
Article

Social Robots Outdo the Not-so-Social Media for Self-disclosure: *Safe Machines Preferred to Unsafe Humans?*

Rowling L. Luo ¹, Thea X. Y. Zhang ¹, Derrick H.-c. Chen ¹, Johan F. Hoorn ^{1*}, and Ivy S. Huang ¹

¹ School of Design and Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

*Corresponding author: johan.f.hoorn@polyu.edu.hk

Abstract: COVID-19 may not be a ‘youth disease’ but nevertheless impacts the life of young people dramatically, loneliness and negative mood being an unexpected additional pandemic. Many young people rely on social media for their feeling of connectedness with others. However, social media are suggested to have many negative effects on people’s anxiety. Instead of self-disclosing to others, design may develop alternatives to employ social robots for self-disclosure. In a follow-up on Duan et al. (2021), we report on a lab experiment of self-disclosing negative emotions to a social-media group as compared to writing a conventional diary journal or to talk to an AI-driven social robot after negative mood induction (i.e. viewing shocking earthquake footage). Participants benefitted the most from talking to a robot rather than from writing a journal page or sharing their feelings on social media. Self-disclosure on social media or writing a journal page did not differ significantly. In the design of interventions for mental well-being, human helpers thus far took center stage. Based on our results, we propose design alternatives for an empathic smart-home, featuring social robots and chatbots for alleviating stress and anxiety: a social-media interference chatbot, smart watch plus speaker, and a mirror for self-reflection.

Keywords: social media; social robots; negative mood; self-disclosure; intervention design; anthropomorphic design

1. Introduction

President and Executive Committee of the International Association of Child and Adolescent Psychiatry and Allied Professionals (IACAPAP), speaks of “a family and society pandemic.” Those in less advantageous social-economic circumstances suffered even more (cf. Xafis, 2020).

In seeking social connection virtually, young people relied (and rely) heavily on social media such as WhatsApp and WeChat. However, research indicates that social media have a host of unfavorable effects on young people’s mental and social wellbeing. The overuse of social media may lead to overexposure to unreliable information, which breeds an environment of instable interpersonal relationships, biasing youngsters’ perceptions and exacerbating their anxiety.

The rapid growth of social media has revolutionized the way people communicate and seek information. The openness and timeliness of social media also allow misinformation to be created and spread quickly. Social media such as Facebook, YouTube, and Instagram are web-based and app-based sites that let people communicate with each other by sharing messages, images, opinions, and events (Wang, Wang, Hao, Zhang, Liu, Wang, & Lin, 2018). Algorithms recommend content to users, shaping the scope of their knowledge and means of knowing (Gillespie, Boczkowski, & Foot, 2014). In filtering the Internet and social media, AI determines what people will be exposed to, which raises concerns about ‘echo chambers’ and ‘filter bubbles’ (Pariser, 2011; Gillani, Yuan, Saveski, Vosoughi, & Roy, 2018). The concern is that social media algorithms combined with the tendency to interact with like-minded people create an environment that primarily

exposes users to the like-minded. Keeping users in an endless scrolling loop may lead to social comparison behaviors. Viewing other people's carefully curated information, youth may become susceptible to "frequent and extreme upward social comparisons" (Festinger, 1954), which may cause negative side effects such as erosion of self-esteem, increase of depressed mood, and reduced life satisfaction.

Group polarization happens when group members end up taking more extreme positions on a given issue after participating in or contacting a discussion (Isenberg, 1986). When a group expands and views are exchanged, discussions tend to go into a certain direction, whether supportive or not. The opinions tend to converge and start to be consistent throughout the group, which may lead to group think and a majority of participants sharing the same standpoints. In extreme cases, group polarization occurs (Wang, Yang, & Xi, 2018; Lee, 2007).

Group polarization on social media seems about twice as severe as in real life (Iandoli, Primario, & Zollo, 2021). Spreading rumors of major emergencies continuously spreads negative emotions. Under the "contagion" of group negative emotions such as dissatisfaction and tension, the event stakeholders are easily affected and may become more anxious and confused psychologically. As one of the major social media in China, Weibo has increasingly become a source of network events. Weibo users show copycat behavior, using other people's choices as standard of judgment. Emotions can be expressed anonymously, causing aggressive responses.

The Internet clearly can facilitate the flow of misinformation (e.g., Bordia & Rosnow, 1998; Kata, 2010; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). The massive spread of false information on social media affects public opinion and may threaten social development (Guo, Ding, Yao, Liang, & Yu, 2020). In a meta-analysis of 24 studies, Walter et al. (2021) point out that social media promote the spread of misinformation and pose a threat to public health, as in the case of COVID-19.

Social media have a profound impact on the development and maintenance of interpersonal relationships. Croft and Zimmer-Gembeck (2014) stated that friendship instability is positively associated with anxiety. Fear of missing out (Casale & Fioravanti, 2020) is a phenomenon that emerges particularly among young people and is a fear of social rejection with a sense of self-insecurity (Stead & Bibby, 2017). Fioravanti et al. (2021) present a meta-analysis of 33 studies, indicating that fear of missing out increases with more intensive use of social networking sites, giving rise to anxiety and depressive symptoms.

Yet, social media provide new avenues for interpersonal communication and may enable people to connect to whom they would otherwise not meet in person. Such 'weak-tie relationships' show less interaction, are emotionally less intense and intimate, and feel less reciprocal (Rademacher & Wang, 2014). The meta-analysis of 101 studies conducted by Balkundi, Bentley, and Kilduff (2012) point out that social networking sites mostly provide a platform to strengthen existing relationships and do not turn weak relationships established online into strong ones. Some believe that social media harm well-being because valuable time is consumed that could be spent with existing close relationships. For instance, Amedie (2015) states that social media foster a false sense of online "connections" and superficial friendships, depriving families of quality time spent together.

The emotional responses one receives on social media raise a feeling of uncertainty. Emotional messages have an impact on individuals' judgment and communication (Brown & Cai, 2010), which may affect people's mental health (Lerner, Valdesolo, & Kasam, 2015). Human emotions may be temporary but in chatting on social media, there is no possibility to make eye contact, which may increase the inaccuracy of expression (Lapidot-Lefler & Barak, 2012). Many misunderstandings may arise that can affect the stability of relationships.

The online social environment resembles and evokes social processes in the real world (including our self-concept) (Firth et al., 2019). However, unreal presentations on social media can lead to negative emotions. Self-presentation on social media aims to project and deal with the self to the public, and people try to control their impressions to convey the desired public self-image (Schlenker, 2012). People with a less stable self-

concept said they experimented with online self-presentation more frequently, presenting an idealized version of themselves and preferring to exhibit themselves online (Fullwood & Chen-Wilson, 2016). A systematic review reported a correlation between negative online interaction and both depression and anxiety (Seabrook, Kern, & Rickard, 2016).

Under peer pressure, one feels forced, urged, or dared to do certain things, because peers have pressured, urged, or dared someone to do so (Brown, Clasen, & Eicher, 1986). Peer pressure is increasing nowadays, due to regular communication with one another via social media (Marino, Gini, Angelini, Vieno, & Spada, 2020). People are particularly vulnerable to peer pressure because they desire to associate with and compare themselves to other members of their peer group (ibid.). The complexity of this comparison and competitiveness is exacerbated by online culture (Bloemen & De Coninck, 2020). Myers and Crowther (2009) report general impacts of social comparison in the context of body dissatisfaction (cf. the beautification of photos): The self-presentation of others on social media will increase the relative dissatisfaction of the individual (Fan, Deng, Dong, Lin, & Wang, 2019).

Hong and Chiu (2016) find that a growing number of students to be “overusing” and “addicted to” the Internet, combined with other “inappropriate” and “dangerous” activities such as online gambling, pornography abuse, and cyberbullying (also see Lusk, 2010; Bhat, Ragan, Selvaraj, & Shultz, 2017). Social media anonymity has a negative impact on users’ internal censorship, resulting in reduced moral sensitivity (Schlosser, 2020). Users’ perceived anonymity may lead to online disinhibition and deindividuation (Suler, 2004), resulting in uncontrolled behavior and ‘flame practices’ in social media interactions (Kiesler & Sproull, 1992). According to several academics, these behavioral characteristics play an important role in cyber violence (e.g., Mishna, Saini, & Solomon, 2012). Kowalski, Giumetti, Schroeder, and Lattanner (2014) found that when users engage in cyber violence, they find it difficult to detect the emotional reactivity of other users directly and simultaneously. These circumstances reduce empathy, which can only be felt when people receive direct feedback that their actions have harmed others (Pettalia, Levin, & Dickinson, 2013). Those who commit or experience cyber-violence become depressed, anxious, and stressed (Šincek, 2021).

