Let's Work Together: Integrating Human Support with Conversational Agents

YI-CHIEH LEE, University of Illinois at Urbana-Champaign, USA
NAOMI YAMASHITA, NTT Communication Science Laboratories, Japan
YUN HUANG, University of Illinois at Urbana-Champaign, USA

Chatbots have the potential of eliciting users' deep self-disclosure and mediating human-to-human interaction. However, little is known about how people would interact differently with a human-supported chatbot than when talking to a chatbot alone. We design chatbots with (CH) and without (CO) human support to deliver suggestions for people to practice journaling skills. We conduct a study to investigate the effects of the two chatbot designs and each participant used their chatbot for four weeks. Our results show that the CH participants perceived a higher level of engagement than the CO participants when they received suggestions for practicing journaling skills and also engaged in deeper self-disclosure. However, after finishing the journaling-skill training session, the CO participants were more willing to keep practicing the suggested journaling skills than the CH group. The COVID-19 pandemic has heightened the challenges of providing healthcare services and social connection. Our in-progress research proposes a human-support chatbot system to explore effective designs to facilitate users' social connection and well-being and through conversational user interfaces.

CCS Concepts: • **Applied computing** → *Psychology*; • **Human-centered computing** → **User studies**.

Additional Key Words and Phrases: Chatbot, Self-disclosure, Human Support, Conversational User Interfaces

ACM Reference Format:
Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Let's Work Together: Integrating Human Support with Conversational Agents. 1, 1 (July 2020), 13 pages. https://doi.org/submission

1 INTRODUCTION

Conversational agents (chatbots) are regarded as one of the promising technologies and are increasingly applied in many domains. Because chatbots provide a fast, convenient, and low-cost communication channel, both scholars and practitioners are keen to develop effective chatbots to address the challenges of providing healthcare services and improving people's well-being. A growing body of research demonstrates how chatbots can be useful for maintaining healthy lifestyles [18, 24], collecting daily health information to share with healthcare professionals [14, 21], and guiding people to improve their general well-being [17, 20].

Despite the success of healthcare chatbots, there are still a number of challenges to overcome. For example, research points out that people easily become disengaged from using a chatbot

Authors' addresses: Yi-Chieh Lee, University of Illinois at Urbana-Champaign, USA, ylee267@illinois.edu; Naomi Yamashita, NTT Communication Science Laboratories, Japan, naomiy@acm.org; Yun Huang, University of Illinois at Urbana-Champaign, USA, yunhuang@illinois.edu.

XXXX-XXXX/2020/7-ART $15.00
https://doi.org/submission

[33, 43], hampering them from long-term interventions. In addition, people may overtrust solutions suggested by chatbots which could be clinically inappropriate [12, 47, 48]. One way to overcome such challenges is to take advantage of human support. By integrating human support with chatbot-based intervention, we may be able to make the best use of both human-based and chatbot-based interventions. Indeed, studies have shown that the integration of human support with chatbot interactions could promote user engagement [33]. In a recent study, Lee et al. proposed a mediator chatbot that promotes deep self-disclosure from users and delivers the information to a human mental-health professional [21].

However, more research is clearly needed on how individuals might respond differently to interaction with a healthcare chatbot alone vs. one with a human supporter. To help fill this gap, this preliminary study with 35 participants deployed two chatbot designs, both of which delivered training in journaling skills, which is common practice in boosting users' general well-being [44] and improving self-reflection [13]. The first version guided the participants itself, while the second integrated human support into its interaction to guide the participants. Over a period of four weeks, we tracked changes and differences in how each version impacted its users' responses to and perceptions of the chatbot system, as well as their levels of compliance with system requests.

This study makes the following contributions. First, it constitutes pioneering empirical work on the effects of integrating human support into human-chatbot interaction to deliver suggestions. And second, by using a longitudinal study method, our research highlights some of the new design opportunities and challenges that arise when chatbot systems with and without human support are used for long-term interventions. Lastly, the COVID-19 pandemic has heightened the challenges of providing healthcare services at all levels, so this research represents a timely contribution to society as well as having the potential to make an impact on the study of human-CA interaction. Our in-progress research extends this human-support chatbot model to explore effective designs to facilitate users' social connection and behavior change through conversational user interfaces.

2 RELATED WORK

2.1 Conversational Agents for Wellbeing

Chatbots are gaining considerable attention in the field of healthcare. Research has shown that chatbots can assist users in tracking their behaviors (e.g., [18, 24]) and feelings (e.g., [10]), which could further be used to solicit social support and self-reflection. Also, many studies designed chatbots to enhance users' well-being [42], notably by encouraging healthier habits or ways of thinking [31], such as better eating habits [24], exercise [18], ways of coping with stress [31], and self-compassion [20]. Though certain features of chatbots give them advantages over humans when supporting mental well-being, research has also shown that people tend to apply the social norms of human relationships to their interactions with computer agents. This tendency, known as the Computers Are Social Actors (CASA) paradigm [28], has informed the design of many computer agents [22, 34, 41]. For example, Ravichander et al. [34] found that reciprocity occurred in human-chatbot interactions that a chatbot's self-disclosure encouraged people's self-disclosure. Similarly, recent work by Lee et al. [22] showed that, over time, people revealed deeper thoughts and more feelings about sensitive topics (e.g., social and sexual relationships, experiences of failure, and the causes of their stress and anxiety) to a chatbot that engaged in deeper self-disclosure (e.g., revealing thoughts, feelings, and personal experiences) than with chatbots that either did not self-disclose, or disclosed less.

