Pre-print of full 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. The final authenticated version of the paper is available online at https://doi.org/10.1007/978-3-030-68288-0_6

36 Questions to Loving a Chatbot: Are People Willing to Self-Disclose to a Chatbot?

E. A. J. Croes [1] and M. L. Antheunis [1]
[1] Tilburg University, Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition
E.A.J.Croes@tilburguniversity.edu

Abstract. The aim of the current study is to determine if people are willing to self-disclose to a chatbot to the same extent as to a human interlocutor and to examine the role of four underlying processes, namely trust, social presence, anonymity, and shame. These aims were tested among 100 participants by means of an experiment with three conditions (chatbot, a human via CMC, or a human face-to-face). In all conditions, participants were asked nine questions to stimulate self-disclosure, which varied in terms of intimacy. The results revealed that participants had the most trust in a face-to-face interaction partner and felt the most social presence face-to-face. However, they felt most anonymous in the chatbot condition. Both trust and anonymity significantly mediated the effect of condition on self-disclosure. The findings of this study have important implications for the implementation of social chatbots for psychotherapy to support people with mental health problems.

Keywords: Self-Disclosure, Human-Chatbot Communication, Social Chatbots.

1 Introduction

In order to give help to an increasing number of people that is suffering from mental health problems, chatbot applications such as Woebot and Wyse are on the rise. Chatbots are conversational programs designed to show humanlike behavior by mimicking text- or voice-based conversations [1, 2]. These so-called mental health chatbots are designed to be a sort of a virtual companion to its users and monitor the user's mood, by guiding them in disclosing their emotional state [3]. Hence, these chatbots should be able to give some support, are cost-effective, can have many interactions at the same time, are always available and have infinite patience. The increasing use of these chatbots created to improve people's emotional well-being illustrates the need in society for such chatbots. It is, therefore, important to better understand the potential of these chatbots in the mental health care.

Crucial for the potential success of mental health chatbots is the user giving personal information to the chatbot. Chatbots have several affordances that may stimulate intimate self-disclosure, such as 24/7 accessibility, anonymity, and its non-judgmental nature [4]. However, there are also reasons to believe that these chatbots may hinder self-disclosure of its users. The chatbot has several communication problems (no interaction memory, limited conversational skills) and due to a lack of Theory of Mind [5] and

emotional intelligence, the chatbot can be perceived as distant and less reliable, which hampers intimate self-disclosure [6].

There are four important processes that may explain why people self-disclose and that may determine the success of chatbot therapy, namely trust, social presence, shame, and anonymity. The first is a patient's trust in their conversation partner. Only when an individual develops a trusting bond with his/her conversation partner or therapist, will he/her feel comfortable enough to self-disclose and experience the sequential benefits of therapy. Research shows that self-disclosure is closely linked to increased closeness, liking and trust in text-based chatbot interactions [7]. Furthermore, the more personalized a chatbot is able to communicate, the more people trust the chatbot. Trust is one of the most important factors, along with empathy, in establishing a strong bond with someone [8].

Social presence, defined as the degree of salience of another person in an interaction [9] , is also found to enhance self-disclosure. Social presence is believed to be highest in communication environments that allow for the transmission of verbal and nonverbal cues. The social presence theory (SPT) posits that the inability to transmit nonverbal cues in conversation impairs impression formation. Specifically, social presence is believed to enhance involvement in an interaction, which results in more psychological closeness [10]. Although research shows that people are able to experience social presence in reduced-cues environments, it is widely accepted that especially nonverbal, visual cues enhance social presence. It therefore remains unclear whether people are able to experience social presence when conversing with a chatbot, and whether this will enhance self-disclosure.

As noted above, an important affordance of chatbot communication, especially text-based chatbots, is anonymity, which may, in turn, stimulate self-disclosure. When communicators feel less identifiable in interactions, they may become less concerned with social evaluation, which may lead to more intimate disclosures [11]. Specifically, feeling anonymous can be important when sharing sensitive issues [12]. In mediated interactions, like chatbot communication, communicators need less social skills to communicate and may feel more in control of the interaction, which enhances a sense of anonymity [13]. Feeling anonymous can make it easier to manage the information one shares about oneself and can lower the threshold to share intimate information. Chatbot communication may evoke the ultimate sense of anonymity, as people are conversing in a mediated environment to a non-human entity [14]. This may stimulate self-disclosure more so than in a computer-mediated environment.

Finally, perceived shame may also determine whether people feel comfortable enough to self-disclose. Self-disclosure is a risky process because it can entail an element of secrecy [15]. Certain information people disclose about themselves, especially in psychotherapy, could be embarrassing if shared widely. With every self-disclosure, individuals risk disconfirmation, invalidation and even ridicule. It may be that people experience less shame when interacting with a chatbot, compared to a human interlocutor. After all, the chatbot is non-judgmental which reduces people's fear to self-disclose potentially sensitive information [4]. Chatbots cannot be offended and will never get tired of listening to someone's problems, which allows people to talk freely without being judged [14]. Hence, chatbot communication may reduce perceived shame, which may enhance self-disclosure.

