This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

Sharing our Emotions with Robots: Why do we do it and how does it make us feel?

Guy Laban, Emily S. Cross

***Abstract*—Self-disclosure and the social sharing of emotions facilitate social relationships and can positively affect people's well-being. Nevertheless, individuals might refrain from engaging in these interpersonal communication behaviours with other people, due to socio-emotional barriers, such as shame and stigma. Social robots, free from these human-centric judgements, could encourage openness and overcome these barriers. Accordingly, this paper reviews the role of self-disclosure and social sharing of emotion in human-robot interactions (HRIs), particularly its implications for emotional well-being and the dynamics of social relationship building between humans and robots. We investigate the transition of self-disclosure dynamics from traditional human-to-human interactions to HRI, revealing the potential of social robots to bridge socio-emotional barriers and provide unique forms of emotional support. This review not only highlights the therapeutic potential of social robots but also raises critical ethical considerations and potential drawbacks of these interactions, emphasising the importance of a balanced approach to integrating robots into emotional support roles. The review underscores a complex but promising frontier at the intersection of technology and emotional well-being, advocating for careful consideration of ethical standards and the intrinsic human need for connection as we advance in the development and application of social robots.**

***Index Terms*—Human–Robot Interaction, Affective Robots, Affective Computing, Self-Disclosure, Social Sharing of Emotion**

I. Introduction

Interpersonal communication plays a crucial role in establishing and maintaining relationships [1], helping us to regulate emotions [2, 3, 4, 5], and enhancing our emotional well-being [6, 7, 8]. It is a process by which people create, interpret, and transmit social messages, including self-disclosure and the social sharing of emotion, which can significantly impact emotional health [7, 8]. Accordingly, interpersonal communication is an essential part of human social life and behaviour, playing a significant role in the affective dynamics of social relationships. Social robots are designed to interact with humans by following human social norms and behaviours [9], and research is highlighting how it is becoming increasingly possible to position these artificial agents as participants in interpersonal communication [10]. Given the fundamental role of self-disclosure in human relationships, understanding how these interpersonal communication dynamics unfold between humans and robots is essential for understanding how we might foster emotional connections and build enduring social relationships with these agents. Furthermore, given the benefits of self-disclosure and emotional sharing for supporting psychological health [6, 7, 8], understanding how these behaviours manifest during interactions with robots holds potential for helping us to assess their capabilities to effectively support users' emotional well-being in meaningful ways.

Therefore, this review paper is dedicated to exploring the nuances of self-disclosure and social sharing of emotion within the context of human–robot interaction (HRI), emphasising the implications for emotional well-being and relationship building between humans and robots. If social robots are to be designed to effectively engage human users on a social level, these machines will need to engage in socially meaningful communication with their human interactants. Considering the role of interpersonal communication behaviours like self-disclosure and the social sharing of emotion in establishing social relationships, we argue that acts of self-disclosure and emotional sharing by human users toward social robots are crucial aspects of establishing human–robot affective social connections. Hence, this review seeks to dissect how the dynamics of self- disclosure, traditionally studied within human-human interactions, translate into the realm of interactions with social robots, especially in settings where they are deployed to provide care, emotional support, and companionship. We aim to introduce a foundational theoretical framework to guide future research on self-disclosure and emotional sharing with robots, supported by empirical evidence from the field.

Here we review the social and behavioural mechanisms of self-disclosure and emotional sharing in HRI. We introduce the psychological theories underlying these behaviours, their extension into HRI settings, and the socio-emotional barriers that may influence these behaviours when interacting with other humans and social robots. Beyond discussing how psychological theories of interpersonal communication and self-disclosure can be applied to the context of HRI, we review relevant empirical research in this field. Accordingly, we explore how social robots and other artificial agents might overcome some of the socio-emotional barriers people experience when engaging in self-disclosure to others, and how interactions with these mechanical agents might emotionally support people. In doing so, we highlight the ways in which engaging in self-disclosure and emotional sharing with social robots can promote humans' emotional well-being. We further address additional considerations of this type of emotional engagement

[1]Guy Laban is with the Department of Computer Science and Technology of the University of Cambridge, UK and with the School of Psychology and Neuroscience, University of Glasgow, Glasgow, UK. guy.laban@cl.cam.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

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

within HRI. We discuss safety and ethical considerations of self-disclosure to robots, as well as the downsides of such interactions, including concerns and identifying potential risks for further discussion. We consider the role of culture in such interactions, reflecting on the need for more inclusive and diverse research to properly study and understand interpersonal communication behaviours like self-disclosure and the social sharing of emotions when interacting with social robots. Finally, we reflect on how ongoing advances in AI will affect the way we communicate with robots, especially how we share emotions and self-disclose as these agents become increasingly adaptive and (artificially) intelligent.

II. Approach

In this review, we adopted an exploratory and iterative approach. Our aim was to build a comprehensive understanding of the dynamics of self-disclosure in HRI by introducing key theories in the field and presenting empirical HRI studies. We began by examining the theoretical scope of self-disclosure and social sharing of emotion, drawing on established frameworks from psychology, communication theory, and social cognition to understand how self-disclosure functions in human relationships. These theoretical insights provided a foundation for our review, helping to define the key dynamics of this social and behavioural phenomenon.

To identify the most relevant studies, we employed a snowballing approach [11, 12], following references from seminal papers across the field of HRI. To complement our conceptual understanding of this phenomenon, we also incorporated studies from related fields such as virtual agents and conversational user interfaces. As we explored the literature, we refined our focus to cover three central themes: why individuals disclose personal information to robots, how such interactions can be sustained, and their impact on emotional well-being. This iterative process allowed us to remain open to emerging topics and adjust our search as new lines of inquiry emerged.

We prioritised peer-reviewed journal articles and conference papers, but also included grey literature when the insights were particularly relevant to our questions. Given the emerging and multidisciplinary nature of the field, rather than applying strict inclusion/exclusion criteria, we focused on studies directly engaging with the dynamics of self-disclosure and emotional sharing. By synthesising empirical studies, we examined how robots—through their ability to provide confidentiality, establish rapport, and offer emotional support— are being deployed to facilitate socially and emotionally meaningful interactions. We also critically evaluated potential biases, limitations, and ethical concerns in the design and deployment of social robots across diverse social and cultural contexts. While our approach allowed for a broad exploration of the topic, it is important to acknowledge that the lack of strict inclusion/exclusion criteria might affect replicability. However, given the emerging and interdisciplinary nature of the field, we believe that the flexibility of this method was crucial for capturing the breadth of existing work.

III. Why do we self-disclose, and how does it make us feel?

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 [13, 14]. Self-disclosure can be perceived as a complicated and dynamic social process that facilitates relationships and improves bonding [15, 16]. People tend to disclose thoughts and feelings with others, especially when experiencing unique and/or challenging life events [17]. Self-disclosure thus serves an evolutionary function of strengthening interpersonal relationships but also produces a wide variety of health and well-being benefits, including coping with stress and traumatic events and eliciting help and support from those one discloses to [18, 19, 20].

One of the most important factors and facilitators of self-disclosure is *reciprocity*, a response or reaction from a conversation partner that can encourage or attenuate the level of disclosure [16] and the relationship in general [15]. Reciprocity is an important interpersonal dynamic for regulating self-disclosure [21] and can be expressed with different verbal and non-verbal behaviours that convey involvement in an interaction [22, 23]. Throughout a conversation, both parties implicitly interpret and react to each other's disclosures, attributing values to the breadth and depths of the disclosures for balancing and regulating their own disclosures and eventually achieving an equilibrium of reciprocal self-disclosure. When equilibrium is not achieved and the level of self-disclosure does not correspond with expectations, it can damage the relationship. Therefore, self-disclosure can feel involuntary or unnatural, and it could be perceived as invasive, abnormal, uncomfortable, and at times, even unethical [15, 16, 24].

However, people often self-disclose for different emotional reasons that extend social norms and the establishment and maintenance of interpersonal relationships. People are influenced by their moods and daily events and tend to self-disclose positive information when experiencing a positive mood or positive events [25]. For example, the concept of *capitalization* explains the interpersonal process of disclosing positive events to close others, which has been linked to individual and relationship well-being [26]. When engaging in capitalization, people disclose positive information to enhance their level of positive affect, and in turn experience lower emotional distress and increased intimacy [27]. This has been studied further, demonstrating that the sharing of positive moods and life events has shown positive benefits to both the sender and the receiver, and has been associated with increased daily positive affect and life satisfaction [17], increased relationship satisfaction and feelings of trust [17, 28, 29], increased self-esteem and decreased loneliness [30], and decreased negative affect [27]. Interestingly, previous studies suggest that even when experiencing negative life events (e.g., coping with a chronic illness like cancer), engaging in capitalization and sharing positive emotions and events with intimate partners enhances relationship well-being independently of sharing bad news [e.g., 29]. People might also use self-disclosure to express and monitor negative emotions when experiencing negative

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

moods, as well as when being distressed, anxious, and fearful [4, 31, 32]. For example, when engaging in *co-rumination* people extensively discuss and revisit problems, speculating about problems and self-disclosing particularly negative feelings [33]. Another similar behaviour would be *emotional venting*, talking about negative emotions and events to feel better and 'getting it off the chest' [34]. Engaging in these behaviours is significantly associated with cognitive traits related to anxiety [35], and while people do not immediately recover from their emotional experiences, they report more subjective benefits compared to people who do not engage in self-disclosing about negative emotions and events [36]. Furthermore, it has been shown that engaging in co-rumination is positively linked with friendship closeness, perceptions of relationship quality, and even greater job satisfaction [37].

The tendency to self-disclose due to personal experiences and emotions and not due to typical reciprocal norms can be further explained by the *social exchange theory* [38, 39], addressing that relationships are formed through the interplay of cost and reward while comparing alternatives. With self-interest and interdependence as the basic features of an interaction, two entities hold a certain value and develop a relationship following the exchange of value. For subjective self-interests (i.e., psychological, emotional, social or economic needs), an exchange is perceived as a social behaviour with a potential (e.g., social, emotional or economic) outcome [39, 40, 41, 42, 43]. Therefore, people with certain needs, positive or negative, would use self-disclosure to receive a certain reward or achieve a desired outcome from an interaction party [44]. Previous studies report that when experiencing illness, people often prefer to disclose and talk about it with people that add positive value to their experience (e.g., other patients, a psychologist, a consultant, a physiotherapist, a family medical doctor, and even their best friend), rather than family members that might get worried and transfer negative emotion to them [45]. Thus, people might self-disclose for the exchange of a variety of reasons, some might be materialistic or subjective, like fame, popularity, novelty and curiosity [41, 42, 43]. Yet, in interpersonal settings, this could be seeking the recognition of one's emotion, empathy, advice, recommendation, or just aspiring to be heard. In fact, previous research highlights the importance of feeling listened to, and how it might affect different factors of emotional well-being like feelings of loneliness [46] and perceptions of burden [47].

