Journal of Communication ISSN 0021-9916

ORIGINAL ARTICLE

Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot

Annabell Ho[1], Jeff Hancock[1], & Adam S. Miner[1,2,3]

1 Department of Communication, Stanford University, Stanford, CA 94305, USA
2 Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305, USA
3 Clinical Excellence Research Center, Stanford University, Stanford, CA 94305, USA

Disclosing personal information to another person has beneficial emotional, relational, and psychological outcomes. When disclosers believe they are interacting with a computer instead of another person, such as a chatbot that can simulate human-to-human conversation, outcomes may be undermined, enhanced, or equivalent. Our experiment examined downstream effects after emotional versus factual disclosures in conversations with a supposed chatbot or person. The effects of emotional disclosure were equivalent whether participants thought they were disclosing to a chatbot or to a person. This study advances current understanding of disclosure and whether its impact is altered by technology, providing support for media equivalency as a primary mechanism for the consequences of disclosing to a chatbot.

Keywords: Self-Disclosure, Communication and Technology, Chatbot, Conversational Agent, Computers as Social Actors, Well-Being, Conversational AI, Digital Assistant, Human-machine communication.

doi:10.1093/joc/jqy026

Self-disclosure, or revealing personal information to someone else (Archer, 1980), generates a wide variety of beneficial outcomes. The beneficial nature of self-disclosure is amplified when the listener responds with support and validation, rather than ignoring or blaming the discloser (Shenk & Fruzzetti, 2011). Previous research from communication and psychology has predominantly examined the effects of intimate disclosure between two people, but recent technological advances have changed the scope and possibilities of disclosure beyond human partners for intimate topics ranging from end-of-life planning (Utami, Bickmore, Nikolopoulou, & Paasche-Orlow, 2017) to mental health treatment (Miner, Milstein, & Hancock, 2017). As chatbots, or computer programs that can simulate human-human conversation, begin to take part in intimate

Corresponding author: Annabell Ho; e-mail: asuh@stanford.edu

conversations, an important question is raised: what are the downstream psychological effects of disclosure when the partner is a computer and not another person?

Past work reveals three kinds of benefits that can accrue from intimate disclosures to a supportive human conversational partner. First, disclosure can impact the immediate emotional experience of a discloser by reducing stress arising from negative experiences (Martins et al., 2013), diminishing anxiety (Tam et al., 2006), and increasing negative affect in the short term (Greenberg & Stone, 1992), which ultimately results in long-term psychological improvement (Kelley, Lumley, & Leisen, 1997). Second, disclosure met with support can improve relational outcomes, enhancing relational closeness and intimacy (Altman & Taylor, 1973; Sprecher, Treger, & Wondra, 2013). Third, disclosure can also improve psychological outcomes deeply rooted in individuals' self-image, such as experiencing greater self-affirmation and a restored sense of worth after intimate disclosure (Creswell et al., 2007).

The identity of a conversation partner, as a human or computer, matters. Previous work has found that the mere perceived identity of the partner as computer or human has profound effects, even when actual identity does not (Fox et al., 2015; Lucas, Gratch, King, & Morency, 2014). Perceived identity is critical to understand, especially from a theoretical perspective, because it gives rise to new processes, expectations of the partner, and effects that do not arise when the partner is always assumed to be human, as in previous work. This could alter disclosure processes and outcomes in fundamental ways. For example, people often avoid disclosing to others out of a fear of negative evaluation. Because chatbots do not think or form judgments on their own, people may feel more comfortable disclosing to a chatbot compared to a person, changing the nature of disclosure and its outcomes (Lucas et al., 2014). On the other hand, people assume that chatbots are worse at emotional tasks than humans (Madhavan, Wiegmann, & Lacson, 2006), which may negatively impact emotional disclosure with chatbots. Despite the importance of a partner's perceived identity, it is unclear whether similar or different outcomes will occur when people disclose to a perceived chatbot instead of another person.

As the conversational abilities of chatbots quickly improve (Zhang et al., 2018) and public interest grows (Markoff & Mozur, 2015; Romeo, 2016), it is critical to understand the emotional, relational, and psychological outcomes of disclosing to a chatbot. Extant research provides three theoretical frameworks that suggest different potential outcomes. First, a theoretical emphasis on perceived understanding suggests that disclosure will only have a beneficial impact when the partner is believed to have sufficient emotional capacity to truly understand the discloser, which chatbots inherently cannot. We refer to this as the perceived understanding framework. Second, research on conversational agents and disclosure intimacy, in contrast, suggests that disclosure will be even more beneficial with a chatbot than a human partner, because chatbots encourage more intimate disclosure. We refer to this as the disclosure processing framework. Third, a media equivalency approach suggests that the effects of disclosure operate in the same way for human and chatbot partners. We refer to this as the computers as social actors (CASA) framework.

Perceived understanding framework

According to the theoretical model of perceived understanding (Reis, Lemay, & Finkenauer, 2017), feeling truly understood, or that the partner "'gets' [disclosers] in some fundamental way," brings emotional, relational, and psychological benefits. This occurs because feeling understood creates a sense of social belonging and acceptance, activates areas in the brain associated with connection and reward, and enhances personal goal pursuit (Reis et al., 2017, p. 1).

According to this model, the positive effects of feeling understood are mediated by the extent to which disclosers perceive that they are understood. In other words, disclosers need to believe the partner understands them before the positive impact of feeling understood can take place. Feeling truly understood in this way goes beyond mere recognition of the discloser's utterances, but arises when disclosers feel partners understand core aspects of who they are as a person and how they experience the world (Reis et al., 2017; Reis & Shaver, 1988).

In the case of a chatbot, disclosers know that a chatbot is a computer program that cannot understand them on this deeper level. The chatbot's responses may be seen as pre-programmed and inauthentic, preventing disclosers from feeling truly understood. They may not, then, experience the positive emotional, relational, and psychological effects of feeling understood. On the other hand, individuals know that another person has the capability of truly understanding them in a way a chatbot cannot, especially who they are and how they experience the world. This may increase perceived understanding, resulting in more positive outcomes for disclosure with a human partner.

> Perceived Understanding Hypothesis: Because of increased perceived understanding, emotional, relational, and psychological effects will be greater when disclosing to a person than to a chatbot.

