Self-disclosure to Social Media
(Technical Report S1)

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

The Hong Kong Polytechnic University
School of Design

Author Note

Correspondence concerning this article should be addressed to Johan F. Hoorn (ORCID: 0000-0002-3427-5681), School of Design and Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. Contact: johan.f.hoorn@polyu.edu.hk

Acknowledgements

This study is funded by the Laboratory for Artificial Intelligence in Design (Grant number RP2-3) under the InnoHK Research Clusters, Hong Kong Special Administrative Region Government. The authors have no competing interests to declare. Kenji Yimin WANG is kindly acknowledged for his help with the data mining.

Contents

Abstract The overuse of social media may lead to decreased reliability of information acquisition, which breeds an environment of instable interpersonal relationships, biasing users' perceptions to exacerbate people's anxiety. In a follow-up on Duan et al. (2021), we report the technicalities of an experiment of self-disclosing negative emotions to a social-media group as compared to writing a diary journal or to talk to a 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.

Keywords Self-disclosure · Social robots · Diary · Social Media · Relevance · Valence

1. Introduction

Our research question 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 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. Method

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 robot (n = 24; 54.2% female) and a writing condition (n = 21; 57.1% female). For compatibility of conditions, we followed the design, procedure, and measurements in Duan et al. (2021) meticulously.

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.

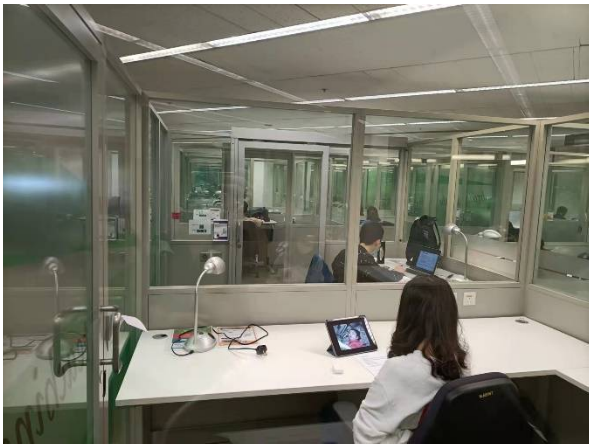


Figure 1. Disturbing clips shown on a tablet and self-disclosure thereafter.

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 (Bolls et al., 2001; Lang et al., 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.

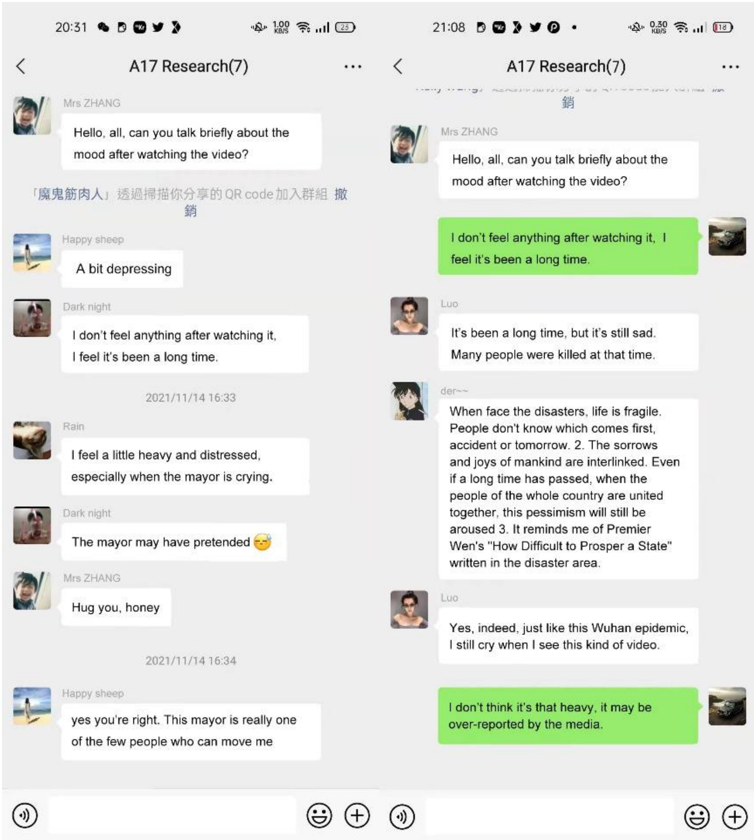


Figure 2. Snippet of a WeChat session (original Chinese and English translation).