Hence, it appears that engaging in self-disclosure can support emotional well-being via the ability to provide and receive support and improve mood and offer a comfortable setting for concealment, sharing feelings, and regulation of emotions [3, 4, 5, 48, 49, 50], and can have a positive impact on mental and physical health [51]. The *map of interpersonal regulation* by Zaki et al. [5] explains that people might use self-disclosure as different *intrinsic regulatory processes* (i.e., being the sender in a self-disclosure relationship) that can have different goals which are response-dependent or independent. When engaging in *intrinsic response-dependent regulation* one might engage in self-disclosure to a conversation partner when seeking feedback that will support their regulatory attempts, like seeking an emphatic response or confirmation [5, 52]. Previous research stresses that seeking support and concealment via disclosure can have positive effects on people's mood and helps them to cope with emotional events [e.g., 48, 50, 53]. When engaging in *intrinsic response-independent regulatio*n via self-disclosing to others, one will seek a channel for disclosure regardless of a potential response or feedback. Accordingly, the mere act of disclosure contains certain psychological components that can affect regulatory success. When sharing with others just for the sake of disclosing emotions and feelings, one might be engaged in appraising their own emotions and experiences and damping the intensity of the emotional experience [5]. This sort of strategy is also known as *affect labelling*, a simple emotional regulation technique aimed at explicitly expressing emotions, or in other words - putting feelings into words [54, 55, 56]. People use self-disclosure for emotional introspective processes, self-reflecting on their emotional experiences, as well as past behaviours and actions [57]. These sorts of self-disclosure behaviours are found to be highly useful for coping with emotional distress [54, 55, 56, 57, 58, 59] and is a meaningful act of mindfulness [60]. A similar example is James Pennebaker's *writing disclosure paradigm* [see 61, 62] which helps people to facilitate their emotions when writing about their own experiences. Pennebaker's paradigm [61, 62] has been validated and found to have short- and long-term effects, including reduced blood pressure, improved mood, and reduced depressive symptoms, as well as long-term positive physical outcomes such as improved memory, improved work performance, and more [63].

IV. When and why might we avoid self-disclosing to others?

Various socio-emotional factors might affect the extent to which people self-disclose, and even enhance self-disclosure avoidance. In organizational settings, for example, people tend to avoid self-disclosure and emotional expression in general for practical reasons, thinking that it might lead to a negative evaluation by colleagues with some potential to lose control over the situation [64, 65]. This reflects the general society-wide perspective of self-disclosure as a sign of weakness, exhibitionism, or mental illness, especially when these disclosures are of intimate content [66, 67]. Furthermore, while self-disclosure tends to facilitate relationships, in a variety of contexts (e.g., healthcare, psychotherapy, and within organisations) individuals might try to avoid the establishment of new intimate relationships with others by talking about themselves [66]. 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 [68]. 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 [69]. This can be evidenced when patients are asked to disclose information to healthcare providers such as medical doctors [70, 71, 72], 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 [73].

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

TABLE I
PSYCHOLOGICAL DRIVERS AND BARRIERS OF SELF-DISCLOSURE, AND THEIR IMPLICATIONS FOR HUMAN–ROBOT INTERACTION

Dimension	Main Insights from Psychological Theory	Implications for HRI
Why We Self-Disclose	**Emotional Relief**: Helps individuals cope with stress, trauma, or emotional challenges.	Robots can provide a non-judgmental outlet for sharing emotions, promoting emotional relief.
	Reciprocity: Mutual sharing deepens social bonds.	Robots may lack traditional reciprocal feedback, and it may be achieved through social exchange.
	Relationship Building: Fosters intimacy and trust between individuals.	Human–robot relationships could be established through acts of self-disclosure to robots, with the robots providing emotional benefits in a social exchange.
	Interpersonal Emotional Regulation: Helps individuals manage and process emotions.	Robots may assist in intrinsic emotion regulation, where the robot's feedback is not essential but its role as a listener can encourage introspection.
When We Avoid Self-Disclosure	**Fear of Judgment**: Individuals may avoid disclosure to avoid being judged negatively.	Robots are often perceived as non-judgmental, which can encourage more open self-disclosure.
	Lack of Reciprocity: Disclosure may be inhibited if the other party does not reciprocate the sharing.	Reciprocity may be perceived differently in interactions with robots, being more one-sided and achieved through a social exchange of emotional benefits.
	Stigma and Shame: Socio-emotional barriers, such as shame, may inhibit disclosure.	Robots could help mitigate shame and stigma-related barriers by providing a sense of confidentiality.
	Reading Expressed Emotions as Social Information: Initial emotional expressions in human interactions are often impulsive and provide social cues that may discourage individuals from self-disclosure.	Robots control their expressions via computing, mechanics, and design, ensuring consistent, neutral, or positive responses. This makes them less likely to exhibit impulsive emotional reactions, which may encourage self-disclosure.

The *Emotions as Social Information* (EASI) *Model* [74] suggests that *emotional expression* [i.e., verbal or non-verbal behaviour that communicates an emotional state or attitude and can be intentional or unconscious; 75] serves a social communication function. It proposes that emotions are not just personal experiences, but also convey information about the individual's internal state and intentions to others. The model proposes that emotions are universal and can be recognized by others through facial expressions, body language, and vocal cues, allowing for social interaction and coordination. More specifically, the model explains that emotional expression may affect the observers' behaviour by triggering inferential processes and/or affective reactions in them [76]. In the context of self-disclosure, 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. Hence, self-disclosure avoidance could be used as an emotion regulation technique for avoiding the emotional expression of others [77], especially when experiencing low mood [49] or feeling insecure [78]. Despite social norms of displaying affect [79], 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 [80]. Therefore, it could be that *robots* and *artificial agents* which are automated non-human entities that can control their expressions via computing, mechanics, and design, and are objectively perceived as objects [81], could avoid some of the socio-emotional barriers to self-disclosure.

V. WHY WOULD WE SELF DISCLOSE TO ARTIFICIAL AGENTS LIKE SOCIAL ROBOTS?

Beyond the positive benefits of self-disclosure to emotional well-being that were mentioned above [e.g., see 25], self-disclosure and verbal interpersonal communication, in general, are key features for the success of social robotic health interventions [82]. Emotional well-being interventions are dependent on open channels of communication where individuals can self-disclose their needs and emotions, from which a listener (i.e., a robot, for the scope of this review) can identify stressors and respond accordingly [83, 84]. This is particularly important for self-help autonomous systems like social robots, as human behaviour and emotions are analysed and synthesized by machines from human output, to respond and react appropriately by extracting salient information and identifying patterns and emotional states [85]. Nevertheless, engaging a robot in a reciprocal conversational interaction is a complex technical task, and when self-disclosing to social robots, human users would rarely experience an equilibrium of reciprocal disclosures [16, 24]. Following mind perception theory [see 86, 87] and the typical users' dissonance from social robots [the gap between users' high expectations, often inspired by robots' design or by science fiction, and the robots' actual, limited capabilities, which frequently leads to disappointment when robots fail to meet human-like social interaction expectations; see 10, 88], it can be assumed that the expectation for a reciprocal verbal engagement in HRIs when self-disclosing to a social robot might negatively affect people's experience, potentially limiting the depth and breadth of their disclosures and verbal engagement with the robot.

However, when engaging in self-disclosure towards social robots (and other artificial agents) it is theorised that people are more likely to embrace different benefits of this behaviour as

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

a form of social exchange [see 38, 42, 44] and might ignore the lack of traditional reciprocity that is expected between humans [16, 89]. Hence, reciprocity in this context is thought to take place as an act of exchange. Another similar (yet, more media-centric) theoretical approach for explaining users' willingness to self-disclose to social robots is the *uses and gratification theory* [90]. This theory posits that media users are not passive consumers, and instead turn to specific media according to the immediate gratification they receive from it. Thus, in the context of social robots, users might turn to social robots for a variety of rewards they might receive when self-disclosing to robots and other artificial agents. One example is online users' willingness to self-disclose personal information to artificial agents like chatbots, and other online algorithms in online marketing and e-commerce settings [91], to receive personal recommendations [92, 93]. However, similarly to self-disclosures between humans [see 25, 31], people might also engage in self-disclosure with artificial agents and social robots due to social and emotional reasons and not only for economic reasons. There is a substantial body of literature using embodied and disembodied artificial agents for eliciting self-disclosure in a variety of settings, reporting that self-disclosing to artificial agents positively affects people's feelings and emotional well-being [see reviews and meta-analysis 94, 95, 96, 97]. For example, in a recent study, 115 participants shared emotional experiences with an artificial agent who provided either emotional or cognitive support messages. The results of the study suggest that regardless of the type of support, self-disclosing emotions to the artificial agents fostered participants' emotional relief. After talking to the agents, participants felt better and expressed feeling closer to the agent and their desire to interact with it again [98]. Another study employed an emphatic disembodied conversational agent (i.e., a chatbot) to engage in verbal (text-based) interactions with 128 socially excluded participants, showing that interacting with the emphatic agent improved their mood [99].

Following the EASI model [74, 76], a number of socio-emotional factors exist for which people would prioritize engagement in self-disclosure to artificial agents and social robots that extend from the positive affect experienced when engaging in self-disclosure.

VI. Confidentiality as a self-disclosure facilitator

One such factor is *confidentiality* and the lack of judgment in interactions with artificial agents. Confidentiality is a substantial factor in this context as it is associated with increases in reporting disclosure in human-human self-disclosure [100], specifically in disclosures about sensitive matters, like emotional well-being and mental health [101]. Previous studies reported that people were more open to virtual agent interviewers than human interviewers in clinical interviews, demonstrating more willingness to disclose information about highly sensitive topics that can be associated with shame and stigma, or might just be considered sensitive [e.g., 102, 103]. For example, a study by Utami and colleagues [104] explored the reactions of older adults when having "end-of-life" conversations with a virtual agent. The study's results show that all study participants were comfortable discussing with the agent about death anxiety, last will and testament, providing compelling evidence for the potential utility of artificial agents in these complex socio-emotional domains. In a study by Lucas et al. [102], participants (N = 239) were led to believe that the artificial agent was controlled by a human or by automation during mental health-related interviews. Participants who were led to believe that they were talking with an automated agent (compared to an agent operated by a human) reported lower fear of self-disclosure, lower impression management, displayed their sadness more intensely and were rated by observers as more willing to disclose. In another study, 203 students rated the sensitivity of different interview topics and indicated their preferences to disclose sensitive and personal information to a human or to an artificial agent. The study reports that there is a preference to disclose to an artificial agent when topics are more sensitive and are likely to evoke negative self-admissions. More specifically, participants stated that they would feel more comfortable discussing sensitive topics with an artificial agent because it could not judge them [103]. An additional study provided supporting evidence for it, showing that when engaging in mental health interviews, a sample of 55 students disclosed more sex-related symptoms to an artificial agent rather than to a real human expert [105]. A recently conducted study with 22 participants further supports this with objective evidence and reports preliminary results stating that despite self-disclosing more (in terms of quantities) to an artificial agent that was introduced as a human (compared to an artificial agent that was introduced as a machine), the disclosures to the agent that was introduced as a machine were significantly more sentimental, and the agent was found to be perceived as friendlier [106].

