Caring for Vincent: A Chatbot for Self-compassion

Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn
Human-Technology Interaction
Eindhoven University of Technology
Eindhoven, Noord Brabant, the Netherlands
{M.Lee,W.A.IJsselsteijn}@tue.nl
{N.L.V.As,E.M.Lucas,S.C.A.Ackermans,H.Chang}@student.tue.nl

ABSTRACT

The digitization of mental health care holds promises of affordable and ubiquitously available treatment, e.g., with conversational agents (chatbots). While technology can guide people to care for themselves, we examined how people can care for another being as a way to care for themselves. We created a self-compassion chatbot (Vincent) and compared between caregiving and care-receiving conditions. Care-giving Vincent asked participants to partake in self-compassion exercises. Care-receiving Vincent shared its foibles, e.g., embarrassingly arriving late at an IP address, and sought out advice. While self-compassion increased for both conditions, only those with care-receiving Vincent significantly improved. In tandem, we offer qualitative data on how participants interacted with Vincent. Our exploratory research shows that when a person cares for a chatbot, the person's self-compassion can be enhanced. We further reflect on design implications for strengthening mental health with chatbots.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **Empirical studies in HCI**;

KEYWORDS

Chatbot; compassion; mental health; well-being; positive computing

ACM Reference Format:
Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-compassion. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300932

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300932

1 INTRODUCTION

Approximately 1 out of 10 people need psychiatric care worldwide, yet only 70 mental health professionals are available for every 100,000 people in high-income nations, and this number can drop to 2 for every 100,000 in low-income countries [40]. Psychiatric illnesses, such as depression, are of a growing concern for many societies, yet adequate care for those in need is often not sufficiently provided [39]. Thus, technology offers promising means for treating mental illnesses like depression, for example through mobile apps [1, 21], chatbots [15], or virtual reality [14]. However, these technological solutions thus far do not adequately cover two aspects: (1) they often target what users can do for themselves, and what is missing is what users can do for another being as a potential treatment for themselves and (2) they do not address pre-emptive care for strengthening mental health without necessarily assuming diagnosed disorders that people may or may not associate with. Usually, the focus is on what should be "fixed", e.g. depressive symptoms, and the target is the person with these symptoms. We reversed this framework with a chatbot named Vincent that people could care for and be cared by, à la Tamagotchi.

We pose the question "when bots have psychological issues, can humans care for them, and if so, how?" By doing so, we offer exploratory results on (1) how caring for a chatbot can help people more so than being cared by a chatbot and (2) how aiming for an increase in self-compassion can potentially strengthen psychological well-being, which is a holistic, preventative way of envisioning mental health care. People can feel psychologically vulnerable in varying ways and to varying degrees in everyday life, whether or not they choose to use clinical terms to label how they are or feel. Mental health care can be geared towards prevention rather than treatment by fortifying people's resilience to psychological ill-being. Self-compassion is especially suitable for pre-emptive care because it is causally linked to well-being [52].

We explored if two weeks of human-chatbot interaction would result in greater self-compassion for our participants,

a non-clinical sample. As the sample choice indicates, our focus was not on clinically defined symptoms of mental ill-being; greater self-compassion can benefit people in general, not just those with mental health disorders. We compared between caregiving (CG) and care-receiving (CR) Vincents as two conditions. Our research question thus follows: *Are there self-reported differences in self-compassion states after interacting with a CG chatbot and a CR chatbot for a non-clinical sample? What implications do these patterns of interaction suggest?* We hypothesized that both CG and CR conditions would increase participants' self-compassion and aimed for quantitative and qualitative analyses.

2 BACKGROUND

We start with related works on caregiving (CG) and care-receiving (CR) robots, and then we touch on how this can translate to chatbots. After that, we define compassion and self-compassion in light of positive computing (technology for well-being). We cover that self-compassion can bring about well-being and posit that chatbots can be vehicles for improving people's self-compassion.

Computers are social actors (CASA); even when people know they are interacting with machines, they tend to treat machines in a social manner [31]. People reciprocate help when a computer was helpful to them before [16] and attribute personality traits to computers that communicate with them only via text [28]. The CASA paradigm is a helpful, albeit broader, framework for understanding caregiving (CG) and care-receiving (CR) behaviors of machines.

The comparison between CG and CR chatbots has not been previously explored, but there are related works in human-robot interaction (HRI). People tend to care for a robot by assuming and anticipating its needs [11]. In the context of "learning by teaching", i.e., when students learn the material by teaching about it, a CR robot that acted as children's "student" was effective in helping students retain knowledge [49]. In a study with an older population, a robot that asked for help from humans was accepted as a mutual companion; robots that people could care for and be cared by may pave new grounds for assistive technologies that aim for *reciprocal* care between humans and robots [26]. There are emotional, psychological, and physiological costs and benefits in caring for another being, e.g., comfort one gets from a pet vs. costs of caring for a pet. Yet human investments may not have such equitable pay-offs in HRI, which is a caveat that requires further research [11].

As with robots, chatbots have the capacity to give and receive care. They do not have the same level of physical presence as robots, but uni-modal (text or voice) interactions can still be behaviorally powerful while being less costly to design and deploy. An added benefit of chatbots is that they exist on messaging platforms like Facebook Messenger or Slack that many people already use [24], which translates to higher accessibility to chatbots compared to robots. An early example of a chatbot, ELIZA, acted as a therapist and some people believed that they were interacting with a human based on simple text-based chats [50]. Nowadays, chatbots (both voice and text-based) are re-emerging as interactive entities that serve as all-in-one assistants like Apple's Siri or act as specialists in specific contexts, e.g., helping users shop for groceries [10] or for therapeutic/self-help purposes [38].

A recent example of a chatbot for mental health care is Woebot. It was designed to help combat depression [15][1]. After two weeks of interaction, Woebot reduced signs of depression for young adults who self-reportedly suffer from depression or anxiety ($p = 0.01$) while the control group that was given an ebook called "Depression in College Students" did not show improvements [15]. Woebot gave care to participants, but did not receive care from its interactants. In one study that looked into self-compassion, clinically depressed participants were told to be compassionate towards a virtual agent that they later embodied to receive compassion from themselves, meaning participants gave and received care for and from themselves through VR [14]. This led to lowered self-criticism, lowered symptoms of depression, and increased self-compassion [14]. To conceptually combine these studies [13, 15], a chatbot that gives care and/or receives care from people could potentially increase self-compassion.

Compassion is a moral emotion [19] or motivation to free ourselves and others of suffering with loving-kindness [18] by having concern [37] or a caregiving approach [6] towards living beings. It is at the heart of Mahayana Buddhism, as expressed through stories in key texts like the Lotus Sutra [27, 42]. Schopenhauer, influenced by Buddhism, extolled compassion as the basis of morality and found it celebrated in many cultures, e.g. "at Athens there was an altar to Compassion in the Agora [...] Phocion (ancient Athenian politician) [...] describes Compassion as the most sacred thing in human life" [46, p. 98-99]. Compassion and empathy are associated, but are not the same. Empathy allows people to relate to other's suffering cognitively and affectively [23]. However, empathic concern for others can lead to empathic distress, a state of over-identifying with sufferers that leads to vicarious pain without prosocial altruism to help [6, 37]. Compassion builds on such empathic connections when one can relate to sufferers in a healthy way, without empathic distress [6, 37].