Disclosure processing framework

A perspective we call the disclosure processing framework emphasizes the advantages that non-human partners may provide compared to human partners. This framework suggests that people will disclose more to chatbots and subsequently experience more positive outcomes. Fears of negative judgment commonly prevent individuals from disclosing deeply to other people. Worries about being rejected, judged, or burdening the listener restrain disclosure to other people, obviating potential benefits (Afifi & Guerrero, 2000). Disclosure intimacy, however, may increase when the partner is a computerized agent rather than another person, because individuals know that computers cannot judge them (Lucas et al., 2014). Computerized agents reduce impression management and increase disclosure intimacy compared to human partners in situations in which fears of negative evaluation may be prominent (e.g., when asked potentially embarrassing questions; Kang & Gratch, 2010; Lucas et al., 2014). If this occurs in all situations, and not just situations that heighten fears of judgment, people may disclose more intimately to a chatbot than they would to a person.

The more intimately individuals' disclosures are to a chatbot, the greater the psychological benefits they may accrue, compared to disclosing less intimately to another person. According to Pennebaker's (1993) cognitive processing model, a key component of the link between cognitive changes and beneficial outcomes is the process by which disclosing what was formerly undisclosed eliminates negative affect and processing and induces reappraisal. Pennebaker & Chung (2007) argue that putting words to these negative emotions and thoughts changes their nature from affective to cognitive. This switch to a cognitive nature reduces the intensity and power of the negative emotion (Lieberman et al., 2007). Forming a narrative of the situation facilitates new insights and eliminates rumination over what was previously confusing or bothersome (Lepore, Ragan, & Jones, 2000; Pennebaker & Chung, 2007). This, paired with supportive responses from the partner, results in emotional, relational, and psychological benefits (Jones & Wirtz, 2006; Pennebaker, 1993).

> Disclosure Processing Hypothesis: Due to greater disclosure intimacy and cognitive reappraisal, emotional, relational, and psychological effects will be greater when disclosing to a chatbot than to a person.

CASA framework

The Computers as Social Actors (CASA) framework predicts a third possibility. According to this framework, people instinctively perceive, react to, and interact with computers as they do with other people, without consciously intending to do so (Reeves & Nass, 1996). This tendency is so pervasive that it is a foundational component of theoretical thinking about interactions between humans and computerized agents, to the extent that it is thought to be "unlikely that one will be able to establish rules for human-agent/robot-interaction which radically depart from what humans know from and use in their everyday interactions" (Krämer, von der Pütten, & Eimler, 2012) [*sic*]. This framework suggests that disclosure processes and outcomes will be similar, regardless of whether the partner is a person or a chatbot.

A plethora of studies have found that people form perceptions of computerized agents and humans in the same way, even though people consciously know that computers are machines that do not have human personalities. For instance, people perceive a computerized agent to be as inspired, strong, or afraid as another person is (von der Putten, Krämer, Gratch, & Kang, 2010). This occurs not just when the partner is actually a computer, but also when the partner is believed to be a computer (von der Putten et al., 2010). The tendency for individuals to judge and react to computers as they do to other people has been observed across different kinds of computerized agents, from embodied conversational agents to robots and text-only chatbots (Eyssel & Hegel, 2012; Krämer et al., 2012; Reeves & Nass, 1996).

Across many different types of social situations, people also behave and interact with computers in ways that are common in human-human interactions, applying social norms derived from experiences with other people to interactions with

computers. Individuals, for instance, are more cooperative towards a computer on the same "team" compared to a computer on a different team, reciprocate a computer's disclosures in similar ways to how they reciprocate another person's disclosures, and behave in more polite ways to a computer if that computer asks questions about its own performance compared to when a different computer asks about its performance (Moon, 2000; Nass, Fogg, & Moon, 1996; Nass, Moon, & Carney, 1999).

This tendency is assumed to occur because of mindlessness (Langer, 1992). When not motivated to carefully scrutinize the veracity of every message and its source (Liang, Lee, & Jang, 2013), we react to the people appearing on screens as if they were present in real life because we have "old brains" (Reeves & Nass, 1996, p. 12) that have not caught up with the new media we now encounter. We mindlessly apply social scripts to those interactions, behaving in ways that are similar to how we would behave to another person, even while consciously aware that the computer is not a person. Work using fMRIs shows how deeply this tendency runs. The neural patterns activated by emotional stimuli are similar whether people observe a robot or a person (Rosenthal-von der Pütten et al., 2014).

Unlike the other two frameworks, this framework suggests that the processes that lead to disclosure benefits, such as perceived understanding, disclosure intimacy, and cognitive reappraisal, should not differ depending on the perceived identity of the conversation partner. Rather, these processes will unfold in the same way, regardless of whether the partner is a chatbot or a person. Individuals may thus disclose to a chatbot as they do to another person, engage in cognitive reappraisal to the same degree, and feel just as understood, regardless of whether the partner is a chatbot or a person.

> Equivalence Hypothesis: Perceived understanding, disclosure intimacy, and cognitive reappraisal processes from disclosing to a partner will lead to equivalent emotional, relational, and psychological effects between chatbot and person partners.

The present study

In this study, we examined the effects of partner identity (chatbot vs. human) on disclosure outcomes to determine which of the three hypotheses is best supported. In an experiment, participants were told they would have an online chat conversation with either a chatbot or a person. We employed a standard method within human-computer interaction, termed a Wizard of Oz[1] (WoZ) method (Dahlbäck, Jönsson, & Ahrenberg, 1993), in which participants were told that the partner is a computer, when in actuality a hidden person behind the scenes (the "wizard") is the one interacting with participants. In other words, some participants were told that they would have a conversation with a chatbot and some were told they would have a conversation with a person, but in all cases, the partner was a person.

The WoZ procedure was used rather than employing a real chatbot because of current chatbots' conversational limitations, and because perceptions of the

partner's identity can impact subsequent effects, while actual identity does not (Lucas et al., 2014). For simplicity, we will refer to the conversations in which participants were told the partner was a chatbot as disclosing "to a chatbot" or that the partner "was a chatbot" and conversations in which participants were told the partner was a person as disclosing "to a person" or that the partner "was a person."

Disclosure was manipulated to either be emotional or factual (Morton, 1978; Reis & Shaver, 1988). Factual disclosures contain objective information about the discloser (e.g., "I biked to school today") whereas emotional disclosures involve emotions and feelings (e.g., "I feel really upset that she could do that to me"). Compared to factual disclosure, emotional disclosure has been found to result in more beneficial outcomes (Pennebaker & Chung, 2007).

Methods

Power

The closest equivalent to our procedure is the expressive writing procedure, in which college students emotionally or factually disclose in writing and psychological outcomes are assessed (Smyth, 1998). This effect size is $r = .35$ (Smyth, 1998). A power analysis with a significance level of .05, powered at 80% (see Cohen, 1992), required at least 23 participants per condition to find an effect, for a total of 92 participants after exclusions.

