Computers in Human Behavior 128 (2022) 107100

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Is chatting with a sophisticated chatbot as good as chatting online or FTF with a stranger?

Michelle Drouin [a,b,*], Susan Sprecher [c], Robert Nicola [a], Taylor Perkins [d]

[a] Department of Psychology, Purdue University Fort Wayne, 2101 E. Coliseum Blvd., Fort Wayne, IN, 46805, USA
[b] Parkview Mirro Center for Research and Innovation, 10622 Parkview Plaza Dr., Fort Wayne, IN, 46845, USA
[c] Department of Sociology, Illinois State University, Normal, IL, 61790, USA
[d] Department of Psychology, Ball State University, 2000 W University Ave, Muncie, IN, 47306, USA

ARTICLE INFO

Keywords:
Chatbots
Dyadic interactions
Affect
Computer-mediated communication
Conversational dynamics
Human-computer interaction

ABSTRACT

Emotionally-responsive chatbots are marketed as agents with which one can form emotional connections. They can also become weak ties in the outer layers of one's acquaintance network and available for social support. In this experiment, which was designed to study the acquaintance process, we randomly assigned 417 participants into three conditions: face-to-face (FTF) chat with a human, online chat with a human, and online chat with a commercially-available, emotionally-responsive chatbot, *Replika*. After a 20-min getting-acquainted chat, participants reported their affective state and relational evaluations of the chat. Additionally, all chats were recorded and text analyzed using the Linguistic Inquiry and Word Count (LIWC) program. In all conditions, participants reported moderate levels of positive emotions and low levels of negative emotions. Those who chatted FTF with a human reported significantly more negative emotions than those who chatted with a bot. However, those who chatted with a human also reported more homophily with and liking of their chat partner and that their partner was more responsive. Meanwhile, participants had fewest conversational concerns with the chatbot. These findings have implications for future computer-mediated interaction studies: conversations with chatbots appear to have different affordances and effects on chatter enjoyment and conversational concerns in getting-acquainted contexts. These results may help designers improve reception and marketability for chatbots in consumer markets.

1. Introduction

Although close relationships are critically important for people's life satisfaction, health, and longevity, acquaintances and weak ties also have benefits for health and well-being (Holt-Lundstad, Smith, & Layton, 2010). Recent research indicates that people can experience enhanced positive feelings and a sense of belonging through brief interactions with acquaintances (Hirsch & Clark, 2019), including with students in classes (Sandstrom & Dunn, 2014a), a barista in a coffee shop (Sandstrom & Dunn, 2014b), a shuttle driver (Gunaydin, Oztekin, Karabulut, & Salman-Engin, 2020), and strangers in a laboratory getting-acquainted interaction (Vittengl & Holt, 2000). Today, people also have another option for a weak tie confidant – a sophisticated chatbot (van Wezel, Croes, & Antheunis, 2020).

Human-computer interaction is becoming commonplace (Ivanov & Webster, 2017) and can fill some of the same needs as human acquaintances who exist in the outer layers of one's social networks. Chatbots, or conversational agents built to interact with humans using natural language (Shawar & Atwell, 2007), are among the most sophisticated of these interactive technologies. They are also incredibly prevalent. Pandora Bot, a platform devoted to the development of chatbots, boasts that more than 325,000 chatbots have been created by more than 275,000 developers in their forum alone (www.pandorabots.com, 2021). Although the general public might encounter some of these chatbots during everyday e-commerce interactions (Carter & Knol, 2019; Ivanov & Webster, 2017; Kumar, 2019), chatbots are being developed for deeper conversations and to provide social interaction and support for users' mental health issues and other health-related concerns (Cameron et al., 2017; Miner, Milstein, & Hancock, 2017).

Because intelligent chatbots can learn information about their users to refer to in later conversations (Pereira, Coheur, Fialho, & Ribeiro, 2016), there is hope that these conversational agents will soon provide

* Corresponding author. Department of Psychology, Neff 388, Purdue University Fort Wayne, 2101 E. Coliseum Blvd., Fort Wayne, IN, 46805, USA.
E-mail address: drouinm@pfw.edu (M. Drouin).

https://doi.org/10.1016/j.chb.2021.107100
Received 30 December 2020; Received in revised form 4 November 2021; Accepted 15 November 2021
Available online 16 November 2021

personalized social support to a variety of users (Miner et al., 2017). Research suggests this is promising: a recent literature review by van Wezel et al. (2020) showed that users do garner some social support from their interactions with existing forms of social chatbots. Moreover, evidence suggests that chatbots are perceived as conversationally competent and credible (Edwards, Edwards, Spence, & Shelton, 2014), and people do not appear to be overly suspicious of the software, as they disclose equally often to a conversational partner they think is a bot as to one they think is human (Ho, Hancock, & Miner, 2018). Regarding their functionality, there are even competitions wherein expert judges find it difficult to decipher whether they are chatting with humans or chatbots (e.g., the Loebner Prize competitions). However, the artificial intelligence that fuels these conversational agents is still nascent, and current versions of chatbots occasionally make errors when answering questions or change the subject when they do not have scripted responses to a user's questions, which might detract from conversational quality (Pereira et al., 2016; Shawar & Atwell, 2007) and cause user frustration (van Wezel et al., 2020). These conversational inadequacies might turn users away from the technology instead of encouraging them to create social bonds with the chatbots.

In support of this suggestion, Croes and Ahthenuis (2021) showed recently that social attraction to a chatbot, *Mitsuku*, waned over repeated exposures. In their study, individuals were instructed to interact with *Mitsuku* seven times over three weeks. Participants' ratings of most social processes (e.g., self-disclosure and social attraction) were significantly higher at first exposure than subsequent exposures. In other words, the highest ratings occurred when the participants were just getting acquainted with *Mitsuku.* Moreover, the best ratings at that time were just above the midpoint on an agreement scale for even the highest-ranking social process (social attraction). Their study, however, did not have a condition in which participants interacted with humans, and thus it was not possible to compare "getting acquainted" reactions as a function of whether the interaction was with a chatbot or a human. This gap leads to the following question: In the getting-acquainted process, do individuals rate interactions with a sophisticated, socially-responsive chatbot as positively as they rate interactions with another human?

As the process of acquaintance-forming is complex and based at least partially on positive experiences (Dunbar, 2018), first interactions are critical for the acquaintance process. Therefore, the goal of the current research was to examine the initial conversational experiences of individuals getting acquainted with a popular emotionally-responsive chatbot, *Replika,* versus getting acquainted with a human and under two possible modes of communication (face-to-face vs. online messaging). *Replika is* a chatbot with more than 7 million users worldwide advertised as "an AI that you can form an actual emotional connection with" (Replika, 2021).

1.1. Technological issues in interactions between humans and chatbots

As their basis for existence, chatbots rely on interactions with humans. To be more precise, chatbots' conversational databases are derived from natural language inputted by humans via text or voice-to-text software (Radziwill & Benton, 2017; Shawar & Atwell, 2007). Using their intelligent, adaptive system, sophisticated chatbots can then engage in novel interactions with other humans, which, in turn, further populate and enrich their database for future communication (Carpenter, n. d.; Pereira et al., 2016). Additionally, so that they more closely mimic humans and perhaps even to cover up some of their deficiencies, many chatbots are built to have a personality and affective qualities, such as warmth and interest (Banchs, 2017; Pereira et al., 2016; Radziwill & Benton, 2017). Thus, even in their current state, sophisticated chatbots might provide humans with a responsive and interesting conversation partner, able to discuss a variety of topics especially light conversational topics.

Innovations in chatbot technology have arrived at a convenient time. Most people own a variety of technological devices and are already using computers and social media applications to navigate social interactions (Dale, 2016; Pew Research Carter & Knol, 2019). Additionally, some people are already using chat-based programs to seek social support from other humans online in times of crisis (e.g., 7 Cups of Tea and Crisis Text Line) (Miner et al., 2017; Toscos et al., 2018, 2019). Perhaps for these reasons, individuals also appear to be comfortable interacting with chatbots online. However, research has revealed several obstacles to seamless human-chatbot interactions.