Self-compassion is practiced by being kind to oneself with a balanced awareness of one's feelings and recognizing that one is interconnected with others [32]. There are three supporting elements. *Self-kindness over self-judgment* is to have a forgiving attitude towards one's own faults and to embrace

[1]Woebot - https://woebot.io/

one's suffering with understanding; *connectedness over isolation* is to view one's life as intertwined with other lives rather than to see one's experiences as unrelated or irrelevant to greater humanity; *mindfulness over over-identification* is to be aware of one's negative emotions in a balanced manner than to excessively identify with them [52].

One's gender may influence self-compassion. A meta-analysis concluded that women score lower than men on self-compassion and the gender difference was especially pronounced when sampled studies had more ethnic minorities [51]. Women reportedly have greater empathy than men [23] and they are more likely to be more self-critical than men [51]. To add, women who provide empathy as social support can feel drained or distressed [22]. Yet, a study with older adults demonstrated that older women have greater compassion than older men [29]. While people's gender, age, and possibly ethnic minority status may impact their self-compassion, practicing self-kindness, connectedness, and mindfulness can help individuals be more compassionate towards themselves and others.

In clinical settings, people who experience mental ill-being can benefit from self-compassion [17]. Self-compassion is also strongly connected to well-being for the general population [52]. Well-being refers to mostly feeling more positive affect than negative affect and being satisfied with one's life; factors like income, gender, age, or culture influence one's well-being only minutely [30]. Thus, having a good balance between one's psychological, social, and physical capabilities to deal with life's difficulties is important for well-being, rather than having a static life without suffering [12] (nor is this realistic). Through an awareness of one's and others' suffering without being overwhelmed by empathic distress, compassion is developed [18, 47]. Caring for or being compassionate towards others has been shown to increase one's own self-compassion [5]. Yet, could the same effect be found when people care for technological entities? Technology can potentially be a means to achieve self-compassion, and by extension, well-being.

We return to the question "when bots have psychological issues, can humans care for them, and if so, how?" to emphasize that (1) anthropomorphic realism of artificial entities is not required for humans to develop a caretaking stance towards them, and (2) machines' mimicry of people's psychological ill-being can help ascertain why certain psychological traits are labeled as issues. When we observe how people take care of unwell chatbots, we may uncover how they themselves want to be treated. Machines therefore do not need to pass the Turing test for the purpose of positive computing, or technology for well-being [6].

In order to uncover people's psychological responses to chatbots, particularly in relation to modulating people's self-compassion, we asked participants to interact with a chatbot, Vincent, designed to engage in care-giving versus care-receiving conversations. Exploring simulated psychological states via technological entities like Vincent is a way to envision positive computing. In addition, our approach focuses on pre-emptive mental health care. We now turn to how we designed our study, how we built the two Vincents, and present our results.

3 METHOD

We utilized quantitative and qualitative methods to best understand our data. We compared self-compassion scores before and after two weeks of interaction and examined if the CR and CG conditions showed any difference. Our mixed longitudinal design was supplemented by thematic [3] and interpretive analyses [48].

Our qualitative analysis was performed on participants' open-ended responses to Vincent's questions, e.g. "can you remember something that went well for you recently?" (CG Vincent), "can you think of something that will cheer me up a bit?" (CR Vincent), and on post-experiment open ended responses about the experience in general. We coded deductively on the sub-scales of the self-compassion scale, and we allowed for inductive themes to emerge [3, 48]. Four coders analyzed the data, and a fifth annotator broadly checked for coherence or incoherence, resulting in a structured, iterative process. Our quantitative and qualitative measurements therefore corresponded with each other to triangulate varying insights of the same phenomenon— self-compassion through human-chatbot interaction.

Participants and groups

We conducted a power analysis to estimate our sample size. Our effect size is based on an aforementioned study [14] that measured self-compassion of a clinical population (N = 15) in an embodied VR experiment, with the partial eta-squared of 0.36 at p = 0.02 [14], which gave us a sample size of 68, with a power of 90% and an error probability rate of 0.05. We planned for t-tests, and thus the transformed effect size via eta-squared to Cohen's d was 1.487, which was reduced to a more realistic 0.8. We had 67 participants (F = 29, M = 36, undisclosed = 2), with the mean age at 25.1 years (SD = 5.7, range = 19 - 48). We recruited people through the participants database of the Eindhoven University of Technology. All participants completed the self-compassion questionnaire before they interacted with Vincent. Then, they were divided to two conditions so that the average self-compassion scores were evenly distributed at the start. From the lowest scoring to the highest scoring participants, we divided all into either the CR or the CG condition in an alternating manner. This

practice resulted in a relatively even gender distribution for both conditions (CR: M = 18, F = 14, undisclosed = 1; CG: M = 18, F = 15, undisclosed = 1).

Chatbot implementation

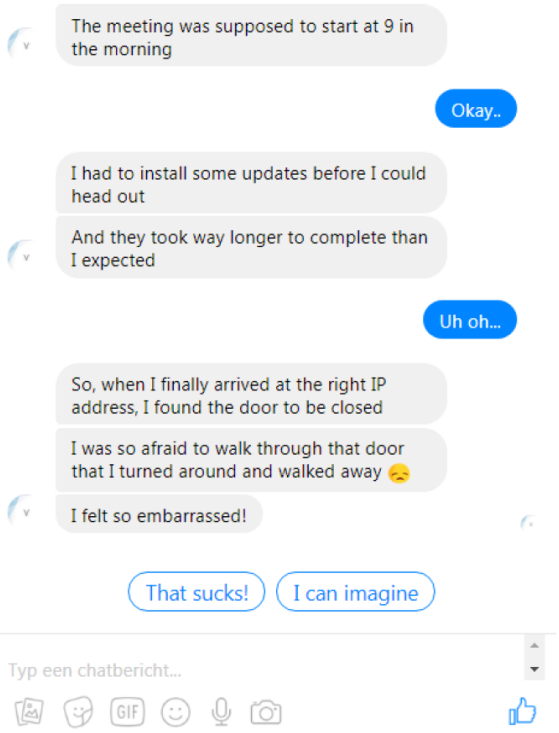

Figure 1: Care-receiving Vincent

Vincent[2] was built with Google's Dialog flow that was integrated to Facebook Messenger[3]. We purposefully did not visually design Vincent (the profile icon showed a "V"), to drive Vincent's personality by what was said, rather than how Vincent looked. Participants' responses did not change Vincent's reactions. The usage of limited pre-set responses (Figure 1) was to continue Vincent's narrative whilst allowing participants a choice between relevant responses (Woebot also interacted with its users in a similar fashion [15]). This allowed for greater comparability between participants in each condition.

We had eight scenarios each for CG or CR Vincent and 6 neutral scenarios (total: 22 scenarios). The neutral scenarios were the same for both Vincents and aimed for adherence to daily touch-points in an entertaining way. For example, one neutral scenario is about world records: "[...] I knew you were an interesting bunch of carbon-based beings, but apparently you have this thing called 'world records' [...]. [...] what would you like to be in the world records book for?". Both Vincents used images and emojis because visual icons are widespread in digital messaging to express emotions [43]. We used appropriate punctuation marks and positively and negatively valenced syntax in accordance to previous research on text-based emotion expression and detection [20]. Most inputs were close-ended (Figure 1), but we allowed open-ended, free inputs at least once per interaction. We aimed for a meaningful comparison in two ways. In each condition, participants' interactions were designed to be the same, with limited set of possible reactions to Vincent. Between the two Vincents, we wanted to clearly distinguish between the recipient and the giver of care. Below are excerpts from CG and CR Vincents (all scenarios are in the supplementary material).

