Opening Up to Social Robots: How Emotions Drive Self-Disclosure Behavior

Guy Laban[1], Arvid Kappas[2], Val Morrison[3], and Emily S. Cross[4]

***Abstract*— Self-disclosing to others can benefit emotional well-being, but socio-emotional barriers can limit people's ability to do so. Self-disclosing towards social robots can help overcome these obstacles as robots lack judgment and can establish rapport. To further understand the influence of affective factors on people's self-disclosure to social robots, this study examined the relationship between self-disclosure behaviour towards a social robot and people's emotional states and their perception of the robot's responses as comforting (i.e., being emphatic). The study included 1160 units of observation collected from 39 participants who conversed with the social robot Pepper (SoftBank Robotics) twice a week for 5 weeks (10 sessions in total), answering three personal questions in each session. Results show that perceiving the robot's responses as more comforting was positively related to self-disclosure behaviour (in terms of disclosure duration in seconds, and disclosure length in number of words), and negative emotional states, such as lower mood, and higher feelings of loneliness and stress, were associated with higher rates of self-disclosure towards the robot. Additionally, higher rates of introversion significantly predicted higher rates of self-disclosure towards the robot. The study reveals the meaningful influence of affective states on how people behave when talking to social robots, especially when experiencing negative emotions. These findings may have implications for designing and developing social robots in therapeutic contexts.**

I. Introduction

Self-disclosure is a communication behaviour aimed at introducing and revealing oneself to others, and it plays a key role in building relationships between two individuals [1], [2]. The ability to self-disclose personal information and emotions is a fundamental aspect of human communication [3], [4], enabling individuals to establish social connections, build relationships, and receive emotional support [5]–[7]. However, many people find it challenging to self-disclose their innermost thoughts and feelings, particularly when interacting with others face-to-face. This difficulty is especially pronounced in situations where individuals feel vulnerable, stressed, or lonely. The social context of disclosure might position the speaker in a fragile place, requiring certain adaptability and considering the social consequences of the disclosure, including the judgment of others [8]. This is highly present due to the fear of shame and stigma when engaging in self-disclosure and sharing personal, and maybe even sensitive matters [9]. This might disrupt the facilitation of interpersonal relationships due to the lack of reciprocity (see [3], [4], [10]), but it can also have health implications when patients might avoid disclosing information to healthcare providers such as medical doctors [11], or when engaging in psychotherapy and being requested to share sensitive information. Patients might draw back and hold to that information due to the fear of being judged and viewed negatively [12].

This often occurs because people tend to treat emotions as social information and not just as personal experiences (see Emotions as Social Information (EASI) model; [13]). When in need to self-disclose to others, people tend to read their conversational partner's emotional expression and social cues (including facial expressions, body language, and vocal tone), seeking emotional feedback for coordinating the social interaction and regulating the inferential processes and/or affective reactions for which they are self-disclosing emotions [13]. In other words, people might have a greater likelihood of self-disclosing when they believe that the person they are disclosing to is likely to provide them with the emotional feedback that they seek, but may avoid self-disclosure if they believe that the person they are interacting with is not likely to provide them with the emotional feedback that they need. This may be because the person is perceived as uninterested, unapproachable, or untrustworthy, or when perceiving a conversational partner's emotional expressions as judgmental, negative, or even threatening. Despite social norms of displaying affect [14], emotional responses to stimuli (like emotional expressions) are often the initial impulsive reaction of a social being. They can happen without thorough perceptual and cognitive processing and are more certain and faster than cognitive evaluations [15]. Therefore, it could be that *social robots* which are automated non-human entities that can control their expressions via computing, mechanics, and design, and are objectively perceived as objects [16], could avoid some of the socio-emotional barriers to self-disclosure [17].

Social robots, autonomous machines that interact and

The authors gratefully acknowledge funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie to ENTWINE, the European Training Network on Informal Care (Grant agreement no. 814072), the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement no. 677270 to EC), and the Leverhulme Trust (PLP-2018-152 to EC).

[1]Guy Laban is with the School of Psychology and Neuroscience, University of Glasgow, Glasgow, UK guy.laban@glasgow.ac.uk

[2]Arvid Kappas with the School of Business, Social and Decision Sciences, Constructor University, Bremen, Germany akappas@constructor.university

[3]Val Morrison with the Department of Psychology, Bangor University, Bangor, UK v.morrison@bangor.ac.uk

[2]Emily S. Cross is with the School of Psychology and Neuroscience, University of Glasgow, Glasgow, UK, the MARCS Institute for Brain, behaviour and Development, Western Sydney University, Sydney, New South Wales, Australia, and the Department of Humanities, Social and Political Sciences, ETH Zürich, Zürich, Switzerland emily.cross@ethz.ch

communicate with humans or other agents by following social behaviours and rules relevant to the specific role for which they have been designed [18], are gradually being introduced in health and care settings [19]. A growing evidence base documents how social robots might function as autonomous tools to support psychological health interventions [20] and both mental [21], [22], and physical health [23], [24]. These robotic agents can take on various forms and shapes and are gradually being deployed across various health and well-being settings because of their ability to function autonomously or semi-autonomously in physical and social spaces alongside humans [19]. Social robots are often designed to simulate social interactions [25], establish rapport [26], recognize emotions [27], and respond to users in a naturalistic and empathetic manner (e.g., [28]–[30]). Due to social robots' social features [16], [31], animate qualities [32] and physical and social embodiment [33], previous studies provide evidence for how social robots might be useful for encouraging humans to self-disclose information and emotions (e.g., [34]–[36]) and provide a sense of companionship to individuals who could use the support of socially-savvy artificial agents.

