Article

Self-disclosure to a Robot: Only for Those Who Suffer the Most

Euphie Y. Duan [1], Ji M. Yoon [1], Edison Z. Liang [2] and Johan F. Hoorn [1,2,*]

[1] School of Design, The Hong Kong Polytechnic University, Hong Kong, China
[2] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
* Correspondence: johan.f.hoorn@polyu.edu.hk; Tel.: +528-2766-4509

Abstract: Social robots may become an innovative means to improve the well-being of individuals. Earlier research showed that people easily self-disclose to a social robot even in cases where that was unintended by the designers. We report on an experiment of self-disclosing in a diary journal or to a social robot after negative mood induction. The off-the-shelf robot was complemented with our in-house developed AI chatbot and could talk about 'hot topics' after having it trained with thousands of entries on a complaint website. We found that people who felt strong negativity after being exposed to shocking video footage benefited the most from talking to our robot rather than writing down their feelings. For people less affected by the treatment, a confidential robot chat or writing a journal page did not differ significantly. We discuss emotion theory in relation to robotics and possibilities for an application in design (the emoji-enriched 'talking stress ball'). We also underline the importance of - otherwise disregarded - outliers in a data set that is of a therapeutic nature.

Keywords: self-disclosure; social robots; diary, emotion theory; relevance; valence

1. Introduction

Since the outbreak of the COVID-19 pandemic, there is an upsurge of interest in social isolation, loneliness, and depression. People living alone, people with low social-economic status, as well as quite unexpectedly, youngsters and students are at risk of loneliness (Bu, Steptoe, & Fancourt, 2020a [96]; Bu, Steptoe, & Fancourt, 2020b [97]). In the United States, lockdowns and social distancing measures were associated with increased levels of loneliness, while loneliness correlated highly with depression and suicidal ideation. Loneliness remained high even after distancing measures were relaxed (Killgore, Cloonan, Taylor, Miller, & Dailey, 2020 [98]). In the United Kingdom, a country severely impacted by the pandemic, people with COVID-19, likely developed psychiatric disorders and were lonelier, particularly women, adolescents, and young adults (Li, & Wang, 2020 [99]). In Hong Kong, where our current study took place, COVID-19 even led to "alarming levels of psychiatric symptoms" with loneliness playing a disadvantageous role (Tso & Park, 2020 [100]). A number of interventions may help reduce the feeling of loneliness during social isolation, among which are mindfulness exercises, lessons on friendship, robot pets, and programs that facilitate making social contact (Williams, Townson, Kapur, Ferreira, Nunn, Galante, ... & Usher-Smith, 2021) [101].

It is also media exposure to negative information such as war and disasters that may lead to negative psychological outcomes, in particular feelings of anxiety (Hopwood & Schutte, 2017) [29]. It seems that in the developed countries, depression, stress, and anxiety increase among the youth as a result of intensive media use. According to Twenge, Joiner, Rogers, and Martin (2018) [86], for example, cases of (attempted) suicide among adolescents have gone up since 2010 in the US, which would be linked to heavy media usage.

To improve individuals' mental well-being, a considerable number of studies focuses on the reduction of negative emotions. Emotions are a characteristic human phenomenon (Frijda, 2007) [22] and have a huge impact on individuals' lives, including judgment and decision-making in different contexts, which in turn, influence people's mental well-being (Lerner, Li, Valdesolo, & Kassam, 2015 [40]). Numerous empirical studies underscore that negative emotions are harmful to individuals' health and well-being (Croyle & Uretzky, 1987 [11]; Levine & Edelstein, 2009 [41]; Mayne, 2001 [48]; Thayer & Ruiz-Padial, 2006 [85]; Lerner, Li, Valdesolo, & Kassam, 2015 [40]; Pressman, Gallagher, & Lopez, 2013 [60]; Consedine & Moskowitz, 2007 [8]; Mayne, 1999 [47]; Kubzansky & Kawachi, 2000 [35]).

However, emotions are a complex matter, involving feelings, physiology, cognition, expression, and behaviors (Sonnemans & Frijda, 1995 [83]; Frijda, 2007 [22]). In appraisal theory (Lazarus, 1991 [39]; Frijda, 2007 [22], 2008), appraisal as a component of emotion is considered both a cognitive (Parkinson, 1995 [54]; Lazarus, 1991 [39]) and an unconscious process (Frijda, 2007) [22], which influences emotions through valence judgments.

Valence by many psychologists is regarded as a force that attracts individuals to pleasant objects or repels them from unpleasant ones (Scherer, 2005 [74]; Frijda, 2007 [22]; Russell, 2003 [66]; Shuman, Sander, & Scherer, 2013 [77]). Valence is considered a building block of different emotions (Barrett, 2006 [3]).

Around the core affect of 'pleasant' versus 'unpleasant,' other aspects of 'valency' are the perception of affective qualities, sensitivity to the stimulus (Russell, 2003) [66], goal conduciveness, emotion antecedent and consequent (Frijda, 2007) [22], coping potential, self-congruency, familiarity, contentment, self-worthiness, and moral goodness (Shuman, Sander, & Scherer, 2013) [77]. Emotions transpire when something is appraised as positive or negative in a certain context. Hence, a positive or negative emotion is the result of the valence appraisal-process.

Another core concept in emotion theory is the relevance of an event to goals, needs, and concerns of the individual (Scherer, 2001) [73]. Relevance is an individual's response to events appraised as impacting his or her concerns (Frijda, 1986 [21], 1998, 2007 [22]; Scherer, 2013 [75]). It reflects the personal meaning attached to an event or object. Relevance directs how grave, severe, or urgent something is felt to be. Scherer (2013) [75] suggests that relevance can be explained in the sense of "an event having significant and demonstrable bearing on the well-being of the individual."

Relevance influences the appraisal of an event, which may or may not trigger one's emotions. As relevance has significant and demonstrable bearing on the well-being of the individual, the stimulus that has such direct bearing must be salient or of high priority when the event occurs (Scherer, 2013) [75]. In our case, self-disclosure to a robot is less relevant to an individual's concerns when s/he feels okay than when stressed out.

Another aspect of emotion that has direct bearing to our robot case is that in the early stages of appraisal, the novelty of a stimulus also generates emotions (Lazarus, 1991 [39]; Scherer, 1984 [72], 2013 [75]). Novelty is closely related to the perception of affective qualities. Scherer (2013) [75] proposes to conceive of novelty in terms of suddenness, unfamiliarity, and unpredictability. Yet, novelty does not seem to be as unanimously present as valence and relevance. Novelty does not last forever, wares out, and at one point is emotionally not a 'surprise' anymore (cf. Fontaine, Scherer, & Soriano, 2013) [24]. In our experiment, then, novelty will not be the focus of interest but a control variable while examining the effect of different media on the reduction of negative affect.

As our mind and body are closely related, researchers indicated that negative emotions are associated with bad subjective health and mental well-being (Pressman, Gallagher, & Lopez, 2013 [60]; Consedine & Moskowitz, 2007 [8]; Mayne, 1999 [47]; Kubzansky & Kawachi, 2000 [35]). One of the popular coping approaches to reduce the level of negative emotions involves emotion-disclosure interventions in which individuals reveal their thoughts and feelings through self-disclosure (Farber, 2006) [18]. Smyth (1998) [82] forwards that such disclosure interventions are effective for both males and females to reduce the level of negative emotions. Pennebaker (1985) [55] suggests that self-disclosure reduces the psychological work of actively having to inhibit emotions and thoughts about negative events (e.g., trauma), which reduces the stress.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

To self-disclose, talking with a psychiatrist but also journal writing is one of the methods widely adopted in psychotherapy. A variety of studies examined journal writing to reduce distress (Alford, Malouff, & Osland, 2005 [1]; Hemenover, 2003 [26]; Horneffer & Jamison, 2002 [27]; Ireland, Malouff, & Byrne, 2007 [30]). Journal writing has beneficial effects particularly for college students (Frattaroli, 2006) [19]. Writing as an intervention would transfer the nonverbal memories into a verbal form that helps reorganize the memories, resulting into stress reduction (Pennebaker & Francis, 1996 [58]; Dalton & Glenwick, 2009) [13].

The meta-analysis by Frisina, Borod, and Lepore (2004) [23] indeed found that writing improved health outcomes (d = .19). However, the effect was stronger for physical outcomes (d = .21) than for psychological outcomes (d = .07) (ibid.). In accordance, Pascoe (2016) [53] states that the effectiveness of writing to reduce the level of negative emotions is but limited and needs further study. The most beneficial form of writing seems to include large numbers of positive emotion words and a moderate number of negative emotion words. Participants who used too many or too few negative emotion words benefited less from a writing intervention (Pascoe, 2016) [53], so that writing may be contra-indicated for individuals with, for instance, alexithymia, who are unable to express emotions (Lumley, 2004) [44]. Moreover, studies conducted by Pennebaker (1993) [56], Pennebaker and Beall (1986) [57], and Murray and Segal (1994) [50] point out that the physical presence of a therapist is what moderates the negative emotions rather than the writing itself.

Problem is that worldwide, mental-health workers, therapists, and psychiatrists are in short supply (World Health Organization, 2018) [94]. Luckily, however, and quite unexpectedly, since the release of the Rogerian chatbot-therapist ELIZA (Weizenbaum, 1966) [92], people nowadays do not merely share their secrets with fellow humans but also with their Apple Siri voice agent (e.g., Saffarizadeh, Boodraj, & Alashoor, 2017) [70] as well as with conversation and companion robots (e.g., Hoorn, Konijn, Germans, Burger, & Munneke, 2015) [28]. Perhaps, then, that social robots may be an 'AI-in-Design' alternative to practice emotion-disclosure interventions with. Provided that they work well, of course.

In that respect, Wada, Shibata, Saito, Sakamoto, and Tanie (2005) [89] showed that social robots alleviate adverse emotions such as loneliness and stress. Measured on a geriatric-depression scale as well as a 'face scale,' the level of depression of participants significantly decreased after interaction with a social robot (ibid.). Jibb, Birnie, Nathan, Beran, Hum, Victor, and Stinson (2018) [31] found that talking to a robot reduced the level of distress among children who underwent cancer treatment. Dang and Tapus (2015) [14] found that social robots can assist humans during emotion-oriented coping, using a stress-eliciting game played together with a robot. Cabibihan, Javed, Ang, and Aljunied (2013) [6] show evidence that robots work well for autistic children, improve their adaptive behaviors (e.g., Robins, Amirabdollahian, Ji, & Dautenhahn, 2010 [65]; Kozima, Nakagawa, & Yasuda, 2005 [33]; Vanderborght et al., 2012 [87]; Tapus et al., 2012 [84]), and even invite self-disclosure in adolescents with autism spectrum disorder (Kumazaki, Warren, Swanson, Yoshikawa, Matsumoto, Takahashi, & Kikuchi, 2018 [34]). Social robots also may increase the mental well-being of older adults through perceived emotional support and interaction (Pu, Moyle, Jones, & Todorovic, 2018) [62].