However, several limitations of chatbot-based interventions remain, and in certain situations, chatbot-based interventions may be less beneficial than those provided by humans [26, 27, 33]. For example, although many healthcare interventions require long-term engagement [6], people may easily become disengaged from the use of self-guided systems, due to loss of motivation and/or failure to incorporate those systems' recommendations into their daily lives [33]. In addition, an investment model shows that purely computer-based interventions are often much less effective than hybrid ones with some expert human input [9], in part because the latter tend to inspire their users to execute a higher proportion of their intervention requests. Finally, Howard et al. [12] has suggested that some people may trust robots too much, due to over-optimism about the viability of the solutions they will suggest, and that this trust becomes a source of risk if robots make clinically suboptimal or inappropriate suggestions. In light of the advantages and disadvantages of human-based and chatbot-based interventions for well-being, it would seem to make sense to integrate them in a way that makes the best possible use of both.

2.2 Integration of Human and Chatbot-based Approaches

Mostly, human support has taken place via communication channels external to the chatbot system, such as phone calls, text messages, and email [1]. Few researchers have suggested deeper integration of human support into chatbot interventions [33, 36, 37]. Indeed, Schueller et al. [36, 37] reviewed prior studies of integrating human support into behavioral intervention technologies and suggested the concepts to guide deeper integration by capturing the trade-offs between client benefits and the available human resources. In particular, Schueller et al.'s findings imply that chatbot intervention and human support should be integrated in a seamless manner, rather than delivered through separate media.

Based on the foregoing review, and in answer to Schueller et al.'s [36, 37] calls for seamless integration of the support provided by chatbots and humans, our research integrates human support with chatbot intervention through a single communication channel, and measures how this affects the social dynamics of the chatbot intervention. In light of prior studies' findings that users tend to build a sense of companionship with chatbots, and disclose more to them than to humans in parallel situations, it is not clear if and how adding a human supporter into a chatbot intervention will change its social dynamics and overall user experience, perhaps including its outcomes [15, 22, 26]. In addition, according to the social presence theory [19], it is suggested that users' perceived immediacy and intimacy affect their engagement in the computer-mediated mode of communication. Thus, seamlessly integrating human presence in a human-chatbot interaction might affect users' experience of one another.

In this research, to explore the effects of integrating human support into a chatbot system, we designed two chatbot designs - with and without human support. The aim of the chatbot was to guide its users to learn new skills. We conducted a four-week study deploying two chatbot conditions, with and without a human supporter (coach), to provide guidance to improve the users' journaling skills. The suggestions delivered to the users in both chatbot conditions were adopted by pre-existing journaling materials (e.g., gratitude journals [8, 39] and expressive writing [3]). The evaluation of how each design influenced its users' experience and journaling behavior was guided by the following two research questions.

- **RQ1**: How do people respond differently to the chatbot with (CH) and without (CO) a human supporter involved in the conversation?
- **RQ2**: How do the effects of interacting with the chatbot differ between the CH and CO conditions?

3 METHOD

Fig. 1 shows our chatbot interface. Because of its similarity to commercially available messenger applications, the participants readily learned how to use it. They were allowed to give free-text

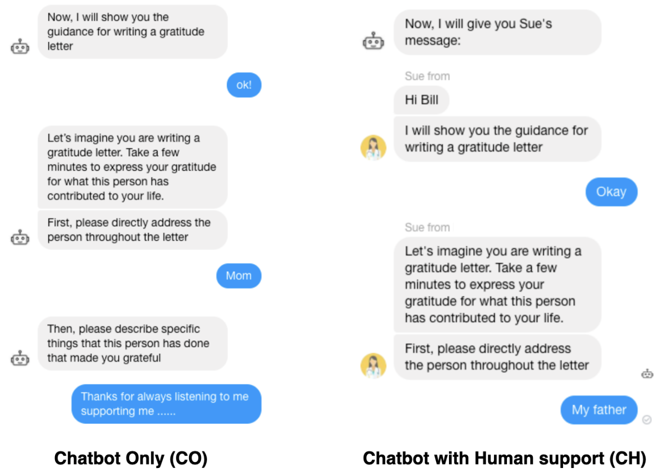

Fig. 1. Our chatting interface. Left: The chatbot gives a suggestion/guidance to the user (CO group), Right: The chatbot introduces a human coach, and the coach gives a suggestion/guidance to the user (CH group).

responses to almost all of the system's questions, though sometimes, it provided a few options for them to choose from. For example, when the chatbot asked about a user's mood, it would also show him/her a list of words that could be used when answering. For the chatbot's appearance, we adopted a bot figure (Fig. 1, Left). However, for the group with human support, when the human coach said something, a female figure named Sue appeared to visually mark which suggestions were from the coach (Fig. 1, Right).