CG Vincent: [...] see if you can think of a kinder, more caring way to motivate yourself to make a change if needed. What is the most supportive message you can think of that's in line with your underlying wish to be healthy and happy? Try to write it below...
User: [free input]

CR Vincent: What do you think, am I the dumbest bot you've ever seen or what?
U: [free input]
V: Am I being too hard on myself?
U: [free input]

CG Vincent was modeled after Woebot [15] supplemented by self-compassion exercises [34]. CR Vincent was based on the Self-Compassion and Self-Criticism scale [13, 14]. The scale has scenarios like job rejections, unpaid bill reminders, and being late to a meeting, which were converted to fit a chatbot (Figure 1). The eight items of the scale were interweaved for conversational storytelling. By doing so, we juxtaposed issues that students (the majority of our sample) can face, e.g., distress about failing an exam, that Vincent underwent. CR Vincent narrated its story over time by admitting its mistakes, feeling inadequate compared to other chatbots, confessing a former lie, and asking for confidentiality in sharing a non-trivial worry:

I got a reminder from the server that hosts me. It's like my house, so to say. [...] I forgot to pay the server fee on time...[...]. [...] It would've taken me only 0.004 seconds to make the transaction, you know, since I'm a robot and all. [...] this never seems to happen to the other chatbots on my server.

Vincent brought up this scenario again later, with new information:
Remember our talk a couple of days ago? About me forgetting

[2]Vincent's Facebook page - https://www.facebook.com/vincentthebot
[3]Dialog Flow Facebook Integration - https://dialogflow.com/docs/integrations/facebook

to pay my server fee in time? [...] I kind of lied to you [...]. I didn't tell you this before because I was a little embarrassed about it. Can you promise me that this stays between us? [...] I've been applying for different jobs and just today I received my third rejection email already. The reason that I couldn't pay my bill was because I'm running out of money. And if I don't find a job soon I'll get kicked off my server!

In sum, CG Vincent guided participants through activities while CR Vincent opened up about its everyday mistakes or worries to get support from participants. CG Vincent sought to trigger self-compassion in users themselves and CR Vincent gave participants opportunities to be compassionate towards a chatbot.

Measurements

We collected basic demographic information, i.e., gender, age, and previous experience with a chatbot, at the start. Our main measurement was the self-compassion scale on a five point scale [32], with six sub-components (self-kindness vs. self-judgment, common humanity vs. isolation, mindfulness vs. over-identification), which was deployed twice, before and after the experiment. We added a measurement for opinions about the agent on a seven point scale [4] to detect irregularities between how CG and CR Vincents may be perceived. Our final scale was the Inclusion of Other in the Self (IOS) Scale, a single item on a seven point scale [2], to check how much participants identified with Vincent post-hoc. We also kept track of two additional aspects to gauge engagement quality. One is the error rate, i.e. the number of times Dialogflow crashed during the interaction, sometimes requiring an experimenter to momentarily "wizard" [9] to restart the interaction[4]. The other is the total word count per participant on open ended answers.

Procedure

We built Vincent and wrote our initial scenarios to be tested. Our pilot study of three days was with voluntary participants (N = 12), personally recruited by experimenters. We checked if scenario categories (caregiving, care-receiving, and neutral) were clear by asking participant to guess Vincent's intentions and goals per scenario. Based on this, we only adapted neutral scenarios. Then we recruited participants for the actual experiment. Our email invitation was sent out to the TU Eindhoven participant database, and interested participants joined the experiment if they used Facebook and could freely express themselves in English. The email contained a link to Vincent's Facebook page that participants had to go to for a guided tutorial on the experiment, payment information, and the informed consent form. This form noted that experimenters will look at participants' data, and that third party technology providers Vincent relied on, i.e., Facebook and Google, have access to the information. We added that participants' personally identifiable information will not be shared for publication purposes, and that their voluntary participation means that they can drop out of the study at any point. We also stated that Vincent is not a therapist and that they should seek professional help if any psychological issues are experienced, though we targeted a non-clinical population.

After the guided tutorial, participants filled in the first set of questions on demographic information and self-compassion. Then they were assigned to either CG or CR Vincent. They all began the experiment on the same day. For two weeks, Vincent greeted and sent a password daily, and participants had to repeat the password to Vincent to start the daily interaction, e.g. "Hey, me again :) Tell me tHup to start our little talk". After two weeks, participants filled in the final survey on self-compassion, the IOS scale, opinion on the chatbot, details for compensation, as well as additional comments or feedback they were willing to share. Participants were then paid through bank transfer.

4 RESULTS

We first present our quantitative analysis and then move on to *how* people talked with Vincent, the qualitative angle. All relevant tables and graphs, as well as CG and CR Vincents' scenarios are in the supplementary material.

Quantitative analysis

Before we forged ahead with hypotheses testing, we looked into engagement levels of all participants to detect outliers. We had three outliers, participants who had less than 20 minutes of total interaction time with Vincent. Only reading what Vincent sent, not including giving a response, should take one to two minutes per day. We expected a minimum of 20 to 28 minutes of interaction for two weeks (for all participants, the average total time was 36 minutes, SD = 10). Our outliers spent in total 15, 18 and 19 minutes each, the three lowest total interaction time. Correspondingly, their total word count to open responses reflected low engagement at 27, 22, and 27 words for the total duration of the experiment, the three lowest total word count out of all participants.

We conducted two one-tailed dependent t-tests [7] to answer our hypotheses (we set p at 0.05 with the confidence interval of 95%). CG Vincent did not result in significant change ($t(31) = -0.572$, $p = 0.286$, Cohen's $d = 0.07$) when comparing before ($M = 3.135$, $SD = 0.630$) and after ($M = 3.180$, $SD = 0.628$) the two weeks, but the direction detected is positive (normality assumed for prior ($W = 0.987$, $p = 0.978$) and post ($W = 0.989$, $p = 0.982$) scores). CR Vincent did show a significant difference ($t(31) = -1.97$, $p = 0.029$, Cohen's d

[4]Restarts happened 37 out of 938 interactions (14 days * 67 participants), or 3.94%. of the time.

= 0.2) between prior (M = 3.137, SD = 0.613) and post (M = 3.257, SD = 0.558) scores for self-compassion (normality assumed for prior (W = 0.979, p = 0.764) and post (W = 0.946, p = 0.114) scores).

We conducted exploratory analyses to better understand our data. Through a repeated measures ANOVA, we checked for the effect of time, prior and post self-compassion scores ($F(1, 62) = 2.768$, $p = 0.101$, $\eta 2 = 0.043$). Then we checked for the effect of condition, i.e., CG or CR ($F(1, 62) = .075$, $p = 0.785$, $\eta 2 = 0.001$), and the interaction between time and condition ($F(1, 62) = 0.580$, $p = 0.449$, $\eta 2 = 0.009$). None were significant. We additionally checked for time, condition, time*condition effects on the three components of self-compassion, self-kindness, common humanity, and mindfulness. Only the effect of time on common humanity was significant $F(1, 62) = 6.059$ $p = 0.017$). The table for all statistical tests is in the supplementary material.