When comparing to self-disclosures to humans, we previously found that people shared more information with a human than with a humanoid social robot [34]. Another study by Bodala and colleagues [37] examined the use of a social robot as a mindfulness coach, in comparison to a human coach. They found that while both coaches were effective, the human coach rated significantly higher than the robotic coach. Yet, a different study by Nomura et al. [38] found that speech interactions with a social robot elicited lower tension compared to interactions with a human agent. The same study [38] showed the benefits of employing social robots for minimising social tension and anxieties, describing that participants with higher social anxiety felt less anxious and demonstrated less tension when knowing that they would interact with a robot as opposed to a human interlocutor. These results are in-line with two studies that found that people in bad mood benefited more from self-disclosing to a robot than participating in writing disclosure using a journal [39] or self-disclosing on social media [40]. Other emotional states might influence people's perceptions and behaviours towards robots. For example, a study with 80 participants reported a set of correlations between self-disclosure behaviour and personality traits, describing a positive correlation between interaction time and extraversion, a negative correlation between conscientiousness and interaction time, and a positive correlation between agreeableness and disclosure length (i.e., the number of words used per disclosure) [41]. Another example includes a previous study with 34 participants that showed that people who reported higher rates of loneliness evaluated the social robot FloBi (via observing a photo) as more anthropomorphic [42]. A similar recent study employing a similar methodology showed similar results that are more explicit towards self-disclosure behaviour. This cross-sectional study with 138 participants showed that there is a correlation between experiencing higher levels of loneliness due to the COVID-19 pandemic and showing a higher willingness to self-disclose to the robot NAO when observing the robot's photo [43]. Nevertheless, both studies examined participants' perceptions and willingness to self-disclose via self-reporting ratings and perceptions when observing pictures of the robots, rather than evaluating objective self-disclosure behaviour via simulating social interactions with a social robot. Accordingly, these results are merely related to the visual stimuli of the social robot's appearance and are missing crucial social information related to the robot's behaviour, social presence, and response to participants' input.

Accordingly, emotional states like mood, loneliness, stress and even traits like extraversion and introversion can influence people's reactions to robots, and the extent to which they self-disclose to these agents. Despite these insights, much remains unknown about the factors that influence self-disclosure to social robots, particularly in real-world settings due to the limited methodologies that have been employed when studying the matter (limited number of observations and using a photo as stimuli). Considering the socio-emotional limitations of engaging in self-disclosure behaviour, and the potential benefits these robotic agents might have for providing a safe and non-judgmental environment for self-disclosure, we were asking:

RQ: What is the relationship between emotional states and self-disclosure behaviour to a social robot?

To answer our research question we conducted a secondary analysis of empirical data from a long-term mediated experiment (see [35]). In a long-term experiment, 39 participants conversed with the social robot Pepper (SoftBank Robotics) twice a week for 5 weeks (10 sessions in total), disclosing to the robot about everyday experiences. We found that participants self-disclosed more to the robot as the sessions progressed, perceiving the robot to be more social, competent and comforting over time. The repeated interactions also led to improved mood (after each session, and over time) and decreased feelings of loneliness [35], [44]. We replicated this study with a sample of informal caregivers, who often experience high levels of emotional distress [45]. Our findings replicated the previous results [36], [46] and showed that caregiver participants felt less lonely and stressed, were more accepting of their caregiving situation, positively reappraised their caregiving situation and experienced reduced feelings of blame towards others [36]. These results demonstrate that people can establish meaningful relationships with social robots and highlight the value of social robot-led interventions with individuals living with considerably difficult life situations. Social robots could potentially elicit rich interactions with stressed individuals over time, acquire relevant information from their disclosures, and support their emotional well-being. Here we are aiming at evaluating the extent of self-disclosure to a social robot (i.e., the duration in seconds, and the length in number of words) due to the presence and variation in different emotional states like loneliness, stress, mood, and comfort, as well as emotional traits like extraversion and introversion,

or the tendency to disclose when in distress. We would like to further understand how subjective feelings and emotions that participants experienced during their participation in the experiment might have been associated with their self-disclosure behaviour towards the social robot Pepper, and can potentially predict the extent to which people self-disclose to social robots.

II. Methods

Consistent with recent proposals [47], [48], we pre-registered the data collection procedure of this study and report for how we determined our sample size, all data exclusions, all manipulations and all measures in the study (see [49]). The pre-registration was specifically documented regarding the primary analysis of the gathered data (see [35]). In addition, following open science initiatives (e.g., [50]), the de-identified data set, stimuli and analysis code associated with this study are freely available online [51]. By making the data available, we enable and encourage others to pursue tests of alternative hypotheses, as well as more exploratory analyses.

A. Data Collection

This study consists of a secondary analysis of data acquired in a long-term online-mediated experiment with 39 participants (M age = 36.41, SD = 12.20, 54% identify as females) reported in [35]. A full description of the experimental design, sample, stimuli, manipulation and procedure can be found in the paper describing the experiment and its initial results [35]. Here we will provide a summary description of the original methods.

All study procedures were approved by the research ethics committee of the University of Glasgow (ethics approval numbers 300200094 & 300200132). All participants provided written informed consent before participating in the study. Participants were recruited via Prolific and were randomly assigned to one of the two discussion topic groups, according to which they conversed with the social robot Pepper (SoftBank Robotics) via Zoom video chats about general everyday topics (e.g., social relationships, work-life balance, health and well-being) for 10 sessions over five weeks. One group's conversation topics were framed within the context of the Covid-19 pandemic (e.g., social relationships during the pandemic, sustaining mental health during the pandemic, etc.), whereas the other group's conversation topics were similar, except no explicit mention of the Covid-19 pandemic was ever made. Each interaction consisted of the robot asking the participant 3 questions (x3 repetitions). The topic of each interaction was assigned randomly before the experimental procedure started, as was the order of the questions. Participants were scheduled to interact with the robot twice a week during prearranged times for five weeks. Considering the sample size, the number of sessions and repetitions, the final number of observations in this study is 1160 (including only cases that were processed correctly).