3.1 Study Design

To observe and understand how the designs affected users' behavior and experience over a period of time, we designed a four-week study. During the study, the chatbot gave the users some suggestions to improve their journaling skills. Among the many interventions, we chose "journaling" because it is a common practice that has been shown effective in enhancing one's well-being [38, 44], boosting mood and reducing anxiety. Since previous research [21, 23, 45] shows that chatbots have the potential to help people improve their mental well-being by eliciting self-disclosure, using a chatbot to learn new journaling skills is a good match. Furthermore, journaling allowed us to observe the users' journaling behaviors by tracing their interaction with the chatbot.

The four-week study (Fig. 2) consisted of three phases: Warm-up, Training and Free-will. In the first of these nine-day segments - **Phase 1: Warm-up** - each chatting task commenced with a *Journaling session*, followed by a *Small-talk session*. This phase was utilized as a warm-up to familiarize participants with chatting with the chatbot and to remove novelty effects.

In **Phase 2: Training**, a *Suggestions session* was added after the journaling and small-talk sessions. In it, either the chatbot or a human supporter (*coach*) gave the participants some suggestions about how to improve their journaling skills and learn new journaling skills. This phase was to investigate users' responses as behavior change (experience) and measure their perceived interaction as well as its effect. On the first day of Phase 2, the chatbot told the CO group: *"From today, I am going to give you some guidance to learn new journaling skills, which could help you (1) gain a better understanding of your own mental-health status; and (2) help you to improve your happiness and well-being."* In the CH group, the same comments were attributed to a coach called Sue, who was introduced by the

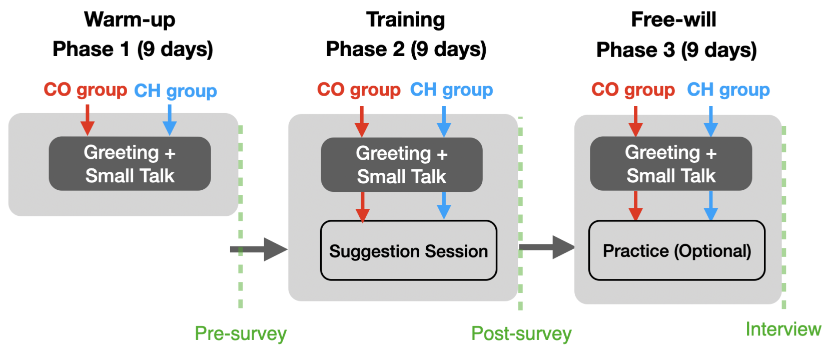

Fig. 2. Study design. The study consisted of three phases: Warm-Up, Training, and Free-will. CO represents the group with Chatbot Only, and CH represents the group with Chatbot with Human support. Human supporter (coach) appeared only during the Suggestion session (in Phase 2) to deliver suggestions. Other chatting sessions were kept the same for both groups.

chatbot as follows: *"I am going to introduce my colleague, Sue, to you. She is a coach to guide you to learn new journaling skills. [...]"* The users were informed that the Sue is a real human supportor.

Lastly, in **Phase 3: Free-will**, no further suggestions were given. Instead, the chatbot encouraged the participants to practice the journaling skills (Fig. 2, *Practice (Optional)*) that had been shared with them during *Suggestion session*, though they could ignore such encouragement without receiving any penalty. This was because the purpose of this phase was to enable the researchers to gauge differences in how our CO and CH chatbot versions affected the participants' journaling practices when following guidance was not required.

3.2 Conversation Design

3.2.1 *Greeting and Small-talk Session.*

Because of journaling's various documented benefits to personal well-being, we designed this type of session – in which users are asked about their moods and the reasons for them – to occur every day. Accordingly, our chatbot prompted the participants to focus their journaling on five topics: their mood, experiences, gratitude, stress, and anxiety. For example, the chatbot asked users: *"What is it and how do you protect yourself from COVID-19?"* and *"Do you feel anxious or stressful because of the COVID-19?"*

Previous research [5, 22, 45] has reported that small-talk (social chat) with a chatbot can improve users' experience of the chatbot system and their performance within it. Our chatbot was designed to engage in self-disclosure, and shared its personal stories in the small-talk sessions. This design was motivated by previous studies' findings that a chatbot's self-disclosure had positive effects on its users' self-disclosure depth [22]; and that mutual self-disclosure could enhance users' positive expectations and motivation, and strengthen the therapeutic bond [11]. The conversational topics used in our small-talk sessions were adopted from previous studies [2, 22] and included feelings, thoughts, and information. The same small-talk topics and responses from the chatbot were received by both the CO and CH groups.

3.2.2 *Suggestion Session.*

The suggestions session was only deployed in Phase 2 (Fig. 2), and both the CO and CH groups received exactly the same suggestions/guidance. The suggestions were adopted from pre-existing journaling materials aimed at improving people's writing skills, with the wider aim of boosting their well-being [8, 44]. Based on the findings of research in positive psychology [38], some interventions can easily be implemented through typing or writing. For example, Gratitude Journaling [8, 25] is an effective skill/tool for the in-depth expression of appreciation to others, as a path to enhancing one's own well-being.