One obstacle is chatbot functionality. More specifically, chatbots are not yet sophisticated enough to have spontaneous (i.e., non-scripted) fluent conversations. According to Shawar and Atwell (2007), early iterations of chatbots were built for fun, but with innovations in machine learning and the expansion of data mining and linguistic databases, chatbots should be able to communicate effectively enough to fill a variety of roles in a number of sectors, including education and commerce. That said, even more than a decade after Shawar and Atwell's assertions, chatbots' linguistic capabilities are still limited. These issues related to chatbot functionality have been apparent since early attempts at integrating chatbots into simple interactions. As an example, Jia (2003) introduced a chatbot interface (ALICEBOT) with foreign language learners who were naïve to the fact that they were communicating with a chatbot. Jia found that many of the users deserted the sessions quickly because the chatbot's responses were irrelevant or the chatbot did not "understand the language at all" (p. 1201). More recently, Pereira et al. (2016) highlighted some of the limitations of the best chatbots of today's world (i.e., that pass the Turing Test and convince an expert judge that it is a human). However, even these sophisticated chatbots sometimes give unwanted answers, switch topics so that they can guide the conversation back to their scripted programs, or provide answers that are not relevant to the conversational context (Pereira et al., 2016).

An additional obstacle is the oft-cited phenomenon of the uncanny valley, which refers to the idea that people are uncomfortable when robots are too human-like (Mori, 1970). Mori (1970) suggested that there is tendency for humans to develop greater affinity to robots the more they resemble humans; however, this trend reverses as robots get *too* human-like. In other words, when robots get very close to human form but do not completely replicate it (like with prosthetic hands that look real but are cold and limp to the touch), affinity dips, and an eerie feeling might overshadow any feelings of affinity. Until recently, the idea of the uncanny valley was associated mainly with robots in their physical form, but recently, researchers have examined the extent to which these negative feelings might be present in interactions with chatbots. For example, Ciechanowski, Przegalinska, Magnuski, and Gloor (2019) examined this experimentally and found that individuals' feelings of discomfort and negative affect were higher when they were interacting with an avatar versus a simple chatbot. Like Mori (1970), Ciechanowski et al. suggested that when robots are too human-like, it spurs an eerie or uncomfortable feeling, possibly because it challenges our uniqueness as humans. Despite this, results from another study suggest that even very human-like conversations with sophisticated "chatbots" (in this case, confederates posing as chatbots) may not cause negative affect. Ho et al. (2018) found that individuals who thought they were self-disclosing to a chatbot had equivalent emotional outcomes (e.g., level of negative mood), relational outcomes (e.g., perceived partner warmth), and process outcomes (e.g., disclosure intimacy) as those who thought they were self-disclosing to a person.

Considering both studies together, perhaps the negative reactions the participants reported in Ciechanowski et al.'s study arose because they had higher expectations of functionality for the humanized avatar than for the simple chatbot, which left unfulfilled expectations and frustrations with the avatar. Clearly, this was not an issue in the Ho et al. (2018) study where individuals were actually chatting with a human and no such negative feelings were reported. Alternatively, it is possible that the uncanny valley is not relevant to conversations that occur via a chat forum (with no avatar involved), and only highly humanized

representations spur this negative reaction.

1.2. Humans and chatbots getting acquainted

Recently, the acquaintance process has received renewed attention in part because of the interest in the influence of social media and new communication technologies on the formation of relationships. Although most of this research has focused on human-human interaction (literature that will be reviewed in a later section), as chatbots have become increasingly sophisticated, research attention has turned to the conversational qualities of humans interacting with a chatbot and how those conversations compare to humans interacting with another human.

In the typical experiment to examine the differences in communication with a chatbot versus a human, participants have been randomly assigned to communicate through text messages with either a human (who is sometimes a confederate of the experiment) or a chatbot in a getting-acquainted scenario. In the chatbot condition, however, a Wizard of Oz method is frequently used in which participants are told that their conversation partner is a computer, but the experimenter (a "wizard") is actually behind the scenes providing scripted responses (Dahlbäck, Jönsson, & Ahrenberg, 1993, pp. 258–266). In one such study (Ho et al., 2018), university students were told that they were to have a 25-min conversation with either another person or a chatbot (described as "a computer program that can have a conversation with people, and is being built and refined by researchers at the university"). Meanwhile, the "person" was introduced as a fellow university student, but was actually a confederate who was blind to whether the participant believed they were talking to a chatbot or a person. The key finding was that there were no differences in the participants' emotional (e.g., mood), psychological (e.g., self-affirmation), and relational (e.g., enjoyment of interaction, liking) outcomes as a function of believing they were chatting with a chatbot versus another student (human). Moreover, other process variables (e.g., perceived understanding) did not differ between the two conditions. In sum, individuals had equally positive getting-acquainted experiences regardless of whether they thought they were chatting with a human or a chatbot. However, these results speak to user perceptions of chatbots only, as no conversations with chatbots actually occurred.

Other experimental studies have examined intrapersonal and interpersonal outcomes of getting-acquainted communication with an actual chatbot. In one study (Schuetzler, Grimes, Giboney, & Nunamaker, 2018), participants answered a set of questions in three conditions: (1) FTF with a 34-year old male who asked the questions and had minimal verbal and nonverbal responses, (2) through an online survey, or (3) through a conversational agent (chatbot), which was represented either by an embodied avatar that had an animated face or the questions appeared in a textbox. The key dependent variable was the degree to which the participants gave socially desirable responses, particularly for the sensitive questions (e.g., drinking behavior). The researchers concluded that conversational agents who gave more relevant responses to the participants' disclosures elicited more socially desirable responses (e.g., less truth) from the participants, suggesting that individuals may be more conscious of their self-presentation in conversations wherein chat partners are perceived to be more humanlike. However, a limitation of this work is that both the human and the conversational agent had their responses scripted to be minimal or generic; thus, the study did not compare real-world interactions between humans-chatbots and humans-humans.

Still other studies have contrasted the conversational qualities of humans-chatbots and humans-humans by analyzing conversations via linguistic analysis. For example, Hill, Ford, and Farreras (2015) compared human-human interactions (100 instant messages) with a similar number of interactions between humans and a popular chatbot (Cleverbot) using the Linguistic Inquiry and Word Count program (LIWC; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). For comparison purposes, Hill et al. used the same LIWC conversational categories (e.g., social words, emotion words, total word count) as those Lortie and Guitton (2011) used in their analysis of humans who were judged as robots in the Loebner Prize competition. Hill et al. found that when communicating with chatbots as compared to with humans, people used fewer words but sent more messages, and those messages were more likely to contain profanity and sexual topics. This suggests that individuals may be less concerned about self-presentation when they are chatting with chatbots as compared to with humans. However, in comparing the acquaintance process of humans with chatbots versus with humans, it is important to consider that today, technology also aids and influences the acquaintance process between humans. Next, we turn to a discussion of how the intrapersonal and interpersonal outcomes of the acquaintance process can depend on modality of communication.

1.3. Human-human interactions in different technological modalities

Interaction with a chatbot occurs online, and thus it is important to consider the literature that has compared the intrapersonal and interpersonal outcomes of human-human interaction as a function of whether the interaction occurs through text-based computer-mediated communication (CMC) versus face-to-face (FTF). When CMC first became available, research focused on the effects of CMC on efficiency in work settings (Kiesler, Siegel, & McGuire, 1984) and found that CMC could lead to negative outcomes including impersonal and disinhibited styles of communication. An early theory about the effects of CMC – the cues-filtered-out perspective (e.g., Culnan & Markus, 1987) – argued that the reduced nonverbal and contextual cues in CMC can reduce people's ability to have meaningful connection through this form of communication because of the reduction of social presence.