VII. Rapport as a self-disclosure facilitator

Nevertheless, beyond supporting the feeling of confidentiality, artificial agents can build a sense of *rapport* by displaying (verbal and nonverbal) social cues of mutual liking, approval, attentiveness and coordination in their communication [107] for engaging users in a more interactive dialogue [108]. For example, a study by Lucas and colleagues [109] employed a virtual agent that affords confidentiality while building rapport to interview active-duty service members about their mental health symptoms when returning from a year-long deployment in Afghanistan. The study results show that participants disclosed more symptoms to a virtual agent interviewer than on the official Post-Deployment Health Assessment (PDHA), and then on an anonymized PDHA. The results of a larger sample experiment with active-duty and former service members reported a similar effect. Another early study by Bickmore, Gruber, and Picard [110] showed in a longitudinal experiment with 33 young adults, that when the artificial agent showed more "relational" skills (behaviours that build and maintain good working relationships over multiple interactions like showing empathy, engaging in more social dialogue, and showing nonverbal immediacy behaviours) participants showed a significant increase in their will to communicate

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

with the agent over time. Interestingly, even when using subtle cues to build rapport it seems to have a meaningful impact. A recent study with 40 participants shows that when artificial agents (voice assistants in this specific study, as the researchers used Amazon Alexa) are using subtle backchanneling cues (e.g., "aha", "go on", "ahm", "I see") it improves human users' perceived degree of active listening, and results in more emotional disclosure (i.e., participants using more positive words in their disclosures) [111].

VIII. Confidentiality and rapport in interactions with social robots

Like other artificial agents (e.g., virtual humans, embodied and disembodied artificial agents), we have some empirical evidence for self-disclosures to social robots following similar principles. There are several studies addressing self-disclosure in child-robot interaction [e.g., 112, 113], but since this paper is focused on adults' interactions with social robots, we will not address these studies here. In terms of rapport, a study by Nakamura and Umemuro [114] found that people might self-disclose more, and their self-esteem grows when self-disclosing to a robot that changes their listening attitude. This is also evidenced in people's perceptions of social robots' communication style in speech interactions, with participants (42 healthy adults) rating a robot that used a human-like communication style as more competent, warmer, and less discomforting, compared to robots employing machine-like communication style [115]. In a previous study, researchers discovered that individuals tend to self disclose more (in terms of the duration of the disclosure and the number of words used) with other humans compared to a humanoid social robot (Nao, by SoftBank Robotics) or a disembodied conversational agent (Google tab voice assistant). However, an interesting finding was that when speaking to the humanoid robot, people tended to mimic its voice by using a higher pitch tone, unlike their interaction with the human agent or the disembodied conversational agent. This suggests that individuals adjust their behaviour when self disclosing to robots, possibly due to the robot's embodiment and its ability to establish a rapport [116]. In a similar experiment involving four sessions of technology-assisted journal writing, participants guided by a social robot, compared to those using a voice assistant, reported higher levels of self-disclosure over time. This was particularly evident in their awareness of thoughts and feelings, suggesting that the rapport established by social robots positively influences the depth and richness of self-disclosure [117]. A recent study suggests that rapport can also be experienced via disclosure reciprocity, as the study found that reciprocal self-disclosure from the robot increased liking in intimate self-disclosure. Nevertheless, the results of the study also report that reciprocal self-disclosure in non-intimate self-disclosure resulted in decreased rates of liking the robot [118]. Similar evidence for the positive role of rapport is evidenced in a study by Duan et al. [119] showing that people in a more negative mood were more likely to benefit from self-disclosing to a robot compared to participating in a writing disclosure to a journal.

An extension to the study address also the gratification of confidentiality when self-disclosing to robots, showing that self-disclosing to a robot was also more effective for those in a negative mood than self-disclosing on social media [120]. Early work by Bartneck et al. [121] shows that in a sample of 44 students, participants were less embarrassed when interacting with a "technical box" than with a social robot "iCat" (Phillips) in medical settings asking participants to undress and disclose relevant personal information (e.g., their weight). Interestingly, the lack of embodiment (of the technical box) made the participants feel less embarrassed in a vulnerable situation. In a recent study, 21 individuals with ASD were requested to answer 10 personal questions asked by three different agents – an android robot (a social robot with a realistic human appearance), a human interviewer, and a written passage on testing paper. The results of the study also highlight the positive role of confidentiality in disclosures to social robots, with the android robot promoting more self-disclosure, especially about the negative topic compared to the human interviewer and the written passage (which also highlights the role of rapport) [122]. The role of confidentiality is also present to a certain extent in a study with participants reporting to benefit the most from disclosing to the social robot Pepper (SoftBank) about their attitudes and opinions, compared to a number of other topics which are less sensitive or personal (e.g., work or study, tastes and interests). Another example of the role of confidentiality when self-disclosing to robots is a study by Nomura and colleagues [123] that provides evidence for the benefits of employing social robots for minimising social tensions and anxieties. The study found that participants with higher social anxiety tended to feel less anxious and demonstrate lower tensions when knowing that they would interact with robots in opposition to humans in a service interaction. In addition, the study suggests that an interaction with a robot elicited lower tensions compared to an interaction with a human agent, regardless of one's level of social anxiety. Therefore, the level of participant embarrassment in response to the android robot seemed to be lower compared to that in the human interviewer condition.

IX. Self Disclosure to Social Robots for supporting Well-being

Like the study by Nomura et al. [123], several other studies used self-disclosure with a social robot as a therapeutic activity. A previous study employed the social robot Sota to perform a conversational stress-coping intervention aimed at encouraging participants (31 adults) to self-disclose to the robot about their worries. Accordingly, the robot was programmed to further ask about the problem presented by the participant to encourage them to self-reflect about it and provoke some emotional response. The study found that self-disclosing to the robot positively affected participants' moods and reduced their anger [124]. Another study employed a non-humanoid social robot acting in a responsive way to human users' self-disclosures in two experiments. In the first experiment with 102 participants, they found that the robot's responsiveness increased the willingness to use it as a companion in stressful

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

situations, and in the second experiment with 74 participants, they found that interacting with a responsive robot improved self-perceptions during a stress-generating task [125].

Examining the effects of self-disclosure on emotional well-being in repeated interactions (including long-term interventions) is crucial, and it shouldn't be limited to the analysis of single interactions alone. A study by Dino et al. [126] employed conversational-based cognitive behavioural therapy (CBT) using a social robot ("Rayen") for older adults (N = 4), meeting the robot twice a week for about an hour for four weeks. The results demonstrate that the individual subjects progressed through the sessions, their average sentence length increased, sharing more positive words, reporting for a more positive mood and some improvements in mental health symptoms. Overall, participants reported being satisfied with verbally interacting and self-disclosing to the robot. It is important to consider that these results are limited due to the restrictive sample size and are mostly an indication of the usability of the developed system reported in the paper. 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 general everyday experiences. The study found that participants self-disclosed more to the robot as the sessions progressed, perceiving the robot to be more socially competent and comforting over time. The repeated interactions also led to improved mood (after each session, and over time) and decreased feelings of loneliness [127]. The study's experimental design was replicated with a sample of informal caregivers [128, 129], who often experience high levels of emotional distress [130]. The findings of the study replicated the previous results [127] 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 [129]. 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.

It is of note that several studies report that the effects of self-disclosure are conditioned to different factors, like personalities, psychological tendencies, or emotional states. For example, one study with 81 participants found that participants with a higher tendency to anthropomorphise attributed higher levels of mind to the social robot Nao (SoftBank Robotics) in self-disclosure interactions [131]. A 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 a robot [132]. Another 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) [133]. Nevertheless, a secondary analysis of [127] entails that people who report for higher scores of introversion, as well as when experiencing negative emotions, are more likely to disclose more to a social robot (i.e., in terms of the disclosures duration and length) [134]. In terms of gender, a study found individual differences in self-disclosure patterns, with Japanese men preferring to disclose more to humans than robots, whereas women showed no significant difference in their willingness to disclose to either humans or robots, highlighting gender-based variations in emotional openness with robotic agents [135].

X. Safety and Ethical Considerations

The introduction of social robots in social settings, particularly in interactions that involve self-disclosures, raises important safety and ethical considerations. As social robots are gradually being integrated into various social and health contexts, it is crucial to carefully examine the potential downsides and address concerns related to privacy, trust, and the preservation of human connection.

Privacy is a primary concern when individuals engage in self-disclosure interactions with social robots [see 136, 137]. These interactions often involve sharing personal and sensitive information. It is imperative to ensure that the data collected by social robots during these exchanges are handled securely and confidentially. Implementing robust data encryption, storage protocols, and access controls is essential to safeguard users' privacy and prevent unauthorized access to personal information. Additionally, clear guidelines and regulations need to be established to govern the use, storage, and protection of such data, taking into account legal and ethical considerations.

When considering situations where sensitive information disclosed to robots may need to be shared with external parties, such as revelations of abuse or suicidal ideation, ethical challenges revolve around maintaining user trust while fulfilling obligations to intervene in critical situations. As Dietrich et al. [138] emphasise, robots must adopt adaptive privacy management strategies to ensure that sensitive disclosures are handled according to ethical guidelines. These strategies could involve tiered escalation protocols, where information is escalated only when safety concerns arise. Comparing this to Petronio's Communication Privacy Management (CPM) theory, which assumes that individuals calculate the perceived costs and benefits of disclosing private information [139, 140], we see both a similarity and distinction: CPM assumes users make rational assessments before sharing sensitive information. In contrast, within the interpersonal dynamics of sharing with robots, we assume that when people have emotional needs that can be fulfilled by a robot, they may overcome privacy considerations, potentially ignoring the consequences in favour of receiving emotional support.

We can gain further insights into the complexity of using robots as both communication partners and mediators from other previous studies. As one example, Levinson et al. [141] highlight that privacy concerns are relationally situated, with users (especially adolescents) expressing heightened sensitivity to who accesses their data. This emphasises the need

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

TABLE II
KEY INSIGHTS AND EVIDENCE ON SELF-DISCLOSURE BEHAVIOURS AND THEIR EFFECTS IN HRI RESEARCH

Concept	Insights	Evidence from Research
Confidentiality as a Driver for Self-Disclosure	• Users disclose more to robots due to perceived confidentiality, especially on sensitive topics. • Reduced perceived judgment in robot interactions increases willingness to disclose.	See references [120, 121, 122, 123, 133].
Rapport as a Catalyst for Self-Disclosure	• Social cues from robots foster rapport, increasing disclosure willingness. • Using relational skills over multiple interactions increases users' desire to disclose.	See references [114, 115, 116, 117, 118, 119, 127].
Self-Disclosure for Emotional Support	• Self-disclosure to social robots offers emotional support and helps reduce stress through emotional relief. • Repeated self-disclosure to robots over time improves emotional well-being and helps with coping during distress.	See references [123, 124, 125, 126, 127, 128, 129].
Individual & Subjective Factors Impacting Self-Disclosure	• Self-disclosure may vary based on factors such as personality traits and emotional states.	See references [131, 132, 133, 134, 135].

for transparency and user control over how robots manage disclosed information. As Erel et al. [142] point out, a robot's social influence is shaped by how well it meets user expectations. When robots exceed these expectations—particularly in handling sensitive information—the trust and influence they exert can be amplified, making clear ethical management essential. Gillet et al. [143] add another dimension with their Interaction-Shaping Robotics (ISR) framework, where robots act as mediators in human relationships. In this role, robots may inadvertently influence how sensitive information is shared, reinforcing the need for ethical protocols to prevent unintended disclosures in group and triadic settings.