Pepper was placed in front of a web camera (Logitech, 1080p), connected to the experimenter's computer (see Figure 1). Behind Pepper was a white wall and a flowerpot with a green plant (see Figure 2). Pepper communicated with participants in this study via the Wizard-of-Oz (WoZ) technique controlled by the experimenter via a PC laptop. All pre-scripted questions and speech items were written and coded in the WoZ system, with the experimenter controlling Pepper by pressing buttons on a PC laptop. Accordingly, the procedure followed a clear pre-programmed protocol where the experimenter did not need to speak or type anything during the interaction, but only pressed the relevant keys to trigger the required or appropriate text delivery via Pepper. Pepper communicated using a cheerful, high-pitched voice, and expressive and animated body language that corresponded to the spoken content and Pepper's physical capabilities. Each interaction was guided by Pepper as a semi-structured interview discussing non-sensitive topics regarding general everyday experiences. Each interaction followed the same order, starting with greetings followed by 3 questions (x3 repetitions). The participants were instructed to have a short conversation with Pepper, following Pepper's lead in the interaction and answering Pepper's questions. Participants were instructed that no time limit was applied for the interactions. They were further encouraged to participate in the interactions the way they saw fit - speaking as little or as much as they wished. In addition, participants were instructed that there were no correct or incorrect answers, and they were encouraged to provide honest answers according to what they felt comfortable with. The task followed the following structure and order:

Fig. 1. The lab settings, including the robot Pepper in front of a web camera, while the experimenter in the back is controlling the robot using the Wizard-of-Oz technique.

- Short greetings (e.g., "Hi there, how are you doing?").
- One pre-defined general question about the participant's day, week, or weekend, to build rapport (e.g., "how was your weekend? Did you do anything interesting?").
- An opening statement introducing the topic of the question (e.g., "I am about to ask you about your social life").
- Two pre-defined, non-sensitive questions that correspond to the topic that was randomly allocated to the interaction.

The questions and topics in the study were influenced by

[1] and [52] as an elicitation technique aiming to capture participants' subjective experiences regarding ten everyday topics (Work, Leisure and Passions, Finances, Relationships, Social Life, Mental Health, Physical Health, Personality, Goals and Ambitions, & Routine and Daily Activities; see [35]). Further detailed information regarding the manipulation and the task, including the task's structure and content, can be found in [35].

Participants were paid a total of £3 for every 30 minutes of participation or participation session if it lasted less than 30 minutes. Participants who completed all 10 sessions were paid an extra £20 after their final interaction. Participants were reminded that their participation was voluntary and they were given the contact information of the main researcher and experimenter should they wish to follow up with any further questions. After completing the study, participants received a comprehensive debriefing message in Prolific (forwarded by Prolific to their associated email address), providing further information about the study, the deception that was used (i.e., the experimenter was using the WoZ approach for communicating with participants to make it look like the robot was responding autonomously), and were again given the contact information of the main researcher and experimenter should they wish to follow up with any further questions or feedback.

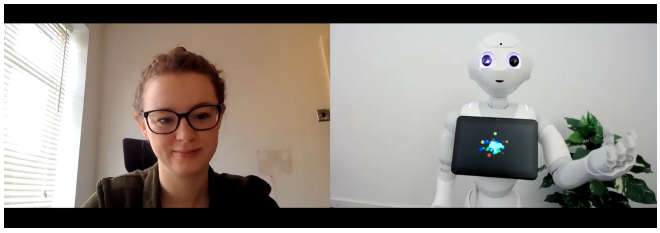

Fig. 2. The interaction from the eyes of the participants and the experimenter. The participants were exposed only to the robot Pepper via zoom chats.

B. Measurements

a) Disclosure duration: Duration of speech in seconds from each recording was extracted and processed using Parselmouth [53], a Python library for Praat [54].

b) Disclosure length: The volume of disclosure in terms of the number of words per disclosure. The recordings were automatically processed using the IBM Watson speech recognition engine, applying the British telephony model. To ensure capturing all utterances within each disclosure we amplified the audio files with 7 decibels and slowed the audio file's pitch. The number of words per disclosure was extracted from the text using a simple length command in Python.

c) Mood: To capture participants' mood change from their interactions with Pepper, participants reported their mood before and after the interaction with Pepper using the Immediate Mood Scaler (IMS-12; see [55]). IMS-12 includes 12 items of polarized moods, ranging from 1 (for negative moods) to 7 (for the equivalent positive moods). For this secondary analysis, a mean reliable scale was constructed for participants' mood before the interaction (M = 5.35, SD = 1.16, Cronbach's α = .96).

d) Comforting responses: To measure the extent to which participants perceived Pepper's responses as comforting the comforting response scale was adapted (see [56]). The scale includes 12 self-reported items rated on a seven-point scale, ranging from 1 (I strongly disagree) to 7 (I strongly agree). Accordingly, a mean scale was constructed (M = 5.50, SD = .89) which was found to be reliable (Cronbach's α = .91).

e) Loneliness: Each session participants were requested to report their feelings and thoughts of loneliness from the last three days using the short-form UCLA loneliness scale (ULS-8; see [57]). The scale includes 8 items rated on a seven-point scale, ranging from 1 (not at all) to 7 (all the time). Accordingly, a mean scale was constructed (M = 2.86, SD = 1.28) which was found to be reliable (Cronbach's α = .90).