Our chatbot system was designed to facilitate that prior studies' guideline [25, 32, 35, 39] to build users' journaling skills. For instance, to aid acquisition of the "gratitude journal" skill, the chatbot first explained the benefits of having that skill, and asked its users to think of someone to whom they were grateful. Then, it gave the instructions *"Let's imagine you are writing a gratitude letter. Take a few minutes to express your gratitude for what this person has contributed to your life. Please directly address the person throughout the letter."* Next, the chatbot encouraged the participants to give more specific and detailed information: *"Describe specific things that this person has done that made you grateful"* and *"Describe how this person's behavior has affected your life."* Finally, the chatbot asked the participants to wrap up, as follows. *"End the letter in a way that identifies it is from you."*

As such, the chatbot's role was to keep guiding users toward the next step, while giving instructions intended to stimulate deep disclosure of their thoughts and feelings. As briefly noted above, the only difference between the CO and CH groups was that the latter's instructions - though identical to those provided directly by the chatbot to the CO group - were acknowledged to have been written by a human coach, and the chatbot was portrayed as her intermediary. The participants would see the coach's message separately from the chatbot's messages, under their own chatting partner image (Fig. 1, Right). Again, the participants were allowed to skip any instruction they did not want to follow, and the length of each response was not restricted. Only one skill (i.e., gratitude journaling [25, 39], expressive writing [32, 35], or the best possible self exercise [30, 40]) was delivered to any participant on a given day, and the participants learned that skill over three consecutive days.

3.3 Participants

In this early stage study, the 35 participants we recruited ranged in age from 20 to 29 (M = 24.74), and none of them reported having any mental illness or history of receiving psychotherapy. All were graduate or undergraduate students. We divided them into two groups of approximately equal size: the CO group (Participant (P)1-17), which received suggestions from the chatbot, and the CH group (P18-35), which ostensibly received suggestions from a human coach. The final composition of CO was 10 females and seven males, and of CH, 11 females and seven males. All participants were familiar with Facebook Messenger or other similar messenger platforms. After the four-week period of interacting with the chatbot, 34 participants attended face-to-face individual interviews, each of which lasted between 45 and 60 minutes. At the end of the study, all participants were invited to a face-to-face interview. This research was reviewed and approved by our institutional review board. This study was conducted from Jan. 2020 to Feb. 2020 in Japan, the time which was at early stage of COVID-19 pandemic.

3.4 Measurement

3.4.1 Conversation Logs.

We collected all participants' chatbot conversation logs, and compared them across the CO and CH groups. To assess how closely the participants followed the system's suggestions, we focused on changes in the depth of their self-disclosure, which the acquisition of journaling skills has been shown to deepen. Previous research has also used depth of self-disclosure as a metric of users' quality of responses to a chatbot [46] and of their trust in a chatbot system [21]. We had two raters code the log data independently according to the three categories proposed by Barak and Gluck-Ofri [4], i.e., information, thoughts, and feelings, each of which is further subdivided into three levels according to the sensitivity of the information disclosed. To analyze how the two chatbot versions, CO and CH, had influenced the participants' depth of self-disclosure, word count, and responsiveness to journaling suggestions, we performed a mixed-model analysis of variance (ANOVA), followed by a Tukey's test of significant difference. For this purpose, we treated the chatbot's suggestions as random effect; groups (CO and CH) and experimental day (Phase 2: 9 days) as independent variables; and self-disclosure levels as the dependent variables.

3.4.2 *Survey.*

The participants were asked to self-report their level of engagement with the chatbot, as such engagement could have influenced their responsiveness to journaling suggestions, and thus their performance when using the system. Additionally, prior studies have suggested that human support probably increases users' engagement, and that loss of engagement is a common obstacle to self-guided treatment [27]. The 12 measurement items for our surveys' engagement dimension were adapted from prior literature [29]. We conducted mixed-model ANOVA to analyze the survey results, with the dependent variables being self-reported engagement, trust, intimacy, self-reflection, and self-awareness. The two independent variables were group membership - i.e., of CO vs. CH - and time (Fig. 2): i.e., before Phase 2 (Pre-survey) vs. after Phase 2 (Post-survey).

3.4.3 *Interview.*

Our semi-structured interview protocol focused on the participants' chatbot experiences, including their daily practices of using the system, their engagement, and their impressions of the chatbot. Follow-up questions covered if/how their attitudes and impressions had changed since the start of the experiment. To capture differences in how the interviewees responded to the chatbot's (or coach's) specific guidance, we asked them to describe their feelings about those suggestions, including if they felt they were worth following; if they felt comfortable about receiving them; what they learned after following/ignoring them; and if such feelings changed over time. Thematic content analysis, which involves iteratively reviewing and labeling interviewees' responses with emerging codes, was applied to all the interview data by two raters working independently.

4 FINDINGS

4.1 Responses to the Two Chatbots (RQ1)

To explore RQ1, regarding how users responded differently to versions of the same chatbot with (CH) and without (CO) a human supporter (coach) in the same communication channel, we analyzed conversation logs from the suggestions sessions during the training period (Phase 2). Since the suggestions given in either chatbot setting required the participants to disclose and reflect more about themselves than had been the case in Phase 1, we measured the quantity and depth of participants' self-disclosure. First, we calculated the word count of participants' responses and compared the differences using mixed-models ANOVAs. Results showed that the CH (M = 159.41, SD = 16.61) group's mean overall word count was significantly higher than that of the CO (M = 118.31, SD = 17.62) group ($F(1, 33) = 6.60$, $p < .01$). There was no significant main effect of experiment day, and no interaction effects.