In sum, the aim of this study is to experimentally test if people are willing to self-disclose to the same extent as they are willing to disclose to a person either face-to-face (FTF) or via online chat. Furthermore, this study aims to investigate the validity of four potential mediators (i.e., trust, social presence, anonymity, and shame) that may account for the effects of interaction partner (text-based chatbot, human FTF, human online chat) on self-disclosure. In doing so, this study contributes to the existing computer-mediated communication (CMC), interpersonal communication, and chatbot communication literature, by determining whether the processes previously found in CMC that facilitate self-disclosure, also play a role in human-chatbot communication. Previous research has compared either CMC and FTF communication12, or CMC and human-chatbot communication4, while the present study compares all three interaction modes to get a more comprehensive picture of what facilitates the self-disclosure process. Potentially, talking to a non-human interaction partner may result in a safer and more anonymous environment, which can enhance self-disclosure. However, chatbots may also decrease a sense of social presence, which can impede self-disclose. It is therefore important to investigate which processes stimulate people to self-disclose, depending on both the interaction environment and the interaction partner.

2 Theoretical Background

2.1 Social Chatbots

While functional, customer service chatbots work very well in a specific domain, social chatbots are designed to keep people company and build an emotional connection with its users [16]. In doing so, it is crucial that the chatbot is able to detect emotions and respond in an adequate way [17]. Social chatbots attempt to connect with their users by asking questions, gathering information and keeping the conversation going so that people want to keep interacting with the chatbot [16]. A profound advantage of social chatbots over more functional chatbots is that they are able to recognize emotions in social interactions and are able to respond in an empathic way [18].

Social chatbots can thus be employed in therapeutic settings and there is an increasing scientific interest in whether chatbots are capable of offering good quality support [19, 20, for instance]. Research shows that people evaluate chatbot therapy as less valuable and enjoyable compared to regular human-human therapy [19]. In contrast, other research shows that social chatbots are able to offer decent and effective support to individuals in behavioral therapy [20]. Furthermore, research shows that a social chatbot is an accessible and effective tool to support people in psychotherapy [21]. Moreover, research also shows that social chatbots may be effective to help people with depression and to ensure people with psychological problems to not deteriorate [22]

2.2 Self-Disclosure in Chatbot Interactions

Self-disclosure, one of the most important conditions of effective therapy, is defined as the act of revealing personal information to others [23]. The act of disclosing personal information involves risk and vulnerability on the part of the discloser, which increases

the likelihood of mutual bonding. As people disclose more intimate information, they develop stronger relationships [24]. When one person discloses personal information to another, it is likely that this disclosure is reciprocated which means mutual trust and understanding is enhanced. Furthermore, self-disclosure can create a feeling of relief in the discloser [25]. Technological developments, like chatbots, can lower the threshold for people to engage in self-disclosure, which is found more frequently in computer-mediated interactions [26].

Studies reveal that text-based CMC stimulates intimate self-disclosure [26, 27. for instance]. Additionally, research shows that people share more sensitive information in a depersonalized questionnaire [28] and that people tend to shy away from sharing negative emotions about themselves when their interaction partner is visually visible to them [29]. Thus, people are expected to share more intimate information when interacting with a chatbot, compared to when they are conversing with another human. Research supports this contention, as it reveals that people prefer to disclose sensitive or personal information with a chatbot compared to a human interviewer [30]. This is especially the case when the chatbot is involved and shows empathy, as an empathic response ensures someone feels understood, which can enhance the relief they experience [31]. An important advantage of chatbot communication over human communication, is that chatbots are unable to share someone's secrets with other people, which can enhance trust [32]. Chatbot communication can be seen as the ultimate form of anonymity; people are not only visually anonymous, they are unable to see their conversation partner, which can stimulate self-disclosure [28]. Previous research shows that CMC stimulates self-disclosure [26], but for the reasons outlined above, human-chatbot interactions may result in *more* self-disclosure. Thus, we formulate the following hypothesis:

H1: People self-disclose more to a chatbot, compared to a human in CMC and a human FTF.

2.3 Trust in a Social Chatbot

There are a number of social processes that may facilitate self-disclosure and help develop a bond between a person and the therapist. An interpersonal bond is formed through mutual understanding and acceptance and includes elements of honesty, safety and trust [33]. As relationships become deeper and more intimate, communicators become more involved and interpersonal trust develops [34]. Intimacy is strongly related to trust and source credibility. Qualitatively better interactions with a strong sense of comfort are those that promote higher levels of credibility and trust, which enhances the likeability of the interaction partner [35]. Furthermore, greater trust is linked to more self-disclosure [36].