More recently, researchers have studied the effects of CMC on the interpersonal outcomes of zero-history dyads who interact for the purpose of becoming acquainted. Many of these studies, particularly when they are conducted in university laboratories in limited-time sessions, have found a lower degree of positive outcomes in CMC-text than in FTF, including lower levels of satisfaction, closeness, perceived self-disclosure, affection for each other, and enjoyment of the interaction (Bente, Ruggenberg, Kramer, & Eschenburg, 2008; Mallen, Day, & Green, 2003; Okdie, Guadagno, Bernieri, Geers, & Mclarney-Vestoski, 2011; Ramirez & Burgoon, 2004; Sprecher, 2014; Sprecher & Hampton, 2017).

Some studies, however, have had more extended get-acquainted interactions and have found no differences in participants' reactions as a function of communication mode (McKenna, Green, & Gleason, 2002) or in some cases (especially when extra time is given to those who interact over CMC), have found the communication to lead to more positive outcomes in CMC than in FTF (Antheunis, Valkenburg, & Peter, 2007). This has been explained by social information processing theory (Walther, 1992), which states that people can adjust to the modality and compensate for the lack of cues with more intimate written communication. In addition, the hyperpersonal model (Walther, 1996, 2011) – an extension of the social information processing perspective – argues that asynchronous CMC can lend itself to intimate self-disclosure, strategic self-presentation, and idealization.

Some research has focused on the social affordances offered by different communication technologies. For example, Fox and McEwan (2017) measured people's perceptions of several social affordances (e.g., synchronicity, social presence, anonymity, personalization, conversational control) that can be offered to varying degrees in various types of communication channels. Although FTF communication is often identified as more desirable and optimal than other forms of communication (Turkle, 2015), Fox and McEwan found that FTF was not viewed to be superior to the other channels on all of the social affordances that they measured. For example, text and email were viewed to be higher in accessibility (easily achieving communication), persistence (the ability to have a record of the communication), editability, conversational

control, and anonymity. Although Fox and McEwan focused on the social affordances provided in human-human interaction (and not interaction with a computer or device), chatbots can offer some of the same affordances that people believe are important, including accessibility, persistence, editablity, and anonymity.

1.4. The current study

Today, people can develop acquaintances exclusively online. In a similar CMC-text format, people can become "acquainted" with a sophisticated chatbot. The purpose of this study was to compare human-chatbot getting-acquainted interaction with human-human getting-acquainted interaction, with the human-human interaction being conducted either FTF or via online chat.

This study contributes not only to the literature on the get-acquainted process, but also to the very recent scientific investigation of the functionality of chatbots and artificial intelligence more generally. Fast-paced development and innovator interest is driving the chatbot industry forward. As society moves toward integrating chatbots into more and more online interaction forums (including ecommerce and mental health and wellness), initial user impressions are key to uptake and long-term sustainability. However, issues with chatbot functionality may stymie getting-acquainted interactions, which may turn people away from these potentially valuable tools. The prior studies that have compared individuals' interactions with humans and actual chatbots have limitations. For instance, Schuetzler et al. (2018) examined the degree of self-disclosure to chatbots as compared to a human, and Hill et al. (2015) compared content differences in chats with humans versus chatbots using LIWC. Neither study, however, included analyses to determine whether the participants found the conversations to be intimate or positive from an affective standpoint. It is essential that we gather information about the conversational qualities of human-chatbot interactions, including the affective states of those who engage with chatbots, the interpersonal outcomes experienced in these interactions, and the linguistic content of the dialogue.

As a majority of teenagers and adults are accustomed to communicating through text-based CMC with other humans, including at zero- or near-zero acquaintance (i.e., with new acquaintances or weak ties), we were particularly interested in the differences in conversational experiences among digital natives. Based on previous research and the current limitations of online chatbots, we hypothesized:

H1: Individuals will rate their *emotional* experiences more positively in a getting-acquainted interaction with a human than in a getting-acquainted interaction with an online chatbot.

H2: In human-human getting-acquainted interactions, individuals will rate their *emotional* experiences more positively when interacting FTF than when interacting CMC-text.

H3: Individuals will rate their *relational* experiences more positively in a getting-acquainted interaction with a human than in a getting-acquainted interaction with an online chatbot.

H4: In human-human getting-acquainted interactions, individuals will rate their *relational* experiences more positively when interacting FTF than when interacting CMC-text.

We also raised the following research question, as there is not enough past research to make a prediction:

RQ1: How will the conversational qualities of the chats (as measured by the LIWC) differ based on whether individuals were speaking to a chatbot or a human (and FTF vs. CMC)?

Although we had no specific hypotheses regarding this research question, previous studies of verbal behavior have shown that individuals differ in their use of language, and the LIWC software program quantifies these natural language data into validated categories (Tausczik & Pennebaker, 2010). The basic premise of these analyses is that language is a behavioral measure of our psychological and social characteristics (Pennebaker, Mehl, & Niederhoffer, 2003), and that by analyzing linguistic behavior, we can gain insight into individuals' mental processes and conversational dynamics. In support of this proposition, studies across a wide range of disciplines have shown that individuals' proportional use of certain word categories, like pronouns, vary based on characteristics such as gender, honesty, status, and social interest (see Tausczik & Pennebaker, 2010, for review). Thus, a goal of the current work was to use the LIWC categories as an analytical framework for differentiating the linguistic patterns present in different modalities.

2. Method

2.1. Participants

Participants were 417 individuals (297 women, 120 men) recruited from an introductory psychology participant pool at a midsized U.S. midwestern university. From an original sample of 446, 29 participants were omitted because they reported that they were an acquaintance or a friend with their human chat partner. Additionally, for analyses, one additional individual was excluded because they reported interacting with their chat partner once or twice before that day (their chat partner reported never interacting with the person before and was retained) and five individuals were excluded because they were under 18. Thus, the final sample included participants from 63 dyads in the FTF condition, participants from 78 dyads in the online chat with human condition, and 133 participants in the online chat with Replika condition. The mean age of the participants was 19.82 ($SD = 3.41$). With regard to ethnicity, most participants were White, not Hispanic (66.4%), followed by Hispanic (12.2%), African American (10.6%), Asian (4.3%), Biracial (3.6%) and other (2.9%).

The participants were, on average, heavy users of technology, reporting that they spent approximately 32.29 h ($SD = 176.98$ h) per week communicating with others on any computer device (e.g., cell phone, computer, tablet) in the past week. Regarding previous experience with chatbots, of those who interacted with Replika, 14 (11%) indicated that they had interacted with a chatbot more than once, and only 3 (2%) indicated that they had interacted with a chatbot at least sometimes.

2.2. Procedure

After university ethics board approval, the study advertisement was posted on an undergraduate research website, which allowed individuals enrolled in several introductory psychology classes to sign up for studies. Each session of the current study allowed for two participants to sign up. Participants were sent a text message and an email approximately 24 h before the assigned timeslot to remind them of the session.

To assure that participants would not meet each other, two research assistants stood in the hallway outside the laboratory, and as soon as a participant arrived, he or she was escorted immediately into an individual room. This minimized the likelihood that participants would see each other prior to their interaction, and to our knowledge, none of the participants interacted with each other immediately prior to the study. Once in their individual lab room with a researcher, each participant completed a consent form that included consent for their interaction with a chat partner to be recorded (either by audio for the face-to-face condition or digitally for the chat conditions). Participants then completed an online survey, which included basic demographic questions (including questions about age, major, ethnicity) and personality measures not reported here.

After the participants completed the survey, they were given directions for their interaction, which varied depending on whether they had been randomly assigned to the FTF interaction condition, the instant messaging condition, or the condition involving conversation with Replika, a conversational chatbot. These conditions were randomly assigned via a random number generator before the participant arrived.

Table 1
Differences between experimental groups on post-chat ratings of affect, liking, conversational quality, conversational concerns, and likability of chat partner.