f) Perceived Stress: Participants were requested to report their feelings and thoughts of periodic stress from the past month using the perceived stress scale [58]. The scale includes 10 statement items rated on a seven-point scale, ranging from 1 (never) to five (very often). A mean scale was constructed (M = 3.30, SD = 1.03) which was found to be reliable (Cronbach's α = .89).

g) Extraversion-Introversion Personality Trait: Participants were asked to rank their personality in terms of extraversion-introversion on a scale of 1 (Not at all) to 9 (Very applicable) on the 8 extraversion items of the Mini-Markers Big Five personality scale [59]. A mean scale was constructed (M = 5.61 SD = 1.42) which was found to be reliable (Cronbach's α = .86).

h) Disclosure Distress Index: Participants were asked to rank the extent to which they tend to self-disclose and conceal with others when in distress on a scale of 1 (Does not describe me at all) to 7 (Describes me extremely well) using the Disclosure Distress Index [60] that includes 12 items. A mean scale was constructed (M = 4.05 SD = 1.49) which was found to be reliable (Cronbach's α = .96).

III. Results

A. Self-Disclosure Duration

1) Correlations: Pearson correlations were used to examine the relationship between emotional states and self-disclosure duration to the robot (in seconds), with log-transformed data of self-disclosure duration to improve normality. The sample size was N = 1160. Results showed weak positive correlations between perceptions of the robot's comforting responses (R = .25, $p < .001$), feelings of loneliness (R = .24, $p < .001$), as well as feelings of stress (R = .08, $p < .001$) and self-disclosure duration. Mood before the interaction with the robot (R = -.19, $p < .001$), extraversion-introversion personality trait score (R = -.20, $p < .001$), and the tendency to disclose in distress (R = -.07, p = .011) had weak negative correlations with self-disclosure duration.

2) Multiple Regression: Multiple linear regression was used to test if emotional states significantly predicted participants' self-disclosure duration when interacting with the social robot Pepper. The outcome variable, self-disclosure duration, was log-transformed to improve normality. The overall regression was statistically significant, R = .43, F(6, 1153) = 42.53, $p < .001$, explaining 17.7% ($adj.R^2$ = .177) of the variance in participants' self-disclosure duration.

Results indicated that perceptions of the robot's comforting responses (β = .13, SE = .01, t(1153) = 10.78, $p <$ 0.001), feelings of loneliness (β = .06, SE = .01, t(1153) = 6.28, $p < 0.001$), mood before the interaction with the robot (β = -.05, SE = .01, t(1153) = -4.74, $p < 0.001$), feelings of periodic stress (β = -.05, SE = .01, t(1153) = -5.39, p < 0.001), and extraversion-introversion personality trait (β = -.05, SE = .01, t(1153) = -5.29, $p < 0.001$), were all significant predictors of self-disclosure duration to the robot, after controlling for each other. The tendency to disclose in distress was not a significant predictor of self-disclosure duration to the robot, β = .01, SE = .01, t(1153) = .89, p = 0.372.

When back transformed, the significant predictors can explain changes in self-disclosure behaviour towards the robot. One increase in perceptions of the robot's comforting responses predicts a 13% increase in self-disclosure duration to the robot, holding all other predictors constant. One increase in reported feelings of loneliness predicts a 6% increase in self-disclosure duration to the robot, holding all other predictors constant. One increase in the reported mood before the interaction with the robot predicts a 5% decrease in self-disclosure duration to the robot, holding all other predictors constant. One increase in reported feelings of periodic stress predicts a 5% decrease in self-disclosure duration to the robot, holding all other predictors constant. Finally, one increase in extraversion personality trait predicts a 5% decrease in self-disclosure duration to the robot, holding all other predictors constant.

These results indicate that perceptions of the robot's comforting responses had the strongest positive influence on self-disclosure duration to the robot ($b*$ = .30), followed by feelings of loneliness ($b*$ = .22), extraversion-introversion personality trait ($b*$ = -.20), feelings of periodic stress ($b*$= -.17), and finally mood before the interaction ($b*$ = -.15).

B. Self-Disclosure Length

1) Correlations: Pearson correlations were used to examine the relationship between emotional states and self-disclosure length to the robot (in number of words), with log-transformed data of self-disclosure length to improve normality. The sample size was N = 1160. Results showed weak to moderate positive correlations between perceptions of the robot's comforting responses (R = .30, $p < .001$), feelings of loneliness (R = .22, $p < .001$), as well as feelings of stress (R = .08, p = .005) and self-disclosure length. Mood before the interaction with the robot (R = -.17, $p < .001$), and extraversion-introversion personality trait score (R = -.11, p $< .001$) had weak negative correlations with self-disclosure duration. The correlation between the tendency to disclose in distress and self-disclosure length to the robot was not significant (R = -.01, p = .404).

2) Multiple Regression: Multiple linear regression was used to test if emotional states significantly predicted participants' self-disclosure length when interacting with the social robot Pepper. The outcome variable, self-disclosure length, was log-transformed to improve normality. The overall regression was statistically significant, R = .42, F(6, 1153) = 42.18, $p < .001$, explaining 17.6% ($adj.R^2$ = .176) of the variance in participants' self-disclosure length.