The question is, if individuals are able to trust a chatbot as much as, if not more than, a human interlocutor. Research shows that chatbots with consistent personalities are seen as more predictable and, thus, more trustworthy [7]. In contrast, chatbots with unpredictable attitudes can create a strong sense of discomfort in its users. In addition, chatbots that show emotion are generally perceived as more likeable and trustworthy, compared to non-emotional chatbots [8]. Furthermore, empathic chatbots are also seen as more trustworthy and supportive than non-empathic chatbots [38]. Moreover, importantly, since chatbots are machines and not people, it may be easier for individuals

to trust that the information they share with the chatbot will not 'leak' into the real world [14]. Therefore, we expect the following:

H2: People have more trust in a chatbot as an interaction partner, compared to a human in CMC and a human FTF, which, in turn, results in more self-disclosure with the chatbot.

2.4 Social Presence in Chatbot Interactions

Social presence is closely related to interpersonal trust and is defined as "the feeling of being with another in a mediated environment" [38, p. 14]. The more cues, and especially visual cues, a communication environment offers, the more social presence interactants experience [9]. This suggests that text-based chatbot interactions would evoke less social presence than communicating with another human in CMC and FTF communication (the richest form of communication). In fact, research has shown that when people talk to someone whom they believe is another person they experience more social presence, compared to when they believe their interaction partner is a robot [39].

Social presence is believed to have many positive psychological effects and leads to more involved communicators and more intimacy [40]. The more social presence people experience, the closer they feel [41].Furthermore, social presence is believed to make messages exchanged between people more intimate and emotional. This suggests that social presence enhances intimate self-disclosure. Thus, we expect the following:

H3: People feel less social presence in human-chatbot interactions, compared to human-human interactions via CMC and FTF interactions, which, in turn, results in less self-disclosure in human-chatbot interactions.

2.5 Perceived Anonymity in Chatbot Interactions

Research shows that CMC channels provide controllability and anonymity, which allows individuals to express themselves more freely and honestly compared to FTF communication [42]. Chatbot communication may be even more anonymous compared to CMC as a chatbot is an artificial interaction partner, will not share any information with other people and will keep your secrets [14]. Self-disclosures will thus never be revealed to the outside world, which can make people feel safe to share intimate information. As a chatbot does not have feelings, people may find it easier to open up. Furthermore, this anonymity can lower the threshold for people to share intimate information, which is especially relevant in a mental health setting. Confidentiality and privacy are important aspects of human-chatbot interaction as these aspects foster trust.

Thus, anonymity may enhance self-disclosure, especially in interactions with a chatbot, which is why we expect the following:

H4: People feel more anonymous in human-chatbot interactions, compared to human-human interactions via CMC and FTF interactions, which, in turn, results in more self-disclosure in human-chatbot interactions.

2.6 Perceived Shame in Chatbot Interactions

Finally, when people feel ashamed to self-disclose they may refrain from doing so. As said, chatbot communication may lower the threshold to self-disclose, which may be due to the fact that people do not experience much shame. A chatbot is a non-judgmental listener, which means people experience less fear of being judged and, hence, less shame when they self-disclose [4]. Furthermore, chatbots cannot be offended, will never get tired of listening to someone's problems, foster a safe environment to vent and will not respond to someone in a negative way. As said, self-disclosure is a risky process, which may involve information people are embarrassed to share. When talking to a chatbot, people may feel safer to disclose and less ashamed, which is why we pose the following hypothesis:

H5: People experience less shame in human-chatbot interactions, compared to human-human interactions via CMC and FTF interactions, which, in turn, results in more self-disclosure in human-chatbot interactions.

3 Method

3.1 Participants

In total, 150 participants participated in this experiment of which 66 males (44%) and 84 females (56%). Participants were, on average, 21.97 years old (*SD* = 3.15). The majority of the participants indicated that their highest level of education was university (76.7%), followed by a university of applied science degree (12%) and a high school degree (7%). Regarding chatbot experience, most of the participants indicated that they communicated with a chatbot several times a year (63.3%), followed by 23.3% who indicated that they have never communicated with a chatbot before. 12% of participants interacted 1-2 times per month with a chatbot and only 1 participant communicated 2-3 times a week with a chatbot. The majority of chatbot interactions were for the purpose of customer service (72.7%), followed by online shopping (19.3%) and fun or entertainment (8.7%). We deliberately chose for a (largely) student sample as students are technology savvy and, therefore, confident and competent in the use of digital devices.

3.2 Design

The experiment consisted of three conditions with a between-subjects design. Participants were randomly assigned to one of the experimental conditions. The condition were: a human-chatbot condition, a CMC condition and a FTF condition. In the human-chatbot condition, participants chatted with a chatbot via text. In the CMC condition, participants chatted with another human being via a text-based chat. In the FTF condition, participants sat in a room at a table across from a female confederate with whom they had a conversation.