Given these insights, a pressing need is emerging for multi-user privacy protocols in shared environments [see 144], especially when a robot's mediation might lead to unintended disclosures. Robots should be equipped with adaptive privacy frameworks that balance confidentiality and privacy with ethical responsibilities, ensuring they act not only as facilitators of self-disclosure, but also as guardians of sensitive information and conduits for further (human) intervention when necessary. This approach should include setting clear boundaries and establishing transparent communication with users, and helping to navigate the ethical complexities of robot-mediated disclosures with care and accountability. It is also important to recognise that robots are merely objects that function using algorithms, and granting them the responsibility to make decisions about protecting individuals based solely on their disclosures opens up a much bigger and thornier moral, legal, and philosophical debate that extends beyond the scope of this paper, but certainly warrants dedicated discussion [see 145, 146, 147, 148].

Another ethical consideration is the development of artificial trust and reliance on social robots when self disclosing to them. While social robots can exhibit human-like behaviours and establish rapport with users, it is important to acknowledge that they are still machines and lack genuine emotions and empathy [81]. Overreliance on social robots for emotional support via self-disclosure and interpersonal interactions may lead individuals to neglect or undervalue human connections. Therefore, it is crucial to emphasize that social robots should be seen as complementary tools rather than substitutes for human interaction. Consequently, it is necessary to give more thought to the role of social robots as companions. Future studies should focus on exploring and examining how social robots can improve, foster, and facilitate social connections, not only with robots but with other humans and social agents (e.g., pets) for improving emotional well-being. This approach should move beyond merely providing artificial attachment, which could potentially evoke negative emotional consequences overtime.

Transparent communication with users is also an ethical imperative in the introduction of social robots in social settings, especially when sharing thoughts and feelings with these agents. The studies discussed in this review highlight the importance of clear communication and users' understanding of the capabilities and limitations of social robots. This also reflects on the methodology employed in HRI research. Openly discussing the purpose of data collection, the algorithms used, and the decision-making processes of social robots can foster trust and mitigate concerns related to the ethical use of this emerging technology. Transparent communication helps manage user expectations, prevents potential misunderstandings, and avoids the formation of false beliefs about the nature of the interaction [see 149]. Additionally, the potential for unintended consequences and biases in interactions with social robots must be carefully examined [150]. It is important to address biases in the design and deployment of social robots, ensuring that they promote inclusively, diversity, and equal treatment of all users. Thorough research and testing are necessary to identify and address any biases that may arise from the

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

deployment and design of social robots, especially when these are aimed at simulating and capturing affect via self disclosure.

XI. The downsides of social interactions with social robots

Although the studies addressed in this review present compelling evidence regarding the potential of social robots to support emotional well-being by facilitating self-disclosure, it is important to acknowledge the downsides of relying solely on robots in social settings for such engagements.

One limitation of social robots is their limited ability to understand and respond to the nuances of human emotions [10]. Human emotions are often complex and multifaceted, requiring empathy, intuition, and contextual understanding to be effectively addressed [see 151]. Social robots may struggle to provide the same level of emotional support, empathy, and understanding that humans can offer [152]. Thus, there is a risk of individuals receiving superficial or inadequate emotional support from social robots, which may not fully address their needs. Relying solely on social robots for self-disclosure interactions may inadvertently isolate individuals from genuine human connections. Previous studies show that individuals in emotional need (e.g., higher rates of loneliness [134], social anxiety [123], introversion [134], and negative emotions in general [119, 134]) show higher tendency to share and confide in robots. While social robots can answer some of the needs for social connection and provide a sense of comfort and companionship, they cannot replace the depth of emotional connection and understanding that human relationships can offer [153]. Overreliance on social robots may lead to a reduction in meaningful human interactions, potentially resulting in feelings of loneliness, social disconnection, and a decline in overall well-being. Thus, it is advisable to consider safe and regulated deployment of social robots for interpersonal communication and prioritize the implementation of socially assistive robots in public settings (such as community centres, local care homes, schools, universities, etc.) rather than introducing them directly into users' homes to prevent unregulated emotional attachment. Furthermore, future studies are encouraged to empirically investigate relationships with social robots in longer periods of time, to have a better understanding of the long-term emotional impact of interpersonal interactions with these agents. Additionally, it should be considered that the introduction of social robots in social settings for self-disclosure should not contribute to *dehumanization* [see 154]. Human touch, nonverbal cues, and the presence of another person can convey a sense of warmth, empathy, and understanding that may be challenging for social robots to replicate [10, 155, 156, 157]. The absence of these human elements in interactions may lead to a sense of detachment or impersonal experiences, particularly in vulnerable individuals who require genuine human connection.

Another concern is that social robots might be utilized for eliciting information and even persuade people who are vulnerable or that are in endangered circumstances. Vulnerable populations, including children, and individuals with mental health conditions, disabilities, or trauma histories, require special attention to ensure their well-being, safety, and protection. For example, previous studies found that children, both preschoolers and older children, were willing to share sensitive information with humanoid robots. In one study, preschool children were as comfortable sharing a secret with a robot as they were with an adult [112], while in another study, older children showed few differences in reporting bullying incidents between human and robotic interviewers [158]. Other studies show that robots preformed better than traditional screening instruments when eliciting information from children about their mental health [159]. However, relying on robots for children's self-disclosure presents ethical concerns, as discussed previously, robots lack genuine empathy and understanding. There is also a risk of children disclosing personal information without fully grasping the consequences or having adequate privacy protection. While social robots may offer support and a comfortable space for self-disclosure, it is essential to address the potential downsides and ethical implications. These include the limitations of social robots in understanding the complex emotional needs of vulnerable individuals, the risk of excessive reliance on robotic support without fostering genuine human connections, and the potential for privacy breaches or exploitation. Ethical considerations involve obtaining informed consent, ensuring privacy and confidentiality, providing culturally sensitive and inclusive interactions, addressing power dynamics, and avoiding harm or discrimination. Moreover, engaging multidisciplinary teams of healthcare professionals, psychologists, educators, and representatives from relevant communities is crucial for navigating the responsible and ethical use of social robots in self-disclosure interactions with vulnerable populations. Therefore, ethical considerations arise when employing such technology for eliciting information in these settings, questioning whether using autonomous agents like social robots for eliciting information aligns with our current moral standards.

XII. The role of culture in self-disclosure to social robots

Culture plays a significant role in shaping communication styles, social norms, and expectations regarding self-disclosure. Different cultures may have varying levels of comfort and willingness to disclose personal information to others [160, 161], including social robots [see 162]. Cross-cultural studies have shown that individuals from collectivist cultures, such as East Asian cultures, tend to exhibit lower levels of self-disclosure compared to those from individualistic cultures, such as Western cultures [163]. Consequently, the way individuals from different cultural backgrounds engage in self-disclosure to social robots may vary. Therefore, it is important to acknowledge that self-disclosure behaviours in interactions with social robots might differ, depending also on cultural contexts and norms [see 161, 164]. Future research should consider conducting comparative studies across different cultures and languages to gain a comprehensive understanding of the cultural nuances and their impact on self-disclosure to social robots.

Moreover, language plays a vital role in communication, affecting the nature and depth of self-disclosure [165]. Self-

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

TABLE III
ETHICAL, PRACTICAL, AND CULTURAL CONSIDERATIONS IN HUMAN-ROBOT SELF-DISCLOSURE

Consideration	Key Points	Implications for HRI
Ethical and Safety Considerations	**Privacy Concerns**: Sharing personal information with robots may lead to data privacy risks.	Robots must incorporate robust data protection and encryption protocols to ensure user safety and trust.
	Sharing Sensitive Information: Ethical obligations may require sharing sensitive disclosures with external parties (e.g., revelations of abuse or suicidal ideation).	Robots should include tiered privacy management strategies to balance user confidentiality with safety, escalating information only in cases of critical concern.
	Artificial Trust: Overreliance on robots for emotional support could lead to diminished human interactions.	Robots should be used as complementary tools, not substitutes for human connections.
	Transparency: Users should clearly understand robots' limitations and capabilities.	Robots need to communicate their functionalities clearly to manage user expectations and mitigate ethical concerns.
The Downsides of Social Interactions with Social Robots	**Limited Emotional Understanding**: Social robots may struggle to appropriately process complex emotions and provide adequate support.	Robots should be clear about their limitations in emotional intelligence to prevent unmet expectations and emotional disengagement.
	Risk of Isolation: Overreliance on robots could reduce meaningful human interactions, leading to social disconnection.	Robots should facilitate and not replace human interactions, especially in long-term emotional engagements. Deployment in public settings should be prioritised over intimate settings, where attachment and dependency are more likely to occur.
	Potential for Over-Sharing: Vulnerable populations may disclose too much personal information due to perceived safety.	Safeguards must be put in place to protect users from over-sharing sensitive information and ensure appropriate use of disclosures.
The Role of Culture in Self-Disclosure	**Cultural Norms and Comfort Levels**: Individuals from collectivist cultures tend to disclose less compared to those from individualist cultures.	Researchers and practitioners should take cultural habits into account for the successful deployment of robots in self-disclosure contexts.
	Impact of Language: Language proficiency and linguistic nuances can shape the depth of self-disclosure.	Robots should provide multi-language support and recognize linguistic nuances to enhance self-disclosure across languages.
Implications of AI Advancements on Self-Disclosure Dynamics	**Enhanced Conversational Adaptivity**: LLM-powered robots can respond more fluidly, encouraging richer self-disclosures.	LLMs offer more personalized communication, but robots must balance adaptivity with maintaining user trust in confidentiality.
	Balance Between Rapport and Confidentiality: The more responsive and contextually adaptive the robot becomes, the more it risks being perceived as intrusively aware, potentially diminishing the sense of confidentiality.	Maintaining a balance between the conversational rapport achieved by LLMs and preserving the sense of confidentiality is crucial as robots become more adaptive.
	Impact on Emotional Complexity: LLM-powered robots may diminish the emotional depth of interactions due to pragmatic responses.	Robots should be designed to simulate empathetic, context-aware responses without compromising the emotional depth of disclosures.
	Impact on Non-Verbal Communication: As users might anticipate more sophisticated non-verbal cues due to advanced speech capabilities, failure to meet aligned expectations can reduce their willingness to disclose personal information.	Designing robots with adaptive non-verbal communication skills, such as gaze and gestures that match and align with advanced speech capabilities, is crucial to foster self-disclosure without causing frustration.

disclosure patterns of research participants may be influenced by their cultural and linguistic backgrounds, as language proficiency and linguistic nuances impact the extent to which individuals are comfortable expressing their emotions and personal experiences. Language can shape the availability of words and expressions to convey specific emotions or experiences, potentially influencing the depth and breadth of self-disclosure [165, 166].