Participants

Participants were recruited from university research participation websites and flyers posted around a university campus. A total of 128 participants took part in the study (68.75% women; M age = 22), and received either course credit or financial compensation. Participants were excluded if technical issues prevented them from participating in the chat conversation for the full amount of time ($n = 5$), if they failed to follow directions ($n = 3$), or if the research assistant noted that there was a major issue in the conversation (e.g., the participant took an extremely long time to respond, such that the conversation was very short; $n = 7$).

Participants were also excluded if they were suspicious of the partner's identity or the study's purpose. A funneled debriefing procedure revealed that some participants guessed the purpose of the study ($n = 2$), and that some were suspicious about whether the partner was a person or a chatbot ($n = 13$). This level of suspicion was lower or on par with other studies employing WoZ methods (e.g., Aharoni & Fridlund, 2007; Hinds, Roberts, & Jones, 2004), and our total sample size after exclusions ($N = 98$) met the threshold from the power analysis.

Training

Before the procedure, three undergraduate research assistants, who acted as confederates in the study and interacted with participants, were trained on how to provide validating responses. Training was overseen by a clinical psychologist. Research

assistants were trained to validate the participant's emotion or experience by expressing understanding and acceptance, to help participants articulate their emotions, and to provide sympathy and empathy. They were also trained to avoid minimizing the participant's experience, criticizing or blaming the participant, and providing unsolicited advice (Burleson, 2003; Jones & Wirtz, 2006).

Procedure

This procedure was reviewed and approved by the university Institutional Review Board. For the study, a 2 (disclosure type: emotional vs. factual) x 2 (perceived partner: chatbot or person) experimental design was employed. Participants were randomly assigned to be told they would have a conversation with either another person or a chatbot. In case participants did not know what a chatbot was, it was described as "a computer program that can have a conversation with people, and is being built and refined by researchers at [university]." The purported other person was introduced as "a student studying Communication at [university]."

Participants were also randomly assigned to engage in an emotional conversation, in which they were asked to discuss a current problem they were facing and their deepest feelings about it, or a factual conversation, in which they were asked to describe their schedule for the day and the upcoming week in an objective way, without mentioning their feelings or emotions (Birnbaum et al., 2016; Gortner, Rude, & Pennebaker, 2006).

Participants then engaged in an online chat conversation for 25 minutes with a trained confederate, who was blind to whether participants thought they were talking to a chatbot or to a person, but not to whether participants were asked to engage in emotional or factual disclosure. In the emotional conversations, the confederate provided the participant with validating responses and asked questions to probe and to elicit more disclosure. In the factual conversations, the confederate responded positively and with interest (e.g., "Great!") and asked follow-up questions, such as, "And how many minutes did it take to walk there?" The confederate was instructed not to ask about the participants' feelings or emotions in these conversations.

The chat interface utilized was Chatplat, an online chat platform enabling real-time chat conversations between confederates and participants. After the conversation ended, participants completed two questionnaires, purportedly unrelated to each other, containing our self-report measures. The chat window was embedded within an online survey tool (Qualtrics), which contained all of our instructions and self-reported measures (for an image of the chat window, see Supplementary Material).

Outcome measures

Three types of outcomes were measured.

Relational measures

We measured how "friendly" and "warm" the partner was (Birnbaum et al., 2016). Responses ranged from "extremely" (5) to "not at all" (1), and were combined to

form an index of partner warmth (α = .85). Participants were also asked how much they "like" the partner (Petty & Mirels, 1981). Responses spanned from "like a great deal" (7) to "dislike a great deal" (1). We also measured how much participants "enjoyed" the interaction (Shelton, 2003). Responses ranged from "a great deal" (5) to "not at all" (1).

Emotional measures

Two questions from the Comforting Responses Scale (Clark et al., 1998) were used to measure immediate emotional experiences. Participants reported whether they "feel more optimistic now that I have talked with the chatbot/person" and whether they "feel better" after the conversation. Responses ranged from "strongly agree" (7) to "strongly disagree" (1) and were combined into a "feeling better" index (α = .89). In addition, we assessed negative mood after the interaction by asking how "distressed," "upset," "guilty," and "afraid" participants felt (PANAS; Watson, Clark, & Tellegen, 1988). Responses ranged from "extremely" (5) to "not at all" (1), and were combined to form an index of negative mood (α = .76).

Psychological measure

We used a common measure of self-affirmation: defensiveness towards health risk information. When faced with threatening information, individuals engage in defensive tendencies, such as dismissing the information as irrelevant. Self-affirmation reduces this defensive tendency, such that the more people have been affirmed, the lower their defensiveness (Sherman, Nelson, & Steele, 2000). We presented participants with an article explaining the cognitive impairments that can result from sleeping less than nine hours per day, especially for college students. Participants were then presented with the following questions from Sherman and colleagues (2000): how important they think it is that college students get nine or more hours of sleep per night ("extremely" [5] to "not at all" [1]); to what extent the participant personally should get nine hours of sleep ("strongly agree" [7] to "strongly disagree" [1]); and how likely it is that the participant personally will increase the amount of sleep he or she receives ("extremely" [5] to "not at all" [1]). These were normalized into z-scores and combined into an index of defensiveness (α = .73). Higher scores indicate lower defensiveness, or higher self-affirmation.

Process measures

Three processes that have been hypothesized to take place in disclosure and lead to emotional, relational, and psychological outcomes were also measured.

Perceived understanding

Two items were chosen from the Working Alliance Inventory (Horvath & Greenberg, 1989) to reflect the sense that someone else "gets it" and the comfort of being fully understood (Reis et al., 2017): "The chatbot/person and I understood each other" and "I felt uncomfortable when talking to the chatbot/person" ("strongly agree" [7] to "strongly disagree [1]). Despite low reliability (α = .49),

these two items lie on the same subscale or factor (Horvath & Greenberg, 1989), so they were combined into an index.

Disclosure intimacy

Transcripts were first coded into utterances, defined as one complete sentence or phrase (Guetzkow, 1950). A second coder coded 20% of the transcripts into utterances. This yielded a Cohen's kappa of .75. There were 3,259 utterances in total. The average length of an utterance was 13 words and, on average, participants used 33.26 utterances.