3.3 Procedure and Materials

After signing informed consent, all participants first filled out a pretest questionnaire with demographic questions along with questions about the participants' personality and their experience with chatbots. Next, participants were randomly asked to have a conversation with either a chatbot, a human via CMC or a human FTF. In the chatbot condition (1) participants chatted with a chatbot via Facebook. The chatbot that was used in this study was built in Flow.ai, which is a platform to design chatbots. In order to allow unlimited word entry, the chatbot was connected to Facebook, using an application programming interface (API). Two Facebook accounts were created. One to connect the chatbot, and one for the participants to communicate with the chatbot. The screen name of the participants' Facebook was 'Participant'. The screenname of the chatbot was Chatbot TU and the name the chatbot introduced itself with was 'Robin', as this is a gender-neutral name. The chatbot did not have a profile picture; the profile picture was a standard Facebook icon, shown when users do not upload a profile picture. Participants talked to the chatbot via Facebook Messenger. They were instructed beforehand that they would be talking to a chatbot. The interaction took place on a computer in the lab, which was already signed into the account. Participants did not receive any credentials related to the account, nor did they have to sign in to Facebook themselves. In the CMC condition (2), participants communicated via the chat function in Skype. To do so, an anonymous Skype-account was created. The name of the account for the participant was 'Participant' and the person they were chatting to was also called 'Robin', a confederate in the experiment. The procedure of this condition was similar to the chatbot condition; the only difference was that the participants were aware that they were communicating with a real human being instead of a chatbot. In the FTF condition (3) participants sat in a room at a table across from a female confederate who greeted them and started the conversation by asking the first question.

All three conditions were question and answer sessions, in which participants were asked nine questions by the interaction partner, in order to provoke self-disclosure. Participants were instructed to only answer the question, without asking any questions back. They were also told that their interaction partner would not respond to their answers. The questions were derived from previous research [43] which used 36 questions as a means to generate closeness (*36 questions to love*). The 36 questions were divided into three sets of twelve questions. For the current experiment, three questions from each set were used. In all three conditions, the same nine questions were asked. After participants had answered all nine questions, the conversation ended and they were led to a room to fill out a posttest questionnaire with measurements for the variables of interest for this study.

3.4 Self-Report Measures

All self-report measures were measured on a 5-point Likert Scale (1 = totally disagree, 5 = totally agree). First, self-disclosure was measured with the following statements [44] (1) 'During the conversation I was able to share personal information about myself', (2) 'During the conversation I felt comfortable sharing personal information' (3)

'During the conversation it was easy to share personal information', (4) 'During the conversation I felt that I could be open' (M = 3.50; SD = .96; α = .91).

Trust was measured by means of four statements [45]: (1) 'My conversation partner was honest', (2) 'My conversation partner was trustworthy', (3) 'My conversation partner was understanding', and (4) 'My conversation partner had good intentions' (M = 3.16; SD = 0.62; α = .70).

Social presence was measured with seven statements [46, 47]: (1) 'During the conversation I was able to respond to the reactions of my conversation partner', (2) 'During the conversation I felt that I was face to face with my conversation partner', (3) 'During the conversation I felt as if I was in the same room as my conversation partner', (4) 'During the conversation my conversation partner came across as "real"', (5) 'During the conversation I felt that I could really get to know my conversation partner', (6) 'During the interview I felt that I was having a conversation with a social being', (7) 'During the conversation I felt that I was having a conversation with an intelligent being' (M = 2.60; SD = 0.93; α = .88.)

The anonymity scale consisted of seven items [48, 49, 50] : (1) 'During the conversation I felt like my conversation partner did not know me', (2) 'During the conversation I felt like my conversation partner recognized me' (*reverse-coded*), (3) 'During the conversation I felt like my personal identity was not visible to my conversation partner', (4) 'During the conversation I felt anonymous', (5) 'During the conversation I felt unrecognizable', (6) 'During the conversation I did not feel identifiable', and (7) 'During the conversation I felt like I could share more about myself because my conversation partner did not know me' (M = 2.99, SD = 0.77; α = .79).

Perceived shame was measured by means of four statements [51]: (1) 'I experienced shame during the conversation', (2) 'During the conversation I worried about what my conversation partner thought of me', (3) 'During the conversation I concealed who I truly am', and (4) 'During the conversation I sometimes felt ashamed about what I shared with my conversation partner' (M = 2.32, SD = 0.94; α = .66).

4 Results

To investigate the effect of condition on self-disclosure and whether this can be explained by trust, anonymity, social presence or shame, a mediation analyses was performed in SPSS using the procedures developed by Preacher and Hayes (PROCESS) [52]. In the analysis, 'condition' was entered as a predictor for self-disclosure and trust, anonymity, social presence and shame were entered as mediators. For the analysis, the categorical 'condition' variable was recoded into two dummy variables, namely for the 'chatbot condition' (1 = chatbot, 0 = CMC and FTF) and the 'CMC condition' (1 = CMC, 0 = chatbot and FTF). The FTF condition served as the reference group.

The first hypothesis proposed that people disclose more information to a chatbot, compared to a human in CMC and a human FTF. The analysis revealed that the effect of both the chatbot condition (b = -.01, SE = 0.22, p = .950) and the CMC condition (b = .19, SE = 0.20, p = .349) on self-disclosure was not significant. Therefore, H1 was not supported.