Considering these factors, it is necessary to recognize the challenges in generalizing findings of self disclosure research in HRI settings beyond their cultural and linguistic contexts. Researchers should aim to include participants from diverse cultural backgrounds, with varying languages, to investigate how cultural norms and linguistic factors influence self-disclosure to social robots. While research samples are often limited, researchers studying self disclosure with social robots should aspire to replicate their research, and reproduce it in variety of cultural settings and conducting interaction encompassing several languages. By conducting cross-cultural studies and including participants from different cultural backgrounds and languages, researchers can deepen our understanding of how cultural and linguistic factors shape self-disclosure to social robots. This will enable us to develop social robots that are culturally sensitive and adaptable, facilitating effective and meaningful interactions across diverse populations.

XIII. IMPLICATIONS OF AI ADVANCEMENTS ON SELF-DISCLOSURE DYNAMICS IN HRI

Recent advancements in artificial intelligence, particularly in large language models (LLMs), have enhanced robots' capacity for contingent linguistic and non-linguistic responses [167]. These capabilities have transformed the way robots engage in conversations, allowing for potentially more fluid, empathetic, and context-aware exchanges during HRIs. In particular, LLMs can contribute to self-disclosure by enabling robots to offer responses that are more emotionally attuned and context-sensitive, encouraging users to share more openly and

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

richly with adaptive and personalised communication. Thus, with a new generation of increasingly adaptive and responsive robotic interactions and architectures, we should expect to see more (and increasingly successful) deployment of robots for facilitating self-disclosure, responding appropriately, and fulfilling many of promises communicated in this paper (which mostly focuses on research performed with robotic agents that lack sophisticated generative AI).

However, we should also consider the potential challenges that these advances may introduce when it comes to maintaining the balance between rapport and confidentiality in self-disclosure. As LLMs make robots increasingly conversationally adaptive, the delicate balance between rapport and confidentiality may be disrupted. The more responsive and contextually adaptive the robot becomes, the more it risks being perceived as intrusively aware, potentially diminishing the sense of confidentiality that is required for fostering open disclosure. This heightened awareness and adaptivity may prompt users to question how their disclosures are being processed or scrutinised, leading to more guarded interactions. A significant concern is whether this increased conversational adaptability might reduce the richness of self-disclosures. On one hand, LLM-powered robots may now handle context so well that users feel less need to explain their emotions in depth, which could limit the complexity and introspection of their self- disclosures. In contrast, the enhanced adaptivity may also foster a sense of being understood, which could encourage users to share more openly and deeply, trusting that the robot can handle the emotional complexity of what they are disclosing.

The increased adaptability of robots powered by LLMs, while allowing for personalised interaction, might also raise questions about how users perceive the robot. A study by Spitale et al. [168] highlights that users felt that robots could effectively adapt to their input, but there was also a tendency for users to feel that they were adapting to the robot in return. This raises an important question about whether greater robot adaptivity could limit users' natural self-disclosure, as they may feel less need to "*mentalise*" or project social attributes onto the robot. Without the cognitive effort involved in engaging deeply with a less adaptable (or intelligent) machine, users' engagements might lack the cognitive reconstruction taking place when they must actively interpret and respond to the robot's limited verbal correspondence. Moreover, participants in a recent study [169] have indicated that while LLM-powered robots excel in building connections and handling deliberative tasks, their responses can sometimes appear too verbose or overly pragmatic, thus disrupting the fluidity of interaction. This overly pragmatic nature of responses might reduce the perceived emotional depth needed for meaningful self- disclosure, as robots risk coming across as too factual and less empathetic. This lack of perceived empathy may detract from the robot's ability to foster deep emotional engagement. Furthermore, participants in the same study [169] addressed their concerns about potential biases and stereotyping in the robot's adaptive responses. Users might perceive this adaptivity as a double-edged sword—creating deeper engagement but at the cost of increased concern about the security and neutrality of their interactions.

Since robots' speech capabilities are advancing substantially due to the adaptation of LLMs, users may have higher expectations in terms of robots' non-verbal communication capabilities. A study by Kim et al. [170] found that users expect more sophisticated non-verbal cues (such as gaze, gestures, and body language) to accompany the robots' enhanced conversational abilities. Any lack of such non-verbal responses can result in user dissatisfaction or a reluctance to disclose personal information. This mismatch between expectations and reality could hinder engagement, leading to frustration and a breakdown in the trust essential for self-disclosure. In addition, when the robot's speech capabilities falter or become illogical due to natural errors with LLMs, this has the potential to cause anxiety and confusion [170].

In summary, the introduction of LLMs in robotic speech is continuing to introduce many new opportunities and challenges for future research regarding self-disclosure and the sharing of emotions during HRIs. While these technical advances hold promise for supporting deeper, more empathetic exchanges, they also present challenges in balancing the need for rapport with the preservation of users' sense of confidentiality. As robots become more adaptive, their role in fostering emotional expression may shift, with potential impacts on how much users are willing to disclose and how deeply they engage with robotic agents.

XIV. Conclusions

The exploration of self-disclosure dynamics in interactions with social robots illuminates a promising avenue for augmenting human emotional well-being amidst socio-emotional challenges such as shame and stigma. The propensity for individuals to seek out social robots for self-disclosure, driven by the perceived confidentiality and reduced fear of judgment, suggests a unique therapeutic potential in these interactions. The insights discussed in this paper, rooted in theory and supported by empirical evidence, underscores the complexity of human–robot communication, revealing that while confidentiality fosters openness, the cultivation of rapport is equally critical for meaningful exchanges with social robots. Therefore, the nuanced interplay between perceived confidentiality and the simulation of social cues by robots creates a fertile ground for richer, more gratifying self-disclosures, contrasting with human interactions that might be hindered by social apprehensions. This delicate balance, wherein individuals perceive robots as non-judgmental yet socially responsive entities, paves the way for discussions on sensitive matters with minimal fear of reprisal, aligning with the theoretical propositions of social exchange theory [16, 24, 89].

Importantly, the empirical evidence addressed in this review paper also suggest that the benefits of self-disclosure to social robots extend beyond mere relief, offering avenues for emotional well-being interventions. The capacity of social robots to simulate understanding and provide feedback can evoke feelings of being heard and understood, essential components of emotional support and emotion regulation. Thus, the integration of social robots into emotional well being and health intervention frameworks emerges as a compelling proposition.

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

In the future, as AI continues to advance, LLM-powered robots have the potential to significantly enhance self-disclosure by offering more adaptive, context-sensitive, and emotionally attuned responses. However, to fully leverage the benefits of LLMs in fostering deeper engagement, it will be essential to carefully balance the development of rapport with ensuring users' confidence in the confidentiality of their disclosures. Looking forward, the potential of social robots in facilitating self-disclosure invites a broader consideration of their roles within societal and ethical boundaries. As we navigate the evolving landscape of HRIs, the implications for privacy, autonomy, and the essence of human relationships warrant thorough examination. Additionally, the influence of cultural norms and individual differences on self-disclosure patterns to robots offers a rich domain for future research, promising insights into the global applicability and customization of robotic interventions.

In synthesizing these perspectives, our discussion here highlights a critical insight: the interaction with social robots, when navigated with sensitivity to human emotions and ethical standards, holds significant promise for supporting individuals in navigating the complexities of their emotional lives. As we continue to study the intricacies of these interactions, the horizon of possibilities for enhancing human well-being through the judicious application of social robots broadens, marking an exciting frontier in the intersection of technology, psychology, and social science.

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] C. R. Berger, "Interpersonal Communication: Theoretical Perspectives, Future Prospects," *Journal of Communication*, vol. 55, no. 3, pp. 415–447, 9 2005. [Online]. Available:

[2] A. L. Barthel, A. Hay, S. N. Doan, and S. G. Hofmann, "Interpersonal Emotion Regulation: A Review of Social and Developmental Components," *Behaviour Change*, vol. 35, no. 4, pp. 203–216, 12 2018.

[3] J. A. Coan, "The Social Regulation of Emotion," in *The Oxford Handbook of Social Neuroscience*, J. Decety and J. T. Cacioppo, Eds. Oxford University Press, 9 2012, pp. 615–623. [Online]. Available:

[4] B. Rimé, "Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review," *Emotion Review*, vol. 1, no. 1, pp. 60–85, 1 2009. [Online]. Available:

[5] J. Zaki and C. W. Williams, "Interpersonal emotion regulation," *Emotion*, vol. 13, no. 5, pp. 803–810, 10 2013. [Online]. Available:

[6] S. G. Hofmann and S. N. Doan, *The social foundations of emotion: Developmental, cultural, and clinical dimensions.* American Psychological Association, 7 2018.

[7] C. Segrin, "Communication and Personal Well-Being," *Encyclopedia of Quality of Life and Well-Being Research*, pp. 1013–1017, 2014. [Online]. Available:

[8] C. Segrin and M. Taylor, "Positive interpersonal relationships mediate the association between social skills and psychological well-being," *Personality and Individual Differences*, vol. 43, no. 4, pp. 637–646, 9 2007.

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

[10] 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. [Online]. Available:

[11] C. Wohlin, M. Kalinowski, K. Romero Felizardo, and E. Mendes, "Successful combination of database search and snowballing for identification of primary studies in systematic literature studies," *Information and Software Technology*, vol. 147, p. 106908, 7 2022.

[12] Adrian Sayers, "Tips and tricks in performing a systematic review," *British Journal of General Practice*, 2007.

[13] 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.

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

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

[16] V. J. Derlega, M. S. Harris, and A. L. Chaikin, "Self-disclosure reciprocity, liking and the deviant," *Journal of Experimental Social Psychology*, vol. 9, no. 4, pp. 277–284, 7 1973.

[17] S. L. Gable, H. T. Reis, E. A. Impett, and E. R. Asher, "What Do You Do When Things Go Right? The Intrapersonal and Interpersonal Benefits of Sharing Positive Events," *Journal of personality and social psychology*, vol. 87, no. 2, pp. 228–245, 2004.

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

[19] 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. [Online]. Available:

[20] 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. [Online]. Available:

[21] L. A. Hosman, "The evaluational consequences of topic reciprocity and self-disclosure reciprocity," *Communication Monographs*, vol. 54, no. 4, pp. 420–435, 1987.