An intimacy scheme based on Altman and Taylor (1973) was adapted. A similar scheme has been used in studies involving disclosure to perceived computer agents and people (e.g., Kang & Gratch, 2010). The scheme used three levels of intimacy: low (coded 1), including objective facts about the situation (e.g., "then she told him what I had said"); medium (coded 2), including attitudes, thoughts, and opinions about the situation (e.g., "I feel like I don't have any control anymore"); and high (coded 3), consisting of explicitly verbalized emotions and affect only (e.g., "I am angry"). Utterances that were not disclosure (e.g., "thank you") were coded as 0. A second coder coded 20% of the utterances, yielding a Cohen's kappa of .85. Each utterance received a score from 0 to 3. Scores for disclosure only (1–3) were averaged and then normalized according to the number of disclosure utterances each participant gave in the conversation, such that each participant received one overall intimacy score.[2]

Cognitive reappraisal

The way words are used in language have been found to reveal the psychological mindset, tendencies, or attitude of the speaker (Tausczik & Pennebaker, 2010). Numerous studies have demonstrated that the specific words used in disclosures are tied to beneficial outcomes. The more causal, insight-related, and positive emotion words used in disclosive texts, the greater the improvements in physical and psychological health (e.g., Pennebaker & Francis, 1996). Usage of these words has been found to reflect cognitive reappraisal and reconstrual processes (Tausczik & Pennebaker, 2010): as disclosers come to think about the situation differently and broaden their perspective, they develop new explanations for why the situation may have happened (causal words), discover new insights as to how to navigate it better (insight-related words), and see the positive in an otherwise negative situation (positive emotion words; Pennebaker & Chung, 2007).

We measured usage of causal, insight-related, and positive emotion words via linguistic inquiry and word count (LIWC), the standard analytic tool used in disclosure and well-being studies. These variables from LIWC are well validated, and have been found to predict positive effects (Pennebaker & Francis, 1996). LIWC both counts the number of instances that words fall into pre-defined categories based on internal dictionaries and provides the percentage of words that fall into each category, preventing variables from being skewed by word count.

Manipulation checks

Disclosure manipulation check

The manipulation check of the emotional and factual conversation conditions included self-report items about the information participants shared (Pennebaker & Beall, 1986): how personal it was ("extremely" [5] to "not at all" [1]), and how much emotion they revealed to their partner ("a great deal" [5] to "not at all" [1]).

WoZ manipulation check

In addition to our self-report measures of suspicion, we examined participants' language style to test whether participants really did believe they were conversing with a chatbot. Past work has found that people assume that computers are limited in their ability to understand human language, and thus speak in simpler ways to computers than to humans, such as by using shorter but more sentences (Branigan, Pickering, Pearson, & McLean, 2010; Hill, Ford, & Farreras, 2015). Simpler and clearer language has also been found to contain fewer conjugations (e.g., "and," "whereas") and non-fluencies (e.g., "hmm" or "umm"; Tausczik & Pennebaker, 2010). Thus, we measured conjugations and non-fluencies through LIWC.

Other differences across conditions

Given the novelty of disclosing to a chatbot, we also included several measures that we did not expect to vary across conditions, but considered important to assess to ensure that the chatbot manipulation did not have unexpected effects beyond those predicted for the disclosure manipulation. It could be that chatbots affect these processes during conversation in ways that humans do not. In addition, we measured individual differences to ensure that the conditions did not differ in underlying levels of variables that could affect the results. But as expected, we did not find differences across conditions on these measures. For full transparency and replicability, we include and explain all of these measures in the Supplementary Material.

Results

Manipulation checks

Disclosure

In the emotional conditions, participants felt that they revealed their emotions more and shared information that was more personal, as compared to participants in the factual conditions ($F[1, 94] = 133.84$, $p < .001$).

WoZ

Participants used clearer and simpler language in the chatbot conditions compared to the person conditions. When talking to a perceived chatbot as opposed to a person, participants used fewer conjugations ($F[1, 94] = 8.30$, $p < .005$, partial $\eta^2 = .08$) and non-fluencies ($F[1, 94] = 9.41$, $p < .003$, partial $\eta^2 = .09$); there were no significant interactions or other main effects for perceived partner on any other measure in the study. This was the case even though the confederates themselves

did not use fewer conjugations or non-fluencies in the chatbot conditions compared to the person conditions ($F[1, 94] = 0.20$, $p = .66$, partial $\eta^2 = .002$; $F[1, 94] = 2.15$, $p = .15$, partial $\eta^2 = .02$).

Overview

The study employed a 2 (disclosure type: emotional vs. factual) x 2 (perceived partner: chatbot vs. person) experimental design. For each dependent variable, an analysis of variance tested the effects of between-subject factors disclosure type and perceived partner and their interaction. A correlation matrix of all dependent variables can be found in Table 1.

Participants who disclosed to chatbots experienced as many emotional, relational, and psychological benefits as participants who disclosed to a human partner. Consistent with previous research, effects were stronger for emotional disclosure versus factual disclosure. This occurred regardless of partner (see Figure 1). This pattern of significant effect of type of disclosure, but no effect of partner, occurred across all three types of outcome measures (emotional, relational, psychological) as well as our three process measures (perceived understanding, disclosure intimacy, cognitive reappraisal). For an overview of the effects, see Table 2. In addition, the processes of perceived understanding, disclosure intimacy, and cognitive reappraisal all correlated as expected with outcomes and, importantly, showed the same pattern across the chatbot and person partners, supporting the Equivalence Hypothesis.

Outcomes

Emotional outcome

Regardless of whether the partner was perceived to be a chatbot or a person, participants reported improved, immediate emotional experiences after emotional disclosure compared to factual disclosure (see Figure 1). Participants reported a higher negative mood ($F[1, 94] = 9.61$, $p < .001$, partial $\eta^2 = .09$), a common byproduct of emotional disclosure that eventually leads to improvement (Greenberg & Stone, 1992), and that they felt significantly better ($F[1, 94] = 13.36$, $p < .001$, partial $\eta^2 = .12$) after emotional disclosure compared to factual disclosure, with no significant effects of perceived partner and no significant interactions. This pattern of results supports the Equivalence Hypothesis.

Relational outcomes

For both human and chatbot partners, perceptions of relationship quality after emotional disclosure were higher compared to factual disclosures, with no effects of perceived partner or any interactions (see Figure 1). After engaging in an emotional conversation, participants reported feeling that the partner was warmer ($F[1, 94] = 4.91$, $p < .03$, partial $\eta^2 = .05$) and that they enjoyed the interaction more ($F[1, 94] = 5.49$, $p < .02$, partial $\eta^2 = .05$) compared to a factual conversation. In addition, there was a trend towards, but not a significant effect of, liking the partner more after emotional compared to factual disclosure ($F[1, 94] = 2.84$, $p = .09$, partial $\eta^2 = .03$).