The means and standard deviations are displayed in Table 1.

Table 1. Means and Standard Deviations

	FTF	CMC	Chatbot
Dependent variable	*M (SD)*	*M (SD)*	*M (SD)*
Self-Disclosure	3.64 (0.75)	3.50 (1.03)	3.36 (1.07)
Trust	3.58 (0.56)	3.00 (0.52)	2.90 (0.54)
Social Presence	3.49 (0.64)	2.32 (0.66)	2.01 (0.73)
Anonymity	2.67 (0.61)	3.03 (0.73)	3.27 (0.86)
Shame	2.31 (0.88)	2.44 (1.07)	2.21 (0.85)

H2 proposed that trust would be a mediator in the relationship between condition and self-disclosure. Although we did not find a significant direct effect of condition on self-disclosure, we could still find significant mediating effects. First, we found a direct effect of both the chatbot condition (b = -.68, SE = 0.11, p < .001) and the CMC condition (b = -.59, SE = 0.11, p < .001) on perceived trust. As the means in Table 1 show, people had the most trust in a FTF interaction partner, followed by a human in CMC, and finally the chatbot. Furthermore, trust was found to significantly impact self-disclosure, b = .53, SE = 0.14, p < .001. As expected, trust explained a significant portion of the effect of condition on self-disclosure. More specifically, the indirect effect for the chatbot condition was b = -0.36, SE = 0.13, 95% BCa CI [-0.62, -0.13] and the indirect effect for the CMC condition was b = -0.31, SE = 0.12, 95% BCa CI [-0.56, -0.11]. However, contrary to our expectations, both indirect effects were negative. This suggests that trust positively impacted self-disclosure in the FTF condition, but had a negative impact in the chatbot and CMC condition. Therefore, H2 could not be supported.

H3 posed that people feel less social presence in human-chatbot interactions, compared to human-human interactions via CMC and FTF interactions, which, in turn, results in less self-disclosure in human-chatbot interactions. The analysis revealed a direct effect of the chatbot condition (b = -1.47, SE = 0.14, p < .001) and the CMC condition (b = -1.17, SE = 0.14, p < .001) on social presence. As expected, people experienced the most social presence in the FTF condition, followed by the CMC condition and, finally, the chatbot condition (see Table 1 for the means). However, social presence did not significantly impact self-disclosure, b = 0.12, SE = 0.12, p = .312. In addition, the indirect effect of condition on self-disclosure via social presence was not significant either (b = -.18, SE = 0.21, 95% BCa CI [-0.61, 0.21] for the chatbot condition; b = -.14, SE = 0.16, 95% BCa CI [-0.48, 0.17] for the CMC condition). So H3 was only partially supported.

H4 predicted that people feel more anonymous in human-chatbot interactions, compared to human-human interactions in CMC and FTF interactions, which, in turn, results in more self-disclosure in human-chatbot interactions. We found a direct effect of the chatbot condition (b = .60, SE = 0.15, p < .001) and the CMC condition (b = .36, SE = 0.15, p = .015) on perceived anonymity (see Table 1 for the means). People felt most anonymous in the chatbot condition, followed by the CMC condition and, finally, the FTF condition. Additionally, anonymity had an effect on self-disclosure, b = .42, SE = 0.10, p < .001. Furthermore, the indirect effect of condition on self-disclosure via anonymity was also significant. The indirect effect for the chatbot condition was b =

.25, *SE* = 0.09, 95% BCa CI [0.10, 0.46] and for the CMC condition *b* = 0.15, *SE* = 0.07, 95% BCa CI [0.04, 0.31]. This suggests that in both the chatbot and the CMC condition anonymity explained the effect on self-disclosure, and this effect was stronger in the chatbot condition. Thus, H4 was supported.

The final hypothesis proposed that people experience less shame in human-chatbot interactions, compared to human-human interactions via CMC and FTF interactions, which, in turn, results in more self-disclosure in human-chatbot interactions. The analysis showed that the direct effect of the chatbot condition (*b* = -.10, *SE* = 0.19, *p* = .595) and the CMC condition (*b* = .13, *SE* = 0.19, *p* = .478) on perceived shame was not significant. Furthermore, both indirect effects of the chatbot condition (*b* = .02, *SE* = 0.04, 95% BCa CI [-0.06, 0.12]) and the CMC condition (*b* = -.03, *SE* = 0.04, 95% BCa CI [-0.14, 0.06]) on self-disclosure via perceived shame were not significant either. So H5 could not be supported.

5 Discussion

The aim of the current study was to determine (1) if people are willing to self-disclose to a chatbot to the same extent as to a human interlocutor and (2) the role of four underlying processes that may explain this self-disclosure, namely trust, social presence, anonymity, and shame.