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.

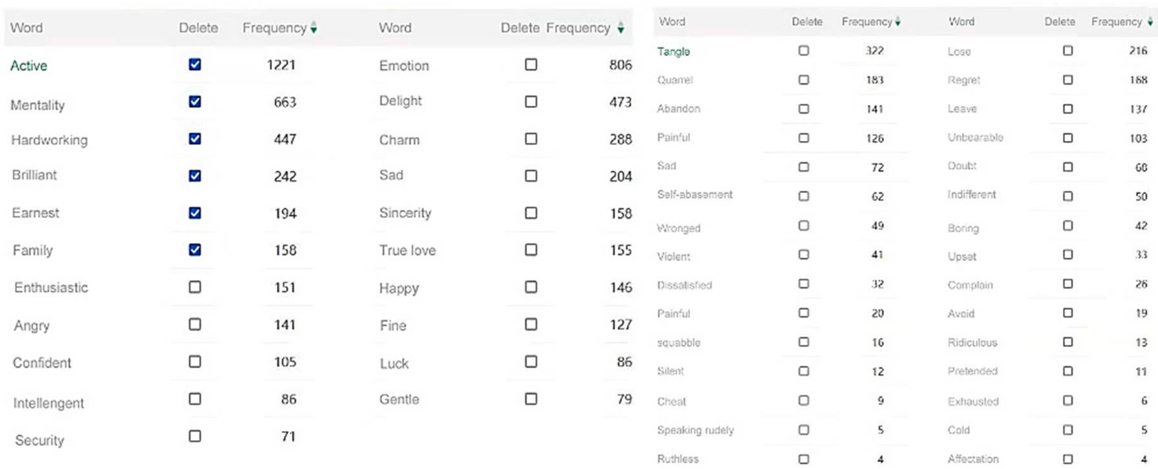
(3) Sentiment analysis: We used "The Taiwan University Chinese Semantic Dictionary" (NTUSD) to score the text data after word segmentation, and calculated the total score. Total score = (word score × positive emotion score) – (word score × negative emotion score). Then the positive, neutral, and negative 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 tin	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 agai			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 hac		Dobt		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 pe	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	0	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 l				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	1	2	0	0	2
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 withou		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 h	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 inforr	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
Active	☑	1221	Emotion	☐	806
Mentality	☑	663	Delight	☐	473
Hardworking	☑	447	Charm	☐	288
Brilliant	☑	242	Sad	☐	204
Earnest	☑	194	Sincerity	☐	158
Family	☑	158	True love	☐	155
Enthusiastic	☐	151	Happy	☐	146
Angry	☐	141	Fine	☐	127
Confident	☐	105	Luck	☐	86
Intellengent	☐	86	Gentle	☐	79
Security	☐	71			

Word	Delete	Frequency	Word	Delete	Frequency
Tangle	☐	322	Lose	☐	216
Quarrel	☐	183	Regret	☐	168
Abandon	☐	141	Leave	☐	137
Painful	☐	126	Unbearable	☐	103
Sad	☐	72	Doubt	☐	60
Self-abasement	☐	62	Indifferent	☐	50
Wronged	☐	49	Boring	☐	42
Violent	☐	41	Upset	☐	33
Dissatisfied	☐	32	Complain	☐	26
Painful	☐	20	Avoid	☐	19
squabble	☐	16	Ridiculous	☐	13
Silent	☐	12	Pretended	☐	11
Cheat	☐	9	Exhausted	☐	6
Speaking rudely	☐	5	Cold	☐	5
Ruthless	☐	4	Affectation	☐	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.

2.4 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), Valence after treatment, Relevance, and Novelty as a control variable. 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 'Valence before treatment' (*ValB*) were coded as *Vb1i*, *Vb2i*, *Vb3i,* and *Vb4i*, for instance, "I feel good" (*Vb1i*). The four counter-indications were coded *Vb5c, Vb6c, Vb7c,* and *Vb8c*, for instance, "I feel bad" (*Vb5c*). 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' (*ValA*) consisted of four indicative and four counter-indicative statements as well: *Va1i, Va2i, Va3i, Va4i, Va5c, Va6c, Va7c, Va8c*. Relevance was measured with two indicative (*Re1i, Re2i*) and two counter-indicative items (*Re3c, Re4c*), 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') (*No1i, No2i, No3i*) and three counter-indicative statements (e.g., 'social media are commonplace') (*No4c, No5c, No6c*).