Table 1 Bivariate Correlations Between Outcome Measures

Outcome measures	1	2	3	4	5
1. Feeling better	–				
2. Negative mood	.19	–			
3. Partner warmth	.44*	.10	–		
4. Enjoyment of conversation	.59*	.10	.59*	–	
5. Liking of partner	.57*	.13	.65*	.63*	–
6. Self-affirmation	.18	−.12	.04	.05	.01

Note: * $p < .001$

Again, the main effect of disclosure condition without any main effect of partner or interaction supports the Equivalence Hypothesis.

Psychological outcomes

Talking with a chatbot instead of a human also did not change whether participants were defensive towards threatening information. Whether talking to a person or to a chatbot, participants in the emotional conditions were less defensive towards changing their behavior after reading threatening health information ($F[1,94] = 8.63$, $p < .004$, partial $\eta^2 = .08$), and thus felt greater self-affirmation, than participants in the factual conditions, with no effect of perceived partner or interaction (see Figure 1). This pattern also aligns with the Equivalence Hypothesis.

Processes

Outcomes of disclosure were similar whether the partner was perceived to be a chatbot or a person, supporting the Equivalence Hypothesis. But did the processes laid out by the Perceived Understanding and Disclosure Processing Hypotheses also take place, and did they differ by type of partner? Recall that according to the Perceived Understanding Hypothesis, disclosure exerts benefits through feeling truly understood by the partner. The Disclosure Processing Hypothesis emphasizes that disclosure is beneficial when intimate disclosure occurs and spurs cognitive reappraisal.

We used a correlational analysis to test that the hypothesized processes were indeed related to the outcomes. We collapsed the six outcome measures into a single standardized outcome index ($\alpha = .70$), given that the outcome measures all exhibited the same pattern, in which emotional disclosure elicited stronger effects than factual disclosure. Across the entire sample, the processes (perceived understanding, disclosure intimacy, and cognitive reappraisal) were significantly associated with the outcome index. In other words, the more participants engaged in these processes, the more beneficial their outcomes were. The strongest association was between disclosure intimacy and the outcomes ($r = .46$, $p < .001$). The linguistic measure of cognitive reappraisal was also significantly associated with outcomes ($r = .37$, $p < .001$), as was understanding ($r = .22$, $p < .03$). These data confirm prior

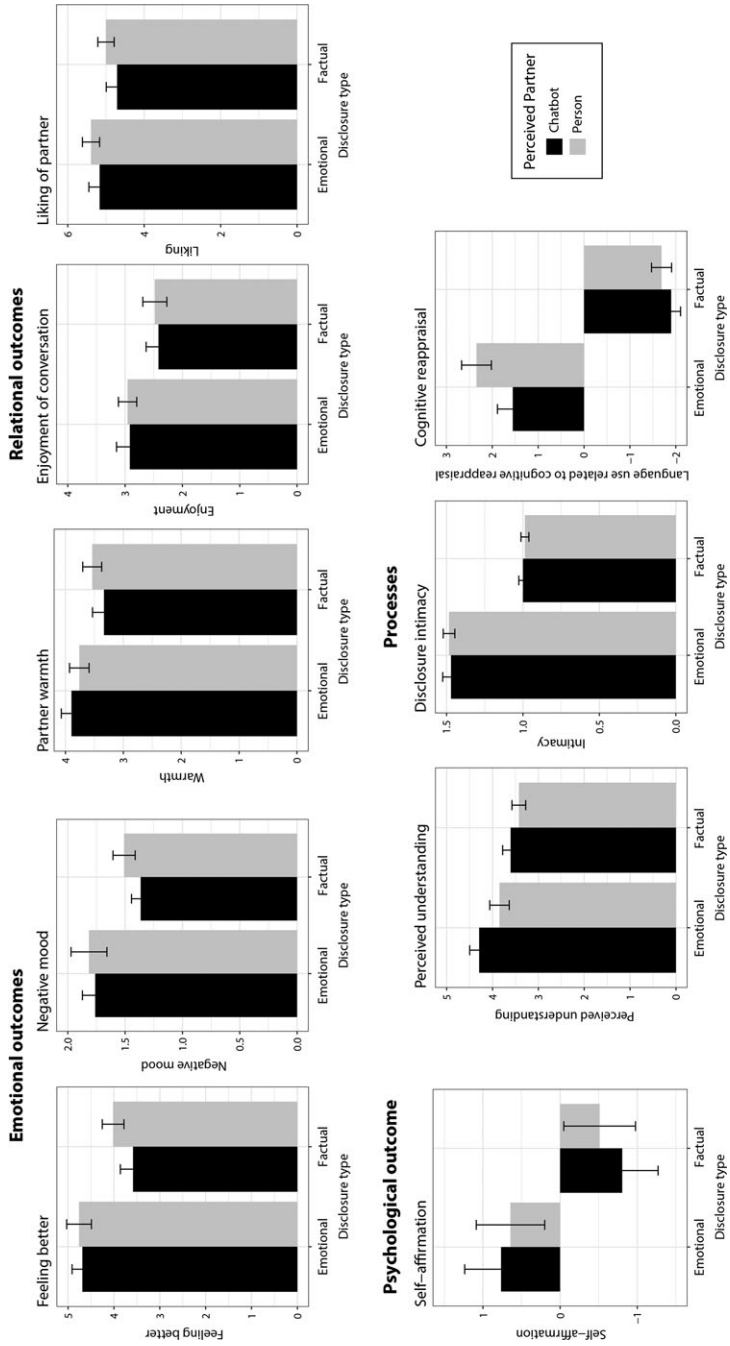

Figure 1 Means of outcome measures and processes by condition.

Table 2 Direction of Main Effects for Each Dependent Variable

Outcome measures	Emotional vs. Factual Disclosure	Chatbot vs. Person
Emotional outcomes		
Feeling better	>	=
Negative mood	>	=
Relational outcomes		
Perceptions of partner warmth	>	=
Enjoyment of the interaction	>	=
Liking	=	=
Psychological outcome		
Self-affirmation	>	=
Processes		
Perceived understanding	>	=
Disclosure intimacy	>	=
Words revealing cognitive reappraisal	>	=

research that intimate disclosure leads to beneficial downstream outcomes through these mechanisms, as the Perceived Understanding and Disclosure Processing Hypotheses suggest (Pennebaker, 1993; Reis et al., 2017).