Our findings revealed no differences between the three conditions concerning self-disclosure. This suggests that people are equally willing to disclose to a human interlocutor and a chatbot. Additionally, we found that trust and anonymity both impacted self-disclosure: when people trusted their interaction partner more and when they felt more anonymous, they self-disclosed more. However, trust was found to be highest in the FTF condition, which is contrary to what we expected. We believed that people would trust a chatbot more than a human interaction partner, as a chatbot will never leak the information you share to other people [14]. Previous research, however, also shows the importance of nonverbal cues in establishing trust [52]. It may thus be that nonverbal cues play a more important role in trust, than the artificiality of the interaction partner.

Furthermore, in line with our expectations, we found that people self-disclosed in the chatbot condition because they felt anonymous. As predicted, chatbot communication creates a sense of ultimate anonymity, more so than communicating with another human using the same modalities. Additionally, based on previous research we know that anonymity is valued because it avoids embarrassment [12]. In the literature, anonymity is central to explain why people self-disclose more in reduced-cues environments and our findings show that chatbot communication evokes more anonymity than CMC communication.

We also found that FTF communication leads to the highest feeling of social presence, which is also what we expected. Although social presence was not found to impact self-disclosure in this study, we did find that visible, co-present interaction conditions evoke the strongest sense of social presence, which is in line with SPT [9] and previous research [53, for instance]. People experience less social presence in chatbot interactions, compared to CMC interactions, which is also in line with previous research which showed that an artificial interaction partner leads to lower expectation of social

presence, compared to a human interlocutor [39].Our findings add to this research and show the importance of perceived humanness in establishing social presence.

We did not find a difference between the conditions regarding shame, which is contrary to what we expected. It seems the chatbot does not create a communication environment in which people feel safer and less embarrassed to self-disclose. It may be that participants felt that the information they shared in all three conditions would be treated confidentially; which is generally the case in scientific research. Hence, they may have felt equally safe to disclose in all three conditions.

5.1 Theoretical and Practical Implications

Our findings have important implications for research and theory on social chatbots and self-disclosure. First, our findings show that people are willing to self-disclose to a chatbot, just as much as to a human interaction partner. This is an important implication, as it shows potential for social chatbots to support people in psychotherapy. One of the most important conditions of successful therapy is self-disclosure and as our study shows that people disclose personal information to a chatbot, this may suggest that chatbots could potentially play a (supporting) role in psychotherapy.

Second, our findings show the importance of anonymity as an underlying explanation as to why people self-disclose in chatbot interactions. First, it shows the importance of anonymity when disclosing personal information, and second it shows that people feel highly anonymous in chatbot interactions. This anonymity, in turn, has positive effects: it lowers people's perceived risks and ensures that they feel safe to self-disclose. Although previous research has highlighted the negative results of anonymity in human-chatbot interactions, such as an increase in profanity [14], our study adds to this by showing that this sense of anonymity can have positive effects as it allows for a safe environment for intimate self-disclosure.

Finally, our findings have practical implications for chatbot developers. Based on our study, we can conclude that FTF communication is still the golden standard regarding social presence and trust. As trust is an important aspect in psychotherapy, and enhances self-disclosure, it is important for developers to create chatbots that come across as trustworthy interaction partners. In the present research, the chatbot was designed to ask questions, which may have impacted our findings. In fact, research shows that reciprocal self-disclosure can increase trust and liking [54]. Therefore, creating a chatbot with visual aspects, to enhance social presence, which is capable of reciprocating the user's self-disclosure, may enhance trust and, in turn, self-disclosure.

5.2 Limitations and Suggestions for Future Research

First, since this study used a chatbot incorporated into Facebook Messenger, this may have affected our findings. Specifically, data collected on Facebook Messenger are subject to Facebook's own privacy policy and this data may be shared with third parties. Although participants conversed using an anonymous account created for the experiment, they may have been cautious with the information they shared. Facebook outlines that it uses data for the improvement of services, especially advertising services, so the

risk to participants was low. Furthermore, the present study opted for a chatbot integrated into Facebook Messenger because of low cost and the fact that people were likely already familiar with the technology. However, future research could attempt to build an independent mobile chatbot application, which means more control over the information collected and less potential vulnerability for the potential release of confidential information.

A second limitation of the present study lies in its design. In all three conditions participants partook in a question and answer session, where the interaction partner did not respond to their disclosures or disclosed anything themselves. As reciprocal self-disclosure may evoke trust and, as a result, self-disclosure it may be interesting for future studies to examine the impact of self-disclosure on the chatbot's end. Is a chatbot capable of self-disclosing in a way that makes the user trust them more? Furthermore, creating a chatbot with an avatar and/or other visual modalities may enhance social presence, which is also something future research could examine. Finally, it may be interesting to further analyse the effects of self-disclosure on people's overall well-being. Although we know, based on the findings in this study, that people self-disclose equally often to a human interlocutor and a chatbot, we do not know what the impact is of this self-disclosure. Do people experience relief after disclosing personal information to a chatbot? Does self-disclosure to an artificial interaction partner improve wellbeing? These are questions that future studies may attempt to answer.