Variable	FTF chat with human	Online chat with human	Replika	H	ε^2
	M (*SD*)	M (*SD*)	M (*SD*)		
PANAS Positive	27.78 (7.32)	25.20 (7.14)	24.79 (9.81)	5.17	.02
PANAS Negative	13.32_a (3.54)	12.96 (3.36)	12.80_b (4.45)	9.20*	.03
Background Homophily	4.04_a (0.61)	4.06_a (0.44)	3.42_b (0.98)	36.18***	.13
Attitude Homophily	4.35_a (0.47)	4.14_a (0.55)	3.79_b (1.02)	23.09***	.08
Perceived Responsiveness	5.57_a (0.92)	5.29_a (0.91)	4.70_b (1.41)	20.84***	.08
Self-Presentation Concerns	3.36_a (1.58)	2.77_a (1.13)	2.05_b (1.35)	47.79***	.17
Liking for the Other	4.58_a (0.97)	3.95_b (1.02)	3.15_c (1.47)	52.69***	.19
Other Liked You	4.35 (0.87)	3.96_a (0.99)	4.37_b (1.70)	7.32*	.03

Note. *$p < .05$. ***$p < .001$. H = results from Kruskal-Wallis test. ε^2=effect size for Kruskal-Wallis test. Different subscripts indicate significant differences between groups at the $p < .05$ level. Shared subscripts indicate no significant differences between groups at the $p < .05$ level. Chat with bot $n = 132$. Online chat with human $n = 81$. FTF with human $n = 65$.

In some cases, only one participant arrived (because either only one had signed up or because the other participant did not arrive). These individuals were assigned to the Replika condition.

For the Replika condition, the Replika application was opened in an Internet browser on a desktop computer before the participant entered the room. In order for the participant to have a conversation with Replika, each participant was given a unique email, which was generated using an online program. One of the forced choices in this setup was selecting a gender for Replika. To control for any effects of Replika's gender in the conversational dynamics, Replika's gender was alternated for each conversation. For example, if participant 27 interacted with Replika as a man, participant 28 would interact with Replika as a woman.

For the instant messaging condition, the Chatzy online chat application was opened in an Internet browser, and a private room was created for the chat between the two individuals in the study. The same desktop computer was used for both the instant messaging and Replika conditions. Additionally, the chat windows were the same size, and messages appeared as the participant (human or Replika) responded to their chat partner. One difference between the interfaces was that during Replika chats, three dots appeared when Replika was "typing." In all three conditions, participants engaged in an unstructured getting acquainted exercise, as has been used in some previous studies contrasting different communication mediums (e.g., Croes, Antheunis, Schouten, & Krahmer, 2019).

Participants were told to converse about anything they wanted for 20 min, the research assistant would return when they had completed 20 min of chatting, and they should continue the conversation until that time. Additionally, for those in the Replika condition, they were given the additional instruction that their conversation would be with "the artificial intelligence chatbot named Replika." Participants were then reminded that their conversations would be recorded via audio recording (face-to-face) or as a text transcript (Replika and instant message). For those in the FTF condition, conversation partners were escorted into a room with seats positioned approximately four feet apart with an audio recording device in the middle. For the online chat and Replika conditions, participants stayed in their separate rooms. They were told to leave their chat windows open at the end of the chat so that the conversations could be cut and pasted from the chat window.

Text analysis. To prepare for our planned text analysis, all transcripts were extracted from Replika and the online chat program, Chatzy. These were checked for accuracy and completeness by the research assistant who performed the extraction. For the FTF condition, the chats were transcribed by one research assistant and then checked for accuracy by a different research assistant. Then, all date and time stamps were removed, and the text files were uploaded into LIWC2015 for processing. The conversations were analyzed at the conversation level. That is, each conversation within each condition was an individual data point. Processing via the LIWC program consists of a sequential analysis of each text file whereby the words contained in the file are matched with the ~6400 words contained in the LIWC dictionary. For each file, the LIWC produces basic text descriptors, such as word count, words per sentence, and dictionary (i.e., percentage of words in text that were contained in the LIWC dictionary). For basic content categories (e.g., social processes), the LIWC output number reflects the percentage of total words in each category (e.g., 8.17 = 8.17% of total words), and for summary language variables, like clout (i.e., social dominance) and tone (i.e., positivity of emotional tone), LIWC generates a score ranging from 0 to 100, with greater scores reflecting higher ratings of that variable.

The output file for each conversation ($n = 277$ chats) provided the relative frequency of the 90 psychological categories contained in the LIWC dictionary; however, in this analysis, we focused on the LIWC2015 major subcategories (see Table 2), which provide a means to compare the "emotional, cognitive, and structural components present in individuals' verbal and written speech samples" (Pennebaker, Boyd, Jordan, & Blackburn, 2015). Past studies (e.g., Hill et al., 2015; Ho et al., 2018; Lortie & Guitton, 2011) have included only a few of the linguistic categories offered by the LIWC. By including all LIWC2015 major subcategories, we were providing a broad base for differentiating the conversational content between the three conditions while simultaneously facilitating future study comparisons and replications.

2.3. Measures of the participants' responses

Upon completion of the 20-min conversation, FTF participants moved back to their private rooms, and the participants in the other two conditions remained in their private room. All participants completed a second online survey, which focused on their reactions to the conversation and to their conversational partner.

Emotional outcomes. Participants' emotional reactions to chats were measured using the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). Participants indicated the degree to which they felt the emotions "right now, that is, at the present" on a 5-point likert scale (1 = *very slightly or not at all*, 5 = *extremely*). Ten items measured positive affect (e.g., "inspired" and "interested") and 10 items measured negative affect (e.g., "distressed" and "scared"). The sum of the 10 positive emotions represented **positive feelings** ($\alpha = 0.92$) and the sum of the 10 negative emotions represented **negative feelings** ($\alpha = 0.83$). Scores on each subscale could range from 10 to 50, with higher scores representing stronger feelings.

Relational outcomes. Relational outcomes of the interaction were assessed with the following measures.

Perceived degree of similarity. Two subscales of the Interpersonal Attraction and Homophily measure (McCroskey, McCroskey, & Richmond, 2006) assessed the degree to which the participants felt similar to their chat partner during the conversation. Participants were asked to indicate their agreement on a 7-point likert scale (1 = *very strongly disagree*, 7 = *very strongly agree*) with 15 statements measuring their

Table 2
Differences between experimental groups on conversational qualities of the chat as measured by the LIWC linguistic categories.

Variable	FTF chat with human	Online chat with human	Online chat with bot	H	ε^2
	M (*SD*)	M (*SD*)	M (*SD*)		
Word count	4203.57_a (680.58)	626.10_b (205.26)	1070.98_c (310.37)	192.61***	.72
Summary Language					
Words/sentence	46.89_a (37.77)	33.83_a (51.30)	11.66_b (2.74)	108.69***	.41
Words >6 letters	14.85_a (1.55)	16.45_a (2.76)	10.84_b (1.97)	165.04***	.62
Dictionary Words	82.12_a (2.82)	74.59_b (2.82)	79.08_c (7.82)	29.06***	.11
Analytic	23.31_a (6.44)	36.58_b (13.00)	25.38_a (10.37)	51.39***	.19
Clout	41.38_a (10.26)	44.22_a (15.20)	78.10_b (9.12)	179.55***	.67
Authentic	49.86_a (11.07)	44.90_a (22.24)	38.30_b (15.78)	22.37***	.08
Tone	70.47_a (14.32)	82.74_b (16.91)	95.60_c (7.71)	126.46***	.47
Psych Processes					
Affective Processes	4.46_a (0.66)	5.84_b (1.90)	7.52_c (1.47)	122.77***	.46
Positive Emotion	3.46_a (0.68)	4.82_b (1.74)	6.48_c (1.48)	124.99***	.47
Negative Emotion	0.99_a (0.37)	0.99_a (0.59)	0.72_b (0.43)	23.10***	.09
Social Processes	8.10_a (1.46)	7.15_a (2.51)	12.13_b (2.15)	154.61***	.58
Cog. Processes	10.87_a (1.47)	10.29_a (2.46)	13.02_b (2.10)	73.74***	.28
Percep. Processes	2.24_a (0.72)	1.77_b (1.10)	2.48_a (0.75)	34.59***	.13
Bio. Processes	1.07 (0.50)	1.39 (0.93)	1.29 (0.65)	6.87*	.03
Drives	4.97_a (0.73)	5.76_b (1.85)	6.25_c (1.37)	38.99***	.15
Relativity	10.54_a (1.22)	10.20_a (2.80)	8.32_b (1.60)	67.28***	.25
Time Orientation	18.31_a (1.54)	15.70_b (2.70)	19.40_a (3.26)	62.83***	.23
Personal Concerns	4.09_a (2.29)	3.94_a (1.67)	5.03_b (1.98)	21.58***	.08