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

[22] M. Argyle and J. Dean, "Eye-contact, distance and affiliation," *Sociometry*, vol. 28, no. 3, pp. 289–304, 1965.

[23] I. J. Firestone, "Reconciling verbal and nonverbal models of dyadic communication," *Environmental psychology and nonverbal behavior*, vol. 2, no. 1, pp. 30–44, 1977.

[24] R. L. Archer and J. H. Berg, "Disclosure reciprocity and its limits: A reactance analysis," *Journal of Experimental Social Psychology*, vol. 14, no. 6, pp. 527–540, 11 1978.

[25] J. P. Forgas, "Affective Influences on Self-Disclosure: Mood Effects on the Intimacy and Reciprocity of Disclosing Personal Information," *Journal of Personality and Social Psychology*, vol. 100, no. 3, pp. 449–461, 3 2011. [Online]. Available:

[26] C. A. Langston, "Capitalizing On and Coping With Daily-Life Events: Expressive Responses to Positive Events," *Journal of Personality and Social Psychology*, vol. 67, no. 6, pp. 1112–1125, 1994. [Online]. Available:

[27] S. L. Gable and H. T. Reis, "Good News! Capitalizing on Positive Events in an Interpersonal Context," *Advances in Experimental Social Psychology*, vol. 42, pp. 195–257, 1 2010.

[28] S. Donato, A. Pagani, M. Parise, A. Bertoni, and R. Iafrate, "The Capitalization Process in Stable Couple Relationships: Intrapersonal and Interpersonal Benefits," *Procedia - Social and Behavioral Sciences*, vol. 140, pp. 207–211, 8 2014.

[29] A. K. Otto, J. P. Laurenceau, S. D. Siegel, and A. J. Belcher, "Capitalizing on everyday positive events uniquely predicts daily intimacy and well-being in couples coping with breast cancer," *Journal of Family Psychology*, vol. 29, no. 1, pp. 69–79, 2015.

[30] H. T. Reis, S. M. Smith, C. L. Carmichael, P. A. Caprariello, F. F. Tsai, A. Rodrigues, and M. R. Maniaci, "Are you happy for me? How sharing positive events with others provides personal and interpersonal benefits," *Journal of Personality and Social Psychology*, vol. 99, no. 2, pp. 311–329, 8 2010.

[31] E. Ignatius and M. Kokkonen, "Factors contributing to verbal self-disclosure," *Nordic Psychology*, vol. 59, no. 4, pp. 362–391, 1 2012. [Online]. Available:

[32] B. Rimé, C. Finkenauer, O. Luminet, E. Zech, and P. Philippot, "Social Sharing of Emotion: New Evidence and New Questions," *European Review of Social Psychology*, vol. 9, no. 1, pp. 145–189, 1 1998. [Online]. Available:

[33] A. J. Rose, "Co–Rumination in the Friendships of Girls and Boys," *Child Development*, vol. 73, no. 6, pp. 1830–1843, 11 2002. [Online]. Available:

[34] E. Zech, B. Rimé, and F. Nils, "Social sharing of emotion, emotional recovery, and interpersonal aspects." in *The regulation of emotion.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2004, pp. 157–185.

[35] L. Carlucci, I. D'Ambrosio, M. Innamorati, A. Saggino, and M. Balsamo, "Co-rumination, anxiety, and maladaptive cognitive schemas: when friendship can hurt," *Psychology Research and Behavior Management*, vol. 11, p. 133, 2018. [Online]. Available:

[36] E. Zech and B. Rimé, "Is talking about an emotional experience helpful? Effects on emotional recovery and perceived benefits," *Clinical Psychology and Psychotherapy*, vol. 12, no. 4, pp. 270–287, 7 2005.

[37] D. L. Haggard, C. Robert, and A. J. Rose, "Co-Rumination in the Workplace: Adjustment Trade-offs for Men and Women Who Engage in Excessive Discussions of Workplace Problems," *Journal of Business and Psychology*, vol. 26, no. 1, pp. 27–40, 3 2011. [Online]. Available:

[38] G. C. Homans, "Social Behavior as Exchange," *https://doi.org/10.1086/222355*, vol. 63, no. 6, pp. 597–606, 5 1958. [Online]. Available:

[39] ——, *Social behavior: Its elementary forms*. Oxford, England: Harcourt, Brace, 1961.

[40] P. Ekeh, *Social Exchange Theory: The Two Traditions*. Cambridge, Massachusetts: Harvard University Press, 1974.

[41] C. J. Lambe, C. M. Wittmann, and R. E. Spekman, "Social Exchange Theory and Research on Business-to-Business Relational Exchange," *Journal of Business-to-Business Marketing*, vol. 8, no. 3, pp. 1–36, 2001.

[42] E. J. Lawler, "An Affect Theory of Social Exchange," *American Journal of Sociology*, vol. 107, no. 2, pp. 321–352, 2001.

[43] E. J. Lawler and S. R. Thye, "Bringing Emotions into Social Exchange Theory," *Annual Review of Sociology*, vol. 25, no. 1, pp. 217–244, 1999.

[44] M. Worthy, A. L. Gary, and G. M. Kahn, "Self-disclosure as an exchange process." *Journal of Personality and Social Psychology*, vol. 13, pp. 59–63, 1969.

[45] G. Herbette and B. Rimé, "Verbalization of Emotion in Chronic Pain Patients and their Psychological Adjustment," *Journal of Health Psychology*, vol. 9, no. 5, pp. 661–676, 7 2004. [Online]. Available:

[46] G. Itzchakov, N. Weinstein, D. Saluk, and M. Amar, "Connection Heals Wounds: Feeling Listened to Reduces Speakers' Loneliness Following a Social Rejection Disclosure," *Personality and Social Psychology Bulletin*, 6 2022. [Online]. Available:

[47] G. Itzchakov, N. Weinstein, and A. Cheshin, "Learning to listen: Downstream effects of listening training on employees' relatedness, burnout, and turnover intentions," *Human Resource Management*, 2022. [Online]. Available:

[48] J. H. Kahn and K. E. Cantwell, "The role of social support on the disclosure of everyday unpleasant emotional events," *Counselling Psychology Quarterly*, vol. 30, no. 2, pp. 152–165, 4 2016.

[49] J. H. Kahn and A. M. Garrison, "Emotional Self-Disclosure and Emotional Avoidance: Relations With Symptoms of Depression and Anxiety," *Journal of Counseling Psychology*, vol. 56, no. 4, pp. 573–584, 10 2009. [Online]. Available:

[50] J. H. Kahn and R. M. Hessling, "Measuring the

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

Tendency to Conceal Versus Disclose Psychological Distress," *Journal of Social and Clinical Psychology*, vol. 20, no. 1, pp. 41–65, 3 2001. [Online]. Available:

[51] V. J. Derlega, B. A. Winstead, R. J. Lewis, and J. Maddux, "Clients' responses to dissatisfaction in psychotherapy: A test of Rusbult's exit-voice-loyalty-neglect model," *Journal of Social and Clinical Psychology*, vol. 12, no. 3, pp. 307–318, 1993.

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

[53] F. Nils and B. Rimé, "Beyond the myth of venting: Social sharing modes determine the benefits of emotional disclosure," *European Journal of Social Psychology*, vol. 42, no. 6, pp. 672–681, 10 2012.

[54] K. Kircanski, M. D. Lieberman, and M. G. Craske, "Feelings Into Words," *Psychological Science*, vol. 23, no. 10, pp. 1086–1091, 8 2012. [Online]. Available:

[55] M. D. Lieberman, N. I. Eisenberger, M. J. Crockett, S. M. Tom, J. H. Pfeifer, and B. M. Way, "Putting Feelings Into Words," *Psychological Science*, vol. 18, no. 5, pp. 421–428, 11 2016. [Online]. Available:

[56] J. B. Torre and M. D. Lieberman, "Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation," *Emotion Review*, vol. 10, no. 2, pp. 116–124, 3 2018. [Online]. Available:

[57] D. I. Tamir and J. P. Mitchell, "Disclosing information about the self is intrinsically rewarding," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 21, pp. 8038–8043, 5 2012. [Online]. Available:

[58] E. Kross, O. Ayduk, and W. Mischel, "When Asking "Why" Does Not Hurt Distinguishing Rumination From Reflective Processing of Negative Emotions," *Psychological Science*, vol. 16, no. 9, pp. 709–715, 12 2016. [Online]. Available:

[59] M. D. Lieberman, T. K. Inagaki, G. Tabibnia, and M. J. Crockett, "Subjective responses to emotional stimuli during labeling, reappraisal, and distraction." *Emotion*, vol. 11, no. 3, p. 468, 6 2011. [Online]. Available:

[60] J. D. Creswell, B. M. Way, N. I. Eisenberger, and M. D. Lieberman, "Neural Correlates of Dispositional Mindfulness During Affect Labeling," *Psychosomatic Medicine*, vol. 690, pp. 560–565, 2007.

[61] J. W. Pennebaker, "Writing about Emotional Experiences as a Therapeutic Process," *Psychological Science*, vol. 8, no. 3, pp. 162–166, 1997. [Online]. Available:

[62] J. W. Pennebaker and S. K. Beall, "Confronting a traumatic event: toward an understanding of inhibition and disease," *Journal of abnormal psychology*, vol. 95, no. 3, pp. 274–281, 1986. [Online]. Available:

[63] K. A. Baikie and K. Wilhelm, "Emotional and physical health benefits of expressive writing," *Advances in Psychiatric Treatment*, vol. 11, no. 5, pp. 338–346, 9 2005. [Online]. Available:

[64] A. Cheshin, "The Impact of Non-normative Displays of Emotion in the Workplace: How Inappropriateness Shapes the Interpersonal Outcomes of Emotional Displays," *Frontiers in Psychology*, vol. 11, p. 6, 2 2020.

[65] F. Steele, *The open organization : the impact of secrecy and disclosure on people and organizations*. Addison-Wesley, 1975.

[66] G. Egan, *Encounter; group processes for interpersonal growth*. Belmont, CA.: Brooks/Cole Pub, 1970.

[67] E. Goffman, *The presentation of self in everyday life.* Oxford, England: Doubleday, 1959.

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

[69] 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. New York, NY, US: The Guilford Press, 2000, pp. 220–242.