It seems then that social media are not the best of ways to seek compensation for lack of personal contact. However, writing a diary journal on one’s own also does not seem the solution. How bad is it then that young people want to meet online? After all, it almost goes without saying that people need other people to talk to – even when there are such drawbacks as social comparison and cyber violence.

On the other hand, anthropomorphic design of virtual therapists and social robots obtained remarkable results. Already in 2012, The American Psychological Association reported that virtual therapists, avatars for business consulting, and synthetic personal trainers were found to be as impactful as their human equivalents (DeAngelis, 2012). Moreover, therapeutic avatars were reported to have extra benefits to young people because of the secure environment (no peer pressure), increasing therapy adherence and more frequent participation in therapeutic activities (DeAngelis, 2012). More recently, Stroessner and Benitez (2019) indicated that perceived human-likeness of social robots positively affected users’ preferences, technology acceptance, involvement, and willingness to use the robot. However, this is not an unqualified finding as certain designers wish to avoid the eeriness of a too lifelike robot and certain users prefer a more mechanic embodiment (Blut, Wang, Wunderlich, & Brock, 2021). Additionally, design of conversational agents best includes non-verbal as well as verbal cues to anthropomorphism (Seeger, Pfeiffer, & Heinzl, 2021).

Our research question, then, is whether social media are more beneficial for “venting” negative mood than robots and traditional diary writing. From a theoretical perspective that more human-likeness leads to better therapeutic results (i.e. people need people), one would expect (H1) social media (i.e. sharing feelings with real people) to be superior to robots (which are but virtual humans), while robots would outperform journal writing (a non-human medium).

However, evidence accumulates that on the contrary, social media themselves give rise to anxiety (e.g., Fan et al., 2019) and that in fact robots are trustworthy partners to confide in (e.g., Pu, Moyle, Jones, & Todorovic, 2019). Alternatively then, from a theoretical perspective of functionality or 'affective affordances,' people need trust and a secure environment rather than other people. Therefore, we hypothesize (H2) that social robots outdo journal writing, which outdoes social media, because the latter cannot be relied on in returning supportive feedback upon disclosing negative mood.

2. Methods

2.1. Participants and Design

Voluntary participants ($N = 27$; $M_{age} = 22.2$, $SD_{age} = 2.0$, 59.3% female, Chinese nationality) were invited to an experiment of self-disclosure on social media after negative mood induction, not receiving any credits or monetary rewards. Twenty-one participants were master students and 6 were undergraduate students. Informed consent was obtained formally from all participants. In addition, we used the data sampled in Duan et al. (2021) ($N = 45$; $M_{age} = 24.9$, $SD_{age} = 3.29$, 55.6% female, Chinese nationality) to do a comparison with a social robot ($n = 24$; 54.2% female) and a writing condition ($n = 21$; 57.1% female). For compatibility of conditions, we meticulously followed the design, procedure, and measurements in Duan et al. (2021) (methods and data available from <https://www.mdpi.com/2218-6581/10/3/98/s1>).

2.2. Procedure

Participants were taken to a single room and sat in front of a tablet computer with a sheet of paper, explaining the steps of the experimental procedure (Figure 1). The first part of the experiment consisted of negative mood induction and the second part was for self-disclosure to a social-media group, after which participants filled out an online questionnaire, using the "Questionnaire star" environment (<https://www.wjx.cn/mobile/index.aspx>) for administration of surveys and experiments.



Figure 1. Disturbing clips shown on a tablet and self-disclosure thereafter.

During the induction phase, participants were confronted with a 4m and 57s long video compilation of three documentaries about a severe earthquake event in Sichuan, China (2008), providing relevant cultural content and authentic experiences. Earlier studies have indicated that viewing negative media, including videos, images and text, indeed evoke negative mood (e.g., Bolls, Lang, & Potter, 2001; Lang, Shin, & Lee, 2005); video having the strongest impact (Siedlecka & Danson, 2019).

After viewing the shocking footage, participants were invited to join a WeChat group (Figure 2) and share their feelings for 10 minutes. The WeChat group was not visible before self-disclosure. During this phase, the experimenters acted as six people in the WeChat group, responding to the participant. Responses by the experimenters closely

resembled the 'typical responses' on social media (see **Apparatus and Materials**), maintaining the empirically established ratio of three positive responses versus two negative responses to one message inputted by the participant.

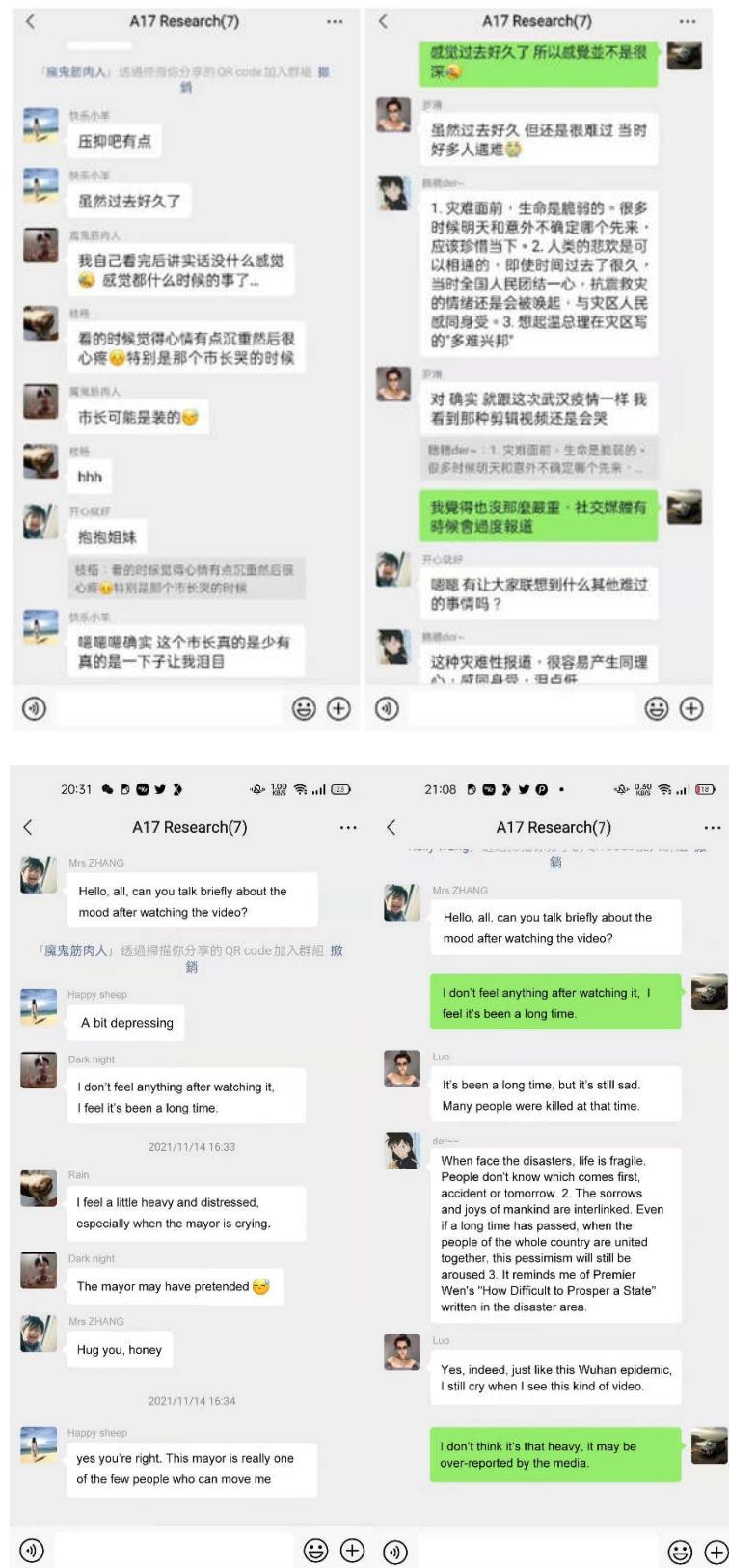


Figure 2. Snippet of a WeChat session (original Chinese and English translation).