Results indicated that perceptions of the robot's comforting responses (β = .16, SE = .01, t(1153) = 12.69 $p <$ 0.001), feelings of loneliness (β = .06, SE = .01, t(1153) = 5.73, $p < 0.001$), mood before the interaction with the robot (β = -.06, SE = .01, t(1153) = -5.10, $p < 0.001$), feelings of periodic stress (β = -.04, SE = .01, t(1153) = -4.39, $p < 0.001$), and extraversion-introversion personality trait (β = -.03, SE = .01, t(1153) = -2.56, p = 0.011), were all significant predictors of self-disclosure duration to the robot, after controlling for each other. The tendency to disclose in distress was not a significant predictor of self-disclosure duration to the robot, β = .01, SE = .01, t(1153) = 1.01, p = 0.311.

When back transformed, the significant predictors can explain changes in self-disclosure behaviour towards the robot. One increase in perceptions of the robot's comforting responses predicts a 16% increase in self-disclosure duration to the robot, holding all other predictors constant. One increase in reported feelings of loneliness predicts a 6% increase in self-disclosure duration to the robot, holding all other predictors constant. One increase in the reported mood before the interaction with the robot predicts a 6% decrease in self-disclosure duration to the robot, holding all other predictors constant. One increase in reported feelings of periodic stress predicts a 4% decrease in self-disclosure duration to the robot, holding all other predictors constant. Finally, one increase in extraversion personality trait predicts a 3% decrease in self-disclosure duration to the robot, holding all other predictors constant.

These results indicate that perceptions of the robot's comforting responses had the strongest positive influence on self-disclosure duration to the robot ($b*$ = .36), followed by feelings of loneliness ($b*$ = .20), mood before the interaction ($b*$ = -.17), feelings of periodic stress ($b*$= -.14), and finally extraversion-introversion personality trait ($b*$ = -.10).

IV. Discussion

The present study investigated the relationship between participants' self-disclosure behaviour (in terms of the disclosure duration and length) towards a social robot, Pepper, and their perception of the robot's comforting responses, negative emotional states, and emotional personality traits. 39 participants conversed with the social robot Pepper twice a week for 5 weeks (10 sessions in total), disclosing to the robot about general everyday experiences.

TABLE I

THE RESULTS OF MULTIPLE REGRESSION ANALYSIS WITH EMOTIONAL FACTORS AS PREDICTORS OF SELF DISCLOSURE DURATION TOWARDS THE ROBOT AND SELF DISCLOSURE LENGTH TOWARDS THE ROBOT.

	Duration				**Length**			
Predictor	*β(SE)*	*b**	*t*	*95%CI*	*β(SE)*	*b**	*t*	*95%CI*
Perceptions of comforting responses	0.13 (0.01)***	0.30	10.78	[0.10,0.15]	0.16 (0.01)***	0.36	12.69	[0.13,0.18]
Feelings of loneliness	0.06 (0.01)***	0.22	6.28	[0.04,0.08]	0.06 (0.01***	0.20	5.73	[0.04,0.08]
Mood before the interaction	-0.05 (0.01)***	-0.15	-4.74	[-0.08,-0.03]	-0.06 (0.01)***	-0.17	-5.10	[-0.08,-0.04]
Feelings of periodic stress	-0.05 (0.01)***	-0.17	-5.39	[-0.06,-0.03]	-0.04 (0.01)***	-0.14	-4.39	[-0.06,-0.02]
Extraversion-introversion	-0.05 (0.01)***	-0.20	-5.29	[-0.07,-0.03]	-0.03 (0.01)*	-0.10	-2.56	[-0.05,-0.01]
Tendency to disclose in distress	0.01 (0.01)	0.03	0.89	[-0.01,0.03]	0.01 (0.01)	0.04	1.01	[-0.01,0.03]
Constant	2.38 (0.07)***	-	11.09	[0.97,1.38]	2.35 (0.07)***	-	32.28	[1.02,1.46]
Model Statistics								
R	0.43***				0.42***			
F(df)	42.53 (6, 1153)				42.18 (6, 1153)			
Adjusted R^2	0.177				0.176			

*Note: * p <0.05, ** p <0.01, *** p <0.001*

Our findings suggest that participants who perceived Pepper's responses as more comforting were more likely to self-disclose more towards the robot. Moreover, this perception was a significant predictor of self-disclosure behaviour, suggesting that the quality of the robot's response is essential in facilitating self-disclosure. Thus, designers of social robots should strive to create robots that can provide appropriate and high-quality responses to users' emotional states and needs, to maximize the effectiveness of the robot as an emotional support tool. Moreover, this key finding may suggest that the robot's ability to exhibit empathic behaviour played a significant role in facilitating self-disclosure. Participants who perceived Pepper's responses as more comforting may have felt that the robot understood and validated their experiences, which increased their willingness to engage in higher rates of self-disclosure behaviour, and potently disclose more personal information (however, this should be assessed with a more thorough and qualitative analysis of the shared content). These findings have important implications for the design and programming of social robots. Specifically, social robots intended for applications such as emotional support, healthcare, and education must be designed to exhibit empathic behaviour. This can include using natural language processing and emotional recognition algorithms to understand users' emotional states and respond appropriately, as well as using nonverbal cues such as facial expressions, body language, and tone of voice to convey empathy and build rapport. By showing understanding, facilitating the feeling of understanding, and responding appropriately to users' emotional states and needs, social robots can provide an effective and safe space for individuals to self-disclose their emotions and needs, which has tremendous implications for emotional support [61] and introducing social robots as interventions in health and care settings.