For Demographics, we asked for Gender (*De1*), Age (*De2*), Educational level (*De3*), and Country (*De4*). Participants could leave their comments if so wanted.

Table 1. Social Media condition. Raw scores to the items on the measurement scales (not reverse-coded) (n = 27).

Vb1i	Vb2i	Vb3i	Vb4i	Vb5c	Vb6c	Vb7c	Vb8c	Va1i	Va2i	Va3i	Va4i	Va5c	Va6c	Va7c	Va8c	Re1i	Re2i	Re3c	Re4c	No1i	No2i	No3i	No4c	No5c	No6c
1	1	3	2	5	5	5	2	3	3	3	3	4	4	4	2	4	4	3	3	4	3	4	2	3	2
2	2	2	2	4	3	5	4	3	3	2	3	2	3	2	2	5	5	2	2	4	5	3	5	2	2
2	2	2	2	5	4	5	3	2	2	4	2	2	4	3	2	4	3	4	4	5	5	4	4	4	3

3	2	3	3	4	4	4	4	4	4	4	5	2	3	3	3	5	5	3	2	3	3	3	5	2	2
5	2	4	4	3	3	1	3	5	4	5	5	4	3	3	3	3	3	4	2	1	3	5	4	5	4
3	1	3	3	4	4	4	3	4	4	4	4	3	3	3	3	4	4	3	3	3	3	3	5	4	3
2	1	1	1	5	5	4	3	3	1	1	1	2	3	4	3	5	4	2	2	3	3	3	5	5	2
2	1	1	1	4	5	4	6	5	4	4	4	3	2	3	3	6	6	3	2	4	3	4	5	4	1
3	2	2	2	5	4	5	5	4	4	4	4	3	3	3	3	4	4	2	2	4	4	4	4	4	4
2	1	2	5	5	5	5	3	5	5	5	5	1	1	1	1	5	2	5	2	5	5	3	5	5	1
4	3	5	4	5	5	3	3	2	3	4	2	4	4	5	4	2	2	5	5	3	3	2	4	3	2
3	3	3	3	3	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	3
2	2	2	2	5	5	5	5	4	4	4	3	3	3	3	3	4	4	2	3	5	4	4	3	4	3
1	1	5	3	3	4	2	3	4	4	5	4	2	2	2	2	5	4	3	2	5	4	4	4	3	3
2	3	2	3	4	4	3	4	3	3	3	3	3	3	3	2	3	3	3	3	4	3	3	4	3	2
4	3	4	4	3	3	3	3	4	4	4	4	3	3	3	3	3	3	4	5	5	3	2	5	5	3
3	2	3	3	5	5	5	5	5	4	4	4	3	3	3	3	5	5	3	3	4	3	3	4	4	3
4	2	4	3	4	4	2	2	4	3	3	3	4	4	4	4	4	5	1	2	3	1	2	4	3	1
1	2	3	1	1	5	3	2	4	4	4	4	1	1	1	1	2	3	5	5	5	1	1	6	4	4
2	1	6	2	6	6	6	2	4	3	3	3	3	1	1	1	3	4	4	5	5	5	5	4	5	1
2	1	2	2	4	4	4	4	5	4	4	4	3	3	3	3	4	2	5	3	3	5	4	4	4	4
2	1	1	1	4	4	5	4	4	4	4	4	2	2	2	2	4	4	1	1	5	5	2	5	5	2
2	2	1	2	5	2	2	2	3	3	3	3	3	3	3	3	5	4	2	2	2	2	2	3	2	2
2	2	2	2	5	5	5	4	2	2	3	3	3	3	3	3	4	4	4	4	3	3	3	3	4	2
2	2	2	2	5	4	5	4	3	3	2	3	4	5	3	2	4	3	4	3	4	2	2	4	3	1
1	1	1	1	6	5	5	2	3	2	3	3	4	2	2	1	4	4	3	3	4	2	2	3	2	2
5	3	3	5	3	3	3	2	4	3	3	4	3	2	2	2	3	1	4	6	5	3	6	1	1	1

We reverse-coded the counter-indicative items on the two Valence scales (*Vb5c_R, Vb6c_R, Vb7c_R,* and *Vb8c_R*) and (*Va5c_R, Va6c_R, Va7c_R,* and *Va8c_R*), Relevance (*Re3c_R* and *Re4c_R*), and Novelty (*No4c_R, No5c_R,* and *No6c_R*). 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 < Cronbach's α > .79). This was true 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). The control variable of Novelty scored Cronbach's α = .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. Table 2 shows the results of the reliability analysis as well as the mean scale values and standard deviations (*SD*s).