In psychotherapy as well, robots may meet the special needs of individuals with cognitive, physical, or social disabilities (Libin & Libin, 2004) [43]. The meta-analysis conducted by Costescu, Vanderborght, and David (2014) [10] indicates that in overall robot-enhanced psychotherapy, robots have medium-sized significant effects on the improvement of behavior but not so much on cognitive and subjective aspects. Yet, individual studies sometimes do show that social robots improve performance on the subjective and cognitive level as well (e.g., Kidd, Taggart, & Turkle, 2006 [32]; Wada, Shibata, Saito, Sakamoto, & Tanie, 2005 [89]; Wada & Shibata, 2007 [88]).

 doi:10.20944/preprints202105.0279.v1

In view of the generally positive therapeutic effects of robots in reducing stress and anxiety, our research question is whether social robots offer an alternative to traditional diary writing to 'let off some steam,' particularly in coming to terms with negatively valenced emotions after violent-media exposure. We expected that social robots would do better than writing down ones feelings because the robot more closely resembles talking to a person (i.e. a virtual therapist) and writing may not be everybody's preferred way of expression. Therefore, we propose (H1) that a social robot that invites self-disclosure from its user decreases the level of negative emotions more than pencil-and-paper approaches do. As a medium (H2), a social robot that invites self-disclosure will be regarded as more relevant to the user's goals and concerns than pencil-and-paper approaches.

2. Materials and Methods

2.1. Participants and Design

After obtaining approval from the institutional Ethical Review Board (filed under HSEARS20200204003), Voluntary participants (N = 45; M_{Age} = 24.9, SD_{Age} = 3.29, 55.6% female, Chinese nationality) were randomly assigned to a between-subjects experiment of self-disclosure after negative-mood induction in a Robot (n = 24; 54.2% female) versus Writing condition (n = 21; 57.1% female). All participants had university training at the master level, except for four doctorate degrees, three bachelors, and one with a diploma degree. Informed consent was obtained formally from all participants. They did not receive any credits or monetary rewards.

2.2. Procedure

Participants were brought in a dimly lit and shielded-off section of the experimenter room and were seated in front of a laptop. The experiment consisted of negative-mood induction and then self-disclosure with one of two media, after which participants filled out an online questionnaire in the Qualtrics environment for administration of surveys and experiments (https://www.qualtrics.com/).

In the induction part, participants were confronted with a 10m6s video compilation of three documentaries about a serious earthquake incident that happened in Wenchuan Sichuan, China in 2008. Viewing negative media, including videos, images, and texts, effectively induces negative emotions with an increasing activation of the aversive system (Bolls, Potter, & Lang, 2001 [4]; Lang, Shin, & Lee, 2005 [37]). In accordance with Siedlecka and Denson's (2019) [78] review that video is the most effective means of mood induction, we prepared a video on the Sichuan earthquake to make the contents culturally related to our participants and bring relevance and realness to the experience.

After the video and 30-40s of instruction, participants either talked to a robot about their experiences during the video or wrote them down on paper. The robot nor writing utensils were visible before self-disclosure. The self-disclosure session took about 10 minutes. The movements of the robot and text input were handled in remote control (Wizard of Oz), the conversation was handled autonomously by our inhouse developed AI chatbot (next section).

After self-disclosure ended, participants filled out a 30-item structured questionnaire (Appendix A), reporting on their assessment of the video clip and talking to the robot or writing the journal page. Appendix A shows the English translation of the Chinese version in the robot condition. Supplementary Materials 1 provides both questionnaire versions, for robot and writing, in Chinese and in English translation. Items on the questionnaire were presented in blocks with pseudo-random sequences of items within blocks, different for each participant. We ended the questionnaire inquiring about demographic information. Upon completion, participants were thanked for their participation and debriefed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

2.3. Apparatus and Materials

2.3.1. Video materials

The video materials for negative-mood induction were 10 minutes and 6 seconds long and were composed of video excerpts from the following three Sichuan earthquake Internet documentaries:

Internet video in memory of the Wenchuan Sichuan earthquake 10th anniversary (cut at 00:02-01:19). Available from https://www.bilibili.com/video/av23087386/

Dazzz2009 (2008, December 31). Internet video record of 512 earthquake in Dujiangyan (cut at 01:20-01:59). Available from https://www.youtube.com/watch?v=Vz0nGbl81fM&list=PLf2PpWDjsx1d6rVUW0 vaGFzhvIr_nRo_8&index=2

Lantian777 (2008, May 16). Internet video 10 minutes after Wenchuan Sichuan earthquake (in full). Available from https://www.youtube.com/watch?v=PI5KL7nvU28

2.3.2. Robot embodiment

For a person to self-disclose, there should be a certain level of trust in the conversation partner, also in human-robot interaction (Saffarizadeh, Boodraj, & Alashoor, 2017 [70]; Martelaro, Nneji, Ju, & Hinds, 2016 [45]; Corcoran, 1988). Therefore, we looked for a small, non-threatening toy-like robot that in appearance would stay far from uncanny effects (cf. Złotowski, Sumioka, Nishio, Glas, Bartneck, & Ishiguro, 2016 [95]). The chatbot part was trained such that superfluous faults in responding were kept to a minimum (cf. Salem, Lakatos, Amirabdollahian, & Dautenhahn, 2015 [71]).

The robot of our choice was a Robotis DARwIn Mini, a 3D printable, programmable, and customizable miniature humanoid robot of 27 cm tall with Bluetooth connection to a laptop. The robot could stand up and move its arms while speaking through an AI chatbot. Technical details about the DARwIn Mini can be found in Supplementary Materials 1. The actions DARwIn could execute during the experiment such as waving and raising its arms were controlled remotely.

2.3.3. Self-disclosure chatbot

The DARwIn Mini cannot speak, therefore, we created our own chatbot, using DARwIn Mini as the humanoid embodiment of our self-disclosure inviting AI chatbot. Next, we give a concise account of the development of both the hardware and software. Supplementary Materials 1 offers further specifications.

<u>Hardware development.</u> Two main components made up the hardware of our chatbot: the core board Raspberry Pi Zero (WH) and an extension board that was connected to the speaker. These two boards we engineered into an integrated circuit. Figure 1 offers an impression of the hardware prototype chatbot.

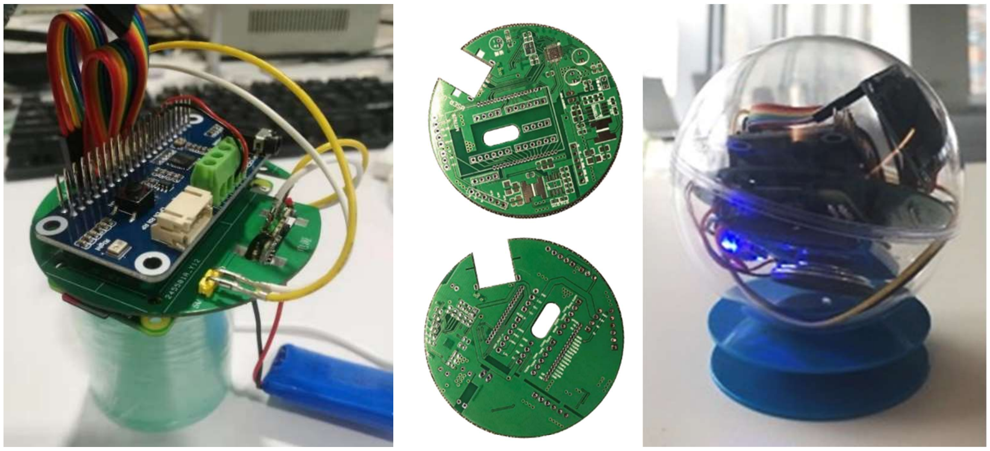

Figure 1. Hardware prototype of the chatbot with in-house-engineered integrated board.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

Software development. To create a chatbot adjacent to the DARwIn Mini, we set up a homepage for test subjects to assess the chatbot system (for details on the chatbot, please refer to Supplementary Materials 1) (www.roboticmeme.com). For website development, we used Semantic UI as the front-end framework (https://semantic-ui.com/) and Node.js as the back-end (https://nodejs.org/en/). We tentatively called the chatbot MEME and invited test subjects to share their secrets with MEME in our test environment. The chatbot on the website had speech recognition in Putonghua, Cantonese, and English, using a Turing robot API. To increase the traffic on our website, we also created an official WeChat account and used Python to run a server in Google Cloud (https://cloud.google.com/). On WeChat, we used Chill chat with the Xiaohuangji corpus for information retrieval.

Ours was a hierarchical chatting system, consisting of three layers: (1) A rule-based layer that focused on certain specific chatting tasks (Eliza.py and regular expressions), (2) an information retrieval-system that searched the answer from a corpus built from Weibo conversations and conversations about movies, and (3) a generation layer that used the general-purpose encoder seq2seq as well as Generative Adversarial Network, a machine-learning tool, to generate a response (https://github.com/google/seq2seq; https://en.wikipedia.org/wiki/Generative_adversarial_network). We adopted the *k*-means algorithm in sentence vector clustering. After many iterations of improvement, the final model could effectively answer a question.

For natural language understanding, we installed a Rasa stack and so made the conversation somewhat more contextualized (https://rasa.com/). For Rasa to estimate what a user means to say, we classified a number of conversational topics that had to do with negative experiences. Therefore, we analyzed the contents of a complaining website and ran a spider program to catch the users' comments, after which we did data mining for hot topics.

For training, we sampled a 2 years' record of almost 500 pages and nearly 10,000 comments. Then we tokenized these utterances and identified the high-frequency items ('hot topics'). An impression of the results is depicted in Figure 2: People worried most about unrequited love, emotions, relationship, family, love, homosexual love, cheating, love crush, self, life, work, making love (sex), being disappointed in live, being the only one, feelings, lost, life, to cheer up, marriage, trouble and worry, loneliness, depression, study, entry exam to university and college, secrets, and love relationships.

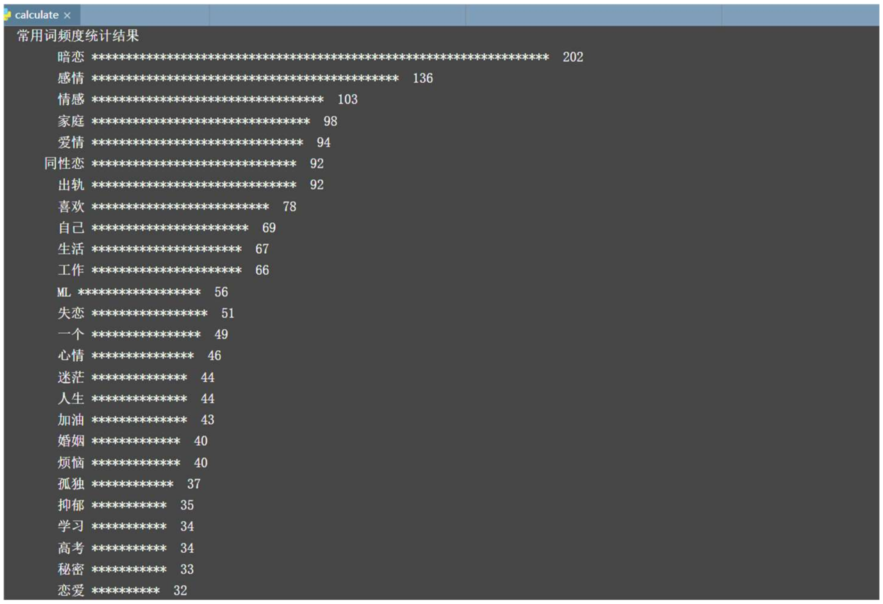

Figure 2. Frequency statistics for hot topics people complained about.