Furthermore, consistent with previous results (e.g., [39], [40], [43]), our results suggest that participants who experienced negative emotional states, such as lower mood before the interaction and higher levels of loneliness, were more likely to self-disclose more towards the robot. These findings suggest that individuals may use social robots as a form of an emotional outlet when experiencing negative emotional states. Notably, both mood and loneliness were significant predictors of self-disclosure behaviour, indicating that these factors may play a crucial role in determining the extent of self-disclosure towards social robots. This key finding suggests that individuals may feel more comfortable opening up to robots when they are experiencing negative emotions. This could be because robots lack social expectations and norms, are non-judgmental due to their electronic brains, and do not express emotions due to their mechanical design. These characteristics can make it easier for people to share their thoughts and feelings without fear of judgment, rejection, shame, or stigma (see [9]). Hence, these results further support the idea that people might feel more inclined to self-disclose towards robots due to their non-judgmental behaviour and appearance, and lack of emotional expression, which makes them an attractive option for those seeking emotional support.

Interestingly, in contrast to Neerincx et al. results [41], our study found that higher rates of introversion were associated with higher rates of self-disclosure towards the robot. This result suggests that introverted individuals may find it easier to self-disclose personal information to a social robot, possibly because of the reduced social anxiety associated with interacting with a machine, which further supports earlier results by Nomura et al. [38]. In line with the previous key finding, introverted individuals may perceive social robots as non-judgmental and trustworthy listeners. These perceptions may further facilitate self-disclosure as introverts may feel more comfortable sharing personal information with a machine that they perceive as non-judgmental. However, the tendency to self-disclose in distress was not found to be a significant predictor of self-disclosure behaviour, further indicating that self-disclosing towards robots might be viewed by participants as a legitimate channel for disclosure despite their tendency not to disclose to others when in distress.

This finding highlights the potential benefits of social robots in providing emotional support to individuals. Unlike human support providers, social robots may not be limited by the availability or the ability to provide immediate support. Additionally, social robots may offer a sense of privacy and confidentiality that may not be available with human support providers.

Finally, we found positive correlations between perceived stress and the self-disclosure behaviour variables (duration and length) in our correlation analysis. However, in our regression models, perceived stress had negative regression coefficients, which suggests that when controlling for the influence of other variables in the model, perceived stress has a negative relationship with rates of self-disclosure towards the robot. The positive correlations between perceived stress and the self-disclosure behaviour variables in the correlation analysis may be partly or fully explained by other variables in the model, such as mood before the interaction or feelings of loneliness. These variables may share variance with perceived stress, which could inflate the correlation between perceived stress and the outcome variable. Accordingly, when situated together in the model, the negative regression coefficient might reflect multicollinearity issues, or the difference in the signs simply reflects random variation around zero (see [62]). This could also mean that perceived stress is not as meaningful in eliciting and predicting self-disclosure behaviour towards robots as other emotional states that were included in the model (e.g., perceptions of the robot's responses as comforting, mood, or feelings of loneliness). Therefore, it is important to interpret regression coefficients in the context of the other variables included in the model, rather than relying solely on the correlations between variables.

V. Conclusions

These findings provide important insights into the emotional factors that influence individuals' self-disclosure behaviour towards social robots. Our results suggest that perceptions of the robot's comforting response, negative emotional states (such as low mood, loneliness, and periodic stress), and emotional personality traits (like extraversion-introversion) play essential roles in facilitating self-disclosure behaviour towards social robots. Our study highlights the potential of social robots as tools for emotional support, particularly for individuals experiencing negative emotional states or those who may find it difficult to self-disclose personal information to others. Our study provides a valuable contribution towards understanding the affective nuances that shape people's behaviour when interacting with social robots, and what might influence individuals to establish meaningful relationships with these robotic agents. We hope that our findings will inspire further research in this area and encourage the development of more effective emotional support tools to enhance mental health and well-being.

ACKNOWLEDGMENTs

The authors gratefully acknowledge funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie to ENTWINE, the European Training Network on Informal Care (Grant agreement no. 814072), the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement no. 677270 to EC), and the Leverhulme Trust (PLP-2018-152 to EC).

References

[1] S. M. Jourard and P. Lasakow, "Some factors in self-disclosure," *The Journal of Abnormal and Social Psychology*, vol. 56, no. 1, pp. 91–98, 1958.

[2] W. B. Pearce and S. M. Sharp, "Self-Disclosing Communication," *Journal of Communication*, vol. 23, no. 4, pp. 409–425, 1973.

[3] I. Altman and D. A. Taylor, *Social penetration: The development of interpersonal relationships*. Oxford, England: Holt, Rinehart & Winston, 1973.

[4] V. J. Derlega, M. S. Harris, and A. L. Chaikin, "Self-disclosure reciprocity, liking and the deviant," *Journal of Experimental Social Psychology*, vol. 9, pp. 277–284, 7 1973.

[5] J. Frattaroli, "Experimental disclosure and its moderators: A meta-analysis," *Psychological bulletin*, vol. 132, no. 6, pp. 823–865, 2006.

[6] P. G. Frisina, J. C. Borod, and S. J. Lepore, "A Meta-Analysis of the Effects of Written Emotional Disclosure on the Health Outcomes of Clinical Populations," *The Journal of nervous and mental disease*, vol. 192, no. 9, 2004.

[7] E. Kennedy-Moore and J. C. Watson, "How and When Does Emotional Expression Help?," *Review of General Psychology*, vol. 5, no. 3, pp. 187–212, 2001.

[8] S. M. Jourard, *Disclosing man to himself*. Van Nostrand Reinhold, 6 1968.

[9] L. Smart and D. M. Wegner, "The hidden costs of hidden stigma.," in *The social psychology of stigma.* (Heatherton. T. F., R. E. Kleck, M. R. Hebl, and J. G. Hull, eds.), pp. 220–242, New York, NY, US: The Guilford Press, 2000.

[10] R. L. Archer and J. H. Berg, "Disclosure reciprocity and its limits: A reactance analysis," *Journal of Experimental Social Psychology*, vol. 14, pp. 527–540, 11 1978.