[70] 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. [Online]. Available:

[71] Naldemirci, N. Britten, H. Lloyd, and A. Wolf, "The potential and pitfalls of narrative elicitation in person-centred care," *Health Expectations*, vol. 23, no. 1, pp. 238–246, 2 2020. [Online]. Available:

[72] C. R. Senteio and D. B. Yoon, "How Primary Care Physicians Elicit Sensitive Health Information From Patients: Describing Access to Psychosocial Information," *Qualitative Health Research*, vol. 30, no. 9, pp. 1338–1348, 3 2020. [Online]. Available:

[73] A. Farber, Barry, *Self-disclosure in Psychotherapy*. Guilford Press, 2006, no. 3. [Online]. Available:

[74] G. A. Van Kleef, "How Emotions Regulate Social Life," *Current Directions in Psychological Science*, vol. 18, no. 3, pp. 184–188, 6 2009. [Online]. Available:

[75] A. F. Shariff and J. L. Tracy, "What are emotion expressions for?" *Current Directions in Psychological Science*, vol. 20, no. 6, pp. 395–399, 12 2011. [Online]. Available:

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

[77] L. B. Rosenfeld, "Self-disclosure avoidance: Why I am afraid to tell you who I am," *Communication Monographs*, vol. 46, no. 1, pp. 63–74, 1979. [Online]. Available:

[78] M. Mikulincer and O. Nachshon, "Attachment Styles and Patterns of Self-Disclosure," *Journal of Personality and Social Psychology*, vol. 61, no. 2, pp. 321–331, 1991. [Online]. Available:

[79] D. Matsumoto, "Cultural similarities and differences in display rules," *Motivation and Emotion*, vol. 14, no. 3, pp. 195–214, 9 1990. [Online]. Available:

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

[81] E. S. Cross and R. Ramsey, "Mind Meets Machine: Towards a Cognitive Science of Human–Machine Inter-

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

actions," *Trends in Cognitive Sciences*, vol. 25, no. 3, pp. 200–212, 3 2021.

[82] G. Laban, V. Morrison, and E. Cross, "Social Robots for Health Psychology: A New Frontier for Improving Human Health and Well-Being," *European Health Psychologist*, vol. 23, no. 1, pp. 1095–1102, 2 2024. [Online]. Available:

[83] H. L. Colquhoun, J. E. Squires, N. Kolehmainen, C. Fraser, and J. M. Grimshaw, "Methods for designing interventions to change healthcare professionals' behaviour: a systematic review," *Implementation Science*, vol. 12, no. 1, p. 30, 2017.

[84] D. Wight, E. Wimbush, R. Jepson, and L. Doi, "Six steps in quality intervention development (6SQuID)," *J Epidemiol Community Health*, vol. 70, no. 5, p. 520, 2016.

[85] 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. Cham: Springer International Publishing, 2020, pp. 233–254.

[86] N. Epley and A. Waytz, "Mind Perception," ser. Handbook of Social Psychology, S. T. Fiske, D. T. Gilbert, and G. Lindzey, Eds. John Wiley and Sons Ltd, 2010.

[87] K. Gray, L. Young, and A. Waytz, "Mind Perception Is the Essence of Morality," *Psychological Inquiry*, vol. 23, no. 2, pp. 101–124, 2012.

[88] B. R. Duffy and G. Joue, "The paradox of social robotics: A discussion," in *AAAI Fall 2005 Symposium on Machine Ethics*, Hyatt Regency, 2005.

[89] L. Becker, "Reciprocity," 1986.

[90] E. Katz, J. G. Blumler, and M. Gurevitch, "Uses and gratifications research," *Public Opinion Quarterly*, vol. 37, no. 4, pp. 509–523, 1 1973. [Online]. Available:

[91] Y. Moon, "Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers," *Journal of Consumer Research*, vol. 26, no. 4, pp. 323–339, 3 2000. [Online]. Available:

[92] G. Laban and T. Araujo, "The Effect of Personalization Techniques in Users' Perceptions of Conversational Recommender Systems," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, 2020. [Online]. Available:

[93] ——, "Don't Take it Personally: Resistance to Individually Targeted Recommendations from Conversational Recommender Agents," in *HAI 2022 - Proceedings of the 10th Conference on Human-Agent Interaction*. New York, NY, USA: Association for Computing Machinery, 2022, pp. 57–66. [Online]. Available:

[94] E. Bendig, B. Erb, L. Schulze-Thuesing, and H. Baumeister, "The Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health – A Scoping Review," *Verhaltenstherapie*, pp. 1–13, 2019. [Online]. Available:

[95] D. Chattopadhyay, T. Ma, H. Sharifi, and P. Martyn-Nemeth, "Computer-Controlled Virtual Humans in Patient-Facing Systems: Systematic Review and Meta-Analysis," *J Med Internet Res*, vol. 22, no. 7, p. e18839, 7 2020. [Online]. Available:

[96] S. Hoermann, K. L. McCabe, D. N. Milne, and R. A. Calvo, "Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review," *J Med Internet Res*, vol. 19, no. 8, p. e7023, 8 2017. [Online]. Available:

[97] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, and J. B. Torous, "Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape," *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, vol. 64, no. 7, pp. 456–464, 7 2019. [Online]. Available:

[98] L. S. Pauw, D. A. Sauter, G. A. van Kleef, G. M. Lucas, J. Gratch, and A. H. Fischer, "The avatar will see you now: Support from a virtual human provides socio-emotional benefits," *Computers in Human Behavior*, vol. 136, p. 107368, 11 2022.

[99] M. de Gennaro, E. G. Krumhuber, and G. Lucas, "Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood," *Frontiers in Psychology*, vol. 10, p. 3061, 1 2020.

[100] C. V. Clark-Gordon, N. D. Bowman, A. K. Goodboy, and A. Wright, "Anonymity and Online Self-Disclosure: A Meta-Analysis," *Communication Reports*, vol. 32, no. 2, pp. 98–111, 5 2019. [Online]. Available:

[101] R. N. McLay, W. E. DEAL, J. A. MURPHY, K. B. CENTER, T. T. KOLKOW, and T. A. GRIEGER, "On-the-Record Screenings Versus Anonymous Surveys in Reporting PTSD," *American Journal of Psychiatry*, vol. 165, no. 6, pp. 775–776, 6 2008. [Online]. Available:

[102] G. M. Lucas, J. Gratch, A. King, and L.-P. Morency, "It's only a computer: Virtual humans increase willingness to disclose," *Computers in Human Behavior*, vol. 37, pp. 94–100, 2014. [Online]. Available:

[103] M. D. Pickard, C. A. Roster, and Y. Chen, "Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions?" *Computers in Human Behavior*, vol. 65, pp. 23–30, 2016. [Online]. Available:

[104] D. Utami, T. Bickmore, A. Nikolopoulou, and M. Paasche-Orlow, "Talk About Death: End of Life Planning with a Virtual Agent," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10498 LNAI, pp. 441–450, 8 2017. [Online]. Available:

[105] K. Yokotani, G. Takagi, and K. Wakashima, "Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design," *Computers in Human Behavior*, vol. 85, pp. 135–145, 2018.

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

[Online]. Available:

[106] G. Warren-Smith, G. Laban, E.-M. Pacheco, and E. S. Cross, "Knowledge cues to human origins facilitate self-disclosure during interactions with chatbots," 2023.

[107] L. Tickle-Degnen and R. Rosenthal, "The Nature of Rapport and Its Nonverbal Correlates," *Psychological Inquiry*, vol. 1, no. 4, pp. 285–293, 1 1990. [Online]. Available:

[108] 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*, 1st ed. New York, NY, USA: Association for Computing Machinery, 9 2021, vol. 1, pp. 433–462. [Online]. Available:

[109] G. M. Lucas, A. Rizzo, J. Gratch, S. Scherer, G. Stratou, J. Boberg, and L.-P. Morency, "Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers ," *Frontiers in Robotics and AI*, vol. 4, p. 51, 2017. [Online]. Available:

[110] T. Bickmore, A. Gruber, and R. Picard, "Establishing the computer–patient working alliance in automated health behavior change interventions," *Patient Education and Counseling*, vol. 59, no. 1, pp. 21–30, 10 2005.

[111] E. Cho, N. Motalebi, S. S. Sundar, and S. Abdullah, "Alexa as an Active Listener: How Backchanneling Can Elicit Self-Disclosure and Promote User Experience," in *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2. New York, NY, USA: ACM, 11 2022, pp. 1–23. [Online]. Available:

[112] C. L. Bethel, M. R. Stevenson, and B. Scassellati, "Secret-sharing: Interactions between a child, robot, and adult," in *2011 IEEE International Conference on Systems, Man, and Cybernetics*, 2011, pp. 2489–2494.

[113] S. R. Nijssen, B. C. Müller, T. Bosse, and M. Paulus, "You, robot? The role of anthropomorphic emotion attributions in children's sharing with a robot," *International Journal of Child-Computer Interaction*, vol. 30, p. 100319, 12 2021.

[114] Y. Nakamura and H. Umemuro, "Effect of Robot's Listening Attitude Change on Self-disclosure of the Elderly," *International Journal of Social Robotics*, 2022. [Online]. Available:

[115] P. S. Dautzenberg, G. M. Vos, S. Ladwig, and A. M. Von Der Putten, "Investigation of different communication strategies for a delivery robot: The positive effects of humanlike communication styles," *2021 30th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2021*, pp. 356–361, 8 2021.

[116] 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. [Online]. Available:

[117] B. Sayis and H. Gunes, "Technology-assisted Journal Writing for Improving Student Mental Wellbeing: Humanoid Robot vs. Voice Assistant," *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 945–949, 3 2024. [Online]. Available:

[118] Y. Mou, L. Zhang, Y. Wu, S. Pan, and X. Ye, "Does Self-Disclosing to a Robot Induce Liking for the Robot? Testing the Disclosure and Liking Hypotheses in Human–Robot Interaction," *International Journal of Human–Computer Interaction*, pp. 1–12, 1 2023. [Online]. Available:

[119] 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, no. 3, p. 98, 7 2021. [Online]. Available:

[120] 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, no. 5, p. 92, 9 2022. [Online]. Available:

[121] C. Bartneck, T. Bleeker, J. Bun, P. Fens, and L. Riet, "The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots," *Paladyn*, vol. 1, no. 2, pp. 109–115, 6 2010. [Online]. Available:

[122] H. Kumazaki, T. Muramatsu, Y. Yoshikawa, Y. Matsumoto, K. Takata, H. Ishiguro, and M. Mimura, "Android Robot Promotes Disclosure of Negative Narratives by Individuals With Autism Spectrum Disorders," *Frontiers in Psychiatry*, vol. 13, p. 1265, 6 2022.

[123] 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. [Online]. Available:

[124] T. Akiyoshi, J. Nakanishi, H. Ishiguro, H. Sumioka, and M. Shiomi, "A Robot That Encourages Self-Disclosure to Reduce Anger Mood," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7926–7933, 10 2021.

[125] G. E. Birnbaum, M. Mizrahi, G. Hoffman, H. T. Reis, E. J. Finkel, and O. Sass, "What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure," *Computers in Human Behavior*, vol. 63, pp. 416–423, 10 2016.

[126] F. Dino, R. Zandie, H. Abdollahi, S. Schoeder, and M. H. Mahoor, "Delivering Cognitive Behavioral Therapy Using A Conversational Social Robot," *IEEE International Conference on Intelligent Robots and Systems*, pp. 2089–2095, 11 2019.

[127] G. Laban, A. Kappas, V. Morrison, and E. S. Cross, "Building Long-Term Human–Robot Relationships: Examining Disclosure, Perception and Well-Being Across Time," *International Journal of Social Robotics*, 2023.