The complete set-up of the self-disclosure AI chatbot is shown in Figure 3. The sing, movie, poem, and weather options were not used in the actual experiment.

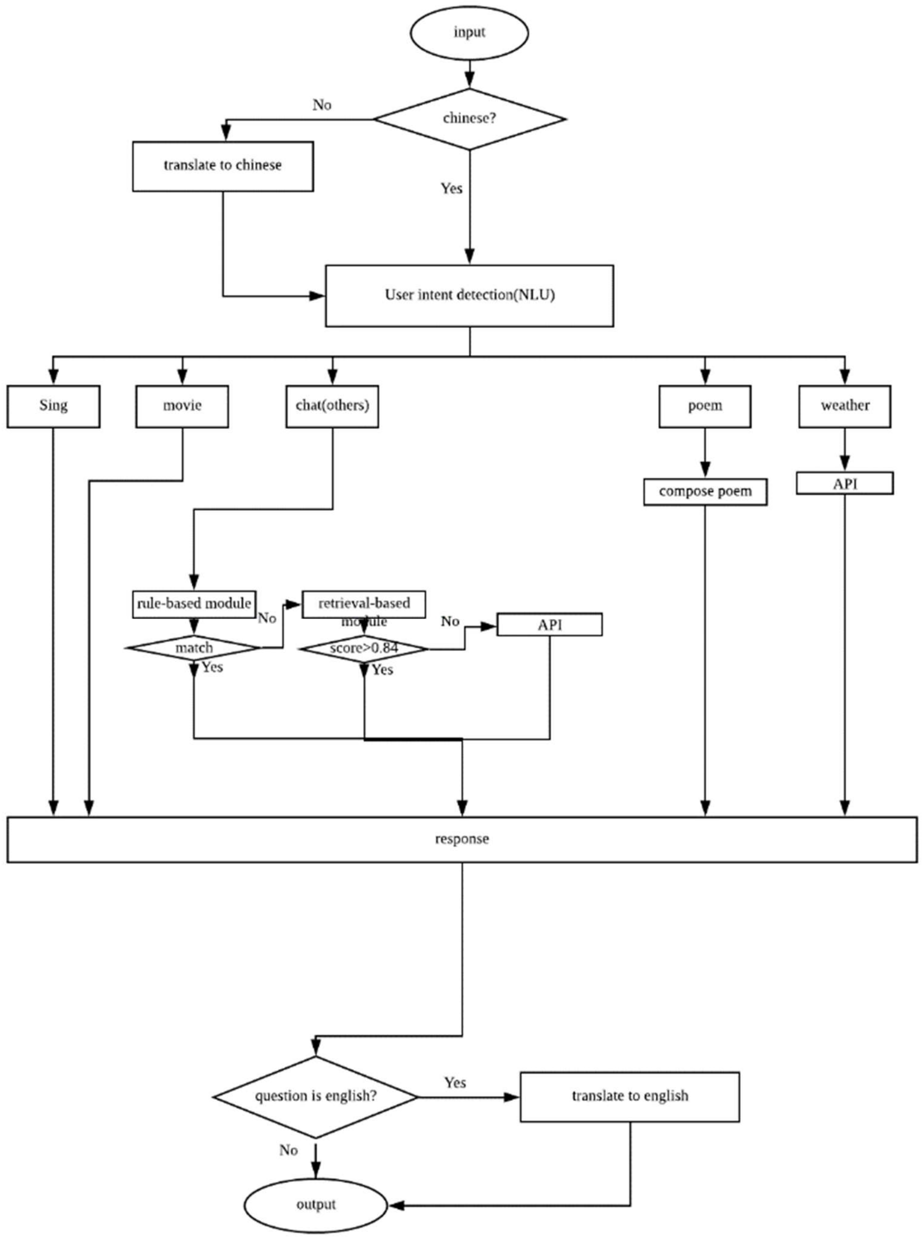

Figure 3. Flowchart for our self-disclosure AI chatbot.

For the experiment, we installed our chatbot system in a voice kit that stood behind the DARwIn Mini. We did not install voice-recognition software because of its inefficiency (i.e. slow and inaccurate). Therefore, a confederate not visible to the participant inputted the participant's utterances. Information processing and replying to the participants was done autonomously by our AI. Figure 4 exhibits the interaction flow.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

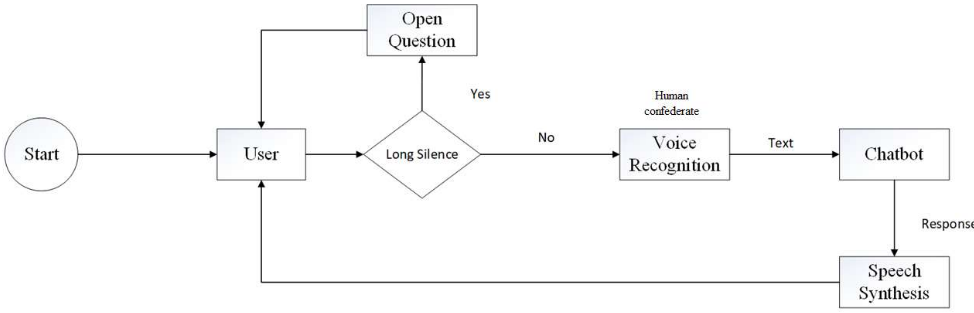

Figure 4. Human-robot interaction flowchart.

Together, the DARwIn Mini standing in front of the voice kit carrying our self-disclosure AI chatbot made up the 'robot condition' in our experiment. Figure 5 shows the final set-up.

Figure 5. DARwIn Mini was placed in front of the voice kit with self-disclosure AI chatbot.

We constructed the conversation following the guidelines in psychotherapy (i.e. Nystul, 2016) [52]. For example, open questions such as "How do you feel about that?" were asked to guide the participants' reflections on their experience. During the conversation, only minimal encouragement like "Yes, I see" was provided by the robot. The open questions that were coded into the chatbot were also posed to participants in the condition of writing, during their instruction. In inviting self-disclosure, the robot basically followed social norms from social penetration theory (Altman & Taylor, 1973 [2]). Based on Nomura and Kawakami (2011) [51], however, the robot did not share secrets with its user and did not (need to) apply reciprocity, although this is an important social

 doi:10.20944/preprints202105.0279.v1

rule in human interaction (cf. Psychopathology Committee of the Group for the Advancement of Psychiatry, 2001) [61]. Thus, the robot was not self-disclosing but invited self-disclosure by asking open questions (Hoorn, Konijn, Germans, Burger, & Munneke, 2015) [28] .

2.4. Measures

For measurement, we worked from a dimensional model of valence and relevance (cf. Smith & Ellsworth, 1985) [81] rather than a categorical model, which classifies emotions by name ('sad' or 'happy'). Emotion words are fuzzy (Frijda, 1998) and appraisal of an event may elicit a variety of emotions (Parkinson, 1995 [54]; Frijda, 2007 [22]). Different people interpret events differently and so different emotions are generated for the same event. If negative emotions are presented only by name, then, consensus among the participants may be low. Because appraisal is a dynamic process (Frijda, 2007 [22]; Carrera & Oceja, 2007 [7]; Russell, 2017 [67]), the possibility exists that an individual has multiple emotions at a single event. It is difficult to list out all possible negative emotions in a questionnaire by name and not measure fatigue effects eventually. Therefore, we assessed the most core concept of valence (positive/pleasant vs negative/unpleasant), which is a more fundamental process compared to aspects of valence that require associative and sometimes conceptual processing (Scherer, 2013) [75].

In self-report instruments such as Positive and Negative Affect Schedule (PANAS), valence is conceived of as two unipolar dimensions (Watson, Clark, & Tellegen, 1988) [91]. Each affective direction would be mediated by an independent neural pathway (e.g., Diener, 1999) [15]. This is in contrast to earlier approaches that maintain a bipolar measurement (for a discussion, see Russell & Carroll, 1999a [68], 1999b [69]). Moreover, two unipolar scales ($0 \rightarrow n$, $0 \rightarrow p$) can at one point stand in a bipolar constellation ($n \leftrightarrow p$) but from a bipolar measurement, one can never return to a unipolar conception. Therefore, we decided for two unipolar dimensions to measure affect and constructed a structured questionnaire with more items on a measurement scale, featuring positive (indicative) items as well as negative (counter-indicative) items. This approach also remedied potential answering tendencies.

We used two versions of a structured questionnaire as appropriate to one of two conditions: Talking with the robot or journal writing on a piece of paper (Appendix A and Supplementary Materials 1). The questionnaire was constructed from emotion literature (e.g., Scherer, 2013 [75]; Frijda, 2007 [22]; Russell, 2003 [63]) and ran four measurement scales: Valence after the movie but before treatment (robot or writing), Valence after treatment, Relevance, and Novelty. Together with the Demographics, Novelty served as a control.

Items were Likert-type statements followed by a 6-point rating scale (1 = strongly disagree, 6 = strongly agree). One half of the items on each measurement scale consisted of four indicative statements and the other half of counter-indications. Blocks of related items were offered in pseudo-random order, different for each participant. Items within blocks also were pseudo-randomly presented to each participant.

The measurement scale 'Valence before treatment' (*ValB*) consisted of four indicative items, for example, "I feel good" and of four counter-indicative items, for example, "I feel bad." We used the same items for measurement of Valence after talking to the robot or writing on paper but adjusted the wording to the situation. Thus, 'Valence after treatment' (*ValA*) also had four indicative and four counter-indicative items. Relevance of robot or writing to goals and concerns (i.e. personal emotion regulation) was measured with two indicative items (e.g., '… is useful') and two counter-indicative items (e.g., '… is worthless').

To control for a possible confound of the robot as a novel means to regulate emotions, the Novelty scale was composed of three indicative items (e.g., '… is new') and three counter-indicative items (e.g., '… is commonplace').

Demographics included information about the participant's Gender, Age, Education level, and Country. At the end of the questionnaire, participants could leave their comments.

Then we conducted reliability analysis on our measurement scales (for elaboration, see Supplementary Materials 1). For the variables of theoretical interest, all measurement scales with all items included, achieved good to very good reliability in the first run (Cronbach's $\alpha \geq .82$). This was so for the separate subscales of Valence (4 items each) and for their combination (*ValB* and *ValA*, 8 items each) as well as for Relevance (4 items). After repair, the control variable of Novelty (5 items) had Cronbach's $\alpha = .77$.

To test discriminant validity, we performed Principal Component Analysis with Varimax rotation on Valence-after (*ValA*), Relevance, and Novelty. Indicative items formed a positive-Valence subscale as the counter-indicative items clustered into a negative-Valence subscale. Items on the Relevance scale neatly fell in line as intended. Novelty showed some spread over both Valence and Relevance. However, because this was a control variable, we kept the scale intact and will observe in the Results section its tendency to coalesce with variables of theoretical interest.

We then calculated the means (*M*) across the items on a scale and performed an outlier analysis for Valence (before and after), Relevance, and Novelty. We found that participant 9 was an outlier in *MValB*, participant 39 in *MValA*. Participant 5 and 21 were outliers for *MValAi*. Participants 39, 27, 38, and 33 were outliers in *MValAc* (see Supplementary Materials 1). There were no outliers in *MNov*, *MRel*, *MValBc*, and *MValBi*. We will perform our effects analysis with and without those outliers.