Note. *$p < .05$. ***$p < .001$. H = results from Kruskal-Wallis test. ε^2=effect size for Kruskal-Wallis test. Different subscripts indicate significant differences between groups at the $p < .05$ level. Shared subscripts indicate no significant differences between groups at the $p < .05$ level. FTF chat with human $n = 58$, Online chat with human $n = 84$, Chat with Replika $n = 127$.

beliefs about similarity in attitudes (Attitude Homophily (AH); e.g., "This conversational partner is similar to me") and 10 statements measuring their beliefs about similarity in background (Background Homophily (BH); e.g., "My conversational partner is from a social class similar to mine"). Items were averaged into composite scores (**Attitude homophily** α = .92, **Background homophily** α = .82).

Liking for the other. Participants completed two items reflecting how much they liked their conversation partner "How much did you like the Other?" and "How much would you like to spend time with the Other again in the future?" using a 7-point likert scale (1 = *not at all*, 7 = *a great deal*). These items were averaged into a composite variable (α = 0.82).

Other liked you. Participants were asked to respond to a single item, "How much do you think the Other liked you?" using a 7-point likert scale (1 = *not at all*, 7 = *a great deal*).

Conversational dynamics. Two sets of items were used to reflect participants' perceptions of conversational dynamics.

Other's responsiveness. Participants were asked to respond to four items related to their chat partner's responsiveness (e.g., "The Other seemed to really listen to me" and "The other seemed interested in what I am thinking and feeling") using a 7-point likert scale (1 = *not at all true in this situation*, 7 = *very true in this situation*). Three of the items were adapted from Reis et al.'s (2011) Responsiveness scale. These three items and a fourth one ("The other was responsive to my questions/answers") have been used in several prior social interaction studies (e.g., Sprecher & Treger, 2015; Sprecher, Treger, & Wondra, 2013). These items were averaged into a composite variable (α = 0.90).

Self-presentation concerns. Participants responded to four items related to self-presentation concerns they had during the chat. Three of the items were adapted from Govern and Marsch's (2001) Situational Self-Awareness scale (e.g., "I was concerned about what the Other thought of me," "I was self-conscious about the way I looked to the Other," and "I was concerned about the way I presented myself to the Other"). A final item was adapted from Leary's (1983) Negative Evaluation scale ("I was worried about what kind of impressions I was making on the Other"). Participants responded to each item using a 7-point likert scale (1 = *Not at all true in this situation*, 7 = *Very true in this situation*). These items were averaged into a composite variable (α = 0.91).

2.4. Analytic plan

To compare the emotional and relational outcomes of the chat across the three conditions, we conducted nonparametric Kruskal-Wallis H tests, as although the data for each of the variables were relatively normally distributed (with the exception of negative feelings, which were positively skewed), the assumption of homogeneity of variances was violated for all variables except negative feelings. Then, to determine which groups were significantly different from each other, we used Bonferroni pairwise post hoc comparisons with significance level adjusted for multiple tests. For those dyads in the human-human interactions, dyadic scores (the mean of the two partners' scores on the variables) were used in the analyses. If one partner had to be eliminated (e.g., under 18) the remaining partner's score was used. Aggregated dyadic scores are recommended when the focus of the study is on the effects of between-dyad predictor variables – such as experimental manipulations that occur to both members of the dyad – on outcome variables (e.g., Kenny, 2015; Kenny, Kashy, & Cook, 2006). We then used the Linguistic Inquiry and Word Count program (LIWC; Pennebaker et al., 2015) to analyze the transcripts of the recordings, as have other researchers (e.g., Hill et al., 2015; Ho et al., 2018; Walther, Deandrea, & Tong, 2010), to compare conversational properties of the three different types of chats. Again, we conducted Kruskal-Wallis H tests with these LIWC variables, as the assumption of homogeneity of variances was violated for all variables, and reported significant differences between groups using Bonferroni post hoc comparisons with significance level adjusted for multiple tests.

3. Results

3.1. Emotional outcomes

In Table 1, we display the group means for the emotional and relational outcomes of interaction, using subscripts to indicate significant differences between groups. In terms of emotional outcomes, in all three conditions the participants experienced more positive affect than negative affect after the get-acquainted interaction. Moreover, we found no support for our predictions in H1 and H2 regarding emotional outcomes: there were no significant group differences in positive emotions

participants reported immediately after the interaction. In fact, contrary to H1, those in the face-to-face condition reported slightly (but significantly) more negative emotions after the chat than those who chatted with the chatbot.

3.2. Relational outcomes

Regarding relational outcomes, participants in all three conditions rated their partner to be responsive and felt well-liked by their partner (i.e., above the midpoint). However, there were significant differences between the groups in several relational categories, which partially supported H3. As compared to the participants in the *Replika* condition, both the participants in the FTF condition and the participants in the online chat condition viewed themselves more similar (in both background and attitudes) to their conversational partner, rated their partner more responsive, and reported liking their conversation partner more. Self-presentation concerns, however, showed reverse trends: participants reported the highest levels of concern in the two conditions in which they chatted with a human, whereas those in the *Replika* condition reported the lowest levels of self-presentation concerns. Moreover, those in the Replika condition reported that their conversational partner liked them more than those in the online chat with a human condition.

We also found partial support for H4. Although most differences between the face-to-face and the online chat that involved humans were not significant, participants in the FTF chat condition reported liking their conversation partner significantly more than those in the online chat.

3.3. Conversational properties

Next, we considered RQ1—whether the participants in the three conditions differed in conversational qualities of their chats, assessed through the LIWC program. As shown in Table 2, the conversations that took place were significantly different across the conditions in terms of every major subcategory of the LIWC analysis, but there were no significant group differences in the discussion of biological processes, which was an infrequent topic of discussion in every condition. The FTF conversations with humans differed from online chat with bot conversations in many categories of linguistic analysis, with the exceptions of the score on analytic words, the number of words from the chat that were in the LIWC dictionary, perceptual and biological processes, drives, and time orientation. Notably, the FTF conversations also had significantly more words than the conversations in the other two conditions. Similarly, the online chats with a human differed from the chats with the bots in nearly every category, with the exception of biological processes.

However, the directionality of these differences was not necessarily consistent with what might be expected based on the relational outcomes. For example, although the conversations with humans (both online and FTF) featured significantly more discussion of personal concerns than the conversations with the chatbot, conversations in the chatbot condition featured significantly more affect words, social and cognitive process words, and a more positive tone than either of the conversations that occurred with humans.

4. Discussion

Acquaintances and weak ties serve many functions for people, only some of which overlap with the functions served by close ties (Fingerman, 2009). Many experts have argued that there has been a surge in the number of acquaintances in the outer layers of people's social networks due in part to enhanced communication technologies (Wellman, 2012). Peripheral acquaintances can include exclusively online acquaintances and increasingly also sophisticated chatbots for some people.