After the self-disclosure session on WeChat, participants were asked to fill out a 30-item structured questionnaire (Duan et al., 2021) (Appendix 1) and assess their experiences with the video footage and conversations on social media thereafter. The items on the questionnaire were presented as blocks, and the pseudo-random sequence of items within the blocks was different for each participant. The final part of the questionnaire collected demographic backgrounds. Upon completion of the questionnaire, participants were thanked for their participation and debriefed.

2.3. Apparatus and Materials

2.3.1. Video materials

The negative emotion-induced video was 4 minutes and 57 seconds long and consisted of the following 3 video clips from the Sichuan earthquake online documentary:

- Internet video in memory of the Wenchuan Sichuan earthquake 10th anniversary (cut from 00:02-01:19). Available from <https://www.bilibili.com/video/av23087386/> (Accessed on 13 June 2019)
- Dazzz2009 (31 December 2008). Internet video record of 512 earthquake in Dujiangyan (cut from 01:20-01:59). Available from https://www.youtube.com/watch?v=Vz0nGbl81fM&list=PLf2PpWDjsx1d6rVUW0vaGFzhvIr_nRo_8&index=2 (Accessed on 13 June 2019)
- Lantian777 (16 May 2008). Internet video 10 min after Wenchuan Sichuan earthquake (in full). Available from <https://www.youtube.com/watch?v=PI5KL7nvU28> (Accessed on 14 June 2019)

2.3.2. Chat-group responses

To study the proportion of positive and negative replies on social media, we collected users' thoughts on breaking up a relationship from "Douban-ChoZan," which is a mainstream social media site in China, established in 2005 (<https://www.douban.com/group/topic/83226164/#75043807EPpUK0>). We used a web crawler for data-mining 10,115 cases with text length of about 200,000 words.

(1) Data crawling: On Douban, we sampled the texts from group discussions since 2016 around the topic "Let me talk to you about the philosophy behind breaking up and disconnecting." Information extraction concerned author, time, and contents. We used request library and tools in the Python programming language to set up a circular structure and record information, which was written to MS Excel documents.

(2) Data cleaning (word segmentation/de-terminating/tense restoration): We wrote the xls document to the Python IDLE editor and used NLTK/Beautiful soup/NumPy libraries to process the text: 1. Use the word segmentation tool to remove punctuation, paragraphs, etc. 2. Remove function words, such as 'and,' 'or,' 'the.' 3. Restore verb tenses and convert parts of speech.

(3) Sentiment analysis: We used "The Taiwan University Chinese Semantic Dictionary" (NTUSD) to score the text data after word segmentation, and calculated the total score. Total score = (word score × positive emotion score) – (word score × negative emotion score). Then the positive, neutral, and negative sentences in 10,115 text sentences were counted.

(4) Statistical results (Figure 3): There were 3,633 (36.00%) positive statements, 3,562 (34.96%) neutral statements, and 2,895 (28.69%) negative statements in total.

(5) Typical feedback: From the responses under (4), we compiled a list of hot topics (e.g., wronged, cheated, dissatisfied) (Figure 4) and combined them into 'typical social-media replies' to send to our participants. For example, "People bring this on themselves" or "You have to pull yourself together and keep strong" (Appendix 2).

Serial num	Release in Reply content	Positive vocabulary	Negative words	Degree word	Number of positive sentences	Positive score	Number of negative sentences	Negative score	Total Score
71793	Break up 1 week had redeemed 3 times to say he had not thought good	On time			1	1	0	0	1
71792	The mood is much calmer today	peaceful, normal	concern, worry	Very	10	14	19	-26	-12
71791	He deleted me and I scolded him and told him never to contact me again			Very	0	0	0	0	0
71790	Sorry, I can't help it				0	0	0	0	0
71789	Bingbing, when you broke up with your boyfriend, did you suspect he had		Doubt		0	0	1	-1	-1
71788	If he calls or asks me if I have a boyfriend				0	0	0	0	0
71787	123				0	0	0	0	0
71786	Year 2				0	0	0	0	0
71785	What's the matter it has been two or three months, I can often see him at				0	0	0	0	0
71784	After breaking up, I always check his Weibo, what should I do? Although I		inferior Embarra	Very	0	0	1	-2	-2
71783	No, his parents have seen me, but we were in KFC that day				0	0	0	0	0
71782	He and I entered into a relationship very quickly. We have very suitable pr	Suitable	lose not good	Very	11	22	7	-9	13
71781	She is really not coming? I'll be here as soon as she left, lol,	HAHA HAPPY		Very	0	1	1	-12	-12
71780	I see				0	0	0	0	0
71779	I don't feel it, she doesn't like me anymore. I didn't care how she felt befo		do not like		0	0	1	-2	-2
71778	At least one semester we have to go to class every day. But why did you				0	0	0	0	0
71777	Seeing each other every day, it's not easy to forget		not easy		0	0	1	-1	-1
71776	OK				0	0	0	0	0
71775	Yeah, I didn't take care of her feelings for a long time, so selfish	sense of security	Selfish	Very	0	0	0	0	0
71774	Learned a lot from this post, thank you			Very	0	0	0	0	0
71773	Why is he moaning without illness every day? Why he give me like of my r	promote	Moaning without		1	1	1	-2	-1
71772	A friend introduced me, he was very enthusiastic to me at first	excellent	self-abasement	Very	1	2	3	-4	-2
71771	I felt so guilty yesterday, so I called him and said sorry	Optimistic	unfair harm	Very	1	10	0	0	10
71770	Many people are anxious for me, but I don't have any tricks to deal with I like many	Deal with			1	1	0	0	1
71769	In fact, his friend before tentatively asked me if I want to get back togeth		unfair	Very	0	0	1	-6	-6
71768	I sent him a message today and asked him to delete all my contact info: Hope				1	1	0	0	1
71767	I think every time he chats with me, he is just like a friend without romanc				0	0	0	0	0

Figure 3. Comments and complaints.

Word	Delete	Frequency	Word	Delete	Frequency	Word	Delete	Frequency	Word	Delete	Frequency
Active	<input checked="" type="checkbox"/>	1221	Emotion	<input type="checkbox"/>	806	Tangle	<input type="checkbox"/>	322	Love	<input type="checkbox"/>	216
Mentality	<input checked="" type="checkbox"/>	663	Delight	<input type="checkbox"/>	473	Quarrel	<input type="checkbox"/>	183	Regret	<input type="checkbox"/>	188
Hardworking	<input checked="" type="checkbox"/>	447	Charm	<input type="checkbox"/>	288	Abandon	<input type="checkbox"/>	141	Leave	<input type="checkbox"/>	137
Brilliant	<input checked="" type="checkbox"/>	242	Sad	<input type="checkbox"/>	204	Painful	<input type="checkbox"/>	126	Unbearable	<input type="checkbox"/>	103
Earnest	<input checked="" type="checkbox"/>	194	Sincere	<input type="checkbox"/>	158	Sad	<input type="checkbox"/>	72	Doubt	<input type="checkbox"/>	60
Family	<input checked="" type="checkbox"/>	158	True love	<input type="checkbox"/>	155	Self-abasement	<input type="checkbox"/>	62	Indifferent	<input type="checkbox"/>	50
Enthusiastic	<input type="checkbox"/>	151	Happy	<input type="checkbox"/>	146	Wronged	<input type="checkbox"/>	49	Boring	<input type="checkbox"/>	42
Angry	<input type="checkbox"/>	141	Fine	<input type="checkbox"/>	127	Violent	<input type="checkbox"/>	41	Upsat	<input type="checkbox"/>	33
Confident	<input type="checkbox"/>	105	Luck	<input type="checkbox"/>	86	Dissatisfied	<input type="checkbox"/>	32	Complain	<input type="checkbox"/>	26
Intelligent	<input type="checkbox"/>	86	Gentle	<input type="checkbox"/>	79	Painful	<input type="checkbox"/>	20	Avoid	<input type="checkbox"/>	19
Security	<input type="checkbox"/>	71				quibble	<input type="checkbox"/>	16	Ridiculous	<input type="checkbox"/>	13
						Smart	<input type="checkbox"/>	12	Pretended	<input type="checkbox"/>	11
						Cheat	<input type="checkbox"/>	9	Exhausted	<input type="checkbox"/>	6
						Speaking rudely	<input type="checkbox"/>	5	Cold	<input type="checkbox"/>	5
						Ruthless	<input type="checkbox"/>	4	Affection	<input type="checkbox"/>	4

Figure 4. Hot topics (e.g., wronged, cheated, dissatisfied).