3. Results

3.1. Demographics

We checked the country that participants came from but only participant (31) reported she was from Africa; the rest were from China. Inspection of the scatter plot, however, showed that number 31 was not in the zone of outliers. Therefore, we decided to treat this person as one of the same sample and not treat her differently in the analysis.

Next, we checked whether Age was correlated with the eight dependent variables (Valence-bipolar before and after, Positive and Negative Valence before and after, Relevance, and Novelty). We calculated Pearson bivariate correlations (two-tailed) and found no significant relations. Age did have a near-significant weak negative correlation with Positive-Valence-before ($r = -.27$, $sig. = .08$), indicating that with higher age, people were less positive after viewing the earthquake video.

Then we examined whether Gender was influential for the eight dependent variables. We ran a MANOVA (Pillai's Trace) to check the effect of Gender but we found no significant multivariate effects ($V = .11$, $F_{(7,37)} = .68$, $p = .688$). Yet, Gender did exact a small univariate effect – barely significant – on the experience of Novelty ($F_{(1,41)} = 4.18$, $p = .047$, $\eta_p^2 = .09$). Throughout, females experienced more Novelty ($M = 4.03$, $SD = .83$) than did males ($M = 3.50$, $SD = .87$). However, Novelty was a control variable in our experiment and not of theoretical interest. Therefore, we concluded that Gender did not have a significant effect on the variables theoretically related to our hypotheses.

Among all participants, there were four with doctorate degrees, three with bachelor's degrees and one with a diploma degree. The rest were all master's degrees. We found participant 39 with a doctorate degree also to be one of the outliers to the scale means. Thus, we excluded this participant from the effect analysis of Educational background.

We put the seven participants with a degree other than master in one group and randomly chose seven other participants (who were not outliers) with a master degree in the other group. We performed an independent samples t-test to check whether Education had effect on the eight dependent variables that related to our theoretical hypothesis. We ran this test five times, each time with a different set of masters and found that in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

certain group comparisons, Education did have an effect on some of the theoretical variables (see Supplementary Materials 1). Therefore, we made two data sets, one with all 45 participants (24 in the robot group and 21 in the writing group) and the other with 31 participants (17 in the robot group and 14 in the writing group), excluding the outliers and the participants with a non-master degree as educational background. These separate sets were used to confront our hypotheses with.

3.2. Manipulation check: Emotional effects after negative-mood induction and after treatment

We wanted to control whether any emotion at all was provoked by the shocking video footage of the earthquake and whether the treatment (robot or writing) evoked any change in emotion at all. Or did everything remain at scale value 1 (no emotions reported)?

For $N = 45$, we ran a one-sample t-test (two-tailed) with 1 as the test value to see if any negative (or positive) emotions occurred after mood induction as well as after treatment. For Positive Valence after the earthquake clips, *MValBi* showed that $t = 8.67$, $p < .00001$. For Negative Valence after the earthquake clips, *MValBc* resulted in $t = 16.44$, $p < .00001$. For Positive Valence with $n = 31$, *MValBi* was $t = 7.00$, $p < .00001$. For Negative Valence with $n = 31$, *MValBc* resulted in $t = 15.38$, $p < .00001$. Thus, more negative than positive mood was induced by the clips, as intended.

For $N = 45$, after treatment (robot or writing), Positive Valence *MValAi* obtained $t = 17.83$, $p < .00001$ while for Negative Valence *MValAc,* $t = 10.35$, $p < .00001$. For $n = 31$, Positive Valence *MValAi* was $t = 18.65$, $p < .00001$ and Negative Valence *MValAc,* $t = 9.39$, $p < .00001$. In other words, more positive than negative emotions were felt after either talking to a robot or writing a diary page, as intended.

To check whether before-after effects of treatment actually occurred, we also ran paired-samples t-tests (two-tailed) in both data sets $N = 45$ and $n = 31$. Note that these are no tests of our hypotheses but a mere inspection if anything happened at all.

For the difference between *MValBc* – *MValAc* with $N = 45$, $t = 9.34$, $p < .00001$. For the difference between *MValBc* – *MValAc* with $n = 31$, $t = 9.42$, $p < .00001$, so that we may conclude that participants after treatment became less negative (*MValBc* was significantly larger than *MValAc*).

For the difference between *MValBi* – *MValAi* with $N = 45$, $t = -7.16$, $p < .00001$. For the difference between *MValBi* – *MValAi* with $n = 31$, $t = -7.24$, $p < .00001$, so that we may conclude that participants after treatment became more positive. Whether through a robot or through writing, treatment had effect into the expected direction.

3.3. Effect of Media (robot vs. writing) on Valence and Relevance

To analyze the changes in Valence after talking to a robot or writing a diary page, we computed three mean difference scores: For bipolar Valence, $\Delta Val = MValA - MValB$; for Positive Valence, $\Delta ValP = MValAi - MValBi$; and for Negative Valence, $\Delta ValN = MValAc - MValBc$. In Table 1, ΔVal, $\Delta ValP$, $\Delta ValN$, *MRel*, and *MNov* are shown for the two conditions (robot vs. writing). Top half of Table 1 shows the averages for the entire sample ($N = 45$), the bottom half with suspected cases excluded ($n = 31$).

Table 1. Valence, Relevance, and Novelty for robot and writing.

Variables	Robot			Writing		
	Mean	*SD*	*n*	*Mean*	*SD*	*N*
ΔVal	1.77	1.26	24	1.11	0.81	21
ΔValP	1.75	1.31	24	.89	1.06	21
ΔValN	1.78	1.30	24	1.32	0.84	21
MRel	4.19	.99	24	3.98	1.33	21
MNov	4.10	.86	24	3.42	0.77	21

N = 45

Variables	Robot			Writing		
	Mean	*SD*	*n*	*Mean*	*SD*	*N*
ΔVal	1.98	1.11	17	1.33	0.83	14
ΔValP	1.99	1.08	17	1.05	1.17	14
ΔValN	1.97	1.27	17	1.61	0.76	14
MRel	4.35	.96	17	4.27	1.08	14
MNov	4.13	.95	17	3.53	0.78	14

n = 31

3.3.1. Effects on bipolar Valence and Relevance

Next, we performed a General Linear Model (GLM) Multivariate analysis of Media (2: robot vs. writing) on *ΔVal* and *MRel* (grand mean scores) with *MNov* as a covariate. We did this for N = 45 and n = 31 separately. For an extensive report, see Supplementary Materials 1.

For the data set (N = 45), with Novelty as a covariate, we did not find significant multivariate effects (V = .09, $F_{(2,41)}$ = 1.98, p = .151, $\eta_p{}^2$ = .09). However, we did find multivariate effects for *MNov* (V = .39, $F_{(2,41)}$ = 12.92, p = .000, $\eta_p{}^2$ = .39), which covaried quite strongly with *MRel* ($F_{(1,42)}$ = 25.91, $p < .001$, $\eta_p{}^2$ = .38).

With Novelty excluded from the analysis, the pattern of multivariate effects was similar as before (V = .09, $F_{(2,42)}$ = 2.09, p = .136, $\eta_p{}^2$ = .09). Officially, we should stop our scrutiny here. Yet, when we looked into the main effect of Media on *ΔVal*, we did see that without Novelty, the effect became significant ($F_{(1,43)}$ = 4.23, p = .046, $\eta_p{}^2$ = .09). As a trend, beneath the surface, it seemed that talking to a robot ($M_{\Delta Val}$ = 1.76, SD = 1.25) had a more positive impact on Valence (bipolar conception) than did writing ($M_{\Delta Val}$ = 1.10, SD = .81) after negative mood induction.

For the data set (n = 31), with Novelty as a covariate, Media (robot vs. writing) did not exert any significant multivariate effects on Δ*Val* or *MRel* (V = .09, $F_{(2,27)}$ = 1.32, p = .285, $\eta_p{}^2$ = .09). Novelty (*MNov*) covaried with other variables (V = .38, $F_{(2,27)}$ = 8.33, p = .002) but this was significant for *MRel* alone ($F_{(1,28)}$ = 15.40, p = .001, $\eta_p{}^2$ = .36). With Novelty discarded in the analysis, the pattern of results did not change. Without the outliers, even the shimmer of a positive change in valence caused by robots or writing remained absent.

3.3.2. Effects on Positive Valence, Negative Valence, and Relevance

For N = 45, we ran two GLM Repeated measures of Media (2 conditions) on within-subjects factor (Δ*ValP* vs Δ*ValN*) with *MRel* and *MNov* separately as covariates. We found no significant multivariate effects on unipolar (Δ*ValP* vs Δ*ValN*), not for the interaction with Media (V = .05, $F_{(1,42)}$ = 2.02, p = .162, $\eta_p{}^2$ = .05), not for *MRel* as covariate (V = .02, $F_{(1,42)}$ = .71, p = .406, $\eta_p{}^2$ = .02), and not for *MNov* as covariate (V = .00, $F_{(1,42)}$ = .004, p = .951, $\eta_p{}^2$ = .000).

With *MRel* included, we did find a marginally significant main effect of Media across Δ*ValP* and Δ*ValN* (non-unipolar Valence): $F_{(1,42)}$ = 3.79, p = .058, $\eta_p{}^2$ = .08. With *MNov* included, however, that main effect was not even marginally significant: $F_{(1,42)}$ = 2.04, p = .161, $\eta_p{}^2$ = .05. This pattern of results remained the same without the covariates, except that as before the effect of Media across Δ*ValP* and Δ*ValN* (non-unipolar Valence) became significant: $F_{(1,43)}$ = 4.23, p = .046, $\eta_p{}^2$ = .09.

For n = 31, we again ran two GLM Repeated measures of Media (2 conditions) on (Δ*ValP* vs Δ*ValN*) with *MRel* and *MNov* as a separate covariate, respectively. As before, we found no significant multivariate effects on (Δ*ValP* vs Δ*ValN*) (V = .03, $F_{(1,28)}$ = .78, p = .162, $\eta_p{}^2$ = .03), not for the interaction with Media (V = .09, $F_{(1,28)}$ = 2.63, p = .116, $\eta_p{}^2$ = .09), not for *MRel* as covariate (V = .01, $F_{(1,28)}$ = .30, p = .588, $\eta_p{}^2$ = .01), and not for *MNov* as covariate (V = .004, $F_{(1,28)}$ = .13, p = .725, $\eta_p{}^2$ = .004). Without the emotional outliers, the main effect of Media on the unipolar conception of Valence (Δ*ValP* vs Δ*ValN*) remained absent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

($F_{(1,28)}$ = .3.14, p = .087, η_p^2 = .10). Without the covariates, the pattern of these results did not change.

In all, we saw that the only marginally significant effect we could establish for the theoretical variables was with N = 45, without *MNov* as a covariate, in a bipolar conception of Valence (ΔVal). We wondered, then, how this could be the case since the mood induction and the treatment had been so successful according to t-test (Section 3.2).

3.4. Effect of Media on Valence and Relevance for those who felt most negative

In clinical trials, it is good practice to contrast a control group with a treatment group and measure the effects of a drug or medical device (e.g., Friedman, Furberg, & DeMets, 2010, p. 2) [20]. We attempted the same but now with depressed people (after mood induction), using two different media (robot vs pen-and-paper). However, another approach in clinical research is to try a drug on healthy volunteers versus patient volunteers and this is what we so far failed to recognize: Part of the participants may not have been affected much by the mood induction and therefore did not need treatment or comfort from our robot or journal writing; after all, they were not distressed, they did feel the emotion but were 'immune to the affliction' so the treatment was superfluous, a sub-sample ceiling effect.