Next, we examined whether these processes occurred equally strongly for each type of partner. For example, if one process occurred more strongly with chatbot partners and an opposing process occurred more strongly with human partners, these could cancel each other out and result in equivalent outcomes. This was not the case. We found that all three processes occurred equally strongly for both types of partners. Participants felt more understood after emotional disclosure compared to factual disclosure ($F[1, 94] = 8.94$, $p < .004$, partial $\eta^2 = .09$), but there was no significant effect of partner, nor a significant interaction. Similarly, participants disclosed more intimately ($F[1, 94] = 286.25$, $p < .001$, partial $\eta^2 = .90$) and used more words reflecting cognitive reappraisal (positive emotion words: $F[1, 94] = 56.92$, $p < .00001$, partial $\eta^2 = .38$; causal words: $F[1, 94] = 29.79$, $p < .00001$, partial $\eta^2 = .24$; insight words: $F[1, 94] = 147.89$, $p < .00001$) after emotional disclosure compared to factual disclosure, with no effect of perceived partner and no interaction.

Finally, we examined whether these processes led to downstream outcomes in the same way across human and chatbot partners. Fisher's z transformation tests revealed that the correlations between the processes and outcomes were similar, with no significant differences in strength, between partner conditions. The correlation between understanding and outcomes was .21 for chatbot partners and .27 for human partners, which were not significantly different from each other ($z_{r1-r2} = .299$, $p = .76$). Similarly, the correlation between disclosure intimacy and outcomes was .45 for chatbot partners and .47 for human partners, which did not significantly differ

($z_{r1-r2} = .12$, $p = .90$). Finally, the correlation between cognitive reappraisal and outcomes was .37 for chatbot partners and .38 for human partners, which also were not significantly different ($z_{r1-r2} = .08$, $p = .93$). Thus, the processes were associated with the outcomes to the same strength for human partners as they were for chatbot partners, providing additional support for the Equivalence Hypothesis.

Discussion

This is the first study to compare the downstream effects of disclosing to a chatbot partner with disclosing to a human partner. We found that chatbots and humans were equally effective at creating emotional, relational, and psychological benefits. Consistent with previous research, effects were stronger after emotional disclosure compared to factual disclosure, and we found that this occurred regardless of the partner's perceived identity (see Table 2). We also found that the processes involved in producing these outcomes (perceived understanding, disclosure intimacy, and cognitive reappraisal) were equivalent in strength and affected outcomes similarly, regardless of the perceived partner. While we had sufficient power to detect main effects of the disclosure manipulation, no effects for the partner manipulation were observed in any of our outcome or process measures. This provides support for the CASA framework and its Equivalence Hypothesis, that people psychologically engage with chatbots as they do with people, resulting in similar disclosure processes and outcomes.

Theoretical implications

Our results align with previous CASA work directly comparing perceptions of supposed humans and supposed computer agents and finding no major differences (von der Putten et al., 2010). Our study takes this work a step further by finding that not only are perceptions similar, but disclosure to bots and humans exert similar beneficial outcomes. To our knowledge, this is the first study to find that more wide-ranging downstream effects, such as those relating to immediate emotional experience or self-affirmation, are similar whether the partner is a computer or a person. These findings expand on the work of communication scholars who have either focused on resulting perceptions but not beneficial outcomes of disclosure (Birnbaum et al., 2016; Gong & Nass, 2007) or on beneficial outcomes when the listener is always human, but not when the listener is a computer (Jones & Wirtz, 2006).

The Perceived Understanding Hypothesis and the Disclosure Processing Hypothesis received partial support, in that our data replicated previous research on these mechanisms, but no support for their predictions on how these mechanisms would be either undermined or enhanced by disclosure to a chatbot partner. This suggests that what matters most is not the partner's humanness, but what occurs in the interaction itself. Did processes like disclosure intimacy, cognitive reappraisal, and perceived understanding take place? Did the interaction itself provide the necessary components for the benefits of disclosure to accrue? This aligns with work in

communication that emphasizes the importance of the features of the actual interaction, and how they shape the effects of disclosure (e.g., Barbee & Cunningham, 1995).

Why would disclosers mindlessly respond to chatbot partners in the same way as human partners, even when it is obvious that chatbots inherently cannot truly understand them or negatively judge them? One possible answer comes from the social monitoring system (SMS) model (Pickett & Gardner, 2005). According to the SMS model, people have a drive to belong with others. When this need increases, such as after rejection, a monitoring system is triggered, which motivates people to pay attention to social information that could connote rejection or acceptance. When the need to belong is met, such as after social acceptance or validation (Lucas, Knowles, Gardner, Molden, & Jefferis, 2010), this monitoring system remains dormant and people do not pay close attention to social information. Thus, when receiving validating responses, disclosers' social monitoring system may not attend to the fact that the partner is a computer that cannot reject them or cannot inherently understand them as a person. Then, as the CASA framework describes, they will react mindlessly in the same way to chatbot partners as they would to human partners, even though they consciously know that computers are non-social, pre-programmed machines. Future work should test whether the SMS model operates in this way and further examine the belongingness process.

Future work should also explore the impact of relationship duration and role expectations on disclosure and subsequent effects. Feeling understood at the deepest level is especially important for establishing connection and intimacy in close relationships (Reis & Shaver, 1988). In this study, the perceived understanding mechanism did not exert effects as strong as the other processes, which is not surprising given that the partner in the present study was a stranger, with whom there was no anticipation of future interaction. Feeling understood may exert more powerful effects when an ongoing relationship is established or anticipated to continue than in the context of a single interaction. In addition to relationship duration, conversation partners often are in specific roles, with unique rights and obligations (e.g., Miner et al., 2016). Understanding the perceived roles and expectations of chatbots in sensitive contexts, along with more long-term interactions, merits further investigation.

Disclosure dynamics

Our participants disclosed equally, in terms of intimacy, to a chatbot as to a person. In other recent research, however, participants disclosed more intimately to a perceived computer than to a perceived person (Gratch, Lucas, King, & Morency, 2014; Kang & Gratch, 2010; Lucas et al., 2014).

Why did these studies show more disclosure intimacy with computers while our study did not? One possibility is that disclosure intimacy increases only when fears of negative evaluation are more salient. In Lucas et al. (2014), participants were asked questions that might raise concerns of judgment (e.g., "Tell me about an event or something that you wish you could erase from your memory" or "What's something you feel guilty about?"). Participants were also asked embarrassing questions

in Kang & Gratch (2010), such as, "What is your most common sexual fantasy?" Gratch et al. (2014) and Lucas et al. (2014) asked questions about psychological health, and stigmas about mental illnesses can stir fears of judgment.

