of Socially Anxious People during Intimate Self-Disclosure. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents* (Santa Cruz, CA) *(IVA'12)*. Springer-Verlag, Berlin, Heidelberg, 212–217. https://doi.org/10.1007/978-3-642-33197-8_22

[12] Jan Kolkmeier, Merijn Bruijnes, and Dennis Reidsma. 2017. A Demonstration of the ASAP Realizer-Unity3D Bridge for Virtual and Mixed Reality Applications. In *Intelligent Virtual Agents* (Stockholm, Sweden, 2017-08) *(Lecture Notes in Computer Science, Vol. 10498)*, Jonas Beskow, Christopher Peters, Ginevra Castellano, Carol O'Sullivan, Iolanda Leite, and Stefan Kopp (Eds.). Springer, Cham, Stockholm, Sweden, 223–226. https://doi.org/10.1007/978-3-319-67401-8_26 tex.ids: kolkmeier2017DemonstrationASAPRealizerUnity3D.

[13] Stefan Kopp, Mara Brandt, Hendrik Buschmeier, Katharina Cyra, Farina Freigang, Nicole Krämer, Franz Kummert, Christiane Opfermann, Karola Pitsch, et al. 2018. Conversational Assistants for Elderly Users–The Importance of Socially Cooperative Dialogue. In *AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications*. CEUR, Stockholm, Sweden, 10–17.

[14] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, ACM, Glasgow, Scotland, UK, 702.

[15] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-Disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[16] Ziyang Li, Pei-Luen Patrick Rau, and Dinglong Huang. 2019. Self-Disclosure to an IoT Conversational Agent: Effects of Space and User Context on Users' Willingness to Self-Disclose Personal Information. *Applied Sciences* 9, 9 (2019), 1887.

[17] Juliana Miehle, Ilker Bagci, Wolfgang Minker, and Stefan Ultes. 2019. A social companion and conversational partner for the elderly. In *Advanced Social Interaction with Agents*. Springer, Cham, 103–109.

[18] Youngme Moon. 2000. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of consumer research* 26, 4 (2000), 323–339.

[19] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, New York, NY, USA, 72–78.

[20] Matthew D Pickard, Catherine A Roster, and Yixing Chen. 2016. Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior* 65 (2016), 23–30.

[21] Abhilasha Ravichander and Alan W Black. 2018. An Empirical Study of Self-Disclosure in Spoken Dialogue Systems. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. ACL, Melbourne, Australia, 253–263.

[22] S. Zahra Razavi, Lenhart K. Schubert, Benjamin Kane, Mohammad Rafayet Ali, Kimberly A. Van Orden, and Tianyi Ma. 2019. Dialogue Design and Management for Multi-Session Casual Conversation with Older Adults. In *Joint Proceedings of the ACM IUI 2019 Workshops*. ACM, Los Angeles, CA, USA, 9.

[23] Dennis Reidsma and Herwin van Welbergen. 2013. AsapRealizer in practice–A modular and extensible architecture for a BML Realizer. *Entertainment computing* 4, 3 (2013), 157–169.

[24] Lazlo Ring, Barbara Barry, Kathleen Totzke, and Timothy Bickmore. 2013. Addressing loneliness and isolation in older adults: Proactive affective agents provide better support. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, IEEE, USA, 61–66.

[25] Lazlo Ring, Lin Shi, Kathleen Totzke, and Timothy Bickmore. 2015. Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces* 9, 1 (2015), 79–88.

[26] Daniel Schulman and Timothy Bickmore. 2009. Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th international conference on persuasive technology*. ACM, ACM, New York, NY, USA, 25.

[27] Daniel Schulman, Timothy Bickmore, and Candace Sidner. 2011. An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In *2011 AAAI Spring Symposium Series*. AAAI, Palo Alto, CA, USA, 4.

[28] Brendan Spillane, Emer Gilmartin, Christian Saam, Benjamin R Cowan, and Vincent Wade. 2018. ADELE: Care and Companionship for Independent Aging. In *ICAHGCA@AAMAS*. CEUR, Stockholm, Sweden, 18–24.

[29] Christiana Tsiourti, João Quintas, Maher Ben-Moussa, Sten Hanke, Niels Alexander Nijdam, and Dimitri Konstantas. 2016. The CaMeLi framework—a multimodal virtual companion for older adults. In *Proceedings of SAI Intelligent Systems Conference*. Springer, Springer International Publishing, Cham, 196–217.

[30] Esther van der Drift, Robbert Jan Beun, Rosemarijn Looije, Olivier Blanson Henkemans, and Mark Neerincx. 2014. A Remote Social Robot to Motivate and Support Diabetic Children in Keeping a Diary. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14)*. ACM, New York, NY, USA, 463–470. https://doi.org/10.1145/2559636.2559664

[31] M.A.M. van Kemenade, E.A. Konijn-Hoorn, and J.F. Hoorn. 2015. Robots humanize care: Moral concerns versus witnessed benefits for the elderly. In *Proceedings of the 8th International Conference on Health Informatics (HEALTHINF) Jan. 12-15, 2015. Lisbon, Portugal*, C. Verdier, M. Bienkiewicz, A. Fred, H. Gamboa, and D. Elias (Eds.), Vol. 1. SciTePress, Lisbon, Portugal, 648–653. https://doi.org/10.5220/0005287706480653

[32] Jelte van Waterschoot, Iris Hendrickx, Arif Khan, Esther Klabbers, Marcel de Korte, Helmer Strik, Catia Cucchiarini, and Mariët Theune. 2020. BLISS: An Agent for Collecting Spoken Dialogue Data about Health and Well-being. In *Proceedings of the 12th Language Resources and Evaluation Conference* (Marseille, France, 2020-05). European Language Resources Association, Marseille, France, 449–458. https://www.aclweb.org/anthology/2020.lrec-1.57

[33] Laura Pfeifer Vardoulakis, Lazlo Ring, Barbara Barry, Candace L Sidner, and Timothy Bickmore. 2012. Designing relational agents as long term social companions for older adults. In *International Conference on Intelligent Virtual Agents*. Springer, Springer, Santa Cruz, CA, USA, 289–302.