Therefore, we performed a median split for both data sets N = 45 and n = 31 on the variable *MValBc* (Negative Valence before treatment). In the data set with N = 45, with the outliers included, 23 participants were on the side of feeling most negative. Twelve of them were in the robot condition and 11 in the writing condition.

For n = 31, without the outliers, 17 participants felt most negative, 10 of which talked to a robot after viewing the footage and 7 did the writing. Table 2 provides the means and SDs for ΔVal, $\Delta ValP$, $\Delta ValN$, *MRel*, and *MNov* for talking to a robot or writing a journal page for those participants who felt very negative after watching the earthquake video.

Table 2. Valence, Relevance, and Novelty of the most negatively affected participants in robot and writing condition (n = 40).

Variables	Robot			Writing		
	Mean	*SD*	*n*	*Mean*	*SD*	*N*
ΔVal	2.74	.83	12	1.56	.84	11
ΔValP	2.68	.84	12	1.31	1.16	11
ΔValN	2.79	.96	12	1.77	.75	11
MRel	4.17	1.04	12	4.25	1.31	11
MNov	3.27	.92	12	4.52	.56	11

With emotional outliers: n = 23

Variables	Robot			Writing		
	Mean	*SD*	*n*	*Mean*	*SD*	*N*
ΔVal	2.65	.80	10	1.69	.83	7
ΔValP	2.55	.81	10	1.42	1.21	7
ΔValN	2.75	.95	10	1.96	.78	7
MRel	4.13	.80	10	1.70	.83	7
MNov	3.45	1.02	10	4.49	.64	7

Without emotional outliers: n = 17

3.4.1. Valence as a bipolar scale in high-negative subjects

For n = 23, GLM Multivariate on ΔVal and *MRel* showed that with Novelty (*MNov*) as a covariate, Media (robot vs writing) exerted significant multivariate effects (V = .46, $F_{(2,19)}$ = 8.09, p = .003, η_p^2 = .46). Media had a significant and moderately strong univariate effect on ΔVal ($F_{(1,20)}$ = 8.80, p = .008, η_p^2 = .31) but not on *MRel* ($F_{(1,20)}$ = 2.16, p = .16, η_p^2 = .10).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

MNov also showed significant multivariate effects (V = .47, $F_{(2,19)}$ = 8.42, p = .002, η_p^2 = .47) but on *MRel* alone ($F_{(1,20)}$ = 16.85, p = .001, η_p^2 = .46), not on *ΔVal* (F < 1, p = 459). Novelty made things more relevant.

After removing *MNov* as a covariate, we found that Media still evoked multivariate effects (V = .40, $F_{(2,20)}$ = 6.79, p = .006, η_p^2 = .40), substantiated by a significant and moderately strong effect of Media on *ΔVal* ($F_{(1,21)}$ = 11.51, p = .003, η_p^2 = .35). There was no significant effect on *MRel* ($F_{(1,21)}$ = .03, p = .867, η_p^2 = .001).

With emotional outliers included, then, talking to a robot ($M_{\Delta Val}$ = 2.74, SD = .83) had a more positive impact on Valence (bipolar conception) than did writing ($M_{\Delta Val}$ = 1.56, SD = .84) after negative mood induction. The level of novelty of the medium made things more relevant but neither medium was significantly more relevant than the other. Novelty did not significantly influence the valence result.

For n = 17, without outliers, GLM Multivariate on *ΔVal* and *MRel* showed that with Novelty as a covariate, significant multivariate effects were established (V = .38, $F_{(2,13)}$ = 3.94, p = .046, η_p^2 = .38). There was a marginally significant main effect of Media on *ΔVal* ($F_{(1,14)}$ = 4.07, p = .063, η_p^2 = .23) but not on *MRel* ($F_{(1,14)}$ = 2.23, p = .157, η_p^2 = .14).

Multivariate effects for *MNov* were significant (V = .44, $F_{(2,13)}$ = 5.16, p = .022, η_p^2 = .44), again for covarying with *MRel* ($F_{(1,14)}$ = 10.87, p = .005, η_p^2 = .44) but not with *ΔVal* ($F_{(1,14)}$ = .15, p = .700, η_p^2 = .01).

After removing *MNov* as a covariate, we found that no significant multivariate effects were present any more (V = .30, $F_{(2,14)}$ = 3.04, p = .080, η_p^2 = .30) although 'under the surface' the between-subjects effects showed a significant effect of Media on *ΔVal* ($F_{(1,15)}$ = 5.64, p = .031, η_p^2 = .27) into the expected direction: Robot ($M_{\Delta Val}$ = 2.65, SD = .80) was higher than Writing ($M_{\Delta Val}$ = 1.69, SD = .83). There was still no significant effect of Media on *MRel* ($F_{(1,15)}$ = .074, p = .790, η_p^2 = .005).

It seemed, then, with the outliers dismissed from the data (and less power due to fewer subjects), that the effects tended to disappear. It is for those who suffer the most that robots are most helpful. The novelty aspect of talking to a robot may make the medium more relevant to personal goals and concerns but is not (or for the less affected only marginally) influential for feeling more positive after a chat with a robot about negative experiences.

3.4.2. Positive and Negative Valence as two unipolar scales in high-negative subjects

For n = 23, we ran two GLM Repeated measures of Media (2 conditions) on within-subjects factor (*ΔValP* vs *ΔValN*) with *MRel* and *MNov* separately as covariate. Multivariate tests showed that no significant effects occurred for *ΔValP* vs *ΔValN* (V = .02, $F_{(1,20)}$ = .36, p = .555, η_p^2 = .02). The height of positive and negative valence did not differ. The interaction of (*ΔValP* vs *ΔValN*) with Media also was not significant (V = .04, $F_{(1,20)}$ = .78, p = .387, η_p^2 = .04) nor was *MRel* as a covariate (V = .003, $F_{(1,20)}$ = .06, p = .815, η_p^2 = .003; $F_{(1,20)}$ = 3.78, p = .066, η_p^2 = .16). However, the main effect of Media was significant ($F_{(1,20)}$ = 13.54, p = .001, η_p^2 = .40), showing that robots exerted higher levels of undifferentiated Valence (non-unipolar) than writing on paper. We repeated the test but now with Novelty as the covariate but *MNov* did not significantly contribute to any of the effects.

Then we did the same for the data set of n = 17. We ran two GLM Repeated measures of Media (2 conditions) on within-subjects factor (*ΔValP* vs *ΔValN)* with *MRel* and *MNov* as separate covariates. Multivariate tests showed that no significant effects were obtained for *ΔValP* vs *ΔValN* (V = .008, $F_{(1,14)}$ = .11, p = .749, η_p^2 = .008). Here as well, the height of positive and negative valence did not differ. The interaction of (*ΔValP* vs *ΔValN*) with Media also was not significant (V = .03, $F_{(1,14)}$ = .48, p = .498, η_p^2 = .033) nor was *MRel* as a covariate (V = .000, $F_{(1,14)}$ = .06, p = .936, η_p^2 = .000). Yet, the main effect of Media remained significant ($F_{(1,14)}$ = 5.98, p = .028, η_p^2 = .30). Repeating the analysis with Novelty as the covariate did not change these results (V = .011, $F_{(1,14)}$ = .16, p = .695, η_p^2 = .011) except for the main effect of Media, which now became marginally significant ($F_{(1,14)}$ = 4.07, p = .063, η_p^2 = .23).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

Thus, with outliers excluded, robots still exerted higher levels of undifferentiated Valence (non-unipolar) than writing on paper. With Novelty included, these positive effects became somewhat more pronounced for the less affected.

3.4.3. Exploratory analyses

In the previous, we saw that Novelty mainly affected Relevance, indicating that a medium becomes more relevant the newer it is to those who emotionally are affected but not too much. In Section 3.1, we found in turn that Novelty was affected by Gender. Therefore, we explored the Media × Gender effects on Novelty with Univariate ANOVA for both data sets $N = 45$ and $n = 31$. The research question was if robots were newer to females than to men or v.v.?

With $N = 45$, only the main effects were significant: Robots ($M = 4.10$, $SD = .87$) were perceived as newer than writing ($M = 3.41$, $SD = .77$) ($F_{(1,41)} = 9.50$, $p = .004$, $\eta_p^2 = .19$). This was independent of Gender. Females ($n = 24$, $M = 4.03$, $SD = .83$) experienced more Novelty than did males ($n = 21$, $M = 3.50$, $SD = .87$) ($F_{(1,41)} = 5.98$, $p = .019$, $\eta_p^2 = .13$), irrespective of the medium (Figure 6).

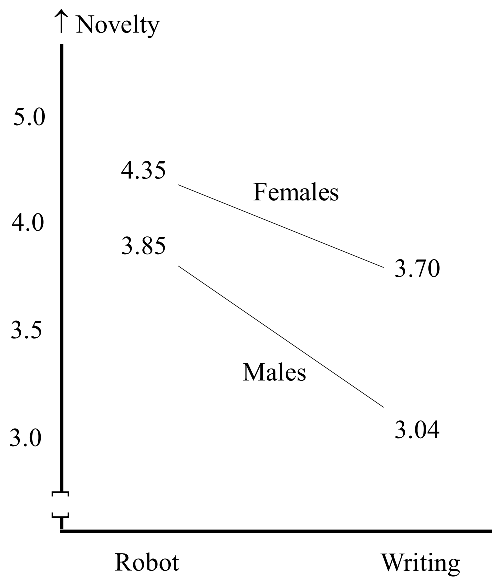

Figure 6. Effects of Gender and Media on Novelty ($N = 45$).

With $n = 31$, only one main effect was significant: Females ($n = 15$, $M = 4.23$, $SD = .74$) experienced more Novelty than males ($n = 16$, $M = 3.51$, $SD = .95$) ($F_{(1,27)} = 5.35$, $p = .029$, $\eta_p^2 = .17$) and medium showed no significant effects any more ($F_{(1,27)} = 2.98$, $p = 0.95$). In sum, females experienced more novelty but not particularly with respect to robots.

4. Discussion and Conclusions

Our manipulation was successful: The video was rated as significantly inducing strong negative mood. Our treatment also was successful: We could demonstrate significant improvement of positive affect and reduction of negative affect after treatment.

We assumed that after negative-mood induction, (H1) a social robot that invites self-disclosure will lower the level of negative emotions more than writing a journal page does. Indeed, our self-disclosure AI chatbot in unison with the DARwIn Mini embodiment made viewers of video recordings of the Wenchuan Sichuan earthquake in China 2008 significantly more positive. This was particularly so for people who were most

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

negatively affected by the video. For those less affected, writing a diary page also sufficed.

In our study, valence should be conceived of as a bipolar dimension. Significant and reasonably strong main effects of robots exerting more positive results than writing were established with bipolar valence, particularly for those with high negativity. Even when we analyzed valence as a within-factor of two levels measured as separate unipolar scales, the significant effects of media happened across positive and negative valence, not to these measures separately. Novelty of the medium (either robot or writing) did not affect the effects on bipolar valence – or occasionally for the less affected.