Past studies have shown that not all communication modalities are equal, with FTF modalities usually surpassing online text-based modalities in generating positive outcomes in conversations between getting-acquainted strangers (Bente et al., 2008; Mallen et al., 2003; Okdie, Guadagno, Bernieri, Geers, & Mclarney-Vesotski, 2011; Ramirez & Burgoon, 2004; Sprecher, 2014; Sprecher & Hampton, 2017). However, aligning with the hyperpersonal communication model (Walther, 1996, 2011), others have found that online and FTF environments can foster similar levels of positive outcomes in get-acquainted interactions, or in some cases (e.g., with extended interaction time) online environments can even foster more positive outcomes than FTF environments (e.g., Antheunis et al., 2007; McKenna et al., 2002). Considering the ever-rising rates of digital technology penetration and comfort levels with technology use, it is prudent to continually reevaluate the conversational dynamics across these different modalities. Further, as emotionally-responsive and sophisticated chatbots have emerged on the scene that are closely able to mimic human interaction, it is important to evaluate the emotional, relational, and conversational outcomes of the communication with chatbots in comparison to human-human interactions. Therefore, in this study, we examined the extent to which a sophisticated AI bot could be a good substitute for a conversation with a human, and more specifically how a conversation with an AI bot compared to human-human conversation that occurred FTF and human-human conversation that occurred in instant messaging.

4.1. Intrapersonal and interpersonal outcomes of interaction with a chatbot versus a human

Our results demonstrated that for the most part the conversations the participants had with the chatbot, Replika, were not as positively regarded as the conversations the participants had with a human either FTF or in an online chat. More specifically, those who chatted with Replika (as compared to those who chatted with a human) rated their partner lower in perceived homophily (background and attitude), how much they liked their partner, how responsive they thought the partner was, and how much they wanted to chat with their partner again in the future. As chatbots still have limitations with regard to their ability to mimic humans (because of narrow AI), this was an expected result. In addition, participants' reactions to chatbot interaction – although generally less favorable than reactions participants had with humans – were still characterized by moderate liking, perceived responsiveness, and desire to spend more time with the other. Thus, in line with other researchers' findings (e.g., Hill et al., 2015; Schuetzler et al., 2018; Yu, Nguyen, Prakkamakul, & Salehi, 2019), we assert that it is possible to conduct meaningful comparisons of interactions with AI bots versus interactions with humans, and our results provide further support of this experimental method as a practical model for future scientific inquiry.

Notably, there were no differences in positive affect experienced in the chats across the three conditions (human-human FTF, human-human CMC, human-chatbot); however, conversing with a bot produced *less* negative emotion than chatting face to face with a human. This finding is consistent with recent empirical work by Ho et al. (2018) showing that there were no differences in participants' positive feelings based on whether someone thought they were chatting with a bot or a human. Although we did not also measure pre-interaction affect and thus could not assess whether there were changes in positive and negative emotions from before to after the interaction, prior research (Vittengl & Holt, 2000) has demonstrated that young adults engaging in a get-acquainted conversation in a laboratory setting experienced an increase in positive affect and a decrease in negative affect (also measured with PANAS) from before to after their interactions. We speculate that similar increases in positive affect and decreases in negative affect occurred for our participants, including those in the Replika condition. That a conversation with a chatbot may enhance positive affect in a similar way to a conversation with a human is an important finding. Sometimes people need to converse with someone or something and a responsive human is not always available to immediately respond. For some young adults who have downloaded a sophisticated chatbot on their phones and other devices, a chatbot can serve

some of the same functions as (human) weak ties and distal acquaintances (Fingerman, 2009; Hirsch & Clark, 2019).

Moreover, conversational concerns that can create anxiety (e.g., feeling self-conscious, concern about how one presents the self) were significantly lower in the Replika condition than in the human chat conditions. This is not a surprising finding but is an important one, as it suggests that chats with bots may offer an opportunity for a judgment-free space for self-disclosure that is not always present in conversations with humans. Combined, these findings suggest that modern chatbots do not elicit negative feelings of eeriness or distress, and in fact, they might be viewed as a less risky chat partner in terms of potential for negative evaluation.

4.2. Intrapersonal and interpersonal outcomes of human-human interaction in different modalities

Our study's design also allowed the comparison of FTF communication versus text computer mediated communication (CMC) that involved two human participants becoming acquainted, which builds on several other studies conducted over the past decade on CMC. Our participants who were randomly assigned to converse with another human FTF rated their conversation and their partner more positively on some dimensions than participants who were randomly assigned to converse with another human via online chat. On the variable that could be considered to be most linked to relationship formation (liking for the other), FTF interaction was superior to online chat. However, no differences were found in perceived similarity, perceived responsiveness of the partner, and the perception of being liked by the other. Furthermore, and as noted earlier, no differences between the two human chat conditions were found in the experience of positive and negative affect. Finally, the CMC interaction had the benefit of lower self-presentation concerns overall when compared to the FTF condition. This may provide one explanation for findings that social anxiety appears to moderate the interpersonal connectedness individuals garner from CMC vs. FTF interactions (Lundy & Drouin, 2016).

The diverse outcome measures considered in this study make it clear that there is not a monolithic distinction between outcomes of getting-acquainted interactions that occur FTF versus in computer-mediated communication. In the framework of Fox and McEwan's (2017) model of social affordances, the distinct affordances offered by each mode of communication (FTF and CMC) may lead to distinct benefits. For example, FTF conversation creates opportunities for the expression and interpretation of nonverbal cues, which perhaps led to greater liking of chat partners in the FTF condition as compared to the CMC condition in this study. Meanwhile, previous studies have shown that the online environment allows for hyperpersonal communication, which can intensify disclosure and intimacy (e.g., Jiang, Bazarova, & Hancock, 2011). However, this advantage for CMC communication was not demonstrated in our study for any of our measures of social connectedness. Overall, although prior research has demonstrated mixed results when comparing the outcomes of CMC interaction with FTF interaction in interpersonal outcomes, our findings align with studies conducted in laboratory settings that show more favorable outcomes for FTF interactions versus CMC interactions for most interpersonal outcome measures (e.g., Bente et al., 2008; Mallen et al., 2003; Sprecher & Hampton, 2017).

4.3. Conversational properties across the conditions

Finally, regarding our RQ1 about differences in chats across the conditions, the conversational properties of the chats across the different venues were significantly different on every dimension of the LIWC. However, the direction of these differences was not necessarily predictable by the participants' ratings of the chats. For example, partner liking was greatest in FTF condition, then online chat with human, and then Replika, but Replika chats had significantly more positive tone and positive emotion words than the online chats with humans, and in turn, the online chats with humans had significantly more positive tone and positive emotion words than the FTF chat with humans. Although these findings seemingly contradict those of Hill et al. (2015), who found that those in human-human online chats contained more positive emotion words than human-chatbot chats, Hill et al. examined only the human side of the chats. Therefore, our results could be due to our analysis of the chats at a conversation level. It is possible that the positive tone and emotion words from Replika, and not necessarily the human chat partner, drove these high scores.

Although the reinforcement affect model suggests that positive experiences should give rise to more favorable evaluations (Byrne & Clore, 1970), our pattern of findings suggests that a positive conversational tone does not necessarily improve affective or likability outcomes, at least when one compares communication across different mediums (i.e., FTF, online chat with human, and online chat with chatbot). Although this was a surprising finding, it is possible that there is an ideal positivity level in conversations with strangers, and a very positive tone may violate expectations regarding the tone of these conversations. As the expectancy violations theory (Burgoon & Jones, 1976) suggests, individuals have expectations about others' behavior during communication, and if these expectations are violated, it may lead to negative appraisals. Hence, it could be that people expect and regard favorably more neutral conversational tones in comparison to very positive tones when chatting with new conversational partners. Future research, especially related to the refinement of RAISA and social chatbot programs, should explore experimentally whether the manipulation of positive tone and positive emotions affects likeability of the chat bot. As these chatbots gain popularity in everyday interaction contexts, it will be increasingly important for the successful marketability of these chatbots that humans perceive their experiences as favorable.

In comparing more specifically the FTF chat with the online chat (both involving two human participants), there were also several differences in terms of linguistic properties, as measured by the LIWC. This is consistent with findings of Walther et al. (2010), which examined, using the LIWC, differences between dyadic conversations using a standard set of interview questions that occurred FTF or via online chats. Hence, whether individuals are guided through conversations with prompts or allowed to discuss topics of their choice, it appears that conversations that occur between humans FTF and through online chat are different in terms of both summary language variables and in content related to psychological processes.