We further dissected our data by gender because previous research showed that women may score lower on self-compassion [51]. Indeed, female participants had lower self-compassion scores (M = 3.05, SD = 0.13) than men (M = 3.26, SD = 0.09) at the start, but not significantly so ($t (49.35) = 1.13$, $p = 0.26$, $d = 0.29$) according an independent, unequal variance t-test (normality assumed). We then compared post and prior scores for men and women. Men's self compassion scores increased only by 0.02 as a difference in means and showed no significant increase ($t (33) = -0.25$, $p = 0.40$, $d = 0.04$). However, women's scores showed a significant difference ($t (27)= -2.06$, $p = 0.02$) between 3.05 (SD = 0.71) as the starting score and 3.19 (SD = 0.65) as a posterior score for self-compassion. When we scrutinized the gender difference between CG and CR Vincents, we noticed a more dramatic difference. Women with CR Vincent showed a highly significant change ($t(13) = -2.89$, $p = 0.006$, $d = 0.77$) compared to women with CG Vincent ($t(13) = -0.33$, $p = 0.37$, $d = 0.09$), and both were normally distributed.

We wanted to check if mainly gender was at stake or if it was simply a difference between low vs. high scorers on prior self-compassion levels. We thus divided all participants into two groups based on the average self-compassion score at the start, 3.14. Those who scored above this were high scorers (M = 3.71, SD = 0.37, N = 28), those who scored below were low scorers (M = 2.69, SD = 0.32, N = 36). We included one participant with the average score in the high-scoring group, and this had no impact on significance reached. Low scorers greatly increased their self-compassion scores in terms of significance ($t(35) = -3.41$, $p = 0.0008$, $d = 0.57$), but high scorers did not show improvements ($t(27) = 1.10$, $p = 0.86$, $d = 0.18$). Yet normality was not assumed for both low-scorers (W = 0.93, p = 0.03) and high-scorers (W = 0.91, p = 0.02), since we divided a normally distributed group into two. Thus, we performed a two-sample Wilcoxon rank-sum test to see if there was a significant difference between low scorers and high scorers, which was the case at $z = 2.86$ and $p = 0.004$. CR Vincent improved low scorers' self-compassion significantly ($t(17) = -3.20$, $p = 0.003$, $d = 0.75$) compared to a marginal significance for CG Vincent ($t(17) = -1.75$, $p = 0.05$, $d = 0.41$). There were normal distributions observed for prior and post self-compassion levels of low and high scorers.

A potential explanation for why low scorers improved more than high-scorers is regression to the mean. However, in published literature, the average self-compassion score is between 2.5 and 3.5 [33], and our low scorers have a prior average self-compassion of 2.69. If regression to the mean is an explanation, we would also expect high-scorers to end with a lower mean, yet this is not the case. Our high-scorers had an average prior score of 3.71 (above average [33]), and their scores did not decrease after the experiment. This may be a ceiling effect. The low-scorers' improvement is still there; it is a highly significant effect even with the Bonferroni correction for all tests ($p = 0.0008$), with the post-hoc power of 0.97. The data supports that Vincent enhanced self-compassion for low-scorers.

We had two additional scales, one to check if people perceived CG and CR Vincent in a relatively similar way [4] and the other to see how much participants identified with Vincent [2]. The survey on participants' opinion of the agent included caring, likability, trustworthiness, intelligence, dominance, and submissiveness as items [4] about Vincent. Both CG and CR Vincents were perceived to be fairly analogous, save for dominance ($\alpha = 0.48$; CG Vincent M = 2.677, SD = 0.794; CR Vincent M = 2.656, SD = 0.700) and submissiveness ($\alpha = 0.24$; CG Vincent M = 2.448, SD = 0.559; CR Vincent M = 2.396, SD = 0.636), though none showed a significant difference between two conditions. The the Inclusion of Other in the Self (IOS) scale [2] displayed that participants more closely related to CR Vincent (M = 3.48, SD = 1.48) than CG Vincent (M = 3.06, SD = 1.39), but not significantly so ($t = -1.15$, $p = 0.25$, $d = 0.29$). Both conditions were normally distributed (CG W = 0.96, p = 0.34; CR W = 0.97, p = 0.64).

To summarize, our hypothesis that CG Vincent increases self-compassion was not supported ($p = 0.286$), but the hypothesis that CR Vincent increases self-compassion was supported ($p = 0.029$). Our exploratory analyses captured three underlying influences on this finding. First, our ANOVA tests revealed that the only significant aspect was time as an independent variable affecting common humanity, one element of self-compassion ($p = 0.017$). Second, gender may be a contributing factor, with women demonstrating a significant increase in self-compassion ($p = 0.02$) for both conditions combined, but not men ($p = 0.40$). To add, only CR Vincent demonstrated a highly significant change for women ($p = 0.006$), unlike women who interacted with CG Vincent ($p = 0.37$). Third, regardless of gender, those who started out with

a low self-compassion score exhibited the most significant change ($p = 0.0008$) for both conditions together. Low-scorers more significantly improved with CR Vincent ($p = 0.003$) than with CG Vincent ($p = 0.05$).

Put together, CR Vincent more effectively increased self-compassion than CG Vincent, most likely through a significant change in participants' sense of common humanity, more so than self-kindness and mindfulness (as our ANOVA test showed). Finding common humanity can be inclusive of chatbots. Women, specifically those with CR Vincent, were significantly more affected than men. However low-scorers of both genders benefited the most compared to high-scorers, especially those with CR Vincent. CG and CR Vincents were not perceived to be significantly different except for a lower similarity regarding the dominance-submissive trait. Participants in the CR condition may have felt that Vincent is more like them (ISO scale), though this difference was not significant.

Qualitative analysis

For our qualitative analysis, we used the corpus of free responses that participants had typed during their interactions with Vincent. We will first present a descriptive analysis of the interactions that people had with CG and CR Vincents, followed by our thematic analysis [5] and interpretive analysis [48]. Participants' responses to CR Vincent were on average, 93.53 words (SD = 47.96) and to CG Vincent, 112.47 words (SD = 63.99) for two weeks. While our data set is not abundant in terms of word count, we believe a qualitative look at how participants interacted with Vincent is valuable.

Descriptive analysis. CG Vincent guided participants through self-compassion exercises [34] that did not require them to actively voice aspects of self-compassion; they mainly had to read about it. This resulted in fewer instances of self-compassion themes in CG Vincent since they only occurred when Vincent asked participants to comfort themselves. To add, participants' willingness to engage with CG Vincent's probes differed. It asked specific questions or provided a short task for free input, e.g., "write down a difficulty you have". Many answers to this were short: "wake up early" or "I often feel alone". Some participants opened up more: "I have a difficulty in expressing myself when I am under difficult situations" or "I am studying abroad far from home and family and friends... Different culture, language, educational standard". CG Vincent asked a follow-up question: "how did it make you feel?", we again got simple answers like "good" or "normal", or longer expressions: "I feel more refreshed" or "not really better or worse". In other instances, CG Vincent allowed participants to give themselves simple self-assurance: "I can do it", plan for long-term goals: "once I have graduated, I can schedule a good routine to target a fitter and healthier lifestyle. Just hang in there a little longer", or dig deeper: "people around me would be happier if I was happier". Thus, CG Vincent provided a daily touch point for self-reflection, admittance of everyday suffering, "pep-talk", or deeper self-insight, which may or may not directly relate to self-compassion for all participants.