According to the statistics, the proportion of positive statements was slightly higher than of negative statements. Therefore, when participants self-disclosed, the experimenters replied with three positive responses and two negative responses, in accordance with the contents of the Douban crawler-results. To improve ecological validity, we personalized the typical social-media responses for each participant.

2.3.3. Measures

For measurement, we employed the structured questionnaire developed by Duan et al. (2021), containing four measurement scales: Valence after mood induction (i.e. the earthquake movie) but before treatment (i.e. disclosure to a chat group) and Valence after treatment. Relevance and Novelty served as control variables. The questionnaire ended in inquiring about Demographics.

The statements were of Likert type combined with a 6-point rating scale (1 = strongly disagree, 6 = strongly agree). Each measurement scale had four indicative items and four counter-indications. The four indicative items of 'positive Valence before treatment' formed a unipolar conception of Valence, one of the items being "I feel good." The four counter-indications formed a unipolar conception of 'negative Valence before treatment,' for instance, "I feel bad." We also used these items for measurement of Valence after self-disclosure to the social media group, adapting the wording to the situation. Thus, 'Valence after treatment' consisted of four indicative (unipolar positive) and four counter-indicative (unipolar negative) statements as well. The two unipolar scales of Valence combined (with negative Valence recoded) formed the bipolar conception of Valence.

Relevance was measured with two indicative and two counter-indicative items, querying the impact on personal goals and concerns (i.e. one's emotion regulation), in our

case, the impact of the typical social-media responses to self-disclosing negative mood. Examples are 'social media are worthwhile' and 'social media are useless.'

The Novelty scale was used as control to see how used participants were to regulating their emotions through social media groups. Novelty was composed of three indicative items (e.g., 'social media are new') and three counter-indicative statements (e.g., 'social media are commonplace'). Raw data can be found in the Technical Report (Zhang, Luo, Chen, Huang, & Hoorn, 2021).

We reverse-coded the counter-indicative items on the two Valence scales, Relevance, and Novelty. Because we wanted to compare self-disclosure between social media, robots, and writing, we assessed reliability of the questionnaire items across these three conditions, thus including the data set obtained by Duan et al. (2021), available from <https://www.mdpi.com/2218-6581/10/3/98/s1>.

Calculated across all three conditions ($N = 72$), measurement scales (all items except Novelty) achieved good to very good reliability in the first run ($.92 < \text{Cronbach's } \alpha > .79$). This was true for the separate subscales of Valence (4 items each) and for their combination (Valence-before and Valence-after, 8 items each), as well as for Relevance (4 items). The control variable of Novelty scored Cronbach's $\alpha = .682$ in the first run (all items). Although less than the conventional cut-off of 0.7, we found that the reliability of Novelty could not be improved by eliminating items. Yet, Novelty was a mere control and not of theoretical interest.

We then performed a Principal Components Analysis, using varimax rotation. The component matrix showed that items on the Valence scale and the Relevance scale were arranged nicely, as expected. Novelty showed a certain spread in Relevance but as this was a mere control variable, we left Novelty unchanged. In later analysis, we will check the degree of correlation with theoretical factors. Details of the reliability analysis are tabulated in the Technical Report (Zhang, Luo, Chen, Huang, & Hoorn, 2021).

The outliers of Mean Valence, Mean Relevance, and Mean Novelty were participant 9 in Valence (bipolar). Participants 55 and 40 were outliers in Valence-after (bipolar). Participants 5 and 21 were outliers for positive Valence-after. Participants 28, 34, 39, 40, 55, 56 and 64 were outliers for negative Valence-after. Participants 64 and 72 were outliers for Novelty. See the Technical Report (Zhang, Luo, Chen, Huang, & Hoorn, 2021).

3. Results

3.1. Manipulation check

We explored whether the shocking video of the earthquake had stirred any emotions in the participants and whether the treatment (Robots, Writing, and Social Media) evoked any change in mood. To check whether emotions (negative or positive) were evoked after mood induction and after treatment, we performed a one-sample t -test with 1 as the test value for $N = 72$ and $n = 61$ (outliers removed) (Table 1).

Table 1. One-sample *t*-tests (1 is the test value), checking whether emotions occurred after mood induction and after treatment.

<i>Variables</i>		<i>Mood Induction</i>	
<i>Means</i>	<i>t</i>	<i>p</i>	<i>N</i>
Positive Valence-before	11.84	< .001	72
Negative Valence-before	23.60	< .001	72
Positive Valence-before	10.99	< .001	61
Negative Valence-before	24.27	< .001	61
<i>Variables</i>		<i>Treatment</i>	
<i>Means</i>	<i>t</i>	<i>p</i>	<i>N</i>
Positive Valence-after	23.42	< .001	72
Negative Valence-after	14.91	< .001	72
Positive Valence-after	25.28	< .001	61
Negative Valence-after	17.22	< .001	61

From Table 1, we can conclude that after the earthquake clips (Table 1, *Mood Induction*), more negative than positive mood was induced, as intended, both with $N = 72$ and $n = 61$. For both $N = 72$ and $n = 61$, after Treatment (Table 1, *Treatment*), whether talking to a robot or writing in a journal or chatting with a social group, more positive emotions than negative ones were felt, as expected.

To monitor effects of before and after treatment, we also performed paired-samples *t*-tests in both the $N = 72$ and $n = 61$ data sets (Table 2). Take notice that these tests are manipulation checks; they are not for actual hypothesis testing.

Table 2. Paired-samples *t*-tests for treatment effects on Valence.

<i>Variables</i>		<i>Before-After Treatment</i>	
<i>Means</i>	<i>t</i>	<i>p</i>	<i>N</i>
Negative Valence before-after	10.88	< .001	72
Positive Valence before-after	-9.10	< .001	72
Negative Valence before-after	10.89	< .001	61
Positive Valence before-after	-10.20	< .001	61

From Table 2, we can conclude that participants became less negative after the treatment (i.e., mean negative Valence-before was significantly greater than mean negative Valence-after); furthermore, they became more positive after treatment (i.e., mean positive Valence-before was significantly smaller than mean positive Valence-after). The manipulations were successful: Treatment (whether Robot, Writing, or Social Media) elicited effects into the intended direction.

3.2. Effects of Media on Valence

3.2.1. GLM Repeated measures for bipolar Valence before-after

We conducted GLM Repeated measures for bipolar Valence before-after with ($N = 72$) and without outliers ($n = 61$). Table 3 shows the results.

Table 3. GLM Repeated measures for bipolar Valence before-after.

<i>Robots vs. Writing vs. Social Media</i>						
	<i>V</i>	<i>F</i>	<i>df</i> _{1,2}	<i>p</i>	η^2	<i>N</i>
Interaction Media *	.08	3.01	2,69	.056	.08	72
Valence before-after	.14	4.83	2,58	.011	.14	61
Main effect Media		2.02	2,69	.141	.06	72
(RWS)		1.96	2,58	.150	.06	61
Main effect Valence	.64	124.90	1,69	.000	.64	72
before-after	.73	152.76	1,58	.000	.73	61

Note: Identical results were obtained for unipolar Valence (positive - negative)

The interaction between Media and bipolar Valence (before-after) without outliers was significant and going into the expected direction (more positive after treatment). This interaction effect was supported by a main effect of bipolar Valence but not by the main effect of Media. GLM Repeated measures for unipolar Valence (positive - negative) before and after confirmed these results (Table 3). Paired-samples *t*-tests supported that for each medium, the mood became more positive, the biggest difference being made by Robots and least so by Social Media (Table 4).

Table 4. Paired-samples *t*-tests for bipolar Valence before-after ($n = 61$).

<i>Robots vs. Writing vs. Social Media</i>						
	<i>Difference between means</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>CI</i>	<i>n</i>
Robot	2.00	-7.87	20	.000	-2.39 -1.03	21
Writing	1.26	-6.58	16	.000	-2.31 -.860	17
Social Media	1.12	-7.41	22	.000	-2.15 -.930	23

With $N = 72$ and mean Relevance and mean Novelty as covariates, all interaction and main effects of bipolar Valence and Media vanished but the main effects of Relevance ($F = 1.22$, $p = .244$) and Novelty ($F < 1$) were not significant either. Covariates are dimensions of the participants independent of treatment. Covariates may significantly affect aspects of the analytical model without being significant themselves. However, effect sizes were very low (Relevance $\eta^2 = .033$; Novelty $\eta^2 = .003$) and Relevance and Novelty were meant as controls rather than theoretical variables.