6 Conclusion

The findings in the present study show potential for social chatbots to support psychotherapy and stand by people with mental health problems. We find that people self-disclose equally often to a chatbot, compared to a human interaction partner and that anonymity plays an important role in why people self-disclose to a chatbot. Furthermore, people do not experience trust and social presence in chatbot interactions, which is where improvements can be made. Finally, more research is needed to determine the potential beneficial effects of self-disclosure in human-chatbot interactions.

References

1. Abdul-Kader, S. A., & Woods, J.: Survey on chatbot design techniques in speech conversation systems. International Journal of Advanced Computer Science & Applications **6**, 72–80 (2015)
2. Vassallo, G., Pilato, G., Augello, A., & Gaglio, S.: Phase coherence in conceptual spaces for conversational agents (pp. 357-371). Hoboken, NJ: John Wiley & Sons (2010)
3. D'Alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., Gleeson, J., & Alvarez-Jimenez, M.: Artificial intelligence-assisted online social therapy for youth mental health. Frontiers in Psychology, **8,** 796 (2017)
4. Mou, Y., & Xu, K.: The media inequality: Comparing the initial human-human and human-AI social interactions. Computers in Human Behavior, **72**, 432-440 (2017)

5. Heyselaar, E. & Bosse, T.: Using Theory of Mind to Assess Users' Sense of Agency in Social Chatbots. In International Workshop on Chatbot Research and Design (pp.158-169). Springer, Cha (2019)
6. Perlman, D. & Fehr, B.: The development of intimate relationships. In D. Perlman & S. Duck (Eds.), Intimate relationships: Development, dynamics, and deterioration (p. 13–42). Sage Publications, Inc. (1987)
7. Chaves, A. P., & Gerosa, M. A.: Single or Multiple Conversational Agents? An Interactional Coherence Comparison. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-13) (2018)
8. Creed, C., Beale, R., & Cowan, B.: The impact of an embodied agent's emotional expressions over multiple interactions. Interacting with Computers, **27**(2), 172-188 (2015)
9. Short, J., Williams, E., & Christie, B.: The social psychology of telecommunications. London: Wiley (1976)
10. Biocca, F., Harms, C., & Gregg, J.: The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In 4th annual international workshop on presence, Philadelphia, PA (pp. 1-9) (2001)
11. Lea, M., Spears, R., & de Groot, D.: Knowing me, knowing you: Anonymity effects on social identity processes within groups. Personality and Social Psychology Bulletin, **27**(5), 526-537 (2001)
12. Walther, J. B., & Boyd, S.: Attraction to computer-mediated social support. Communication Technology and Society: Audience Adoption and Uses, **153188**, 50-88 (2002)
13. Philippot, P., & Douilliez, C.: Impact of social anxiety on the processing of emotional information in video-mediated interaction. In A. Kappas & N. C. Kramer (Eds.), Face-to-face communication over the internet: emotions in a web of culture, language and technology (pp. 127-143). Cambridge: Cambridge University Press (2011)
14. Skjuve, M., & Brandtzæg, P. B.: Chatbots as a new user interface for providing health information to young people. Youth and news in a digital media environment–Nordic-Baltic perspectives (2018)
15. Jourard, S. M.: The transparent self. Van Nostrand Reinhold Company (1971)
16. Shum, H. Y., He, X. D., & Li, D.: From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers of Information Technology & Electronic Engineering, **19**(1), 10-26 (2018)
17. Augello, A., Gentile, M., & Dignum, F.: An overview of open-source chatbots social skills. In International Conference on Internet Science (pp. 236-248). Springer, Cham (2017)
18. Ho, A., Hancock, J., & Miner, A. S.: Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. Journal of Communication, **68**(4), 712-733 (2018)
19. Bell, S., Wood, C., & Sarkar, A.: Perceptions of chatbots in therapy. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-6) (2019)
20. Fitzpatrick, K. K., Darcy, A., & Vierhile, M.: Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Mental Health, **4**(2), e19 (2017)
21. Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M.: Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Mental Health, **5**(4), e64 (2018)
22. Martínez-Miranda, J.: Embodied conversational agents for the detection and prevention of suicidal behaviour: current applications and open challenges. Journal of Medical Systems, **41**(9), 135 (2017)