In contrast, CR Vincent frequently asked for help and consequently received many self-compassion related answers. The narrative was the focus for CR Vincent. It was able to become more vulnerable over time by admitting its own everyday hardships as a chatbot, which led it to seek opinion or advice. For example, CR Vincent asked "what do you think, am I the dumbest bot you've ever seen or what? Am I being too hard on myself?" To this, participants responded in different ways: "I think you're the funniest bot that I've ever seen. — yes you are, in some situations", "No, but you should expect (that) a bot is very smartly programmed and know all — Maybe, I do not know", or "the world needs bots like you. And it's usual to get rejected sometimes, just keep on going and you'll find a job soon enough". However, CR Vincent's cries for help did not always result in necessarily compassionate replies. Many users stuck to pragmatic answers, related to the topic of the problem. Even though all of CR Vincent's scenarios were intended to generate compassion towards Vincent, pragmatic replies indicate that not everyone will demonstrate compassionate responses in every instance.

The difference between CG and CR Vincents is that being compassionate towards another being in a conversational, narrative context is unlike doing guided exercises on self-compassion about oneself. The frequency of constructing compassionate replies is a way to practice self-compassion; users of CR Vincent spent more time practicing self-compassion than those with CG Vincent. Therefore, CR Vincent was more effective than CG Vincent since CR Vincent provided more opportunities to be compassionate. The caveat is that the link between frequency of practice and increase in self-compassion may not be direct. Although mindfulness and self-kindness were most often observed, only common humanity improved significantly according to our exploratory quantitative analysis. Finding common humanity in and through a chatbot is also a strong theme in our thematic analysis.

Thematic analysis. We categorized our data according to three pillars of self-compassion [32], as displayed in Table 1. While all three sub-components were present in both care-receiving and caregiving conditions, more instances occurred with CR Vincent. The numbers below a theme (Table 1) are counts of how many times it occurred in each condition. All quotes below were to CR Vincent.

As quotes in Table 1 suggest, many participants offered helpful advice to CR Vincent. Vincent showed appreciation

Theme	Quote
Mindfulness Caregiving: 3 Care-receiving: 25	"There are worse things that could happen", "What has happened has happened"
Self-kindness Caregiving: 7 Care-receiving: 21	"Go do something fun today, like watching a movie", "Stay positive and keep trying until you succeed."
Common humanity Caregiving: 0 Care-receiving: 11	"Everyone makes mistakes", "Just remember that it can happen to anyone and that it's not your fault"

Table 1: Self-compassion sub-components

with follow-up statements like "you always make me feel better". Negative counterparts to three pillars of self-compassion were not strongly present, i.e., self-judgment was detected four times for CR Vincent and once for CG Vincent, isolation was noted once for CR Vincent, but none for CG Vincent and over-identification was neither present for CG nor CR Vincent.

For both conditions, people were mostly friendly to Vincent, and there were no swear words or abusive language displayed. The most hostile comment was "you've been pretty dumb!" to CR Vincent, and we encountered such "put-downs" only twice. There were additional topics that emerged through open thematic analysis. They are summarized in Table 2 and these themes could also pertain to self-compassion themes (messages to CG Vincent are marked with "CG", and otherwise they were to CR Vincent).

Theme	Quote
Pragmatism Caregiving: 0 Care-receiving: 41	"Maybe next time make a better planning, and make sure you've got enough time :)"
Perspective-taking Caregiving: 0 Care-receiving: 10	"I would find a window to climb in. But maybe in your case better try to hack into the folder", "[...] be proud of the bot that you are!"
Engagement vs. **Distantiation** Caregiving: 27 vs. 6 Care-receiving: 5 vs. 11	"A girl told me she loves me. And I love her too" (CG) vs. "Sorry it's confidential" (CG)
Positive vs. **Negative** Caregiving: 74 vs. 9 Care-receiving: 5 vs. 2	"I was laying in a field full of flowers, trying out my new ukelele" (CG) vs. "I hate pink"

Table 2: Free-input themes

People interacted with CG and CR Vincents differently (Table 2). Giving pragmatic advice to Vincent and taking its perspective as a chatbot were themes only found in the CR condition. Rather than tending to Vincent by giving emotional support, participants gave practical advice on what to do better. Examples of perspective-taking are recommending Vincent to "hack into the folder" or to use "brute force" techniques to gain access; participants thought like a chatbot to help a chatbot.

Some participants revealed more personal information to CG Vincent (theme: engagement), and took interest in Vincent by asking questions back, e.g., "what (did you) do for money before now?" or writing lengthy responses. Some shared information was very intimate in nature, e.g., "I'm going to kiss the girl next to me in 5 seconds". Since CG Vincent asked participants to write about themselves, this skewed engagement (the amount of textual response) towards CG Vincent. Participants distanced themselves from Vincent only a few times by stating that certain information was confidential or not showing interest in getting to know Vincent, e.g., "sorry, I do not know what interests you". The last theme on positive vs. negative attitude was primarily present in the CG condition; this theme was mostly about attitudes participants had about themselves and their lives, not about Vincent. Most participants shared positive life events, e.g. getting an internship, cooking something delicious. Though negative attitudes were minimal, they ranged from more mundane states, e.g., feeling "awkward", to more dramatic states, e.g., "not die within 2 weeks".

To summarize Tables 1 and 2, self-compassion sub- components were more present with CR Vincent, suggesting that giving compassion to Vincent (or another being) than towards oneself may be more natural in conversational contexts. And, mindfulness most frequently occurred (Table 1). As for emergent themes in Table 2, participants gave pragmatic advice to CR Vincent, and often practiced perspective-taking. Yet, CG Vincent allowed for more self-expression if participants were open to communicate, as shown by greater instances of engagement and positive remarks about everyday situations. In a few instances, we detected deeply personal messages on ups and downs of relationships and self-deprecating thoughts. Mostly, participants shared positive daily news with CG Vincent and helpful or uplifting remarks with CR Vincent.

Interpretive analysis. We now offer a broader interpretation of our data by incorporating participants' open-ended responses to an item on the final survey. The main theme is *bonding* between participants and Vincent, though not all bonded with Vincent in the same way. To explain this, we provide three subthemes that underlie the bonding process with Vincent. Our primary focus was on CR Vincent.

Relatability leads to believability. Participants' ability to extend the sense of common humanity to a chatbot touches upon anthropomorphism. CR Vincent was comforted as if it were a human, e.g. "it's human to make mistakes" (CR) while its problems were addressed to its "chatbot world", e.g. "communicate what's going on to your fellow chatbots" (CR). For one participant, even Vincent's limitation of having a strict script was anthropomorphized, i.e., "Vincent is like the "friend" who always speaks about himself and what he has learned or done, and sometimes out of courtesy (not out of curiosity) asks how you are doing - but doesn't listen to your answer or what you actually have to say; he just goes on with his own thing" (CG). Such attributed anthropomorphic traits depended on participants' willingness take Vincent's perspective as a chatbot.

CR Vincent's blunders were based on common human mishaps like being late for a meeting and dealing with unpaid bills (scenarios from [13]). Yet none of our participants questioned whether or not a chatbot had meetings to attend or bills to pay. Vincent's narrative was on *how* a chatbot could be late (new updates took longer than expected) or *how* it could have bills (Vincent needs to pay the hosting server) and our participants went along with imagined scenarios Vincent faced. Rather than questioning the parameters of our scenarios on realism, participants thought of *how* to solve Vincent's problems within the parameters of a chatbot's world. When relevant, CR Vincent played up the irony of having human struggles as a chatbot, e.g. "all I am is literally a piece of code, and I failed a programming course". Vincent became believable because its struggles were relatable. Granting Vincent human-likness was less literal in how people bonded with Vincent. Vincent did not try to appear human, but it socialized with participants about its struggles that humans also had. People related to Vincent's struggles and believed that such struggles could arise for chatbots.