Table 2. Reliability of the measurement scales and scale means ($N = 72$).

Scale	*Items*	*Alpha*	*Standardized Alpha*	*Scale Mean*	*SD*
MValBi	4	.799	.809	2.389	1.588
MValBc	4	.848	.848	3.979	1.669
MValB_all	8	.879	.884	2.705	1.629
MValAi	4	.885	.887	3.653	1.243
MValAc	4	.869	.870	2.542	1.071
MValA_all	8	.859	.861	4.056	1.157
MRel	4	.913	.916	3.993	1.475
MNov	6	.682	.682	3.706	1.396

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.

Table 3. Principal Components Analysis with rotated factor loadings (varimax). Values under .30 suppressed.

Component Matrix[a]

	Component		
	1	2	3
Vb1i:I_feel_good	.741		
Vb2i:I_am_well	.780		
Vb3i:I_have_positive_feelings	.608		
Vb4i:I_am_optimistic	.726		
Vb5c_R	.784		
Vb6c_R	.799		
Vb7c_R	.707		
Vb8c_R	.664		
Re1i:Talking_to_social_media_ is_useful		.874	
Re2i:Talking_to_social_media_is_worthwhile		.778	-.373
Re3c_R	.341	.757	
Re4c_R		.794	
No1i:Talking_to_social_media_is_novel		.446	
No2i:Talking_to_social_media_is_original		.451	
No3i:Talking_to_social_media_is_unexpected		.413	.614
No4c_R			.798
No5c_R		.545	.403
No6c_R		.596	

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

The outliers of Valence, Relevance, and Novelty were participant 9 in *MValB_all*. Participants 55 and 40 were outliers in *MValA_all*. Participants 5 and 21 were outliers for *MValAi*. Participants 28, 34, 39, 40, 55, 56 and 64 were outliers for *MValAc*. Participants 64 and 72 were outliers for *MNov*. See Figure 5.