Then a note on analysis. If we had followed conventional statistical practice, we had eliminated the outliers from our data set and had found no differences between writing and robots in alleviating stress and anxiety. In reporting a null effect, looking at normal distributions only would have missed the upshot that those who are most in need of mental support should not be deprived of a treatment that is more effective than traditional text writing; something that comes closer to a therapist; a social robot that can relieve the shortage of caregivers in the mental care sector.

We also hypothesized (H2) that a social robot that invites self-disclosure is more relevant to goals, needs, and concerns than writing on paper. This was not the case. Although we measured the highest grand mean averages for relevance, whether tested for those high or low on emotional negativity, men or women, relevance did not differ for any of the fixed factors tested and did not significantly contribute to the effects on valence. It was only in unison with novelty that relevance took effect. The novelty aspect of talking to a robot or writing on paper apparently made the medium more relevant to personal goals and concerns. As an extra, women experienced more novelty of the presented medium than men but this was not specific to the robot or the writing condition.

New technologies such as social robots bring various opportunities of discovering new methods to improve an individual's well-being and suggest that such new technologies will alleviate the pressure on current healthcare, such as care for older adults, depressed youth, and groups with special needs (Broadbent, 2017 [5]). Our study focused on social robots helping individuals to improve their mental well-being through self-disclosure. The results suggest that individuals who have a relatively high level of negative emotions benefit the most from robots.

Our results are not consistent with Slavin-Spenny, Cohen, Oberleitner, and Lumley (2011) [79] that the effect of four conditions of disclosure (writing, private spoken, talking to a passive listener, talking to an active facilitator) had about the same effect in reducing negative emotions because after the disclosure session, the negative emotions remained. Our results also run counter with Murray and Segal (1994) [50] that the two procedures (talking and writing) were almost identical in reducing negative affect and in producing adaptive changes in cognition and self-esteem. However, our results are not at odds with Murray, Lamnin, and Carver (1989) [49] who compared writing and talking with a psychotherapist and found that after writing, no increase of positive emotions happened whereas after talking with the therapist, positive emotions increased. Maybe the answer lies in the change of focus: Talking to a (virtual) therapist does not so much decrease negativity as it compensates negativity by increasing positive affect.

Not reducing negativity, however, would go against studies by Murray and Segal (1994) [50], Epstein, Sloan, and Marx (2005) [16], Sloan, Marx, Epstein, and Lexington (2007), Perez, Penate, Bethencourt, and Fumero (2017) [80], which all show that emotional disclosure intervention is effective in reducing the level of negative emotions. Perhaps that the decrease in negativity takes longer than the immediate joy of encountering a (virtual) human. The length of the emotional disclosure session in the said studies was way longer than our 10 minutes. Frattaroli (2006) [19] concluded from a meta-analysis that such sessions usually last for days or weeks.

A limitation of our current study is that participants took the questionnaire but once, after talking with the robot or after writing, rather than after the video as well as after

 doi:10.20944/preprints202105.0279.v1

treatment. Emotions are short-lived, notoriously ephemeral, and decay rapidly after elicitation (Scherer, 2005 [74]; Costa, Cauda, Crini, Tatu, Celeghin, De Gelder, & Tamietto, 2014 [9]; Hassenzahl, 2004) [25]. Self-reports should be administered as soon as possible to certify that the measured emotion is related to the experienced one (Lane & Terry, 2000 [36]; Qiao-Tasserit, Garcia Quesada, Antico, Bavelier, Vuilleumier, & Pichon, 2017 [63]).

Additionally, emotions tend to subside once individuals reflect on them (Mauss, & Robinson, 2009) [46], for example, by filling out a questionnaire. Perhaps next time we should combine self-reports together with physiological reactions and behaviors for triangulation purposes (Erevelles, 1998 [17]; Lang, 1993 [38]; Frijda, 1986 [21], 2007 [22]; Scherer, 2005 [74], Scherer, & Zentner, 2001 [76]).

Our results fall in line with the positive effects robots exerted in HRI studies conducted by Wada, Shibata, Saito, Sakamoto, and Tanie (2005) [89], Jibb, Birnie, Nathan, Beran, Hum, Victor, and Stinson (2018) [30], Dang and Tapus (2015) [14], Kumazaki, Warren, Swanson, Yoshikawa, Matsumoto, Takahashi, and Kikuchi (2018) [34], and Pu, Moyle, Jones, and Todorovic (2018) [62]. In psychotherapy as well, Libin and Libin (2004) [43], Costescu, Vanderborght, and David (2014) [10], Kidd, Taggart, and Turkle (2006) [32], Wada, Shibata, Saito, Sakamoto, and Tanie (2005) [89], and Wada and Shibata (2007) [88] report beneficial effects of robot-enhanced therapy.

With such overwhelming evidence for robot-supported mental well-being, we felt we should set out to put this knowledge into design practice. As is, we are in the process of developing our own robot MEME, a talking stress ball that embodies our self-disclosure AI chatbot. MEME may replace throwing a brick, launching a gas grenade, or suicidal ideation. We hope MEME may help people who feel depressed during social isolation or calms down those who seek violence to settle their disputes in Hong Kong. Figure 7 shows the development steps so far.

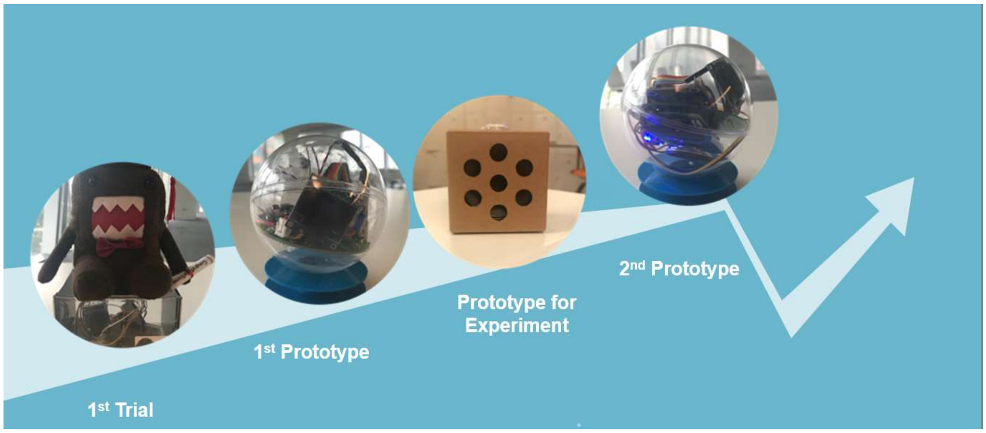

Figure 7. Prototyping the talking stress ball MEME.

MEME is round and covered in silicone rubber for a soft touch, look, and feel. It is portable, pocket-size, and is easy to carry on. The cover can be adapted according to the user's taste. Interaction with MEME takes place through emoji (Figure 8). Many studies in Computer-Mediated Communication assert the importance of emoji in non-verbal interactions to represent a person's affectionate or depressed feelings (e.g., Crystal, 2006 [12]; Rezabek & Cochenour, 1998 [64]; Wolf, 2000 [93]; Li & Yang, 2018 [42]). MEME can be found at www.roboticmeme.com. The logo was designed according to the directions of Walsh, Winterich, and Mittal (2010) [90].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

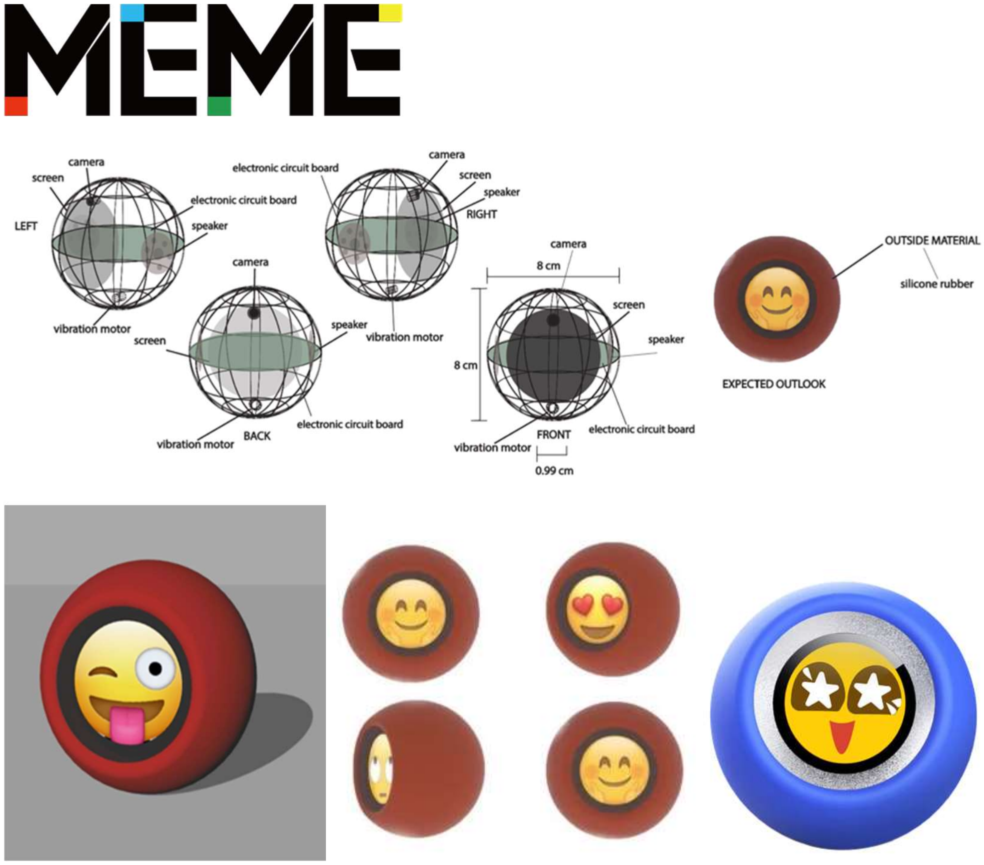

Figure 8. MEME complements the chatbot function with emoticons.

To conclude, the video footage of the 2008 Sichuan earthquake aroused negative emotions, which were mitigated by self-disclosure to a robot or by writing a journal diary page. The choice of medium was indifferent to most participants, both means worked for them. For those who felt extremely bad after the shocking video, however, the medium did make a difference. Those high on negative valence were significantly more positive after talking to a social robot.

Valence in this study should be conceived of as a bipolar scale (i.e. more positive is less negative). Relevance had little to do with these effects and was most susceptible to novelty of the medium. The newer, the more personally relevant the medium seemed. The experience of novelty had little effect on valence and was higher for robots than for writing. Although females experienced more novelty throughout, this had nothing to do with robots as such. Robots seem good candidates to aid people with stress and anxiety problems. We took a shot at such opportunities by designing our own MEME stress ball with an emoji enhanced self-disclosure inviting AI chatbot.

Noteworthy, however, is that the positive effects we observed are valid for the robot as a whole. We should not attribute the positive effects on emotional valence to particular design features such as specific parts of the embodiment or the quality of the chatbot. As one of the participants astutely commented:

> The robot answered my questions in weird ways sometimes, and repeated some questions. I think the unexpected movement of the robot was the best part of the experiment. It affectively changed my mood. Not so much the conversation itself.