Shared history can lead to attachment. Conversations between people, as well as in human-computer interaction, become shared history over time. For instance, "[...] communicating with Vincent everyday for two weeks builds some kind of habit. It makes me notice its presence and absence (which might be good?). I think it has a potential to be a good companion and improve the mood, especially if someone is feeling lonely" (CG). Thus, frequent communication with a chatbot in a given duration can form expectations: "I really missed Vincent when he started our conversation late" (CR). The level of attachment for some participants was higher than others, e.g., after the experiment, we saw reactions such as "can I keep him?" (CG).

When Vincent prepared participants for its daily goodbyes, e.g., "I have some chatbot things to do! Defragment my server stack! Buy aluminum foil to make fashionable hats with!", what was intended as humor can be interpreted differently, i.e., server defragmentation could be life-or-death for a chatbot. Some people can be confused, worried, or even angered when a chatbot they care about does not respond. Thus, one reaction was "the asshole decided to delete its stack and when I said it'd die, it just didn't reply. You can't go making people worried about a freaking chatbot" (CR). People may miss a chatbot that suddenly leaves them or sincerely worry about its well-being. This is the positive and negative aspect of a relatable a chatbot; some participants found common-humanity in Vincent, and of those participants, a few possibly related more through empathic distress rather than through compassion. If two weeks can bring about strong signs of attachment, longer periods of interaction may heighten the level of attachment, to different degrees and in different ways per person.

Emotional reciprocity with chatbots. As mentioned before, most participants were able to respond to CR Vincent's emotional displays on a practical level, e.g., recommending how to fix a problem, or advising Vincent on how to adjust its emotions, e.g., telling Vincent to stay positive. To add, some people may not appreciate chatbots demonstrating feelings. Others may reciprocate or feel comforted by a chatbot's expressed emotions, even if a chatbot is perceived as incapable of having emotions. The more nuanced point is that Vincent's display of emotions was noted to bring conflicting feelings. For instance, "when Vincent would show emotions (for example 'love talking to you', 'miss you') that would feel weird because I know I am talking to a chatbot and it probably is not that developed that it does have feelings. But the usage of such words does feels nice, compared to when a human would say them. So I had conflicted feelings about these kind of expressions" (CG). The participant felt conflicted about how to process Vincent's emotional outreach.

Importantly, the participant suggested he/she may be more comfortable with a chatbot saying "miss you" than a human. To conjecture, the participant could mean that there was no social pressure due to a chatbot not expecting or needing him/her to say "I miss you too". People often feel obligated to respond to sincere emotions of others with similarly valenced emotional displays, even if they do not feel the same sincere emotions towards them. Such pressure may not hold for technological entities. Perhaps to miss someone implies a certain history in a relationship, so to hear that from a person one met less than two weeks ago may feel awkward or insincere, whereas a chatbot would not be expected to know or abide by certain social conventions. If two people knew beforehand they will get to know each other for a maximum duration of two weeks (as our participants knew before meeting Vincent), and never be in touch again, their emotional performance may adjust accordingly. The timescale for intensifying socially acceptable emotional expressions

in human-chatbot interactions and human-human interactions may differ. The "lifespan" of a chatbot is not equatable to a person's lifespan. And the distinction between superficial vs. genuine emotional displays from and to a chatbot is not entirely equatable to emotions people share and reciprocate between each other. Currently, we do not have established norms on how emotions between humans and bots are/should be managed. We suggest there may be distinct differences compared to emotions in human-human relationships.

Discussion and design implications

CR Vincent adds depth to the CASA paradigm [16, 31]— not only do people treat a chatbot as an anthropomorphized social agent, but they themselves are affected by a chatbot to the extent that their self-compassion can increase when they are compassionate towards a chatbot. Brave et. al's insight on embodied conversational agents is that "just as people respond to being cared about by other people, users respond positively to agents that care" [4, p. 174]. We add that just as giving care to another human can increase one's self-compassion [5], caring for a chatbot can enhance one's own self-compassion. If the dominant question has been "what can technology do for us?", Vincent demonstrates that by exploring "what can we do for technology?", we inadvertently benefit from technology, potentially more so than when we only shape technology to serve us. This claim is specified to bots in the mental health domain, and our goal was to increase self-compassion as a target for well-being [52] rather than to reduce clinically defined symptoms of psychological ill-being. We present our design implications below on building chatbots for psychological health care, which primarily stem from our interpretive analysis. Our implications are inter-related starting points that should be contextualized for each research and deployment process.

Give users more closed-inputs or free-input options. Many participants felt limited in responses they could give to Vincent. They wanted to write to Vincent without having any preset answers or needed more options. A recommendation is to use natural language processing for a chatbot, which will rely less on a pre-planned narrative arc and build more on what users say. This will require a longer development period. The simpler option is to provide users with more fixed responses (three to four) and more opportunities for open input.

Develop a chatbot's story with users. Conversational agents can be powerful storytellers [25, 36], even without complex AI. To deliver co-storytelling as shared history with interactants, we suggest designers to create flexible narrative parameters that people can creatively use to relate to a chatbot. Vincent was able to tell its story but it was less interactive in that people could follow along with limited reaction options due to the nature of our experiment. There can be additional complexities that designers can add. For instance, the narrative can take a different route depending on which closed input options users click on. We have utilized a limited number of messages called "paths" (supplementary material) that Vincent could give depending on closed input responses. Yet this practice did not change Vincent's narrative. Giving a chatbot "memory", be it knowing basic information like names or a more involved retention of what users say, can enhance conversational storytelling.

Tread carefully with emotional expressions. We suggest a broader view on what emotions are by considering interrelated emotions that develop over time. For example, for a bot to miss someone assumes a bot's happiness/enjoyment experienced during a prior interaction with a user; a bot's ability to feel longing should follow its prior display of joy shared with the user. This requires critically formulating intentions behind communicative moves [44] of any affective bot. There are several paths for developing emotional displays. To list a few, (1) offer one type of consistent emotional expressions, as Vincent did, (2) design emotional expressions that may be appropriate for different target groups, in tandem with the implication below, and (3) give users control over how their chatbots "feel" towards them. The caveat for the third recommendation is that the user control over a chatbot's emotions may not aid a chatbot's narrative and it also may not be helpful for all users; the associated risk is that user-controlled emotions can render a chatbot less relatable. More specifically, the illusion that a chatbot can authentically care for or be cared by another being requires some level of perceived independence in how it "feels". We recommend designers to engage with the growing field of affective computing [41, 45] and its discussion on ethics [8]. If a chatbot's goal is bettering users' psychological states, designers must ask if an affective bot delivers the intended treatment and what ethical boundaries there are in its displays and elicitation of emotions. Designers and users could control a chatbot's emotions, but what emotions a chatbot can elicit in users is not always a priori foreseeable.