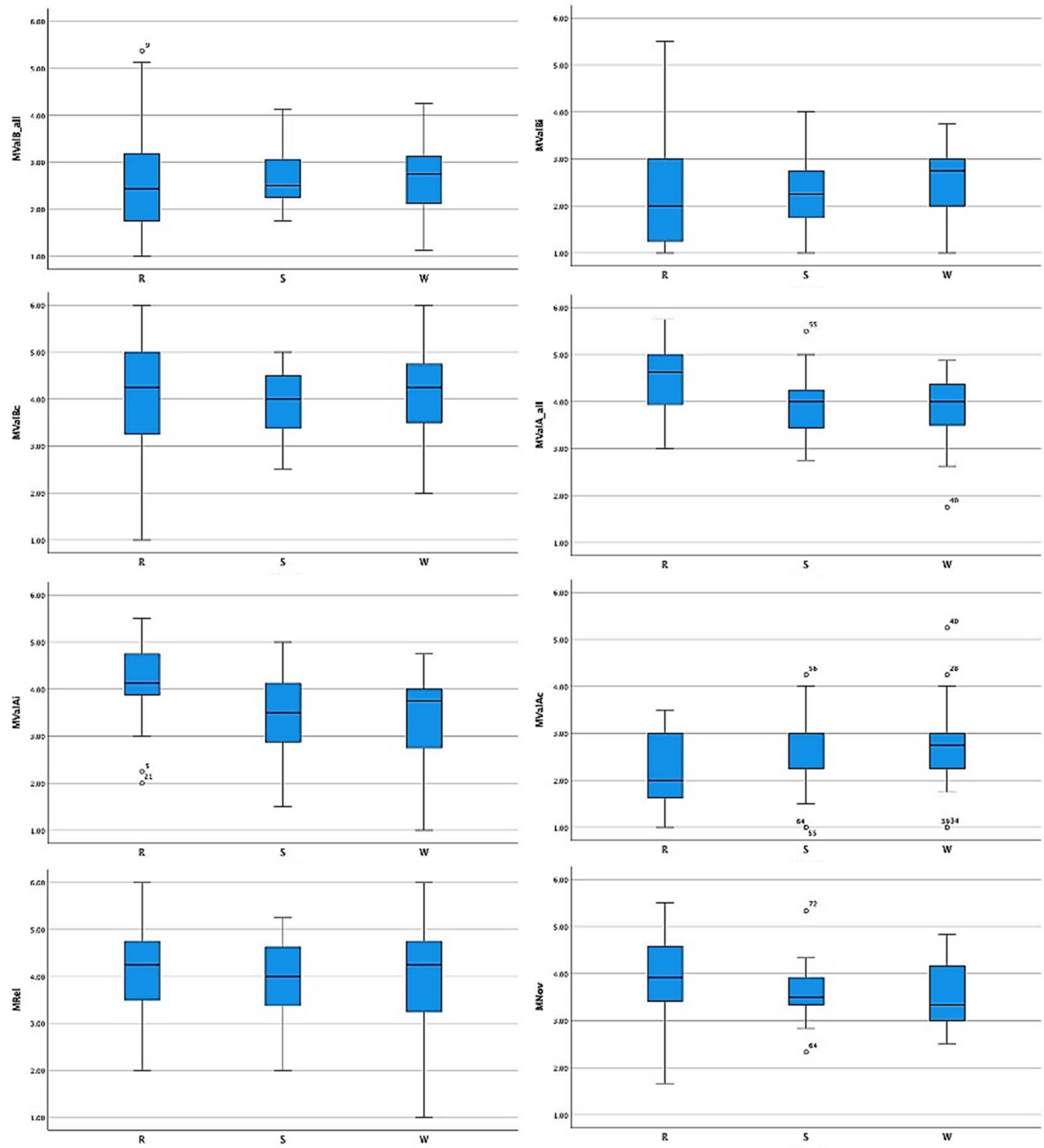


Figure 5. Outlier analysis. R = Robot, S = Social Media, W = Writing.

3. Results

3.1 Manipulation Check: Emotional Effects after Negative Mood Induction and after Treatment

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 4).

Table 4. One-sample t-tests (1 is the test value), checking whether emotions occurred after mood induction and after treatment.

Variables	Mood Induction		
	T	*p*	*n*
MValBi	11.84	< .001	72
MValBc	23.60	< .001	72
MValBi	10.99	< .001	61
MValBc	24.27	< .001	61
Variables	**Treatment**		
	T	*p*	*n*
MValAi	23.42	< .001	72
MValAc	14.91	< .001	72
MValAi	25.28	< .001	61
MValAc	17.22	< .001	61

From Table 4, we can conclude that after the earthquake clips (Table 4, 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 4, 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 5). Take notice that these tests are manipulation checks; they are not for actual hypothesis testing.

Table 5. Paired-samples t-tests for treatment effects on Valence.

Variables	Before-After Treatment		
	T	*p*	*n*
MValBc- MValAc	10.88	< .001	72
MValBi-MValAi	-9.10	< .001	72
MValBc-MValAc	10.89	< .001	61
MValBi - MValAi	-10.20	< .001	61

From Table 5, we can conclude that participants became less negative after the treatment (i.e., *MValBc* was significantly greater than *MValAc*); furthermore, they became more positive after treatment (i.e., *MValBi* was significantly smaller than *MValAi*). The manipulations were succesful: Treatment (whether robot, writing, or social media) elicited effects into the intended direction.

3.2. Effects of Media (Robot vs. Writing vs. Social Media) on Valence

3.2.1 GLM Repeated measures for bipolar Valence before-after

Table 6. GLM Repeated measures for bipolar Valence before-after.

Robots vs. Writing vs. Social Media						
	V	*F*	$df_{1,2}$	*p*	η_p^2	*N*
Interaction	.08	3.01	2,69	.056	.08	72
*Media*Valence before-after*	.14	4.83	2,58	.011	.14	61
		2.02	2,69	.141	.06	72

Main effect Media (RWS)		1.96	2,58	.150	.06	61
Main effect Valence before-after	.64	124.90	1,69	.000	.64	72
	.73	152.76	1,58	.000	.73	61

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

We conducted GLM Repeated measures for bipolar Valence before-after with (N = 72) and without outliers (n = 61) (Table 6). 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 6). 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 7).