4.1.1 *Depth of Self-disclosure.*

With regard to *Information*, there was no significant effect of any factor; i.e., neither CO/CH membership nor experiment day significantly impacted how the participants disclosed information to either version of the chatbot. The group averages for informational self-disclosure across all suggestion questions were M = 1.7, SD = .67 for CO, and M = 1.65, SD = .57 for CH.

Second, in the case of *Thoughts*, our analysis revealed a main effect of group (F(1, 33) = 29.6, $p <$ .001), with the participants in the CH group disclosing more thoughts. However, there was no effect of experiment day, and no interaction effects. The group averages for self-disclosure of thoughts across all suggestion sessions were M = 1.89, SD = .13 for CO and M = 2.32, SD = .12 for CH.

Lastly, there was a significant main effect of group membership on the self-disclosure of *Feelings* (F(1, 33) = 12.12, $p < .05$) - with the members of CH disclosing feelings significantly more than CO members did - but no main effect of experiment day, and no interaction effects. The group averages for self-disclosure of feelings across all suggestion sessions were M = 1.91, SD = .11 for CO and M = 2.22, SD = .13 for CH. In summary, the CH group was found to have given longer responses containing deeper feelings and more thoughts than the CO group during the suggestion sessions of Phase 2.

Here we show sample responses with three levels of self-disclosure from participants. Chatbot Question: *"Do you feel anxious or stressful because of COVID-19?"* These are three of the participant answers: *"Yes, the virus has already spread in Japan, so it may also affect Kyoto."* (coded level 1 Feelings), *"My mother is a nurse, I feel it is kind of dangerous that she is still working in the hospital now."* (coded level 2 Feelings), and *"I feel so bad when I saw all the bad news about this on social media. And, I do feel stressed about this, which makes me not concentrate."* (coded level 3 Feelings).

4.2 Perceived Engagement (RQ2)

The engagement level revealed significant main effects of both group (F(1, 33) = 8.63, $p < .001$) and time (F(1, 33) = 4.76, $p < .05$), but there was no significant interaction effect: with the CH group reporting significantly higher engagement than CO, both groups' engagement levels generally increased after the training period ended.

In the interview, ten **CO group** members said that their engagement with the chatbot increased over time because it sent them useful suggestions and prompted them to accomplish something new every day. For example, one participant (P5, F) said: *"Engagement increased over time because I got used to the chatbot and some of its suggestions were useful. I felt more engaged and it gave me motivation."* Besides, most participants in the CO group felt comfortable about deciding for themselves, on a case-by-case basis, whether to follow the chatbot's suggestions or not follow them. When they did follow them, they generally felt happy and surprised that they had been able to learn something useful from a chatbot. As one put it, *"Although I did not practice those suggestions a hundred percent, I think I am still on track. When you learn something from doing this, you will feel more motivated. So, my next step is to keep practicing them."* (P10, F)

Although practicing new journaling skills seemed to enhance engagement for many participants, four members of the CO group reported that doing so caused them fatigue and annoyance. For example, *"I felt annoyed because some suggestions were time-consuming to carry out. I did not expect that I would have to expend so much effort."* (P17, M) Moreover, those participants expressed discomfort at certain suggestions they received from the chatbot. As one of these interviewees said, *"Sometimes, I felt the chatbot was too bossy, especially when it started to give me suggestions. It was okay when the suggestion seemed useful. But when the suggestion was not useful, or the chatbot prompted me to disclose more, I got a bit annoyed. 'It's just a robot, why is it giving me instructions?' That kind of feeling."* (P3, M)

In addition, three participants in the CO group reported that their engagement gradually decreased, due to loss of interest in the chatbot system, i.e., the novelty effect [46]. One of them said, *"I was more engaged in the beginning. This [chatting with a chatbot every day] was a new thing for me. But as I got used to it, my engagement level dropped."* (P14, M)

In the **CH group**, according to their survey responses, the CH group felt more highly engaged with the chatbot system than their CO group counterparts did. Most of the CH group interviewees reported positive attitudes toward practicing the focal journaling skills, and provided two reasons for this. First, much like the CO group, they felt that the content of the suggestions themselves improved their engagement, for example, *"Although practicing journaling skills was time-consuming, following the suggestions helped me understand myself better. So, though lengthy, the process led me to good results."* (P33, F)

Secondly, more than half (n=11) of the participants in the CH group highlighted the importance of human support, noting that the involvement of a coach increased their willingness to take suggestions seriously. This was because they thought the coach personalized the suggestions for them and would monitor their practicing activities on the chat channel. As one explained: *"The suggestions were from a coach. I thought the coach might see my responses and give me further suggestions. So, I was more careful about my responses for the suggestions."* (P24, M)

Nevertheless, the ostensible involvement of a human coach seemed to negatively affect the engagement of a minority of the CH participants. Three of them noted that it increased their expectations: i.e., that they would receive highly personalized suggestions and feedback. Because our study design did not actually offer such features, these users' engagement with the chatbot was diminished. As one of them put it, *"The coach gave me suggestions, but they were general suggestions. In fact, I wanted to have more personalized guidance. I felt the coach did not pay attention on my responses."* (P19, F) Moreover, because the coach only offered general suggestions, several mentioned feeling disappointed that 'she' could not really give them personalized feedback or suggestions. For instance, *"I was kind of disappointed by the low level of her involvement to customize suggestions."* (P26, F)