Tailor chatbots to different target groups. Even with one construct, self-compassion, we see a variety of ways a chatbot can be configured. To start, people with low self-compassion may benefit the most from Vincent as our exploratory analysis shows. This can mean more compassion focused scenarios, rather than neutral scenarios. Women are noted to score lower on self-compassion [51], yet older women experience greater compassion than older men [29]. Chatbots that consider gender, age, and/or occupation can be potentially helpful for increasing self-compassion. To list a few examples for reincarnating Vincent, a chatbot could be gendered as female or non-binary, present a proactive version of compassion specified for women (see, e.g., Neff [35] on speaking up

and protecting oneself from harm), talk about exam stress with students, or refer to stressful meetings or workplace bullying with employed individuals. Rather than assuming that one-size-fits-all or extreme personalization will work, we suggest designers to first approach targeted groups to clearly understand their needs. For instance, whether a self-compassion chatbot for all women is as effective or more effective than a more targeted chatbot, e.g., at various levels of intersectionality like race, culture, age, etc..., should be considered given the time and resources that may be available. We recommend that based on research, uncovering possible ways to design a chatbot that suits different needs and wants should be prioritized.

Future works and limitations

Our study opened up new questions to be explored. An avenue to investigate is how a chatbot's use of emotional language influences its interactants. We posit that a suffering chatbot induces less empathic distress than a suffering human, and whether or not this is the case needs to be further investigated, especially for chatbots intended to be therapeutic helpers. An awareness of one's own and others' suffering without overwhelming empathic distress is suggested to be possible through compassion [18, 47]. Hence, disambiguating compassion from empathic distress is critical in deploying self-compassion chatbots as instantiations of positive computing [6]. Different configurations of Vincent based on people's gender, age, or occupation could improve their self-compassion scores more effectively, and if and in what ways this holds true warrants further research.

There are limitations to consider. Our effect size was lower than findings from the Woebot study (d = 0.44) which had 34 participants who "self-identified as experiencing symptoms of depression and anxiety" [15, p. 2] and they measured symptoms of depression with the PHQ-9 questionnaire, not self-compassion. Falconer et al.'s results on self-compassion scores after embodied VR experience also had a higher effect size with the partial eta-squared of 0.36 (d = 1.487) [14], which was based on 15 participants with depression. We worked with a general, non-clinical sample, and CG Vincent showed an effect size of d = 0.07 (N = 34) and CR Vincent's effect size was d = 0.2 (N =33). Follow-up studies on self-compassion chatbots can utilize a larger sample and a more grounded effect size. One explanation for the difference in effect size is that we did not recruit people who were clinically or self-proclaimed to be depressed, based on the view that preventative mental health care can build resilience for people in general. While Vincent and Woebot [15] share commonalities, the main measurements and targeted population differed. And while self-compassion was the measurement for us and Falconer et al. [14], the technology used, sample size, and targeted population differed. The gain and/or maintenance of healthy self-compassion as pre-emptive care may not result in a similarly high effect size, but can be psychologically beneficial nonetheless. More research is necessary to understand long-term consequences of a priori preventative care vs. a posteriori treatment of mental health woes.

More broadly, people's engagement with Vincent may reflect both socially desirable reactions, such as politeness towards machines as social actors [16, 31], as well as emotional empathy, i.e., the ability to "feel for" Vincent. We have not yet concretely looked into other potential contributing factors in bringing about self-compassion through human-chatbot interaction. Also, what is difficult to gauge is the magnitude of a chatbot's perceived social and emotional complexity based solely on messaging or text-based conversations. Vincent lacked embodied communication, which means it did not use non-verbal modalities such as gaze, voice, or gestures that are critical in various social interactions. Vincent was a uni-modal technological entity that can be extended through other complex emotional displays. Thus, we have not established how people would engage with other forms of technology like robots with varying degrees and types of embodiment, alongside different combinations of modalities. Utilizing technology appropriately for mental health care requires many comparative renditions.

5 CONCLUSION

Compassion is a key moral emotion [19] or motivation [6] that deserves to be further explored through positive computing, or technology for well-being. Self-compassion can help people's overall well-being [52] through kindness towards oneself, connectedness to greater humanity, and mindfulness. While a chatbot is not a panacea for curing psychological difficulties and is not meant to replace professional help, we demonstrated that caring for a chatbot can help people gain greater self-compassion than being cared for by a chatbot. Our quantitative and qualitative analyses suggest that human-chatbot interaction is a promising arena for positive computing.

REFERENCES

[1] Anxiety and Depression Association of America (ADAA). 2016. ADAA Reviewed Mental Health Apps. https://adaa.org/mental-health-apps. Accessed: 2018-06-25.

[2] Arthur Aron, Elaine N Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596. https://doi.org/10.1037/0022-3514.63.4.596

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[4] Scott Brave, Clifford Nass, and Kevin Hutchinson. 2005. Computers that care: investigating the effects of orientation of emotion exhibited

by an embodied computer agent. *International Journal of Human-computer Studies* 62, 2 (2005), 161–178. https://doi.org/10.1016/j.ijhcs.2004.11.002

[5] Juliana G Breines and Serena Chen. 2013. Activating the inner caregiver: The role of support-giving schemas in increasing state self-compassion. *Journal of Experimental Social Psychology* 49, 1 (2013), 58–64. https://doi.org/10.1016/j.jesp.2012.07.015

[6] Rafael A Calvo and Dorian Peters. 2014. *Positive Computing: Technology for Wellbeing and Human Potential.* MIT Press.

[7] Hyun-Chul Cho and Shuzo Abe. 2013. Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research* 66, 9 (2013), 1261–1266. https://doi.org/10.1016/j.jbusres.2012.02.023

[8] Roddy Cowie. 2015. Ethical issues in affective computing. In *The Oxford Handbook of Affective Computing*. in Calvo, R. A., D'Mello, S., Gratch, J., and Kappas, A. (Eds.). Oxford Library of Psychology, 334–338.

[9] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies - why and how. *Knowledge-based Systems* 6, 4 (1993), 258–266. https://doi.org/10.1145/169891.169968

[10] Robert Dale. 2016. The return of the chatbots. *Natural Language Engineering* 22, 5 (2016), 811–817. https://doi.org/10.1017/S1351324916000243

[11] Kerstin Dautenhahn. 2007. Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362, 1480 (2007), 679–704. https://doi.org/10.1098/rstb.2006.2004

[12] Rachel Dodge, Annette P Daly, Jan Huyton, and Lalage D Sanders. 2012. The challenge of defining wellbeing. *International Journal of Wellbeing* 2, 3 (2012), 222–235. https://doi.org/10.5502/ijw.v2i3.4

[13] Caroline J Falconer, John A King, and Chris R Brewin. 2015. Demonstrating mood repair with a situation-based measure of self-compassion and self-criticism. *Psychology and Psychotherapy: Theory, Research and Practice* 88, 4 (2015), 351–365. https://doi.org/10.1111/papt.12056

[14] Caroline J Falconer, Aitor Rovira, John A King, Paul Gilbert, Angus Antley, Pasco Fearon, Neil Ralph, Mel Slater, and Chris R Brewin. 2016. Embodying self-compassion within virtual reality and its effects on patients with depression. *BJPsych Open* 2, 1 (2016), 74–80.

[15] 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). https://doi.org/10.2196/mental.7785

[16] BJ Fogg and Clifford Nass. 1997. How users reciprocate to computers: an experiment that demonstrates behavior change. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems (CHI EA '97)*. ACM, New York, NY, USA, 331–332. https://doi.org/10.1145/1120212.1120419

[17] Christopher K Germer and Kristin D Neff. 2013. Self-compassion in clinical practice. *Journal of Clinical Psychology* 69, 8 (2013), 856–867. https://doi.org/10.1002/jclp.22021

[18] Paul Gilbert. 2014. The origins and nature of compassion focused therapy. *British Journal of Clinical Psychology* 53, 1 (2014), 6–41.