[11] H. B. Beckman and R. M. Frankel, "The effect of physician behavior on the collection of data," *Annals of internal medicine*, vol. 101, no. 5, pp. 692–696, 1984.

[12] A. Farber, Barry, *Self-disclosure in Psychotherapy*. No. 3, Guilford Press, 2006.

[13] G. A. Van Kleef and S. Côté, "The Social Effects of Emotions," *Annual Review of Psychology*, vol. 73, pp. 629–658, 1 2022.

[14] D. Matsumoto, "Cultural similarities and differences in display rules," *Motivation and Emotion*, vol. 14, pp. 195–214, 9 1990.

[15] R. B. Zajonc, "Feeling and thinking: Preferences need no inferences," *American Psychologist*, vol. 35, pp. 151–175, 2 1980.

[16] E. S. Cross and R. Ramsey, "Mind Meets Machine: Towards a Cognitive Science of Human–Machine Interactions," *Trends in Cognitive Sciences*, vol. 25, pp. 200–212, 3 2021.

[17] G. Laban, "Social Robots as Communication Partners to Support Emotional Health and Well-Being," in *2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 1–5, IEEE, 10 2022.

[18] C. Breazeal, "Toward sociable robots," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 167–175, 2003.

[19] A. Henschel, G. Laban, and E. S. Cross, "What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You," *Current Robotics Reports*, no. 2, pp. 9–19, 2021.

[20] N. L. Robinson, T. V. Cottier, and D. J. Kavanagh, "Psychosocial Health Interventions by Social Robots: Systematic Review of Randomized Controlled Trials," *J Med Internet Res*, vol. 21, no. 5, pp. 1–20, 2019.

[21] G. Laban, Z. Ben-Zion, and E. S. Cross, "Social Robots for Supporting Post-traumatic Stress Disorder Diagnosis and Treatment," *Frontiers in psychiatry*, vol. 12, 2022.

[22] A. A. J. Scoglio, E. D. Reilly, J. A. Gorman, and C. E. Drebing, "Use of Social Robots in Mental Health and Well-Being Research: Systematic Review," *J Med Internet Res*, vol. 21, no. 7, p. e13322, 2019.

[23] R. Feingold Polak and S. L. Tzedek, "Social Robot for Rehabilitation: Expert Clinicians and Post-Stroke Patients' Evaluation Following a Long-Term Intervention," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '20, (New York, NY, USA), p. 151–160, Association for Computing Machinery, 2020.

[24] A. Langer, R. Feingold-Polak, O. Mueller, P. Kellmeyer, and S. Levy-Tzedek, "Trust in socially assistive robots: Considerations for use in rehabilitation," 9 2019.

[25] A. Kappas, R. Stower, and E. J. Vanman, "Communicating with Robots: What We Do Wrong and What We Do Right in Artificial Social Intelligence, and What We Need to Do Better," in *Social Intelligence and Nonverbal Communication* (R. J. Sternberg and A. Kostić, eds.), pp. 233–254, Cham: Springer International Publishing, 2020.

[26] J. Gratch and G. Lucas, "Rapport Between Humans and Socially Interactive Agents," in *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, vol. 1, pp. 433–462, New York, NY, USA: Association for Computing Machinery, 1 ed., 9 2021.

[27] M. Spitale and H. Gunes, "Affective Robotics For Wellbeing: A Scoping Review," in *10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, IEEE, 2022.

[28] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic Robots for Long-term Interaction: Evaluating Social Presence, Engagement and Perceived Support in Children," *International Journal of Social Robotics*, vol. 6, pp. 329–341, 3 2014.

[29] A. M. Rosenthal-von der Pütten, F. P. Schulte, S. C. Eimler, S. Sobieraj, L. Hoffmann, S. Maderwald, M. Brand, and N. C. Krämer, "Investigations on empathy towards humans and robots using fMRI," *Computers in Human Behavior*, vol. 33, pp. 201–212, 2014.

[30] S. Park and M. Whang, "Empathy in Human–Robot Interaction: Designing for Social Robots," *International Journal of Environmental Research and Public Health*, vol. 19, p. 1889, 2 2022.

[31] R. Hortensius and E. S. Cross, "From automata to animate beings: the scope and limits of attributing socialness to artificial agents," *Annals of the New York Academy of Sciences*, vol. 1426, no. 1, pp. 93–110, 2018.

[32] E. S. Cross, R. Ramsey, R. Liepelt, W. Prinz, and A. F. d. C. Hamilton, "The shaping of social perception by stimulus and knowledge cues to human animacy," *Philosophical transactions of the Royal Society of London.Series B, Biological sciences*, vol. 371, no. 1686, p. 20150075, 2016.

[33] R. Hortensius, F. Hekele, and E. S. Cross, "The Perception of Emotion in Artificial Agents," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 4, pp. 852–864, 2018.

[34] G. Laban, J.-N. George, V. Morrison, and E. S. Cross, "Tell me more! Assessing interactions with social robots from speech," *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 136–159, 2021.

[35] G. Laban, A. Kappas, V. Morrison, and E. S. Cross, "Human-Robot Relationship: long-term effects on disclosure, perception and well-being," *PsyArXiv*, 2022.

[36] G. Laban, V. Morrison, A. Kappas, and E. S. Cross, "Coping with Emotional Distress via Self-Disclosure to Robots: Intervention with Caregivers," *PsyArxiv*, 2023.

[37] I. P. Bodala, N. Churamani, and H. Gunes, "Teleoperated robot coaching for mindfulness training: A longitudinal study," *2021 30th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2021*, pp. 939–944, 8 2021.