In addition, the impression of human support caused stress to some of the CH users, who felt the coach was judging their answers. As one of them explained, *"I felt the coach would judge my answers, so I tried to answer the questions as thoroughly as possible, and this made me feel stressed."* (P32, F) While this drove them to implement the suggestions, it also triggered negative feelings when they could not follow the suggestions, for example,*"I felt sorry when I could not follow the suggestions, because I guess the coach put a lot of effort into designing this chatbot to give help me."* (P30, F)

4.3 Prolonging Participants' Practice of Suggestions at the Free-Will Period (RQ2)

To gauge the differences in how our CO and CH chatbot versions affected the participants' journaling practices when following guidance was not required, we first calculated the ratio of participants who practiced the skills per day during the free-will period (Phase 3, Fig. 2). We then compared the mean ratio between the conditions using a t-test. Interestingly, the results showed that more CO participants voluntarily practiced the suggestions than CH participants (M(CO) = .80, SD = .08; M(CH) = .57, SD = .12; t = 4.15, p < .001).

Reasons of Keeping the Practices – Across the **both groups**, the participants reported similar reasons for countinuing practicing the skills during the free-will period. A majority of the sample (n=20) indicated that, because they had benefited from practicing the journaling skills, they felt motivated to keep doing so. As one explained, *"When I followed a suggestion from the chatbot, I was excited to know that even a robot could improve my life! It's not the type of advice we can get in our daily lives, so I keep practicing them."* (P15, M)

Some of the participants emphasized that their relationships with the chatbot also encouraged them to keep practicing the taught skills. As one noted, *"I have been asked about my mood [...] every day, so I gradually felt close to the chatbot. Also, I learned a lot from the chatbot, which gave me a good impression of it. I wanted to keep on using those journaling skills."* (P27, F)

Three of the participants indicated that it was helpful when the chatbot reminded them of the journaling skills learned in the training phase. As one of them put it, *"I forgot about some skills* *quickly. But by reviewing all the journaling skills, my memories were strengthened, and that gave me confidence that I could successfully improve my mental well-being."* (P13, M)

Also, eight participants appreciated the daily prompts and encouragement by the chatbot. One of them said, *"I think it's not a bad thing to be prompted by the chatbot. It would be hard to have a real human reminding me to practice these skills every day. So the chatbot gave me more motivation to continue working on something."* (P33, F) Moreover, many participants were also motivated to follow the system's suggestions by reminders the chatbot sent to them. One of them shared his prior experience with cultivating journaling skills; the interviewee said, *"I once tried to do something similar, but I could not keep doing it on my own. However, when the chatbot encouraged me and sent me reminders, I felt motivated to follow its suggestions. This is quite different from practicing alone."* (P18, F)

Reasons for Quitting to Practice – The participants who did not often practice in the free-will period gave several reasons for this, which we summarize below. First, many of them felt tired of journaling after the training period: as one of them stated, *"I wanted to take a short break from what I had been doing."* (P19, M)

Second, practicing journaling skills was deemed optional in Phase 3, and five participants reported seeing this aspect of our system's design as giving them an excuse to skip it. As one mentioned, *"I feel that if I am given too many alternatives I will just give myself excuses. Especially when I was a little busy with other stuff, I might skip the practice."* (P9, F)

Especially, the **CH group** participants gave group-specific reasons for not practicing in Phase 3. Six of them reported that, because there was no longer a coach monitoring them after Phase 2, they did not feel it was necessary to keep practicing. One said, *"I stopped practicing the system's suggestions after the coach stopped giving new ones, because no-one was monitoring my responses and I just wanted to skip that."* (P32, F)

Lastly, two members of the CH group noted that practicing the same journaling skills they had learned in Phase 2 a second time would not have brought them new insights, and thus, they tended not to practice in Phase 3. As one of them put it, *"I think those journaling skills are still useful, but less useful than when I practiced them the first time."* (P34, M)

5 DISCUSSION AND FUTURE WORK

5.1 The Effect of Time and Effort

This study's nine-day period in which the participants familiarized themselves with the chatbot and built social bonds with it (Phase 1). While our interview findings also echo previous work regarding the important role of time in establishing social relationships between the participants and the chatbot [7, 22], that data also showed that some participants experienced a novelty effect [46]: i.e., that they were highly engaged in chatbot conversation at the very beginning, but that this excitement gradually decreased thereafter. This reveals a key challenge to deploying chatbots in real-life situations [6, 22]. Regardless of whether the effect of the passage of time is positive or negative, however, our findings suggest the importance of longitudinal study designs when testing chatbot technologies.

Furthermore, the interview results seemed to indicate that the CH group members had higher levels of dependency on the chatbot, and this could have diminished their motivation to keep journaling once it became an optional activity. We also observed the higher pressure mentioned by some CH-group members during Phase 2. This could have led to feelings of relief among that group, from the start of Phase 3, that the frequency of journaling could be reduced. Their behavior would be in line with the theory of investment model [6], insofar as their failure to perceive benefits of continued journaling practice decreased their commitment to engage in such practice. However, the CO group of participants had both lower self-disclosure and lower engagement in Phase 2 than their CH-group counterparts which might have led them to perceive more benefits from ongoing practice once Phase 2 had ended. Moreover, the lack of pressure to follow system suggestions that the CO participants perceived could have decreased the perceived costs of doing so, and thus increased their willingness to continue practicing the journaling skills in Phase 3. Thus, this finding suggests that the design of chatbots for longitudinal usage should modulate users' perceived benefit and cost of interacting with a chatbot.