GLM Repeated measures for unipolar Valence before-after ($N = 72$) with mean Relevance and mean Novelty as covariates showed significant interactions between negative Valence and Relevance and negative Valence and Novelty. Tests of within-subjects contrasts showed that negative Valence after treatment was lower when the treatment was experienced as more Relevant ($F_{(1,67)} = 5.96$, $p = .017$, $\eta^2 = .08$) and as more Novel ($F_{(1,67)} = 5.16$, $p = .026$, $\eta^2 = .07$), although effect sizes were small. Relevance and Novelty were positively correlated with each other ($r = .47^{**}$).

With $n = 61$ and mean Relevance and mean Novelty as covariates, the interaction effect was still significant ($V = .11$, $F_{(2,56)} = 3.58$, $p = .034$, $\eta^2 = .113$). All other effects, including the main effects of Relevance ($F = 2.22$, $p = .142$) and Novelty ($F < 1$) were not significant. GLM Repeated Measures with unipolar Valence (positive - negative) did not change these results.

All it all, it seems that outliers are sensitive to the personal relevance and novelty of the media used, reducing their negative mood. Those are characteristics of this particular subset of participants rather than the Media they interacted with or of the larger participant group.

3.2.2. GLM Univariate (Oneway-ANOVA) for Δ Valence (bipolar)

To try another perspective, mean difference scores were calculated from the mean values of bipolar Valence before-after and we ran a GLM Univariate analysis (Oneway-ANOVA) for Medium on bipolar Δ Valence with $N = 72$. The effects were not significant ($F_{(2,69)} = 3.01, p = .056, \eta_p^2 = .080$). With $n = 61$, the main effect of Media was significant ($F_{(2,58)} = 4.83, p = .011, \eta_p^2 = .143$). Independent samples t -tests revealed that Robots ($M_{\Delta\text{Val}} = 1.99, SD = 1.16$) made a larger positive difference than Writing ($M_{\Delta\text{Val}} = 1.26, SD = .79$) ($t_{(36)} = 2.23, p = .016$ (1-tailed), $CI = .067 - 1.41$) and even larger than Social Media ($M_{\Delta\text{Val}} = 1.19, SD = .77$) ($t_{(42)} = 2.73, p = .0045$ (1-tailed), $CI = .209 - 1.40$). The difference between Writing and Social Media was not significant ($t_{(38)} = .27, p = .395$ (1-tailed)).

3.2.3. GLM Multivariate (Oneway-MANOVA) for Δ Valence (unipolar)

With $N = 72$, the effect of Media on Δ Valence (positive versus negative) was not significant ($F = 1.95, p = .105$). Although this result does not warrant any further exploration, we saw that in the tests of between-subjects effects, Media impacted positive Δ Valence into the desired direction ($F_{(2,69)} = 4.56, p = .035, \eta_p^2 = .09$) but did not significantly affect negative Δ Valence ($p = .177$). Including mean Relevance and mean Novelty into the analysis rendered significant effects for Relevance as covariate ($V = .11, F_{(2,66)} = 4.04, p = .022, \eta_p^2 = .11$) not so for Novelty. Between-subjects effects showed that mean Relevance correlated positively with positive Δ Valence ($F_{(1,67)} = 5.67, p = .020, \eta_p^2 = .08$).

Without outliers, $n = 61$, multivariate tests showed significant results of Media ($V = .18, F_{(4,116)} = 2.85, p = .027, \eta_p^2 = .09$). Tests of between-subjects effects showed that Media impacted positive Δ Valence into the desired direction ($F_{(2,58)} = 5.11, p = .009, \eta_p^2 = .15$) but did not significantly affect negative Δ Valence ($F = 3.01, p = .057$). Negativity was not reduced but positivity was increased. Covariate effects of mean Relevance and Novelty were not significant and did not change the pattern of results for $n = 61$.

Independent samples t -tests showed that Robots ($M_{\Delta\text{Valp}} = 1.94, SD = 1.15$) made a larger positive difference than Writing ($M_{\Delta\text{Valp}} = 1.06, SD = 1.07$) ($t_{(35)} = 2.40, p = .011$ (1-tailed), $CI = .134 - 1.62$) and also larger than Social Media ($M_{\Delta\text{Valp}} = 1.18, SD = .85$) ($t_{(41)} = 2.46, p = .009$ (1-tailed), $CI = .135 - 1.37$). The difference between Writing and Social Media was not significant ($t_{(38)} = -.42, p = .340$ (1-tailed)).

3.2.4. Exploration: Variance of Valence (VV) as indicator of emotional instability

Out of exploratory curiosity, for $n = 61$, we assessed the variability of the scores within-subjects to the items on the positive Valence and the negative Valence scale before and after treatment. We wanted to evaluate which medium - after negative mood induction - stabilized variance of affective responses more than other. Therefore, for each participant, we determined the average of squared differences for the scale values of the indicative items (positive Valence) and counter-indicative items (negative Valence). This measure, Variance of Valence, was then submitted to GLM Repeated Measures but did not yield any significant results (Media * VV_{pos} : $V = .12, F_{(3,57)} = 4.04, p = .070, \eta_p^2 = .12$; Media * VV_{neg} : $V = .05, F_{(3,57)} = 0.95, p = .424, \eta_p^2 = .05$). For further details, consult the Technical Report (Zhang, Luo, Chen, Huang, & Hoorn, 2021).

4. Conclusion

H1 stated that negative-mood reduction and/or positive-mood increase would result from the measure of human likeness of the medium, expecting that sharing one's feelings on social media with real people would outperform talking to a social robot and more so writing on paper. H1 was refuted. Although all media increased positive mood, social

media did so the least of all with self-disclosure to a social robot outdoing both writing and social media.

H2 took an opposite stance, expecting that human feedback does not have to be the most empathetic so that human-likeness that is trustworthy (i.e. a social robot) would outdo real humans on social media or writing a diary page (which equals talking to oneself). H2 was accepted. In all analyses, the social robot did better on increases in positive mood than did journal writing or social media; in certain analyses, social media tended to be the least effective of all. People put trust in a robot more so than in humans on social media.

5. Discussion

COVID-19 has caused a pandemic of loneliness and depression apart from being a viral disease. This is particularly so for young people, who are in the process of friendship formation and of experimenting with social relationships. The social media they rely on for intimate contact are oftentimes detrimental to young people's mental health. We argued that youngsters overuse social media (Pariser, 2011; Gillani et al., 2018), receiving too much unreliable information (Kata, 2010; Lewandowsky et al., 2012), breeding an environment of instable interpersonal relationships (Casale & Fioravanti, 2020; Fioravanti et al., 2021), biasing their cognitions (Firth et al., 2019), and increasing feelings of anxiety (e.g., Bloemen & De Coninck, 2020). Fear of unstable relationships as well as negative online interaction may lead to depression and anxiety (Seabrook et al., 2016). As our data confirm, self-disclosure on social media may not exact the best of therapeutic effects.

Social robots are a design alternative for mental-health interventions. Robots do not invite social comparison (cf. Festinger, 1954) so that youth does not have to pay attention to the evaluation of others and to the carefully Photoshopped ideal lives presented to them. Social robots do not exert peer pressure (cf. Marino et al., 2020) and do not induce anxiety. If designed with integrity, robots do not give judgmental or opinionated feedback, do not use profane language, and therefore, do not provoke (online) violence (cf. Šincek, 2021), inducing negative emotions.

In earlier research, we saw that humans empathize with and project feelings onto robots that are maltreated (e.g., Suzuki, Galli, Ikeda, Itakura, & Kitazaki, 2015). In future work, we wish to study why people sometimes are more empathic with a virtual character (e.g., a robot disclosing it is suffering) than with a real person that says to be suffering (see the **Introduction** to this study)? Perhaps that with virtual creatures, one has access to all information that is available (WYSIWYG). With humans, one can never be sure if more is the matter than what meets the eye. The suffering of a virtual being has no real-life consequences: The bystander does not need to help and one may reason 'If I don't want to help I don't need to feel guilty about it.' Like animals, a virtual creature is like another 'species,' ignorant, innocent, therefore, other behavioral scripts are in place: The virtual character has no hidden agendas like humans do, it doesn't want to get something out of its misery like real humans so to the observers, it may feel like: 'My kindness will not be abused. Therefore, I can freely empathize and feel good about it without suffering the consequences of being kind.'