[38] T. Nomura, T. Kanda, T. Suzuki, and S. Yamada, "Do people with social anxiety feel anxious about interacting with a robot?," *AI & SOCIETY*, vol. 35, no. 2, pp. 381–390, 2020.

[39] Y. Duan, M. Yoon, Z. Liang, and J. F. Hoorn, "Self-Disclosure to a Robot: Only for Those Who Suffer the Most," *Robotics 2021, Vol. 10, Page 98*, vol. 10, p. 98, 7 2021.

[40] R. L. Luo, T. X. Y. Zhang, D. H.-C. Chen, J. F. Hoorn, and I. S. Huang, "Social Robots Outdo the Not-So-Social Media for Self-Disclosure: Safe Machines Preferred to Unsafe Humans?," *Robotics 2022, Vol. 11, Page 92*, vol. 11, p. 92, 9 2022.

[41] A. Neerincx, C. Edens, F. Broz, Y. Li, and M. Neerincx, "Self-Disclosure to a Robot "In-the-Wild": Category, Human Personality and Robot Identity," *RO-MAN 2022 - 31st IEEE International Conference on Robot and Human Interactive Communication: Social, Asocial, and Antisocial Robots*, pp. 584–591, 2022.

[42] F. Eyssel and N. Reich, "Loneliness makes the heart grow fonder (of robots) - On the effects of loneliness on psychological anthropomorphism," *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 121–122, 2013.

[43] A. Penner and F. Eyssel, "Germ-Free Robotic Friends: Loneliness during the COVID-19 Pandemic Enhanced the Willingness to Self-Disclose towards Robots," *Robotics*, vol. 11, no. 6, 2022.

[44] G. Laban, A. Kappas, V. Morrison, and E. S. Cross, "User Experience of Human-Robot Long-Term Interactions," in *Proceedings of the 10th International Conference on Human-Agent Interaction*, HAI '22, (New York, NY, USA), pp. 287–289, Association for Computing Machinery, 2022.

[45] T. A. Revenson, K. Griva, A. Luszczynska, V. Morrison, E. Panagopoulou, N. Vilchinsky, and M. Hagedoorn, "The Emotional Experience of Caregiving," in *Caregiving in the Illness Context*, pp. 38–47, Palgrave Pivot, London, 2016.

[46] G. Laban, V. Morrison, A. Kappas, and E. S. Cross, "Informal Caregivers Disclose Increasingly More to a Social Robot Over Time," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI EA '22, (New York, NY, USA), pp. 1–7, Association for Computing Machinery, 2022.

[47] L. D. Nelson, J. P. Simmons, and U. Simonsohn, "Let's Publish Fewer Papers," *Psychological Inquiry*, vol. 23, no. 3, pp. 291–293, 2012.

[48] J. P. Simmons, L. D. Nelson, and U. Simonsohn, "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychol Sci*, vol. 22, no. 11, pp. 1359–1366, 2011.

[49] G. Laban, A. Kappas, V. Morrison, and E. S. Cross, "Socially Assistive Robots in Prolonged Interactions for Measuring Expression in Natural Settings," tech. rep., 2 2021.

[50] M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. P. du Sert, U. Simonsohn, E.-J. Wagenmakers, J. J. Ware, and J. P. A. Ioannidis, "A manifesto for reproducible science," *Nature Human Behaviour*, vol. 1, no. 1, p. 21, 2017.

[51] G. Laban, A. Kappas, V. Morrison, and E. S. Cross, "OSF Repository: Socially Assistive Robots in Prolonged Interactions for Measuring Expression in Natural Settings," 2020.

[52] S. M. Jourard, *Self-disclosure: An experimental analysis of the transparent self*. Oxford, England: John Wiley, 1971.

[53] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 11 2018.

[54] P. Boersma and D. Weenink, "PRAAT, a system for doing phonetics by computer," *Glot international*, vol. 5, pp. 341–345, 2001.

[55] M. Nahum, T. M. Van Vleet, V. S. Sohal, J. J. Mirzabekov, V. R. Rao, D. L. Wallace, M. B. Lee, H. Dawes, A. Stark-Inbar, J. T. Jordan, B. Biagianti, M. Merzenich, and E. F. Chang, "Immediate Mood Scaler: Tracking Symptoms of Depression and Anxiety Using a Novel Mobile Mood Scale," *JMIR Mhealth Uhealth*, vol. 5, no. 4, p. e44, 2017.

[56] R. A. Clark, A. J. Pierce, K. Finn, K. Hsu, A. Toosley, and L. Williams, "The impact of alternative approaches to comforting, closeness of relationship, and gender on multiple measures of effectiveness," *Communication Studies*, vol. 49, pp. 224–239, 9 1998.

[57] R. D. Hays and M. R. DiMatteo, "A Short-Form Measure of Loneliness," *Journal of Personality Assessment*, vol. 51, pp. 69–81, 3 1987.

[58] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of health and social behavior*, vol. 24, no. 4, pp. 385–396, 1983.

[59] G. Saucier, "Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers," *Journal of personality assessment*, vol. 63, no. 3, pp. 506–516, 1994.

[60] J. H. Kahn, B. E. Hucke, A. M. Bradley, A. J. Glinski, and B. L. Malak, "The Distress Disclosure Index: A research review and multitrait–multimethod examination," *Journal of Counseling Psychology*, vol. 59, no. 1, pp. 134–149, 2012.

[61] J. Zaki, "Integrating Empathy and Interpersonal Emotion Regulation," *Annual Review of Psychology*, vol. 71, pp. 517–540, 1 2020.

[62] R. F. Falk and N. B. Miller, *A primer for soft modeling*. Akron, Ohio: University of Akron Press, 1 ed., 1992.