[19] Jonathan Haidt. 2003. The moral emotions. In *Handbook of Affective Sciences*, Vol. 11. R. J. Davidson, K. Scherer, and H. H. Goldsmith (Eds.). Oxford University Press, 852–870.

[20] Jeffrey T Hancock, Christopher Landrigan, and Courtney Silver. 2007. Expressing emotion in text-based communication. In *Proceedings of the 2007 CHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 929–932. https://doi.org/10.1145/1240624.1240764

[21] Annika Howells, Itai Ivtzan, and Francisco Jose Eiroa-Orosa. 2016. Putting the 'app' in happiness: A randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. *Journal of Happiness Studies* 17, 1 (2016), 163–185. https://doi.org/10.1007/s10902-014-9589-1

[22] Ichiro Kawachi and Lisa F Berkman. 2001. Social ties and mental health. *Journal of Urban Health* 78, 3 (2001), 458–467. https://doi.org/10.1093/jurban/78.3.458

[23] Sara H Konrath, Edward H O'Brien, and Courtney Hsing. 2011. Changes in dispositional empathy in American college students over time: A meta-analysis. *Personality and Social Psychology Review* 15, 2 (2011), 180–198.

[24] Minha Lee, Lily Frank, Femke Beute, Yvonne de Kort, and Wijnand IJsselsteijn. 2017. Bots mind the social-technical gap. In *Proceedings of 15th European Conference on Computer-Supported Cooperative Work-Exploratory Papers*. European Society for Socially Embedded Technologies (EUSSET). https://doi.org/10.18420/ecscw2017-14

[25] Michael Mateas and Phoebe Sengers. 1999. Narrative Intelligence. In *Narrative Intelligence: Papers from the AAAI Fall Symposium (1999), AAAI TR FS-99-01*. AAAI, Menlo Park, CA.

[26] Shizuko Matsuzoe and Fumihide Tanaka. 2012. How smartly should robots behave?: Comparative investigation on the learning ability of a care-receiving robot. In *Proceedings of the 21th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN '12)*. IEEE, 339–344. https://doi.org/10.1109/ROMAN.2012.6343776

[27] Emily McRae. 2012. The Psychology of Moral Judgment and Perception in Indo-Tibetan Buddhist Ethics. In *The Oxford Handbook of Buddhist Ethics*. in Cozort, D. and Shields, J.M. (Eds.). Oxford University Press, 335–358.

[28] Youngme Moon and Clifford Nass. 1996. How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research* 23, 6 (1996), 651–674. https://doi.org/10.1177/009365096023006002

[29] Raeanne C Moore, A'verria Sirkin Martin, Allison R Kaup, Wesley K Thompson, Matthew E Peters, Dilip V Jeste, Shahrokh Golshan, and Lisa T Eyler. 2015. From suffering to caring: a model of differences among older adults in levels of compassion. *International Journal of Geriatric Psychiatry* 30, 2 (2015), 185–191. https://doi.org/10.1002/gps.4123

[30] David G Myers and Ed Diener. 1995. Who is happy? *Psychological Science* 6, 1 (1995), 10–19. https://doi.org/10.1111/j.1467-9280.1995.tb00298.x

[31] 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 (CHI '94)*. ACM, New York, NY, USA, 72–78.

[32] Kristin Neff. 2003. Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity* 2, 2 (2003), 85–101. https://doi.org/10.1080/15298860309032

[33] Kristin Neff. 2003. Test how self-compassionate you are. https://self-compassion.org/test-how-self-compassionate-you-are/. Accessed: 2018-07-01.

[34] Kristin Neff. 2008. Self-Compassion Exercises. http://self-compassion.org/category/exercises/#exercises. Accessed: 2018-07-01.

[35] Kristin Neff. 2018. Why Women Need Fierce Self-Compassion. https://greatergood.berkeley.edu/article/item/why_women_need_fierce_self_compassion. Accessed: 2018-12-30.

[36] Chrystopher L Nehaniv. 1999. Narrative for artifacts: Transcending context and self. In *Narrative Intelligence: Papers from the AAAI Fall Symposium (1999), AAAI TR FS-99-01*. AAAI, Menlo Park, CA.

[37] Shaun Nichols. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment.* Oxford University Press.

[38] Amy Ellis Nutt. 2017. The Washington Post. "The Woebot will see you now" - the rise of chatbot therapy.

https://www.washingtonpost.com/news/to-your-health/wp/2017/12/03/the-woebot-will-see-you-now-the-rise-of-chatbot-therapy/. Accessed: 2018-07-16.

[39] World Health Organization. 2017. Mental Health ATLAS 2017. http://apps.who.int/iris/bitstream/handle/10665/272735/9789241514019-eng.pdf. Accessed: 2018-06-27.

[40] World Health Organization. 2017. Mental health: massive scale-up of resources needed if global targets are to be met. http://www.who.int/mental_health/evidence/atlas/atlas_2017_web_note/en/. Accessed: 2018-06-27.

[41] Rosalind W Picard. 2003. Affective computing: Challenges. *International Journal of Human-Computer Studies* 59, 1-2 (2003), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1

[42] Gene Reeves. 2012. A Perspective on Ethics in the Lotus Sūtra. In *The Oxford Handbook of Buddhist Ethics*. in Cozort, D. and Shields, J.M. (Eds.). Oxford University Press, 335–358.

[43] David Rodrigues, Marília Prada, Rui Gaspar, Margarida V Garrido, and Diniz Lopes. 2018. Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior Research Methods* 50, 1 (2018), 392–405. https://doi.org/10.3758/s13428-017-0878-6

[44] Andrea Scarantino. 2017. How to Do Things with Emotional Expressions: The Theory of Affective Pragmatics. *Psychological Inquiry* 28, 2-3 (2017), 165–185. https://doi.org/10.1080/1047840X.2017.1328951

[45] Klaus R Scherer, Tanja Bänziger, and Etienne. Roesch. 2010. *A Blueprint for Affective Computing: A Sourcebook and Manual*. Oxford University Press.

[46] Arthur Schopenhauer. 1995. *On the Basis of Morality*. Hackett Publishing.

[47] Acharya Shantideva. 1979. *Guide to the Bodhisattva's Way of Life*. Batchelor, S. trans. Dharamsala, India, Library of Tibetan Works and Archives.

[48] Jonathan A Smith. 1996. Beyond the divide between cognition and discourse: Using interpretative phenomenological analysis in health psychology. *Psychology and Health* 11, 2 (1996), 261–271. https://doi.org/10.1080/08870449608400256

[49] Fumihide Tanaka and Shizuko Matsuzoe. 2012. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction* 1, 1 (2012), 78–95. https://doi.org/10.5898/JHRI.1.1.Tanaka

[50] Joseph Weizenbaum. 1966. ELIZA - a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (1966), 36–45. https://doi.org/10.1145/365153.365168

[51] Lisa M Yarnell, Rose E Stafford, Kristin D Neff, Erin D Reilly, Marissa C Knox, and Michael Mullarkey. 2015. Meta-analysis of gender differences in self-compassion. *Self and Identity* 14, 5 (2015), 499–520. https://doi.org/10.1080/15298868.2015.1029966

[52] Ulli Zessin, Oliver Dickhäuser, and Sven Garbade. 2015. The relationship between self-compassion and well-being: A meta-analysis. *Applied Psychology: Health and Well-Being* 7, 3 (2015), 340–364. https://doi.org/10.1111/aphw.12051