5.2 Integrating Human Support

We found that the chatbot versions in the CO and CH conditions had specific advantages and disadvantages when it came to delivering guidance. Such advantages could potentially be enhanced, and such disadvantages mitigated, through alterations in how guidance is delivered. For example, although integrating human support and a chatbot into the same communication channel may have led the participants to develop unrealistically high expectations about the coach (i.e., that she would give specific feedback/suggestions based on all activities in the channel), and made some feel pressured to follow her suggestions, such effects might be reduced by having the chatbot deliver suggestions on a human coach's behalf, in its own 'voice.' In other words, if the chatbot served as an intermediary between users and the coach, the former might feel less pressure from, and have more realistic expectations of, the coach than if she delivered her suggestions to them in her 'voice.'

Our study results also show that the participants' perceptions and attitudes toward the chatbot changed as they interacted with it across the different temporal phases of the experiment. This effect was especially salient among the CO group. The participants' discomfort aroused by the chatbot giving suggestions appears to reflect that people tend to form strong early impressions of a chatbot's 'personality' and, while knowing intellectually that chatbots do not actually have personalities, to feel discomfort when chatbots' conversational style and/or content deviates from their original patterns. To mitigate such an effect, it might be useful in future studies to divide the communication channel into multiple chatbots, based on their roles: e.g., chatbot A plays the role of a conversational partner, while chatbot B plays the role of a coach. Another idea to mitigate the effect is to develop a context-aware chatbot that shows compassion with the users when delivering guidance. This idea echoes a prior study [46] which suggested that a chatbot with active listening skills [16] could show compassion with users and elicit higher quality responses and engagement.

Our preliminary research extends the understanding of human-chatbot interaction by exploring how a chatbot could be designed to provide suggestions with integrating human support. Future research could consider integrating synchronized communication channel for the users with a human support which extend the understanding. We hope our findings could make an impact on future design of conversational user interface for improving well-being and social support.

6 ACKNOWLEDGMENTS

This work is partially supported by Grant for Scientific Research (A) 17H00771 from Japan Society for the Promotion of Science (JSPS). We also thank the Faculty Start-up Fund provided by University of Illinois at Urbana-Champaign iSchool (School of Information Sciences) for supporting this research.

REFERENCES

[1] Gerhard Andersson and Pim Cuijpers. 2009. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive behaviour therapy* 38, 4 (2009), 196–205.

[2] Arthur Aron, Edward Melinat, Elaine N Aron, Robert Darrin Vallone, and Renee J Bator. 1997. The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin* 23, 4 (1997), 363–377.

[3] Karen A Baikie and Kay Wilhelm. 2005. Emotional and physical health benefits of expressive writing. *Advances in psychiatric treatment* 11, 5 (2005), 338–346.
[4] Azy Barak and Orit Gluck-Ofri. 2007. Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior* 10, 3 (2007), 407–417.
[5] Timothy Bickmore and Justine Cassell. 1999. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*. 87–92.
[6] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2010. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence* 24, 6 (2010), 648–666.
[7] Timothy W Bickmore, Lisa Caruso, Kerri Clough-Gorr, and Tim Heeren. 2005. 'It's just like you talk to a friend' relational agents for older adults. *Interacting with Computers* 17, 6 (2005), 711–735.
[8] Don E Davis, Elise Choe, Joel Meyers, Nathaniel Wade, Kristen Varjas, Allison Gifford, Amy Quinn, Joshua N Hook, Daryl R Van Tongeren, Brandon J Griffin, et al. 2016. Thankful for the little things: A meta-analysis of gratitude interventions. *Journal of counseling psychology* 63, 1 (2016), 20.
[9] David Daniel Ebert, Dirk Lehr, Filip Smit, Anna-Carlotta Zarski, Heleen Riper, Elena Heber, Pim Cuijpers, and Matthias Berking. 2014. Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial. *BMC Public Health* 14, 1 (2014), 807.
[10] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4, 2 (2017), e19.
[11] Marvin R Goldfried, Lisa A Burckell, and Catherine Eubanks-Carter. 2003. Therapist self-disclosure in cognitive-behavior therapy. *Journal of clinical psychology* 59, 5 (2003), 555–568.
[12] Ayanna Howard and Jason Borenstein. 2019. Trust and Bias in Robots: These elements of artificial intelligence present ethical challenges, which scientists are trying to solve. *American Scientist* 107, 2 (2019), 86–90.
[13] Delaura L Hubbs and Charles F Brand. 2005. The paper mirror: Understanding reflective journaling. *Journal of Experiential Education* 28, 1 (2005), 60–71.
[14] Maia L Jacobs, James Clawson, and Elizabeth D Mynatt. 2015. Comparing health information sharing preferences of cancer patients, doctors, and navigators. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 808–818.
[15] Jinkyu Jang and Jinwoo Kim. 2019. Healthier Life with Digital Companions: Effects of Reflection-Level and Statement-Type of Messages on Behavior Change via a Perceived Companion. *International Journal of Human–Computer Interaction* (2019), 1–18.
[16] Susanne M Jones, Graham D Bodie, and Sam D Hughes. 2019. The impact of mindfulness on empathy, active listening, and perceived provisions of emotional support. *Communication Research* 46, 6 (2019), 838–865.
[17] Christina Kelley, Bongshin Lee, and Lauren Wilcox. 2017. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 629–641.
[18] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection companion: A conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 70.
[19] Nanda Kumar and Izak Benbasat. 2002. Para-social presence and communication capabilities of a web site: a theoretical perspective. *E-Service* 1, 3 (2002), 5–24.
[20] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion *(CHI '19)*. ACM, New York, NY, USA, Article 702, 13 pages. https://doi.org/10.1145/3290605.3300932
[21] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proceedings of the ACM on Human-Computer Interaction* (2020).
[22] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. *ACM CHI Conference on Human Factors in Computing Systems* (2020).
[23] Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI* 4 (2017), 51.
[24] Kai Lukoff, Taoxi Li, Yuan Zhuang, and Brian Y Lim. 2018. TableChat: Mobile Food Journaling to Facilitate Family Support for Healthy Eating. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–28.
[25] Sonja Lyubomirsky, Rene Dickerhoof, Julia K Boehm, and Kennon M Sheldon. 2011. Becoming happier takes both a will and a proper way: An experimental longitudinal intervention to boost well-being. *Emotion* 11, 2 (2011), 391.