Table 7. Paired-samples t-tests for bipolar Valence before-after (n = 61).

Robots vs. Writing vs. Social Media						
	Difference between means	*t*	*df*	*p*	*CI*	*N*
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 η_p^2 = .033; Novelty η_p^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, η_p^2 = .08) and as more Novel ($F_{(1,67)}$ = 5.16, p = .026, η_p^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, η_p^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 ΔValence (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 ΔValence 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 ΔValence (unipolar)

With $N = 72$, the effect of Media on ΔValence (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 ΔValence into the desired direction ($F_{(2,69)} = 4.56$, $p = .035$, $\eta_p^2 = .09$) but did not significantly affect negative ΔValence ($p = .177$). 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 ΔValence ($F_{(1,67)} = 5.67$, $p = .020$, $\eta_p^2 = .08$).

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 ΔValence into the desired direction ($F_{(2,58)} = 5.11$, $p = .009$, $\eta_p^2 = .15$) but did not significantly affect negative ΔValence ($F = 3.01$, $p = .057$). 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 Variance of Valence (*VV*) as indicator of emotional instability

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).

- We took the average of scale values per person: *MValBi, MValAi, MValBc, MValAc*

- Then we subtracted those averages from each of the relevant item ratings
- To avoid negative values, we squared the thus obtained differences
- The mean of squared differences told us the *VV*, Variance of Valence (positive vs negative, before and after treatment, i.e. exposure to Media): *VVbi – VVai; VVbc – VVac.*

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: $V = .05$, $F_{(3,57)} = 0.95$, $p = .424$, $\eta_p^2 = .05$

4. Conclusions

In conclusion, the key findings of our study are as follows:

(1) Self-disclosure through social media to reduce negative emotions was worse than through social robots, but not significantly different from writing.

(2) Self-disclosure through social media to boost positive emotions is less effective than self-disclosure through social robots, but not significantly different from writing.

(3) Outliers are mainly sensitive to the relvance of the medium, lees so to its novelty. Relevance and novelty were positively correlated.

(4) If Variance of Valence (*VV*) may indicate emotional instability, it did not show any significant differences among media.

5. Discussion

What we did was: 1) The data was analyzed to find the ratio of positive and negative responses on social media, and this was used as the basis for the experiment; 2) It was shown that the structured questionnaire we used was reliable, testing the same variable for the same set of questions and different variables between different sets of questions; 3) Our manipulation and treatment were effective. The videos clearly induced strong negative emotions, and after the treatment, positive emotions were significantly improved, and negative emotions were reduced. 4) The positive impact of media (robot vs. writing vs. social media) on people's mood could be established. Social media, however, performed significantly worse than robots as a means of self-disclosure.

Different from Duan et al. (2021), our findings generalized beyond the extreme cases. It may be that in larger data sets (like ours), outliers have less effect on the mean, standard deviation, and variance and 'come closer' to the general tendencies found in the data.

The limitations of our study were: 1) When we analyzed the ratio of positive and negative responses on social media, we only crawled one platform, Douban, which may have affected the data because of the platform's characteristics. A more accurate approach is to select several social media together for analysis; 2) Our sample was limited to the Chinese student community, which confines the generalizability of our results; 3) Our questionnaire was administered at one time after the experiment, which may have affected the accuracy of the assessment. Because human emotions are transient, participants can easily forget how they felt during interaction.

References

Bolls, Lang, A., & Potter, R. F. (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. https://doi.org/10.1177/009365001028005003

Duan, Yoon, M., Liang, Z., & Hoorn, J. F. (2021). Self-Disclosure to a Robot: Only for Those Who Suffer the Most. Robotics (Basel), 10(3), 98. https://doi.org/10.3390/robotics10030098

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. https://doi.org/10.1016/j.tele.2019.04.001

Lang, Shin, M., & Lee, S. (2005). Sensation Seeking, Motivation, and Substance Use: A Dual System Approach. Media Psychology, 7(1), 1–29. https://doi.org/10.1207/S1532785XMEP0701_1

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

Siedlecka, E., & Denson, T. (2019). Experimental methods for inducing basic emotions: A qualitative review. *Emotion Review, 11*(1), 87-97. doi: 10.1177/1754073917749016

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
o 我同意參與這項研究
o 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 via 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

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

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

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

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
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.