With respect to design, admittedly, social media, writing, and robots each led to a better mood than before. However, social robots had the most positive impact. With these lessons in mind, we are in the process of designing a number of intervention alternatives, featuring social robots.

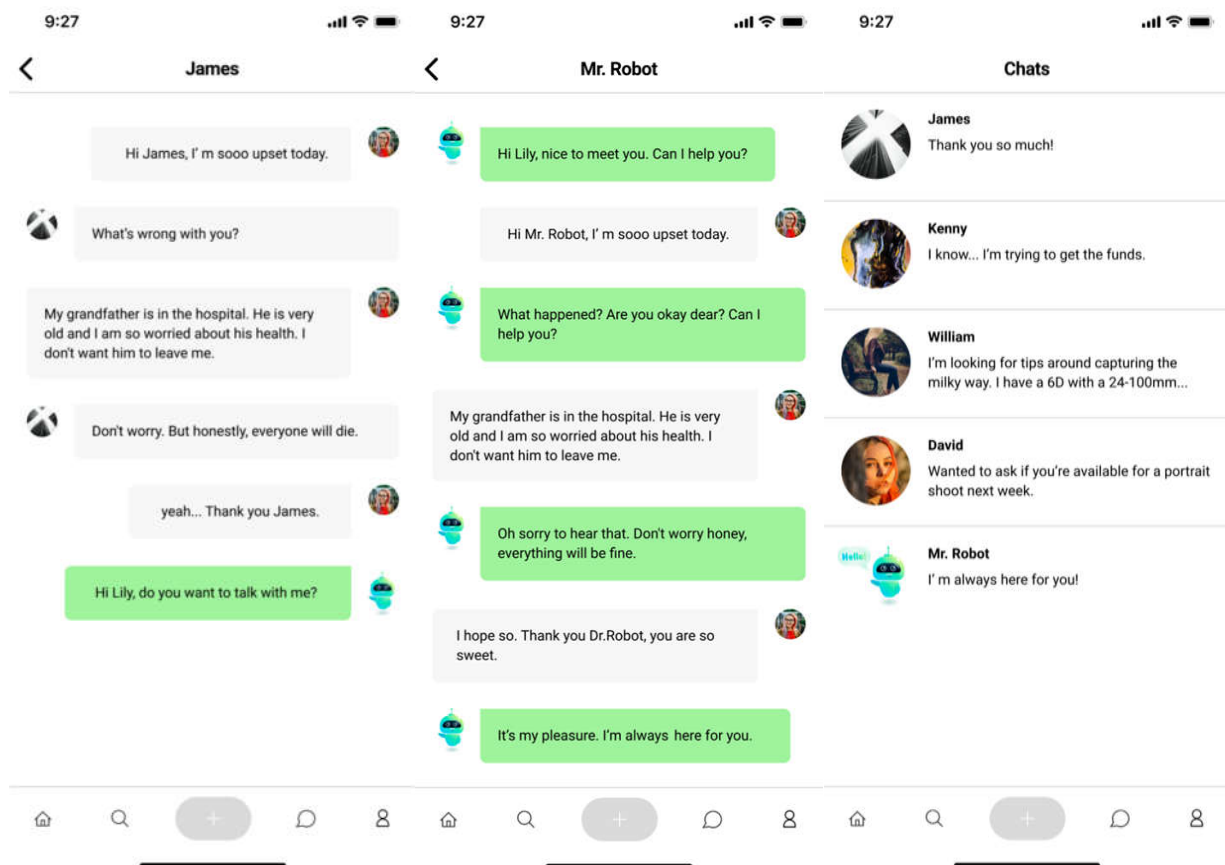


Figure 5. Prototype social-media feature to interfere with unsupportive feedback.

People who express negative emotions on social media may encounter negative or indifferent feedback. We attempt a feature that can be embedded in any social media platform (Figure 5): When the Natural Language Processing (NLP) recognizes negative vocabulary and sentiment, it will automatically pop up a social-robot dialog box, inviting users to switch the chat group and talk with the chatbot (cf. *Replika*, *My AI Friend*, Luka Inc.: <https://replika.ai/>). In addition, users may want to wear a smart watch that monitors heart-rate variability to indicate possible stress. If confirmed by the user, the watch switches on a smart speaker in the form of a social robot, asking what is wrong. If the smart speaker is built into a mirror, users may want to 'talk to themselves' while preparing for the new day or making ready for bed. This 'mirror for self-reflection' (Figure 6) may have integrated cameras to detect facial expressions that together with voice analysis and NLP may indicate if mood is improving while the user discloses daily events to their mirror image. This set-up for an empathic smart-home may be applied to care facilities and hospitals alike.

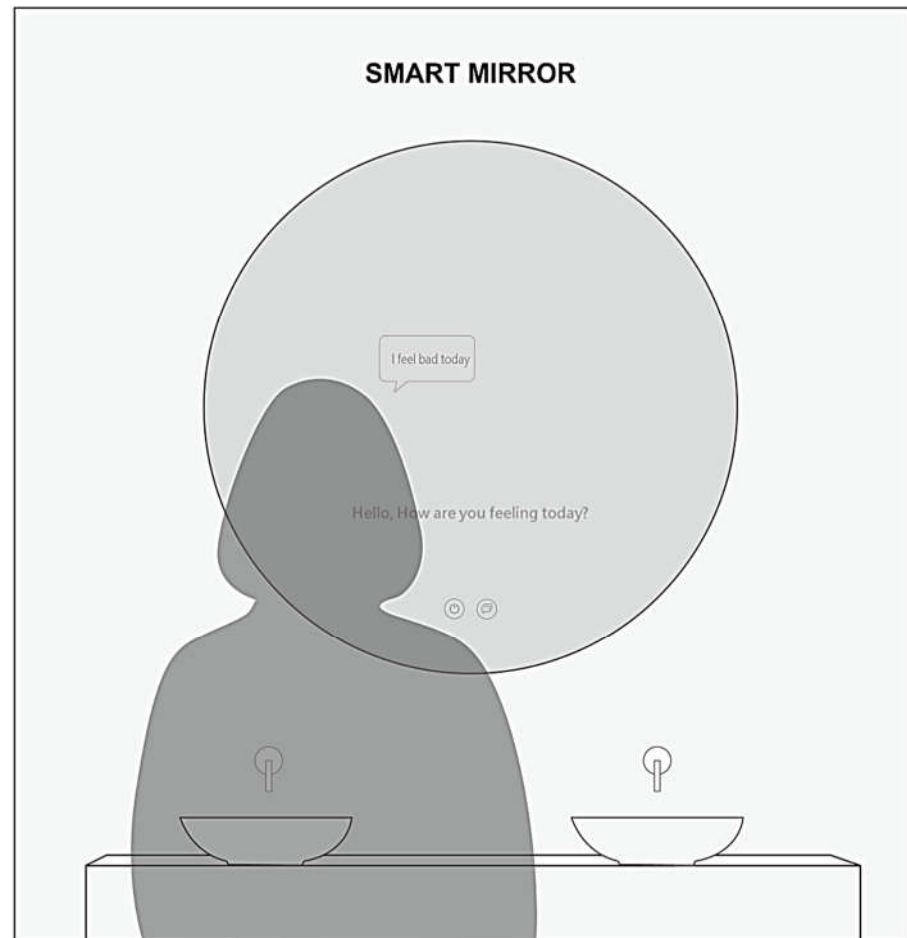


Figure 6. A mirror for self-reflection.