[26] Adam S Miner, Nigam Shah, Kim D Bullock, Bruce A Arnow, Jeremy Bailenson, and Jeff Hancock. 2019. Key considerations for incorporating conversational AI in psychotherapy. *Frontiers in psychiatry* 10 (2019).

[27] David Mohr, Pim Cuijpers, and Kenneth Lehman. 2011. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *Journal of medical Internet research* 13, 1 (2011), e30.

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

[29] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[30] Nansook Park and Christopher Peterson. 2009. Character strengths: Research and practice. *Journal of college and character* 10, 4 (2009).

[31] SoHyun Park, Jeewon Choi, Sungwoo Lee, Changhoon Oh, Changdai Kim, Soohyun La, Joonhwan Lee, and Bongwon Suh. 2019. Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. *Journal of medical Internet research* 21, 4 (2019), e12231.

[32] James W Pennebaker. 2010. Expressive writing in a clinical setting. *The Independent Practitioner* 30 (2010), 23–25.

[33] Juanan Pereira and Óscar Díaz. 2019. Using health chatbots for behavior change: a mapping study. *Journal of Medical Systems* 43, 5 (2019), 135.

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

[35] Lindsey M Rodriguez, Chelsie M Young, Clayton Neighbors, Michelle T Campbell, and Qian Lu. 2015. Evaluating guilt and shame in an expressive writing alcohol intervention. *Alcohol* 49, 5 (2015), 491–498.

[36] Stephen M Schueller, Kathryn Noth Tomasino, Emily G Lattie, and David C Mohr. 2016. Human support for behavioral intervention technologies for mental health: the efficiency model. *management* 21 (2016), 22.

[37] Stephen M Schueller, Kathryn Noth Tomasino, and David C Mohr. 2017. Integrating human support into behavioral intervention technologies: the efficiency model of support. *Clinical Psychology: Science and Practice* 24, 1 (2017), 27–45.

[38] Martin EP Seligman and Mihaly Csikszentmihalyi. 2014. Positive psychology: An introduction. In *Flow and the foundations of positive psychology*. Springer, 279–298.

[39] Martin EP Seligman, Tracy A Steen, Nansook Park, and Christopher Peterson. 2005. Positive psychology progress: empirical validation of interventions. *American psychologist* 60, 5 (2005), 410.

[40] Kennon M Sheldon and Sonja Lyubomirsky. 2006. How to increase and sustain positive emotion: The effects of expressing gratitude and visualizing best possible selves. *The journal of positive psychology* 1, 2 (2006), 73–82.

[41] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376843

[42] John Torous and Laura Weiss Roberts. 2017. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. *JAMA psychiatry* 74, 5 (2017), 437–438.

[43] Ha Trinh, Ameneh Shamekhi, Everlyne Kimani, and Timothy W Bickmore. 2018. Predicting User Engagement in Longitudinal Interventions with Virtual Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 9–16.

[44] Allison Utley and Yvonne Garza. 2011. The therapeutic use of journaling with adolescents. *Journal of Creativity in Mental Health* 6, 1 (2011), 29–41.

[45] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S. Shyam Sundar. 2020. Alexa as Coach: Leveraging Smart Speakers to Build Social Agents that Reduce Public Speaking Anxiety *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376561

[46] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2019. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys. *arXiv preprint arXiv:1905.10700* (2019).

[47] Jin Xu, G Bryant De'Aira, Yu-Ping Chen, and Ayanna Howard. 2018. Robot therapist versus human therapist: Evaluating the effect of corrective feedback on human motor performance. In *2018 International Symposium on Medical Robotics (ISMR)*. IEEE, 1–6.

[48] Jin Xu, G Bryant De'Aira, and Ayanna Howard. 2018. Would You Trust a Robot Therapist? Validating the Equivalency of Trust in Human-Robot Healthcare Scenarios. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 442–447.