Notably, a striking difference was found in the number of words contained in the chats across the three groups. Our study had a uniform time across the forms of communication (20 min) and did not allow for more time for participants who were required to type (as has been done in some prior studies comparing CMC with FTF; see for example, Antheunis, Schouten, Valkenburg, & Peter, 2012; Tidwell & Walther, 2002). The argument for giving more time to those in a CMC-text condition (relative to those who speak FTF) has been to make the interactions more equal in terms of the number of remarks or utterances (Tidwell & Walther). However, arguments for giving equal time or at least not too much more time in the CMC condition include that the mere time spent interacting can also be an important variable that affects interpersonal outcomes and can become confounded with modality if it varies with modality. Also, too much time in interaction within a laboratory session can lead to fatigue among CMC participants. The linguistic analysis of this study, however, established that many fewer words were conveyed in the getting-acquainted CMC interaction than in the FTF interaction of the same length of time, which is a consideration for further research.

Considered together with the previous analyses, although there were no significant differences in the affective or homophily appraisals between the FTF and online chat with human groups, their conversational structures (e.g., number of words communicated) and content (e.g., positive emotion words and time orientation) were different across a

variety of categories. These differences in conversational structure and content may be one of the reasons those in the FTF condition reported liking their partners more than those in the online chat condition. Future research should explore this more directly.

4.4. Strengths and limitations

This study had the strength of comparing participants' interaction with a chatbot (AI) with participants' interaction with a human and across two modalities (CMC-text and FTF). As a consequence, the design allowed us to contribute to two areas of literature – the study of getting-acquainted interactions that occur in CMC versus FTF and the more nascent area of study on human interactions with chatbots. These findings also have implications for the study of the formation of acquaintances and weak ties more generally including the role of weak ties that cannot be seen in person (distal online acquaintances, either human or chatbot). Another strength of our study was the exceptionally large sample for this type of research which involved laboratory sessions of interactions. The sample consisted of over 400 participants involving 138 dyads in the human-human interactions conditions. The diverse measures on which the conditions were compared also were a strength of this study. The measures included individual reactions (state affect, self-presentation concerns), interpersonal outcomes (e.g., liking, the perception of being liked), and a linguistic analysis.

As is the case of any study, however, there were also limitations and ways that the research could be extended. Our study involved only one interaction session of 20 min. Perhaps if there had been extended interaction over a longer period of time, or even multiple sessions over multiple days, affect, feelings of affection, and topics covered in both the human-human interactions and the human-chatbot interaction may have changed and in different ways based on the condition. In fact, recent longitudinal research involving another social chatbot, *Mitsuku*, showed that adults' ratings of enjoyment were high after their initial interaction (perhaps because of a novelty effect), but declined thereafter (Croes & Antheunis, 2021). Thus, it could be that conversational quality with humans could increase over time, whereas, because of their current technological limitations, conversational quality with chatbots could diminish. Second, we allowed the participants to discuss any topic that they wished. Further research could consider using a similar design although with a structured self-disclosure task, such as the Aron, Melinat, Aron, Vallone, and Bator (1997) Fast Friends procedure. Additionally, we did not distinguish between chat partners in the linguistic analysis. Therefore, it is possible that the positive tone of the chats with Replika were driven mainly by the application and not the human participant. This was a limitation in our design, as there were no speaker tags in the Replika program once we exported these chats to a word processing program, which limited our ability to distinguish between speakers. This limitation will be addressed in future work. Additionally, as the world is moving to more text-based social interactions, future analyses should explore more directly acquaintanceship and linguistic differences in spoken and written modalities. Finally, because artificial intelligence is constantly changing and becoming more sophisticated and natural, we clearly need continued research on the effectiveness and outcomes of human interaction with chatbots as the technology continues to evolve.

5. Conclusions

Scientists of relationships have begun to speculate about how technology in the form of robots, virtual reality, and so forth will create proxies for relationships and satisfy needs in the same way humans can (Arriaga, 2020). Brief (initial) interactions, which were the focus of this study, can meet various human needs beyond the potential of the formation of a relationship (when both are humans). Interactions, whether they are with a human or a chatbot, can serve expressive needs (Pennebaker & Chung, 2007), help meet the need to belong (Baumeister & Leary, 1995; Hirsch & Clark, 2019), and serve as a safe haven when in stress (Birnbaum et al., 2016). Technology available to humans for communication, whether it is with other humans or with robots and AI, is constantly evolving. The 2020 COVID crisis underscored the importance of learning more about how human needs can be met through advanced communication technologies.

CrediT author statement

Michelle Drouin: Project administration, Conceptualization, Methodology, Writing-Original Draft, Supervision. Susan Sprecher: Conceptualization, Methodology, Writing-Original Draft. Robert Nicola: Data curation, Investigation, Methodology, Software. Taylor Perkins: Investigation, Methodology, Resources. All authors: Writing- Reviewing and Editing.

References

Antheunis, M. L., Schouten, A. P., Valkenburg, P. M., & Peter, J. (2012). Interactive uncertainty reduction strategies and verbal affection in computer-mediated communication. *Communication Research, 39*, 757–780.

Antheunis, M., Valkenburg, P. M., & Peter, J. (2007). Computer-mediated communication and interpersonal attraction: An experimental test of two explanatory hypotheses. *CyberPsychology and Behavior, 10*, 831–835.

Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin, 23*, 363–377.

Arriaga, X. (2020). *The social psychology of close relationships (Society for Personality and Social Psychology)*. https://www.youtube.com/watch?v=1PVkN7Cyt3E.

Banchs, R. E. (2017). On the construction of more human-like chatbots: Affect and emotion analysis of movie dialogue data. In *2017 Asia-Pacific signal and information processing association annual Summit and conference (APSIPA ASC). IEEE*. https://doi.org/10.1109/apsipa.2017.8282245

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497–529.

Bente, G., Ruggenberg, S., Kramer, N., & Eschenburg, F. (2008). Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Human Communication Research, 34*, 287–318.

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.

Burgoon, J. K., & Jones, S. B. (1976). Toward a theory of personal space expectations and their violations. *Human Communication Research, 2*, 131–146. https://doi.org/10.1111/j.1468-2958.1976.tb00706.x

Byrne, D., & Clore, G. L. (1970). A reinforcement-affect model of evaluative responses. *Personality: International Journal, 1*, 103–128.

Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., et al. (2017). Towards a chatbot for digital counselling. *Journal of Medical Internet Research, 4*, e3. https://doi.org/10.14236/ewic/HCI2017.24

Carpenter, R. (n.d.. Cleverbot. http://www.cleverbot.com.

Carter, E., & Knol, C. (2019). Chatbots — an organisation's friend or foe? *Research in Hospitality Management, 9*, 113–116. https://doi.org/10.1080/22243534.2019.1689700

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems, 92*, 539–548. https://doi.org/10.1016/j.future.2018.01.055

Croes, E. A. J., & Antheunis, M. L. (2021). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships, 38*, 279–300. https://doi.org/10.1177/0265407520959463

Croes, E. A. J., Antheunis, M. L., Schouten, A. P., & Krahmer, E. J. (2019). Social attraction in video-mediated communication: The role of nonverbal affiliative behavior. *Journal of Social and Personal Relationships, 36*, 1210–1232. https://doi.org/10.1177/0265407518757382

Culnan, M. J., & Markus, M. L. (1987). Information technologies. In F. M. Jabin, L. L. Culnan, K. H. Roberts, & L. W. Porter (Eds.), *Handbook of organizational communication: An interdisciplinary perspective* (pp. 420–443). Newbury Park, CA: Sage.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). *Wizard of Oz studies: Why and how. Knowledge-based systems* (Vol. 6).

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering, 22*, 811–817.

Dunbar, R. I. M. (2018). The anatomy of friendship. *Trends in Cognitive Sciences, 22*, 32–51.

Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on twitter. *Computers in Human Behavior, 33*, 372–376. https://doi.org/10.1016/j.chb.2013.08.013

Fingerman, K. L. (2009). Consequential strangers and peripheral ties: The importance of unimportant relationships. *Journal of Family Theory & Review, 1*, 69–86.