In contrast, our participants were not asked to disclose anything particularly embarrassing or stigmatizing, and often chose to discuss matters that they had also disclosed to others. This may be why intimacy of disclosure to bots compared to humans was equivalent, suggesting that people only disclose more intimately to a computer compared to a person when fears of being judged are heightened. Recent work provides evidence for this possibility, showing that differences in socially desirable responses to conversational agents, based on fears of being negatively evaluated by a survey interviewer, occurred only for sensitive questions, not for non-sensitive questions, and that people prefer disclosing to a computer compared to a human on very sensitive topics, but show no preference on mildly sensitive topics (Pickard, Roster, & Chen, 2016; Schuetzler, Giboney, Grimes, & Nunamaker, 2018). It is important for future work to actually investigate whether disclosure intimacy to computers increases compared to humans in all situations, or only in situations in which fears of negative evaluation are especially salient.

Limitations

The study has several limitations. First, all of our participants were affiliated with a large research university. It is unclear whether the general population would exhibit the same effects, given that they may be less familiar with technological innovations than our participants.

Another limitation is that the conversations did not involve reciprocity from the partner. While reciprocity occurs naturally in conversations between two people (Altman & Taylor, 1973), confederates did not reciprocate disclosures from participants with disclosures of their own, to prevent pulling disclosers' attention away from their own feelings (Burleson, 2003). Prior work has shown that disclosure followed by supportive responses without reciprocal disclosure still leads to strong effects and benefits for the participants (e.g., Shenk & Fruzzetti, 2011). Thus, our study did not include reciprocal disclosure, which may explain why we did not find significant effects in terms of liking of the partner. Liking of the partner is shaped more by listening to disclosures than by making disclosures (Sprecher, Treger, Wondra, Hilaire, & Wallpe, 2013), so future studies can examine whether reciprocal disclosure from the partner leads to different results than those found in this study.

Additionally, participants' perceptions of chatbot competence were not measured before the conversation. Though we saw evidence of lower expectations of chatbots' language processing abilities, it is possible that the chatbot's responses were better than participants expected in some other way, inflating perceptions of the chatbot's capabilities and leading to similar outcomes compared to human partners. Future research should address this possibility.

Finally, though prior work has found equivalence across conditions even when the partner is actually a human or actually a computer (e.g., Lucas et al., 2014), it is

possible that there is something extra that humans did or said in our study, undetectable by our current LIWC measures, that led to equivalence in effects. Future work should explore the use of advanced computational approaches to conversation analysis (e.g., Althoff, Clark, & Leskovec, 2016).

Conclusion

This is the first study to examine the psychological impact of disclosing to a partner with a computer identity, such as a chatbot, compared to another person. Across a variety of downstream outcomes, disclosing to a chatbot was just as beneficial as disclosing to another person. Emotional disclosure was more beneficial than factual disclosure because of enhanced perceived understanding, disclosure intimacy, and cognitive reappraisal, with no difference depending on whether the partner was perceived to be a chatbot or a person.

Acknowledgments

This work was supported by award NSF SBE 1513702 from the National Science Foundation, the Stanford Cyber-Initiative Program, and a National Institutes of Health, National Center for Advancing Translational Science, Clinical and Translational Science Award (KL2TR001083 and UL1TR001085). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank the research assistants who helped collect data, as well as the Social Media Lab and the anonymous reviewers for their valuable feedback.

Supplementary Material

Supplementary material are available at *Journal of Communication* online.

Notes

1 This technique derives its name from the 1939 Hollywood film, *Wizard of Oz*, in which the main character and her companions encountered intimidating creatures, such as monsters that were actually controlled by a small human man (the "wizard") behind a curtain (Fleming, 1939).

2 An intimacy scheme with more granularity (e.g., differentiating between disclosures on more superficial topics and disclosures on more personal topics) did not change the pattern of results. Thus, we present results using this three-level scheme for consistency with prior work.

References

Afifi, W. A., & Guerrero, L. K. (2000). Motivations underlying topic avoidance in close relationships. In S. Petronio (Ed.), *Balancing the secrets of private disclosures* (pp. 165–180). Mahwah, NJ: Erlbaum.

Aharoni, E., & Fridlund, A. J. (2007). Social reactions toward people vs. computers: How mere lables shape interactions. *Computers in Human Behavior, 23*(5), 2175–2189.

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics, 4*, 463–476.

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, England: Rinehart & Winston.

Archer, R. L. (1980). Self-disclosure. In D. M. Wegner & R. R. Vallacher (Eds.), *The Self in Social Psychology* (pp. 183–205). New York, NY: Oxford University Press.

Barbee, A. P., & Cunningham, M. R. (1995). An experimental approach to social support communications: Interactive coping in close relationships. *Annals of the International Communication Association, 18*(1), 381–413.

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass, O. (2016). What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior, 63*, 416–423.

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics, 42*(9), 2355–2368.

Burleson, B. R. (2003). Emotional support skills. In J. O. Greene & B. R. Burleson (Eds.), *Handbook of communication and social interaction skills* (pp. 551–594). Mahwah, NJ: Lawrence Erlbaum.

Clark, R. A., Pierce, A. J., Finn, K., Hsu, K., Toosley, A., & Williams, L. (1998). The impact of alternative approaches to comforting, closeness of relationships, and gender on multiple measures of effectiveness. *Communication Studies, 49*, 224–239.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Creswell, J. D., Lam, S., Stanton, A. L., Taylor, S. E., Bower, J. E., & Sherman, D. K. (2007). Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? *Personality and Social Psychology Bulletin, 33*(2), 238–250.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—Why and how. *Knowledge-based Systems, 6*(4), 258–266.

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology, 42*(9), 2213–2230.

Fleming, V. (Director). (1939). *The wizard of oz* [Motion picture]. Los Angeles, CA: Metro Goldwyn Mayer.

Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human-Computer Interaction, 30*(5), 401–432.

Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human Communication Research, 33*(2), 163–193.

Gortner, E. M., Rude, S. S., & Pennebaker, J. W. (2006). Benefits of expressive writing in lowering rumination and depressive symptoms. *Behavior Therapy, 37*(3), 292–303.

Gratch, J., Lucas, G. M., King, A. A., & Morency, L. P. (2014, May). It's only a computer: The impact of human-agent interaction in clinical interviews. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems* (pp. 85–92). Los Angeles, CA: International Foundation for Autonomous Agents and Multiagent Systems.