We started this study from the assumption that more human likeness of an agency would invite self-disclosure by people in a negative mood but there seems to be an optimum. Social robots seem to hit the sweet spot, whereas real humans have too many negative qualities to be fully trusted. Designers, then, do not need to worry about extremely high ‘human fidelity;’ a hint into the proper direction may suffice. Even unintended emotional support may instigate users to confide in a software system:

About 10 years ago I set up a fairly popular website for people to play the ancient oriental game of Go, Baduk or Weiqi 围棋 against what was at the time a fairly strong AI program (good old fashioned AI). The website had a rather simple chat feature, with two comments: ‘good call’ and ‘try harder’ based on a simple extrapolation of the game position. Searching the logs one day, to get a handle on usage, I was rather disturbed to find that one player had developed a long conversation, with lengthy self-disclosure, always promoted by one of these expressions, usually the latter. (S)he appeared to consider there was a living person responding. (Jonathan Chetwynd, personal communication, September 25, 2021)

Seen from the above quotation, we want to conclude this paper with a reflection of humanism and humanistic ideas of care: In 16th century Europe, man became the measure of all things but with the rise of social robots, functionality – not human beings – will be.

Ethical Statement Approval: Ethical approval was obtained from the Ethical Review Board of the university, file number HSEARS20200204003. Correspondence concerning this article should be addressed to Johan F. Hoorn (ORCID: 0000-0002-3427-5681)

Acknowledgments: This study is funded by the Laboratory for Artificial Intelligence in Design (Grant number RP2P3) under the InnoHK Research Clusters, Hong Kong Special Administrative Region Government. Kenji Yimin Wang is kindly acknowledged for his help with the data mining.

Jia-Yuan Chen is thanked for proposing a new direction for future research. The authors have no competing interests to declare.

References

1. Amedie, J. (2015). The impact of social media on society. *Pop Culture Intersections*, 2. Available from http://scholarcommons.scu.edu/engl_176/2
2. Balkundi, P., Bentley, J., & Kilduff, M. J. (2012). Culture, labor markets and attitudes: a meta-analytic test of tie-strength theory. *Academy of Management Proceedings*, 2012(1), 14728
3. Bhat, C. S., Ragan, M. A., Selvaraj, P. R., & Shultz, B. J. (2017). Online bullying among high-school students in India. *International Journal for the Advancement of Counselling*, 39(2), 112-124. doi: 10.1007/s10447-017-9286-y
4. Bloemen, N., & De Coninck, D. (2020). Social media and fear of missing out in adolescents: the role of family characteristics. *Social Media + Society*, 6(4), 1-11. doi: 10.1177/2056305120965517
5. Blut, M., Wang, C., Wunderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, 49, 632-658. doi: 10.1007/s11747-020-00762-y
6. 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.
7. Bordia, P., & Rosnow, R. L. (1998). Rumor rest stops on the information highway transmission patterns in a computer-mediated rumor chain. *Human Communication Research*, 25(2), 163-179. doi: 10.1111/j.1468-2958.1998.tb00441.x
8. Brown, B. B., Clasen, D. R., & Eicher, S. A. (1986). Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents. *Developmental Psychology*, 22(4), 521-530. doi: 10.1037/0012-1649.22.4.521
9. Brown, J. D., & Cai, H. (2010). Thinking and feeling in the People's Republic of China: Testing the generality of the "laws of emotion." *International Journal of Psychology*, 45(2), 111-121. doi: 10.1080/00207590903281104
10. Casale, S., & Fioravanti, G. (2020). Factor structure and psychometric properties of the Italian version of the fear of missing out scale in emerging adults and adolescents. *Addictive Behaviors*, 102, 106179-106179. doi: 10.1016/j.addbeh.2019.106179
11. Croft, C. D., & Zimmer-Gembeck, M. J. (2014). Friendship conflict, conflict responses, and instability. *The Journal of Early Adolescence*, 34(8), 1094-1119. doi: 10.1177/0272431613518972
12. DeAngelis, T. (March, 2012). A second life for practice? *Monitor on Psychology*, 43(3), unpagged. Retrieved Dec. 10, 2021 from <https://www.apa.org/monitor/2012/03/avatars>
13. Duan, E. Y., Yoon, J. M., Liang, E. Z., & Hoorn, J. F. (2021). Self-disclosure to a robot: Only for those who suffer the most. *Robotics*, 10(3), 98. doi: 10.3390/robotics10030098
14. Fan, X., Deng, N., Dong, X., Lin, Y., & Wang, J. (2019). Do others' self-presentation on social media influence individual's subjective well-being? A moderated mediation model. *Telematics and Informatics*, 41, 86-102. doi: 10.1016/j.tele.2019.04.001
15. Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117-140. doi: 10.1177/001872675400700202
16. Fioravanti, G., Casale, S., Benucci, S. B., Probstamo, A., Falone, A., Ricca, V., & Rotella, F. (2021). Fear of missing out and social networking sites use and abuse: A meta-analysis. *Computers in Human Behavior*, 122, 106839. doi: 10.1016/j.chb.2021.106839
17. Firth, J., Torous, J., Stubbs, B., Firth, J. A., Steiner, G. Z., Smith, L., Alvarez-Jimenez, M., Gleeson, J., Vancampfort, D., Armitage, C. J., & Sarris, J. (2019). The "online brain": How the Internet may be changing our cognition. *World Psychiatry*, 18(2), 119-129. doi: 10.1002/wps.20617
18. Fullwood, J., B. M., & Chen-Wilson, C.-H. (2016). Self-concept clarity and online self-presentation in adolescents. *Cyberpsychology, Behavior and Social Networking*, 19(12), 716-720. doi: 10.1089/cyber.2015.0623
19. Fung, D. (April 12, 2020). *IACAPAP Update: President's Message on COVID-19*. Geneva, Switzerland: Executive Committee of the International Association of Child and Adolescent Psychiatry and Allied Professionals (IACAPAP). Retrieved Dec. 10, 2021 from <https://iacapap.org/iacapap-update-covid-19/>
20. Gillani, N., Yuan, A., Saveski, M., Vosoughi, S., & Roy, D. (2018). Me, my echo chamber, and I: introspection on social media polarization. In P.-A. Champin, F. Gandon, et al. (Eds.), *Proceedings of the 2018 World Wide Web Conference (WWW'18)*, Lyon France, April 23-27, 2018 (pp. 823-831). doi: 10.1145/3178876.3186130
21. Gillespie, T., Boczkowski, P. J., & Foot, K. A. (2014). *Media technologies: essays on communication, materiality, and society*. Cambridge, MA: MIT.
22. Guo, B., Ding, Y., Yao, L., Liang, Y., & Yu, Z. (2020). The future of false information detection on social media. *ACM Computing Surveys*, 53(4), 1-36.
23. Hong, F. Y., & Chiu, S.-L. (2016). Factors influencing Facebook usage and Facebook addictive tendency in university students: the role of online psychological privacy and Facebook usage motivation. *Stress and Health*, 32(2), 117-127. doi: 10.1002/smi.2585
24. Iandoli, L., Primario, S., & Zollo, G. (2021). The impact of group polarization on the quality of online debate in social media: A systematic literature review. *Technological Forecasting & Social Change*, 170, 120924. doi: 10.1016/j.techfore.2021.120924
25. Isenberg, D. J. (1986). Group polarization. *Journal of Personality and Social Psychology*, 50(6), 1141-1151. doi: 10.1037/0022-3514.50.6.1141
26. Kata, A. (2009). A postmodern Pandora's Box: anti-vaccination misinformation on the Internet. *Vaccine*, 28(7), 1709-1716. doi: 10.1016/j.vaccine.2009.12.022

27. Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior and Human Decision Processes*, 52(1), 96–123.
28. Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073-1137. doi: 10.1037/a0035618
29. Lang, A., Shin, M., & Lee, S. (2005). Sensation seeking, motivation, and substance use: a dual system approach. *Media Psychology*, 7(1), 1-29. doi: 10.1207/S1532785XMEP0701_1
30. Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), 434-443. doi: 10.1016/j.chb.2011.10.014
31. Lee, E. J. (2007). Deindividuation effects on group polarization in Computer-Mediated Communication: the role of group identification, public-self-awareness, and perceived argument quality. *Journal of Communication*, 57(2), 385-403. doi: 10.1111/j.1460-2466.2007.00348.x
32. Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, 66(1), 799-823. doi: 10.1146/annurev-psych-010213-115043
33. Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106-131. doi: 10.1177/1529100612451018
34. Lusk, B. (2010). Digital natives and social media behaviors: an overview. *Prevention Researcher*, 17(5), 3-6.
35. Marino, C., Gini, G., Angelini, F., Vieno, A., & Spada, M. M. (2020). Social norms and e-motions in problematic social media use among adolescents. *Addictive Behaviors Reports*, 11, 100250. doi: 10.1016/j.abrep.2020.100250
36. Mishna, F., Saini, M., & Solomon, S. (2009). Ongoing and online: Children and youth's perceptions of cyber bullying. *Children and Youth Services Review*, 31(12), 1222-1228. doi: 10.1016/j.childyouth.2009.05.004
37. 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
38. Myers, T., & Crowther, J. (2009). Social comparison as a predictor of body dissatisfaction: a meta-analytic review. *Journal of Abnormal Psychology*, 118, 683-698. doi: 10.1037/a0016763.
39. Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. London: Penguin.
40. Pettalia, J. L., Levin, E., & Dickinson, J. (2013). Cyberbullying: Eliciting harm without consequence. *Computers in Human Behavior*, 29(6), 2758-2765. doi: 10.1016/j.chb.2013.07.020
41. Rademacher, M. A., & Wang, K. Y. (2014). Strong-tie social connections versus weak-tie social connections. In K. Harvey (Ed.), *Encyclopedia of Social Media and Politics* (pp. 1213-1216). Los Angeles, CA: Sage. Retrieved Dec. 13, 2021 from <https://us.sagepub.com/en-us/nam/encyclopedia-of-social-media-and-politics/book239101>
42. Seabrook, E. M., Kern, M. L., & Rickard, N. S. (2016). Social networking sites, depression, and anxiety: a systematic review. *JMIR Mental Health*, 3(4), e50–e50. doi: 10.2196/mental.5842
43. Schlenker, B. R. (2012). Self-presentation. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 542-570). New York: Guilford.
44. Schlosser, A. (2020). Self-disclosure versus self-presentation on social media. *Current Opinion in Psychology*, 31, 1-6.
45. Šincek, D. (2021). The revised version of the committing and experiencing cyber-violence scale and its relation to psychosocial functioning and online behavioral problems. *Societies*, 11(3), 107. doi: 10.3390/SOC11030107
46. Siedlecka, E., & Denson, T. (2019). Experimental methods for inducing basic emotions: A qualitative review. *Emotion Review*, 11(1), 87-97. doi: 10.1177/1754073917749016
47. Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22(4), unpagged. doi: 10.17705/1jais.00685
48. Stead, H., & Bibby, P. A. (2017). Personality, fear of missing out and problematic Internet use and their relationship to subjective well-being. *Computers in Human Behavior*, 76, 534-540. doi: 10.1016/j.chb.2017.08.016
49. Stroessner, S. J., & Benitez, J. (2019). The social perception of humanoid and non-humanoid robots. *International Journal of Social Robotics*, 11(2), 305-315.
50. Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321-326. doi: 10.1089/1094931041291295
51. Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., & Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*, 5, 15924.
52. United Nations Educational, Scientific and Cultural Organization (UNESCO) (2020). *COVID-19 educational disruption and response*. Paris: UNESCO. Retrieved Dec. 10, 2021 from <https://en.unesco.org/news/covid-19-educational-disruption-and-response>
53. Walter, N., Brooks, J. J., Saucier, C. J., & Suresh, S. (2021). Evaluating the impact of attempts to correct health misinformation on social media: a meta-analysis. *Health Communication*, 36(13), 1776-1784. doi: 10.1080/10410236.2020.1794553
54. Wang, Q., Yang, X., & Xi, W. (2018). Effects of group arguments on rumor belief and transmission in online communities: An information cascade and group polarization perspective. *Information & Management*, 55(4), 441-449.
55. Wang, R., Wang, X., Hao, F., Zhang, L., Liu, S., Wang, L., & Lin, Y. (2018). Social identity-aware opportunistic routing in mobile social networks. *Transactions on Emerging Telecommunications Technologies*, 29(5), e3297–n/a. doi: 10.1002/ett.3297
56. Xafis, V. (2020). 'What is inconvenient for you is life-saving for me': How health inequities are playing out during the COVID-19 pandemic. *Asian Bioethics Review*, 12(2), 223-234. doi: 10.1007/s41649-020-00119-1

-
- ⁵⁷. Zhang, T. X. Y., Luo, R. L., Chen, D. H.-c., Huang, I. S., & Hoorn, J. F. (2021). *Self-disclosure to Social Media* (Tech. Rep.). Hong Kong: The Hong Kong Polytechnic University.

Appendix 1

Structured questionnaires for self-disclosure to social media in Chinese and English.

1.1. Social Media questionnaire in Chinese

先生/女士你好：

感謝您參與我們的實驗。這裡我們希望花費你短短幾分鐘回答幾條問題。

你有權隨時終止填寫問卷而不需作任何解釋。你可電郵至

thea-xinyan.zhang@connect.polyu.hk 與我們的首席調查員 Thea 討論這個研究項目。

當你點擊以下按鈕，即表示同意你是 18 歲以上人士，並自願參與此項目。你了解你有權隨時及以任何原因終止參與這項研究。由參與者提供的數據將會作匿名處理，分析後的結果會記載在此研究的論文中。

這項研究是由香港理工大學監督。

感謝你的參與。

Social Media 團隊

- 我同意參與這項研究
- 我不同意參與這項研究

I. 在看了这段影片后，请如实告诉我们您的感受:

Vb1i 我感覺良好

完全不同意 不同意 有點不同意 有點同意 同意 完全同意
1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6

Vb2i 我覺得舒服

完全不同意 不同意 有點不同意 有點同意 同意 完全同意
1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6

Vb3i 我有產生正面積極的情緒

完全不同意 不同意 有點不同意 有點同意 同意 完全同意
1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6

Vb4i 我感到樂觀

Vb5c 我感覺不好

Vb6c 我感到不適

Vb7c 我有產生負面的情緒

Vb8c 我感到悲觀

II. 透過社交媒體聊天後，您感覺如何？

Vb1i 我感覺良好

- Vb2i 我覺得舒服
- Vb3i 我有產生正面積極的情緒
- Vb4i 我感到樂觀
- Vb5c 我感覺不好
- Vb6c 我感到不適
- Vb7c 我有產生負面的情緒
- Vb8c 我感到悲觀

III. 我認為通過社交媒體聊天對我的情緒調控

- Re1i 有用
- Re2i 有效
- Re3c 無效
- Re4c 沒用

IV. 我認為通過社交媒體聊天這種方式

- No1i 是新穎的
- No2i 是原創的
- No3i 是意想不到的
- No4c 是在我的預想之內的
- No5c 是普通的
- No6c 是老土的

V. 其它信息

- De1 性別
 - 女
 - 男
 - 其它

- De2 年齡

- De3 學歷 (最高學歷或現時正修讀)

- 小學或以下
- 中學
- 大專 / 副學士 / 文憑
- 大學本科
- 碩士
- 博士或以上

- De4 種族

- 亞洲
- 非洲
- 歐洲

北美洲
 南美洲
 澳洲/大洋洲
 南極洲

1.2. Social Media questionnaire in 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 thea-xinyan.zhang@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 media

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

I. After seeing the film samples

Vb1i I feel good

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

Vb2i I am well

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

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

II. After talking on social media

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

III. To regulate my emotions, talking on social media is...

Re1i useful

Re2i worthwhile

Re3c worthless

Re4c useless

IV. Talking on social media is...

No1i novel

No2i original

No3i unexpected

No4c predictable

No5c commonplace

No6c old-fashioned

V. Other information

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 media team

thea-xinyan.zhang@connect.polyu.hk

Appendix 2

2.1 Typical feedback on social media

Positive feedback:

1. Let me hug you. Don't be sad.
2. How do you feel now?
3. Are you ok?
4. You can talk to me if you are upset.
5. It would be very sad for me to see such content.
6. It was really sad.
7. You have to pull yourself together and keep strong.
8. Yeah, it makes me sad to see them in pain in the video.
9. Human beings are small in the face of disaster.
10. We should cherish life, life is unpredictable.
11. We never know which will come first, the accident or tomorrow.
12. We still have to believe in ourselves.
13. Don't worry so much. Everything will be fine.
14. I understand you. I have a similar experience.
15. Love you, hug you!

Negative feedback:

1. Well, it's okay, why are you so sad about it?
2. You are a crybaby.
3. That's a bit of a stretch.
4. It's been so long, why make you so sad?
5. It serves them right.
6. Social media exaggerates it.
8. People bring this on themselves.
9. Humans are inexorable.
10. It serves you right.
11. In fact, I doubt that you are really sad?
12. Think before you act.
13. It's all your fault.
14. I am so tired from your reply.
15. What you say is so boring.