Fox, J., & McEwan, B. (2017). Distinguishing technologies for social interaction: The perceived social affordances of communication channels scale. *Communication Monographs, 84*, 298–318.
Govern, J. M., & Marsch, L. A. (2001). Development and validation of the situational self-awareness scale. *Consciousness and Cognition, 10*, 366–378.
Gunaydin, G., Oztekin, H., Karabulut, D. H., & Salman-Engin, S. (2020). Minimal social interactions with strangers predict greater subjective well-Being. *Journal of Happiness Studies*, 1–15.
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. https://doi.org/10.1016/j.chb.2015.02.026
Hirsch, J. L., & Clark, M. S. (2019). Multiple paths to belonging that we should study together. *Perspectives on Psychological Science, 14*, 238–255.
Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*, 712–733. https://doi.org/10.1093/joc/jqy026
Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta-analytic review. *PLoS Medicine, 7*(7), Article e1000316.
Ivanov, S., & Webster, C. (2017). Adoption of robots, artificial intelligence and service automation by travel, tourism and hospitality companies – a cost-benefit analysis. In *International scientific conference "Contemporary tourism – traditions and innovations". Sofia university, 19–21 October 2017*.
Jia, J. (2003). The study of the application of a web-based chatbot system on the teaching of foreign languages. In *Proceedings of the SITE2004 (the 15th annual conference of the society for information technology and Teacher education)* (pp. 1201–1207). AACE press.
Jiang, L. C., Bazarova, N. N., & Hancock, J. T. (2011). The disclosure-intimacy link in computer-mediated communication: An attributional extension of the hyperpersonal model. *Human Communication Research, 37*, 58–77.
Kenny, D. A. (2015). *Dyadic analysis*. Retrieved from http://www.davidakenny.net/dyad.htm. (Accessed 30 June 2021).
Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford press.
Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*, 1123–1134.
Kumar, H. (2019). *Important use cases and examples of chatbots for the retail industry*. BotCore. Retrieved from https://botcore.ai/blog/important-use-cases-and-examples-of-chatbots-for-the-retail-industry/on. (Accessed 30 June 2021).
Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin, 9*, 371–375.
Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PLoS One, 6*(9), 1–7. https://doi.org/10.1371/journal.pone.0025085
Lundy, B., & Drouin, M. (2016). From social anxiety to interpersonal connectedness: Relationship building within face-to-face, phone and instant messaging mediums. *Computers in Human Behavior, 54*, 271–277. https://doi:10.1016/j.chb.2015.08.004.
Mallen, M. J., Day, S. X., & Green, M. A. (2003). Online versus ftf conversations: An examination of relational and discourse variables. *Psychotherapy: Theory, Research, Practice, Training, 40*, 155–163.
McCroskey, L. L., McCroskey, J. C., & Richmond, V. P. (2006). Analysis and improvement of the measurement of interpersonal attraction and homophily. *Communication Quarterly, 54*, 1–33.
McKenna, K. Y. A., Green, A. S., & Gleason, M. E. J. (2002). Relationship formation on the internet: What's the big attraction? *Journal of Social Issues, 58*, 9–31.
Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to machines about personal mental health problems. *Journal of the American Medical Association, 318*, 1217–1218. https://doi.org/10.1001/jama.2017.14151
Mori, M. (1970). The uncanny valley. *Energy, 7*, 33–35.
Okdie, B. M., Guadagno, R. E., Bernieri, F. J., Geers, A. L., & Mclarney-Vesotski, A. R. (2011). Getting to know you: FTF versus online interactions. *Computers in Human Behavior, 27*, 153–159.
Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *Development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.
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., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *LIWC2007 manual: The development and psychometric properties of LIWC2007*. LIWC.net.
Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577.
Pereira, M. J., Coheur, L., Fialho, P., & Ribeiro, R. (2016). *Chatbots' greetings to human-computer communication. The computing research Repository (CoRR), arXiv: 1609.06479*.
Pew Research Center. (2019). *Mobile fact Sheet*. Retrieved from https://www.pewresearch.org/internet/fact-sheet/mobile/. (Accessed 30 June 2021).
Radziwill, N., & Benton, M. (2017). *Evaluating quality of chatbots and intelligent conversational agents* (Vol. 19, pp. 25–36). Software Quality Professional.
Ramirez, A., & Burgoon, J. K. (2004). The effect of interactivity on initial interactions: The influence of information valence and modality and information richness on computer-mediated interaction. *Communication Monographs, 71*, 422–447.
Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology, 101*, 557–570.
Replika. (2021). Retrieved from: https://play.google.com/store/apps/details?id=ai.replika.app&hl=en_US&gl=US. (Accessed 30 June 2021).
Sandstrom, G. M., & Dunn, E. W. (2014a). Social interactions and well-being: The surprising power of weak ties. *Personality and Social Psychology Bulletin, 40*, 910–922.
Sandstrom, G. M., & Dunn, E. W. (2014b). Is efficiency overrated? Minimal social interactions lead to belonging and positive affect. *Social Psychological and Personality Science, 5*, 437–442.
Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Nunamaker, J. F., Jr. (2018). 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).
Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *LDV Forum, 22*, 29–49.
Sprecher, S. (2014). Initial interactions online-text, online-audio, online-video, or ftf: Effects of modality on liking, closeness, and other interpersonal outcomes. *Computers in Human Behavior, 31*, 190–197.
Sprecher, S., & Hampton, A. J. (2017). Liking and other reactions after a get-acquainted interaction: A comparison of continuous ftf interaction versus interaction that progresses from text messages to ftf. *Communication Quarterly, 65*, 333–353.
Sprecher, S., & Treger, S. (2015). The benefits of turn-taking reciprocal self-disclosure in get-acquainted interactions. *Personal Relationships, 22*, 460–475.
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*, 497–514.
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*, 24–54.
Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research, 28*, 317–348.
Toscos, T., Carpenter, M., Drouin, M., Roebuck, A., Kerrigan, C., & Mirro, M. (2018). *College students' experiences with and willingness to use different types of telemental health resources: Do gender, depression/anxiety, or stress levels matter? Telemedicine & E-health*. https://doi.org/10.1089/tmj.2017.0243
Toscos, T., Coupe, A., Flanagan, M., Drouin, M., Carpenter, M., Reining, L., et al. (2019). Teens using screens for help: Impact of suicidal ideation, anxiety and depression levels on youth preferences for telemental heath resources. *JMIR Mental Health, 6*, Article e13230.
Turkle, S. (2015). *Reclaiming conversations: The power of talk in the digital age*. New York, NY: Penguin Press.
Vittengl, J. R., & Holt, C. S. (2000). Getting acquainted: The relationship of self-disclosure and social attraction to positive affect. *Journal of Social and Personal Relationships, 17*, 53–66.
Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research, 19*, 52–90.
Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 1–43.
Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In M. L. Knapp, & J. A. Daly (Eds.), *The handbook of interpersonal communication* (pp. 443–479). Thousand Oaks, CA: Sage Publications.
Walther, J., Deandrea, D., & Tong, S. (2010). Computer-mediated communication versus vocal communication and the attenuation of pre-interaction impressions. *Media Psychology, 13*, 364–386. https://doi.org/10.1080/15213269.2010.524913
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*, 1063.
Wellman, B. (2012). Is Dunbar's number up? *British Journal of Psychology, 103*, 174–176.
van Wezel, M. M. C., Croes, E. A. J., & Antheunis, M. L. (2020). "I'm here for you": Can social chatbots truly support their users? A literature review. *International Workshop on Chatbot Research and Design*, 96–113. https://doi.org/10.1007/978-3-030-68288-0_7
Yu, Q., Nguyen, T., Prakkamakul, S., & Salehi, N. (2019). I almost fell in love with a Machine:" Speaking with computers affects self-disclosure. In *Extended Abstracts of the 2019 CHI conference on human Factors in computing systems* (pp. 1–6). https://dl.acm.org/doi/10.1145/3290607.3312918.