Let this be a reminder to us robot researchers, AI developers, and designers: The robot made funny moves and that cheered this participant up. Not the conversation

 doi:10.20944/preprints202105.0279.v1

about difficult things. Perhaps in the future, in concert with talking stress balls, we should create paper and pens that make sudden funny moves as well.

Supplementary Materials: Technical Report S1: Self-disclosure to a Robot or on Paper.

Author Contributions: Conceptualization, E.Y.D.; methodology, E.Y.D., J.M.Y., and J.F.H.; software, E.Z.L.; validation, E.Y.D., J.M.Y., and E.Z.L.; formal analysis, E.Y.D., J.M.Y., and J.F.H.; investigation, E.Y.D., J.M.Y., and E.Z.L.; resources, E.Y.D., J.M.Y., E.Z.L., and J.F.H.; data curation, J.F.H.; writing—original draft preparation, E.Y.D. and J.F.H.; writing—review and editing, J.F.H.; visualization, E.Y.D. J.M.Y., and J.F.H.; supervision, J.F.H.; project administration, J.F.H.; funding acquisition, J.F.H. All authors have read and agreed to the published version of the manuscript.

Funding: The contribution by Johan F. Hoorn was supported by the project *Negative-mood reduction among HK youth with robot PAL (Personal Avatar for Life)* of the Artificial Intelligence in Design Laboratory in Hong Kong (grant number: AiDLab RP2P3).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Subjects Ethics Sub-committee of the university filed under HSEARS20200204003. The authors declare that they have no conflict of interest.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data can be made available by the corresponding author.

Acknowledgments: This study was a Chinese, Korean, and Dutch collaboration of students in the School of Design and the Dept. of Computing. Zachary Tan is kindly acknowledged for assembling the DARwIn Mini as part of his primary school project on robotics.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Appendix A

Structured questionnaire for self-disclosure to a robot (English translated from the Chinese). Online Resource 1 provides the questionnaire versions for robot as well as for writing, in Chinese and English.

Dear Sir/Madam,

Thank you for your time for our experiment. We would like to ask you to answer a few questions. Answering these questions will only take a few minutes.

You have the right to withdraw at any point during the study, for any reason, and without any prejudice. If you would like to contact the Principal Investigator in the study to discuss this research, please e-mail <name> via <name>@connect.polyu.hk.

By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason. The data provided by the participants of the study will be processed and published anonymously in the results sections of the paper.

This study is supervised by The Hong Kong Polytechnic University.

Thank you for your participation.

With kind regards,
Team Social Robot MEME

I agree to participate in this study
I do not agree to participate in this study

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

Scale Valence-before-treatment

I. After seeing the film samples…

Vb1i I feel good

```
Totally                 Disagree   Agree a              Totally
Disagree   Disagree     a little   little     Agree     agree
1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6
```

Vb2i I am well

```
Totally                 Disagree   Agree a              Totally
Disagree   Disagree     a little   little     Agree     agree
1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6
```

Vb3i I have positive feelings
Vb4i I am optimistic
Vb5c I feel bad
Vb6c I am unwell
Vb7c I have negative feelings
Vb8c I am pessimistic

Scale Valence-after-treatment

II. After talking to the robot…

Vb1i I feel good
Vb2i I am well
Vb3i I have positive feelings
Vb4i I am optimistic
Vb5c I feel bad
Vb6c I am unwell
Vb7c I have negative feelings
Vb8c I am pessimistic

Scale Relevance

III. To regulate my emotions, talking to the robot is…

Re1i useful
Re2i worthwhile
Re3c worthless
Re4c useless

Scale Novelty

IV. Talking to a robot is…

No1i novel

 doi:10.20944/preprints202105.0279.v1

No2i original
No3i unexpected
No4c predictable
No5c commonplace
No6c old-fashioned

Demographics

De1 Gender

Female
Male
Other

De2 Age

De3 What is your highest completed education or current education level?

Primary school or below
Secondary school
Post-secondary school / Associate Degree / Diploma
University undergraduate
Master degree
Doctoral degree or above

De4 Ethnicity

Asia
Africa
Europe
North America
South America
Australia/Oceania
Antarctica

If you have any further questions or remarks about this questionnaire, please let us know.

You can write your feedback below.

Kind regards,

Social Robot MEME
<name>@connect.polyu.hk

References

1. Alford, W., Malouff, J., & Osland, K. (2005). Written emotional expression as a coping method in child protective services officers. *International Journal of Stress Management, 12*(2), 177-187.
2. Altman, I., & Taylor, D. (1973). *Social penetration: The development of interpersonal relationships.* New York, NY: Holt.
3. Barrett, L. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality, 40*(1), 35-55.

4. Bolls, P., Lang, A., & Potter, R. (2001). The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements. *Communication Research*, *28*(5), 627-651.
5. Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*(1), 627-652.
6. Cabibihan, J., Javed, H., Ang, M., & Aljunied, S. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics, 5*(4), 593-618.
7. Carrera, P., & Oceja, L. (2007). Drawing mixed emotions: Sequential or simultaneous experiences? *Cognition & Emotion*, *21*(2), 422-441.
8. Consedine, N. S., & Moskowitz, J. T. (2007). The role of discrete emotions in health outcomes: A critical review. *Applied and Preventive Psychology*, *12*(2), 59-75.
9. Costa, T., Cauda, F., Crini, M., Tatu, M., Celeghin, A., De Gelder, B., & Tamietto, M. (2014). Temporal and spatial neural dynamics in the perception of basic emotions from complex scenes. *Social Cognitive and Affective Neuroscience, 9*(11), 1690-1703.
10. Costescu, C. A., Vanderborght, B., & David, D. O. (2014). The effects of robot-enhanced psychotherapy: a meta-analysis. *Review of General Psychology*, *18*(2), 127–136.
11. Croyle, R. T., & Uretzky, M. D. (1987). Effects of mood on self-appraisal of health status. *Health Psychology, 6*(3), 239-253.
12. Crystal, D. (2006). *Language and the Internet.* Cambridge, UK: Cambridge University.
13. Dalton, J., & Glenwick, D. (2009). Effects of expressive writing on standardized graduate entrance exam performance and physical health functioning. *The Journal of Psychology, 143*(3), 279-292.
14. Dang, T.-H.-H., & Tapus, A. (2015). The role of motivational robotic assistance in reducing user's task stress. *International Journal of Social Robotics, 7*(2), 227-240. doi: 10.1007/s12369-014-0256-9
15. Diener, E. (1999). Introduction to the special section on the structure of emotion. *Journal of Personality and Social Psychology*, *76*(5), 803-804.
16. Epstein, E. M., Sloan, D. P., & Marx, B. (2005). Getting to the heart of the matter: written disclosure, gender, and heart rate. *Psychosomatic Medicine, 67*(3), 413-419.
17. Erevelles, S. (1998). The role of affect in marketing. *Journal of Business Research*, *42*(3), 199-215.
18. Farber, B. (2006). *Self-disclosure in psychotherapy.* New York: Guilford.
19. Frattaroli, J. (2006). Experimental disclosure and its moderators: a meta-analysis. *Psychological Bulletin*, *132*(6), 823-865.
20. Friedman, L. M., Furberg, C. D., & DeMets, D. L. (2010). *Fundamentals of clinical trials.* New York: Springer.
21. Frijda, N. (1986). *The emotions (studies in emotion and social interaction)*. New York: Cambridge University.
22. Frijda, N. (2007). *The laws of emotion.* Mahwah, NJ: Erlbaum.
23. Frisina, P. G., Borod, J. C., & Lepore, S. J. (2004). A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *The Journal of Nervous and Mental Disease*, *192*(9), 629-634.
24. Fontaine, J., Scherer, K., & Soriano, C. (2013). *Components of emotional meaning. A sourcebook.* Oxford, UK: Oxford University.
25. Hassenzahl, M. (2004). Emotions can be quite ephemeral; we cannot design them. *Interactions, 11*(5), 46-48.
26. Hemenover, S. (2003). Individual differences in rate of affect change: studies in affective chronometry. *Journal of Personality and Social Psychology, 85*(1), 121-131.
27. Horneffer, K., & Jamison, P. (2002). The emotional effects of writing about stressful experiences: an exploration of moderators. *Occupational Therapy in Health Care, 16*(2-3), 77-89.
28. Hoorn, J. F., Konijn, E. A., Germans, D. M., Burger, S., & Munneke, A. (2015). The in-between machine: The unique value proposition of a robot or why we are modelling the wrong things. In S. Loiseau, J. Filipe, B. Duval, & J. van den Herik (Eds.), *Proceedings of the 7th International Conference on Agents and Artificial Intelligence (ICAART) Jan. 10-12, 2015. Lisbon, Portugal* (pp. 464-469). Lisbon, PT: ScitePress.
29. Hopwood, T. L., & Schutte, N. S. (2017). Psychological outcomes in reaction to media exposure to disasters and large-scale

violence: A meta-analysis. *Psychology of Violence, 7*(2), 316-327. doi: 10.1037/vio0000056

30. Ireland, M., Malouff, J., & Byrne, B. (2007). The efficacy of written emotional expression in the reduction of psychological distress in police officers. *International Journal of Police Science & Management, 9*(4), 303-311.
31. Jibb, L. A., Birnie, K. A., Nathan, P. C., Beran, T. N., Hum, V., Victor, J. C., & Stinson, J. N. (2018). Using the MEDiPORT humanoid robot to reduce procedural pain and distress in children with cancer: A pilot randomized controlled trial. *Pediatric Blood & Cancer, 65*(9), e27242. doi: 10.1002/pbc.27242
32. Kidd, C., Taggart, W., & Turkle, S. (2006). A sociable robot to encourage social interaction among the elderly. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06) May 15-19, 2006. Orlando, FL* (pp. 3972-3976). Piscataway, NJ: IEEE.
33. Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication care: A case-study in autism therapy. In *IEEE International Workshop on Robot and Human Interactive Communication (ROMAN '05) Aug 13-15, 2005. Nashville, TN* (pp. 341-346). Piscataway, NJ: IEEE.
34. Kumazaki, H., Warren, Z., Swanson, A., Yoshikawa, Y., Matsumoto, Y., Takahashi, H., … & Kikuchi, M. (2018). Can robotic systems promote self-disclosure in adolescents with autism spectrum disorder? A pilot study. *Frontiers in Psychiatry, 9*, 36. doi: 10.3389/fpsyt.2018.00036
35. Kubzansky, L. D., & Kawachi, I. (2000). Going to the heart of the matter: Do negative emotions cause coronary heart disease? *Journal of Psychosomatic Research, 48(4-5)*, 323-337.
36. Lane, A. M., & Terry, P. C. (2000). The nature of mood: Development of a conceptual model with a focus on depression. *Journal of Applied Sport Psychology, 12*, 16-33.
37. Lang, A., Shin, M., & Lee, S. (2005). Sensation seeking, motivation, and substance use: A dual system approach. *Media Psychology, 7*(1), 1-29.
38. Lang, P. (1993). The three-system approach to emotion. In N. Birbaum & A. Ohman (Eds.), *The structure of emotion* (pp. 18-30). Bern, Switzerland: Hogrefe.
39. Lazarus, R. (1991). *Emotion and adaptation.* Oxford, UK: Oxford University.
40. Lerner, J., Li, Y., Valdesolo, P., & Kassam, K. (2015). Emotion and decision making. *Annual Review of Psychology, 66*, 799-823.
41. Levine, L., & Edelstein, R., (2009). Emotion and memory narrowing: A review and goal relevance approach. *Cognition and Emotion, 23*(5), 833-875.
42. Li, L., & Yang, Y. (2018). Pragmatic functions of emoji in internet-based communication---a corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education, 3*, 16. doi: 10.1186/s40862-018-0057-z
43. Libin, A. V., & Libin, E. V. (2004). Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach. *Proceedings of the IEEE, 92*(11), 1789-1803.
44. Lumley, M. (2004). Alexithymia, emotional disclosure, and health: a program of research. *Journal of Personality, 72*(6), 1271-1300.
45. Martelaro, N., Nneji, V., Ju, W., & Hinds, P. (2016). Tell me more designing HRI to encourage more trust, disclosure, and companionship. In C. Bartneck, Y. Nagai, A. Paiva, & S. Šabanović (Eds.), *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI '16) Mar. 07-10, 2016. Christchurch, New Zealand* (pp. 181-188). Piscataway, NJ: IEEE.
46. Mauss, I., & Robinson, M. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209-237.
47. Mayne, T. J. (1999). Negative affect and health: The importance of being earnest. *Cognition & Emotion, 13*(5), 601-635.
48. Mayne, T. J. (2001). Emotions and health. In T. J. Mayne & G. A. Bonanno (Eds.), *Emotions: current issues and future directions* (pp. 361-397). New York: Guilford.
49. Murray, E. J., Lamnin, A. D., & Carver, C. S. (1989). Emotional expression in written essays and psychotherapy. *Journal of Social and Clinical Psychology, 8*(4), 414-429. doi: 10.1521/jscp.1989.8.4.414