Greenberg, M. A., & Stone, A. A. (1992). Emotional disclosure about traumas and its relation to health: Effects of previous disclosure and trauma severity. *Journal of Personality and Social Psychology, 63*(1), 75–84.

Guetzkow, H. (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology, 6*(1), 47–58.

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

Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-Computer Interaction, 19*(1), 151–181.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*(2), 223–233.

Jones, S. M., & Wirtz, J. G. (2006). How does the comforting process work? An empirical test of an appraisal-based model of comforting. *Human Communication Research, 32*, 217–243.

Kang, S. H., & Gratch, J. (2010). Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation and Virtual Worlds, 21*(3–4), 473–482.

Kelley, J. E., Lumley, M. A., & Leisen, J. C. (1997). Health effects of emotional disclosure in rheumatoid arthritis patients. *Health Psychology, 16*(4), 331–340.

Krämer, N. C., von der Pütten, A., & Eimler, S. (2012). Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In M. Zacarias & J. V. de Oliveira (Eds.), *Human-computer interaction: The agency perspective* (pp. 215–240). Berlin, Germany: Springer-Verlag.

Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition, 1*(3), 289–305.

Lepore, S. J., Ragan, J. D., & Jones, S. (2000). Talking facilitates cognitive–emotional processes of adaptation to an acute stressor. *Journal of Personality and Social Psychology, 78*(3), 499–508.

Liang, Y. J., Lee, S. A., & Jang, J. W. (2013). Mindlessness and gaining compliance in computer-human interaction. *Computers in Human Behavior, 29*(4), 1572–1579.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421–428.

Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100.

Lucas, G. M., Knowles, M. L., Gardner, W. L., Molden, D. C., & Jefferis, V. E. (2010). Increasing social engagement among lonely individuals: The role of acceptance cues and promotion motivations. *Personality and Social Psychology Bulletin, 36*(10), 1346–1359.

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors, 48*(2), 241–256.

Markoff, J., & Mozur, P. (2015, July 31). For sympathetic ear, more chinese turn to smartphone program. *New York Times*. Retrieved from https://www.nytimes.com/2015/08/04/science/for-sympathetic-ear-more-chinese-turn-to-smartphone-program.html

Martins, M. V., Peterson, B. D., Costa, P., Costa, M. E., Lund, R., & Schmidt, L. (2013). Interactive effects of social support and disclosure on fertility-related stress. *Journal of Social and Personal Relationships, 30*(4), 371–388.

Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to machines about personal mental health problems. *JAMA : the Journal of the American Medical Association, 318* (13), 1217–1218.

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine, 176*(5), 619–625.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research, 26*(4), 323–339.

Morton, X. L. (1978). Intimacy and reciprocity of exchange: A comparison of spouses and strangers. *Journal of Personality and Social Psychology, 36*, 72–81.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669–678.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology, 29*(5), 1093–1109.

Pennebaker, J. W. (1993). Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour Research and Therapy, 31*(6), 539–548.

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95*, 274–281.

Pennebaker, J. W., & Chung, C. K. (2007). Expressive writing, emotional upheavals, and health. In H. S. Friedman & R. C. Silver (Eds.), *Foundations of Health Psychology* (pp. 263–284). New York, NY: Oxford University Press.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion, 10*(6), 601–626.

Petty, R. E., & Mirels, H. L. (1981). Intimacy and scarcity of self-disclosure: Effects on interpersonal attraction for males and females. *Personality and Social Psychology Bulletin, 7*(3), 493–503.

Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior, 65*, 23–30.

Pickett, C. L., & Gardner, W. L. (2005). The social monitoring system: Enhanced sensitivity to social cues as an adaptive response to social exclusion. In K. D. Williams, J. P. Forgas, & W. von Hippel (Eds.), *Sydney Symposium of Social Psychology series. The social outcast: Ostracism, social exclusion, rejection, and bullying* (pp. 213–226). New York, NY: Psychology Press.

von der Putten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior, 26*(6), 1641–1650.

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. Cambridge, England: Cambridge University Press.

Reis, H. T., Lemay, E. P., & Finkenauer, C. (2017). Toward understanding understanding: The importance of feeling understood in relationships. *Social and Personality Psychology Compass, 11*(3), e12308.

Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process. In S. W. Duck (Ed.), *Handbook of personal relationships: Theory, research and interventions* (pp. 367–389). New York, NY: Wiley.

Romeo, N. (2016, December 25). The chatbot will see you now. *The New Yorker.* Retrieved from https://www.newyorker.com/tech/elements/the-chatbot-will-see-you-now

Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior, 33*, 201–212.

Schuetzler, R. M., Giboney J. S., Grimes G. M., and Nunamaker J. F. (2018, January). The influence of conversational agents on socially desirable responding. In *Proceedings of the 51st Hawaii International Conference on System Sciences.* (pp. 283–292). Manoa, HI: Hawaii International Conference on System Sciences (HICSS). http://hdl.handle.net/10125/49925

Shelton, J. N. (2003). Interpersonal concerns in social encounters between majority and minority group members. *Group Processes & Intergroup Relations, 6*(2), 171–185.

Shenk, C. E., & Fruzzetti, A. E. (2011). The impact of validating and invalidating responses on emotional reactivity. *Journal of Social and Clinical Psychology, 30*(2), 163–183.

Sherman, D. A., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin, 26*(9), 1046–1058.

Smyth, J. M. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Clinical Psychology, 66*(1), 174–184.

Sprecher, S., Treger, S., & Wondra, J. D. (2013). Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships, 30*(4), 497–514.

Sprecher, S., Treger, S., Wondra, J. D., Hilaire, N., & Wallpe, K. (2013). Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology, 49*(5), 860–866.

Tam, T., Hewstone, M., Harwood, J., Voci, A., & Kenworthy, J. (2006). Intergroup contact and grandparent–grandchild communication: The effects of self-disclosure on implicit and explicit biases against older people. *Group Processes & Intergroup Relations, 9*(3), 413–429.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29* (1), 24–54.

Utami, D., Bickmore, T., Nikolopoulou, A., & Paasche-Orlow, M. (2017). Talk about death: End of life planning with a virtual agent. In J. Beskwo, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, & S. Kopp (Eds.), *International Conference on Intelligent Virtual Agents: Lecture Notes in Computer Science*, Vol. 10498 (pp. 441–450). Cham, Switzerland: Springer.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070.

Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*. Retrieved from https://arxiv.org/pdf/1801.07243.pdf