[128] 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*, ser. CHI EA '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 1–7.

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

[Online]. Available:

[129] ——, "Coping with Emotional Distress via Self-Disclosure to Robots: Intervention with Caregivers," *PsyArxiv*, 2023. [Online]. Available:

[130] T. A. Revenson, K. Griva, A. Luszczynska, V. Morrison, E. Panagopoulou, N. Vilchinsky, and M. Hagedoorn, "What Is Caregiving and How Should We Study It?" in *Caregiving in the Illness Context*, T. A. Revenson, K. Griva, A. Luszczynska, V. Morrison, E. Panagopoulou, N. Vilchinsky, and M. Hagedoorn, Eds. London: Palgrave Macmillan UK, 2016, pp. 1–14. [Online]. Available:

[131] F. Eyssel, R. Wullenkord, and V. Nitsch, "The role of self-disclosure in human-robot interaction," *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*, vol. 2017-January, pp. 922–927, 12 2017. [Online]. Available:

[132] 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. [Online]. Available:

[133] 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.

[134] G. Laban, A. Kappas, V. Morrison, and E. S. Cross, "Opening Up to Social Robots: How Emotions Drive Self-Disclosure Behavior," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 8 2023, pp. 1697–1704. [Online]. Available:

[135] T. Uchida, H. Takahashi, M. Ban, J. Shimaya, T. Minato, K. Ogawa, Y. Yoshikawa, and H. Ishiguro, "Japanese Young Women Did not Discriminate between Robots and Humans as Listeners for Their Self-Disclosure -Pilot Study-," *Multimodal Technologies and Interaction 2020, Vol. 4, Page 35*, vol. 4, no. 3, p. 35, 6 2020. [Online]. Available:

[136] C. Lutz and A. Tamò-Larrieux, "Do Privacy Concerns About Social Robots Affect Use Intentions? Evidence From an Experimental Vignette Study," *Frontiers in Robotics and AI*, vol. 8, p. 63, 4 2021.

[137] C. Lutz, M. Schöttler, and C. P. Hoffmann, "The privacy implications of social robots: Scoping review and expert interviews," *Mobile Media & Communication*, vol. 7, no. 3, pp. 412–434, 9 2019. [Online]. Available:

[138] M. Dietrich, M. Krüger, and T. H. Weisswange, "What should a robot disclose about me? A study about privacy-appropriate behaviors for social robots," *Frontiers in Robotics and AI*, vol. 10, p. 1236733, 12 2023.

[139] S. Petronio, "Communication Boundary Management: A Theoretical Model of Managing Disclosure of Private Information between Marital Couples," *Communication Theory*, vol. 1, no. 4, pp. 311–335, 11 1991. [Online]. Available:

[140] ——, *Boundaries of privacy: Dialectics of disclosure.*, ser. SUNY series in communication studies. Albany, NY, US: State University of New York Press, 2002.

[141] L. Levinson, C. Nippert-Eng, R. Gomez, and S. Šabanović, "Snitches Get Unplugged: Adolescents' Privacy Concerns about Robots in the Home are Relationally Situated," *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 423–432, 3 2024. [Online]. Available:

[142] H. Erel, M. Vázquez, S. Sebo, N. Salomons, S. Gillet, and B. Scassellati, "RoSI: A Model for Predicting Robot Social Influence," *ACM Transactions on Human-Robot Interaction*, vol. 13, no. 2, p. 22, 6 2024. [Online]. Available:

[143] S. Gillet, M. Vázquez, S. Andrist, I. Leite, and S. Sebo, "Interaction-Shaping Robotics: Robots That Influence Interactions between Other Agents," *ACM Transactions on Human-Robot Interaction*, vol. 13, no. 1, p. 23, 3 2024. [Online]. Available:

[144] E. M. Schomakers and M. Ziefle, "Privacy Concerns and the Acceptance of Technologies for Aging in Place," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11592 LNCS, pp. 313–331, 2019. [Online]. Available:

[145] A. Adensamer, R. Gsenger, and L. D. Klausner, ""Computer says no": Algorithmic decision support and organisational responsibility," *Journal of Responsible Technology*, vol. 7-8, p. 100014, 10 2021.

[146] M. Coeckelbergh, "Robot rights? Towards a social-relational justification of moral consideration," *Ethics and Information Technology*, vol. 12, no. 3, pp. 209–221, 6 2010. [Online]. Available:

[147] D. L. Gogoshin, "Robot Responsibility and Moral Community," *Frontiers in Robotics and AI*, vol. 8, p. 768092, 11 2021. [Online]. Available:

[148] R. Hakli and P. Mäkelä, "Moral Responsibility of Robots and Hybrid Agents," *The Monist*, vol. 102, no. 2, pp. 259–275, 4 2019. [Online]. Available:

[149] G. Bejarano, F. Li, N. Ruijs, and Y. Lu, "Ethics & AI: A Systematic Review on Ethical Concerns and Related Strategies for Designing with AI in Healthcare," *AI*, vol. 4, no. 1, pp. 28–53, 12 2022. [Online]. Available:

[150] M. Lee, J. Sin, G. Laban, M. Kraus, L. Clark, M. Porcheron, B. R. Cowan, A. Følstad, C. Munteanu, and H. Candello, "Ethics of Conversational User Interfaces," in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 1–7. [Online]. Available:

[151] C. E. Izard, "Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues," *Annual review of psychology*, vol. 60, p. 25, 1 2009.

[152] S. Park and M. Whang, "Empathy in Human–Robot Interaction: Designing for Social Robots," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1889, 2 2022. [Online]. Available:

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3470984

[153] T. J. Prescott and J. M. Robillard, "Are friends electric? The benefits and risks of human-robot relationships," *iScience*, vol. 24, no. 1, p. 101993, 1 2021.

[154] N. Haslam, "Dehumanization: An Integrative Review," *Personality and Social Psychology Review*, vol. 10, no. 3, pp. 252–264, 8 2006. [Online]. Available:

[155] J. Urakami and K. Seaborn, "Nonverbal Cues in Human–Robot Interaction: A Communication Studies Perspective," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 2, pp. 1–21, 3 2023. [Online]. Available:

[156] 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.

[157] 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.

[158] C. L. Bethel, Z. Henkel, K. Stives, D. C. May, D. K. Eakin, M. Pilkinton, A. Jones, and M. Stubbs-Richardson, "Using robots to interview children about bullying: Lessons learned from an exploratory study," in *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*. Institute of Electrical and Electronics Engineers Inc., 11 2016, pp. 712–717.

[159] N. I. Abbasi, G. Laban, T. Ford, P. B. Jones, and H. Gunes, "Robotising Psychometrics: Validating Well-being Assessment Tools in Child-Robot Interactions," in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2024.

[160] V. N. Giri, "Culture and Communication Style," *Review of Communication*, vol. 6, no. 1-2, pp. 124–130, 1 2006. [Online]. Available:

[161] Y. W. Chen and M. Nakazawa, "Influences of Culture on Self-Disclosure as Relationally Situated in Intercultural and Interracial Friendships from a Social Penetration Perspective," *Journal of Intercultural Communication Research*, vol. 38, no. 2, pp. 77–98, 2010. [Online]. Available:

[162] V. Lim, M. Rooksby, and E. S. Cross, "Social Robots on a Global Stage: Establishing a Role for Culture During Human–Robot Interaction," *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1307–1333, 9 2021. [Online]. Available:

[163] H. R. Markus and S. Kitayama, "Culture and the self: Implications for cognition, emotion, and motivation," *Psychological Review*, vol. 98, no. 2, pp. 224–253, 1991.

[164] O. Korn, N. Akalin, and R. Gouveia, "Understanding Cultural Preferences for Social Robots," *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 2, 5 2021.

[165] K. A. Lindquist, A. B. Satpute, and M. Gendron, "Does Language Do More Than Communicate Emotion?" *Current Directions in Psychological Science*, vol. 24, no. 2, pp. 99–108, 4 2015. [Online]. Available:

[166] K. A. Lindquist, J. K. MacCormack, and H. Shablack, "The role of language in emotion: Predictions from psychological constructionism," *Frontiers in Psychology*, vol. 6, no. MAR, p. 444, 4 2015.

[167] C. Zhang, J. Chen, J. Li, Y. Peng, and Z. Mao, "Large language models for human–robot interaction: A review," *Biomimetic Intelligence and Robotics*, vol. 3, no. 4, p. 100131, 12 2023.

[168] M. Spitale, M. Axelsson, and H. Gunes, "VITA: A Multi-modal LLM-based System for Longitudinal, Autonomous, and Adaptive Robotic Mental Well-being Coaching," 12 2023. [Online]. Available:

[169] ——, "Appropriateness of LLM-equipped Robotic Well-being Coach Language in the Workplace: A Qualitative Evaluation," 1 2024. [Online]. Available:

[170] C. Y. Kim, C. P. Lee, and B. Mutlu, "Understanding Large-Language Model (LLM)-powered Human-Robot Interaction," *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 371–380, 3 2024. [Online]. Available:

XV. Biography Section

Guy Laban is a Postdoctoral Research Associate at the Department of Computer Science & Technology of the University of Cambridge, and a member of the Affective Intelligence and Robotics Laboratory (AFAR). Guy pursued his PhD studies in Neuroscience and Psychology as a Marie Skłodowska-Curie Fellow at the School of Psychology and Neuroscience of the University of Glasgow. During his PhD studies, Guy was an Early Stage Research (ESR) member of ENTWINE, the European Training Network on Informal Care, a Marie Skłodowska-Curie Innovation Training Network (ITN) funded by the European Union. Guy's research interests centre on supporting individuals' emotional well-being through interactions with robots. Specifically, Guy investigates how individuals convey their emotions to robots, and how these interactions can be leveraged to support their overall emotional health. Guy studies, develops, and designs robots that facilitate meaningful social exchanges with human users across various contexts.

Emily S. Cross received the Ph.D. degree in cognitive neuroscience from Dartmouth College, Hanover, NH, USA. Recently, she joined ETH Zürich to establish the Professorship for Social Brain Sciences in Zurich, Switzerland. Prior to this, she was based jointly with the MARCS Institute, Western Sydney University, Sydney, NSW, Australia; the Department of Cognitive Science, Macquarie University, Sydney; and the School of Psychology and Neuroscience, University of Glasgow, Glasgow, U.K. She leads a vibrant and diverse research team that uses brain imaging techniques, robots, and complex action training paradigms to explore how experience-dependent plasticity and expertise is manifest across brain and behaviour, with a particular focus on how our experiences and expectations about artificial agents influence our interactions with these agents. Her work has been supported by a range of funders worldwide, including the European Research Council, the Australian Research Council, Research Councils UK, the Dutch Science Foundation, the Alexander von Humboldt Stiftung, the National Institutes of Health, and the Fulbright Commission.

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on November 01,2024 at 19:06:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.