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2021 doi:10.20944/preprints202105.0279.v1

50. Murray, E., & Segal, D. (1994). Emotional processing in vocal and written expression of feelings about traumatic experiences. *Journal of Traumatic Stress, 7*(3), 391-405.
51. Nomura, T., & Kawakami, K. (2011). Relationships between robot's self-disclosures and human's anxiety toward robots. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT'11) Aug. 22-27, 2011. Lyon, France* (pp. 66-69). Piscataway, NJ: IEEE. doi: 10.1109/WI-IAT.2011.17
52. Nystul, M. (2016). *Introduction to counseling: An art and science perspective* (5th). Thousand Oaks, CA: Sage.
53. Pascoe, P. (2016). Using patient writings in psychotherapy: Review of evidence for expressive writing and cognitive-behavioral writing therapy. *American Journal of Psychiatry Residents' Journal, 11*(3), 3-6.
54. Parkinson, B. (1995). *Ideas and realities of emotion.* London: Routledge.
55. Pennebaker, J. W. (1985). Traumatic experience and psychosomatic disease. Exploring the roles of behavioural inhibition, obsession, and confiding. *Canadian Psychology, 26*(2), 82-95.
56. Pennebaker, J. (1993). Putting stress into words: health, linguistic, and therapeutic implications. *Behaviour Research and Therapy, 31*(6), 539-548.
57. Pennebaker, J., & Beall, S. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95*(3), 274-281. doi: 10.1037/0021-843X.95.3.274
58. Pennebaker, J., & Francis, M. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion, 10*(6), 601-626. doi: 10.10801026999396380079
59. Perez, S., Penate, W., Bethencourt, J., & Fumero, A. (2017). Verbal emotional disclosure of traumatic experiences in adolescents: the role of social risk factors. *Frontiers in Psychology*, *8,* 372. doi: 10.3389/fpsyg.2017.00372
60. Pressman, S., Gallagher, M., & Lopez, S. (2013). Is the emotion-health connection a "first-world problem"? *Psychological Science, 24*(4), 544-549.
61. Psychopathology Committee of the Group for the Advancement of Psychiatry (2001). Reexamination of therapist self-disclosure*. Psychiatric Services, 52*(11), 1489-1493.
62. Pu, L., Moyle, W., Jones, C., & Todorovic, M. (2019). The effectiveness of social robots for older adults: A systematic review and meta-analysis of randomized controlled studies. *The Gerontologist, 59*(1), e37-e51. doi: 10.1093/geront/gny046
63. Qiao-Tasserit, E., Garcia Quesada, M., Antico, L., Bavelier, D., Vuilleumier, P., & Pichon, S. (2017). Transient emotional events and individual affective traits affect emotion recognition in a perceptual decision-making task. *PLoS One, 12*(2), e0171375. doi: 10.1371/journal.pone.0171375
64. Rezabek, L. L., & Cochenour, J. J. (1998). Visual cues in Computer-Mediated Communication: Supplementing text with emoticons. *Journal of Visual Literacy, 18*(2), 201-215. doi: 10.1080/23796529.1998.11674539
65. Robins, B., Amirabdollahian, F., Ji, Z., & Dautenhahn, K. (2010). Tactile interaction with a humanoid robot for children with autism: A case study analysis involving user requirements and results of an initial implementation. In C. A. Avizzano & E. Ruffaldi (Eds.), *Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '10) Sept. 13-15, 2010. Viareggio, Italy* (pp. 704-711). Piscataway, NJ: IEEE. doi: 10.1109/ROMAN.2010.5598641
66. Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172.
67. Russell, J. A. (2017). Mixed emotions viewed from the psychological constructionist perspective. *Emotion Review, 9*(2), 111-117.
68. Russell, J. A., & Carroll, J. M. (1999a). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125*(1), 3-30.
69. Russell, J. A., & Carroll, J. M. (1999b). The phoenix of bipolarity: Reply to Watson and Tellegen (1999). *Psychological Bulletin, 125*(5), 611-617.
70. Saffarizadeh, K., Boodraj, M., & Alashoor, T. M. (2017). Conversational assistants: Investigating privacy concerns, trust, and self-disclosure. In Y. J. Kim, R. Agarwal, & J. K. Lee (Eds.), *Proceedings of the International Conference on Information Systems - Transforming Society with Digital Innovation (ICIS '17) Dec. 10-13, 2017. Seoul, South Korea.* Atlanta, GA: Association for Infor-

mation Systems.
71. Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15), March 02-05, 2015. Portland, OR* (pp. 141-148). New York: ACM. doi: 10.1145/2696454.2696497
72. Scherer, K. R. (1984). On the nature and function of emotions: A component process approach. In Scherer, K. R., Ekman, P. (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.
73. Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: theory, methods, research* (pp. 92-120). New York, NY: Oxford University.
74. Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695-729.
75. Scherer, K. R. (2013). The nature and dynamics of relevance and valence appraisals: theoretical advances and recent evidence. *Emotion Review, 5*, 150-162.
76. Scherer, K. R. & Zentner, M. R. (2001). Emotional effects of music: production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: theory and research* (pp. 361-392). Oxford, UK: Oxford University.
77. Shuman, V., Sander, D., & Scherer, K. R. (2013). Levels of valence. *Frontiers in Psychology, 4*, 261. doi: 10.3389/fpsyg.2013.00261
78. Siedlecka, E., & Denson, T. (2019). Experimental methods for inducing basic emotions: A qualitative review. *Emotion Review, 11*(1), 87-97.
79. Slavin-Spenny, O., Cohen, J., Oberleitner, L., & Lumley, M. (2011). The effects of different methods of emotional disclosure: Differentiating post-traumatic growth from stress symptoms. *Journal of Clinical Psychology, 67*(10), 993-1007.
80. Sloan, D. M., Marx, B. P., Epstein, E. M., & Lexington, J. M. (2007). Does altering the writing instructions influence outcome associated with written disclosure? *Behavior Therapy, 38*(2), 155-168.
81. Smith, C. A. & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotions. *Journal of Personality and Social Psychology, 48*(4), 813-838.
82. Smyth, J. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology, 66*(1), 174-184. doi: 10.1037/0022-006X.66.1.17
83. Sonnemans, J., & Frijda, N. (1995). The determinants of subjective emotional intensity. *Cognition and Emotion*, 9(5), 483-506.
84. Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., … David, D. O. (2012). Children with autism social engagement in interaction with Nao, an imitative robot. A series of single case experiments. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 13,* 315-347. doi: 10.1075/is.13.3.01tap
85. Thayer, J. F., & Ruiz-Padial, E. (2006). Neurovisceral integration, emotions and health: an update. *International Congress Series, 1287*, 122-127.
86. Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. *Clinical Psychological Science, 6*(1), 3-17. doi: 10.1177/2167702617723376
87. Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., Dirk, L., & David, D. O. (2012). Using the social robot Probo as a social story telling agent for children with ASD. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 13,* 348-372. doi: 10.1075/is.13.3.02van
88. Wada, K., & Shibata, T. (2007). Social effects of robot therapy in a care house - change of social network of the residents for two months. In *Proceedings of the 2007 IEEE International Conference on Robotics and Automation (ICRA '07) April 10-14, 2007. Roma, Italy* (pp. 1250-1255). Piscataway, NJ: IEEE.
89. Wada, K., Shibata, T., Saito, T., Sakamoto, K., & Tanie, K. (2005). Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In *Proceedings of the 2005 IEEE International Conference on Robotics*

and Automation (ICRA '05) April 18-22, 2005. Barcelona, Spain (pp. 2785-2790). doi: 10.1109/ROBOT.2005.1570535

90. Walsh, M. F., Winterich, K. P., & Mittal, V. (2010). Do logo redesigns help or hurt your brand? The role of brand commitment. *Journal of Product & Brand Management, 19*(2), 76-84.
91. Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070.
92. Weizenbaum, J. (1966). ELIZA - a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36-45.
93. Wolf, A. (2000). Emotional expression online: Gender difference in emotion use. *Cyber Psychology & Behaviour, 3*(5), 827-833.
94. World Health Organization (2018, June 06). *WHO's Mental Health Atlas 2017 highlights global shortage of health workers trained in mental health.* Available from https://www.who.int/hrh/news/2018/WHO-MentalHealthAtlas2017-highlights-HW-shortage/en/
95. Złotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2016). Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy. *Paladyn: Journal of Behavioral Robotics, 7*(1), 55-66. doi: 10.1515/pjbr-2016-0005
96. Bu, F., Steptoe, A., & Fancourt, D. (2020a). Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. Public Health, 186, 31-34.
97. Bu, F., Steptoe, A., & Fancourt, D. (2020b). Loneliness during a strict lockdown: Trajectories and predictors during the COVID-19 pandemic in 38,217 United Kingdom adults. Social Science & Medicine, 265, 113521.
98. Tso, I. F., & Park, S. (2020). Alarming levels of psychiatric symptoms and the role of loneliness during the COVID-19 epidemic: A case study of Hong Kong. Psychiatry research, 293, 113423.
99. Li, L. Z., & Wang, S. (2020). Prevalence and predictors of general psychiatric disorders and loneliness during COVID-19 in the United Kingdom. Psychiatry research, 291, 113267.
100. Killgore, W. D., Cloonan, S. A., Taylor, E. C., Miller, M. A., & Dailey, N. S. (2020). Three months of loneliness during the COVID-19 lockdown. Psychiatry Research, 293, 113392.
101. Williams, C. Y., Townson, A. T., Kapur, M., Ferreira, A. F., Nunn, R., Galante, J., ... & Usher-Smith, J. A. (2021). Interventions to reduce social isolation and loneliness during COVID-19 physical distancing measures: A rapid systematic review. PloS one, 16(2), e0247139.