23. Archer, R. L., & Burleson, J. A.: The effects of timing of self-disclosure on attraction and reciprocity. Journal of Personality and Social Psychology, **38**(1), 120 (1980)
24. Valkenburg, P. M., & Peter, J.: Social consequences of the Internet for adolescents: A decade of research. Current Directions in Psychological Science, **18**(1), 1-5 (2009)
25. Choi, Y. H., & Bazarova, N. N.: Self-disclosure characteristics and motivations in social media: Extending the functional model to multiple social network sites. Human Communication Research, **41**(4), 480-500 (2015)
26. Joinson, A. N.: Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. European Journal of Social Psychology, **31**(2), 177-192 (2001)
27. Antheunis, M. L., Schouten, A. P., Valkenburg, P. M., & Peter, J.: Interactive uncertainty reduction strategies and verbal affection in computer-mediated communication. Communication Research, **39**(6), 757-780 (2012)
28. Tourangeau, R., Couper, M. P., & Steiger, D. M.: Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. Computers in Human Behavior, **19**(1), 1-24 (2003)
29. Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K.: When the interface is a face. Human-Computer Interaction, **11**(2), 97-124 (1996)
30. Bhakta, R., Savin-Baden, M., & Tombs, G.: Sharing secrets with robots?. In EdMedia+ Innovate Learning (pp. 2295-2301). Association for the Advancement of Computing in Education (AACE) (2014)
31. Farber, B. A., Berano, K. C., & Capobianco, J. A.: Clients' Perceptions of the Process and Consequences of Self-Disclosure in Psychotherapy. Journal of Counseling Psychology, **51**(3), 340 (2004)
32. Joinson, A. N., & Paine, C. B.: Self-disclosure, privacy and the Internet. The Oxford handbook of Internet Psychology, 2374252 (2007)
33. Tillmann-Healy, L. M.: Friendship as method. Qualitative Inquiry, 9(5), 729-749.
34. Burgoon, J. K., & Hale, J. L. (1984). The fundamental topoi of relational communication. Communication Monographs, **51**(3), 193-214 (2003)
35. Houser, M. L., Horan, S. M., & Furler, L. A.: Dating in the fast lane: How communication predicts speed-dating success. Journal of Social and Personal Relationships, **25**(5), 749-768 (2008)
36. Gibbs, J. L., Ellison, N. B., & Lai, C. H.: First comes love, then comes Google: An investigation of uncertainty reduction strategies and self-disclosure in online dating. Communication Research, **38**(1), 70-100 (2011)
37. Brave, S., Nass, C., & Hutchinson, K.: Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. International Journal of Human-Computer Studies, **62**(2), 161-178 (2005)
38. Biocca, F., Harms, C., & Burgoon, J. K.: Toward a more robust theory and measure of social presence: Review and suggested criteria. Presence: Teleoperators & Virtual Environments, **12**(5), 456-480 (2003)
39. Lachlan, K. A., Spence, P. R., Edwards, A., Reno, K. M., & Edwards, C.: If you are quick enough, I will think about it: Information speed and trust in public health organizations. Computers in Human Behavior, **33**, 377-380 (2014)
40. Walther, J. B.: Interpersonal effects in computer-mediated interaction: A relational perspective. Communication Research, **19**(1), 52-90 (1992)
41. Walther, J. B., Loh, T., & Granka, L.: Let me count the ways: The interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. Journal of Language and Social Psychology, **24**(1), 36-65 (2005)

42. Valkenburg, P. M., & Peter, J.: Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. Journal of Adolescent Health, **48**(2), 121-127 (2011)
43. Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J.: The experimental generation of interpersonal closeness: A procedure and some preliminary findings. Personality and Social Psychology Bulletin, **23**(4), 363-377 (1997)
44. Ledbetter, A. M.: Measuring online communication attitude: Instrument development and validation. Communication Monographs, **76**(4), 463-486 (2009)
45. Chiou, J. S., & Droge, C.: Service quality, trust, specific asset investment, and expertise: Direct and indirect effects in a satisfaction-loyalty framework. Journal of the Academy of Marketing Science, **34**(4), 613 (2006)
46. Nowak, K. L., & Biocca, F.: The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence: Teleoperators & Virtual Environments, **12**(5), 481-494 (2003)
47. Lee, K. M., Peng, W., Jin, S. A., & Yan, C.: Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. Journal of Communication, **56**(4), 754-772 (2006)
48. Rains, S. A.: The impact of anonymity on perceptions of source credibility and influence in computer-mediated group communication: A test of two competing hypotheses. Communication Research, **34**, 100-125. (2007)
49. Qian, H., & Scott, C. R.: Anonymity and self-disclosure on weblogs. Journal of Computer-Mediated Communication, **12**(4), 1428-1451 (2007)
50. Hite, D. M., Voelker, T., & Robertson, A.: Measuring perceived anonymity: The development of a context independent instrument. Journal of Methods and Measurement in the Social Sciences, **5**(1), 22-39 (2014)
51. Andrews, B., Qian, M., & Valentine, J. D.: Predicting depressive symptoms with a new measure of shame: The Experience of Shame Scale. British Journal of Clinical Psychology, **41**(1), 29-42 (2002)
52. Bohannon, L. S., Herbert, A. M., Pelz, J. B., & Rantanen, E. M.: Eye contact and video-mediated communication: A review. Displays, **34**(2), 177-185 (2013)
53. Werkhoven, P. J., Schraagen, J. M., & Punte, P. A.: Seeing is believing: communication performance under isotropic teleconferencing conditions. Displays, 22(4), 137-149 (2001)
54. Bevacqua, E., Richard, R., & De Loor, P.: Believability and Co-presence in Human-Virtual Character Interaction. IEEE Computer Graphics and Applications, **37**(4), 17-29 (2017)