Robot mirroring: Improving well-being by fostering empathy with an artificial agent representing the self

David Antonio Gómez Jáuregui
Univ. Bordeaux, ESTIA
Institute of Technology
Bidart, France
d.gomez@estia.fr

Felix Dollack
Smart Care Tech Co., Ltd.
Tokyo, Japan
felix.dollack@smartcaretech.co.jp

Monica Perusquía-Hernández
University of Essex
Colchester, United Kingdom
perusquia@ieee.org

***Abstract*—Well-being has become a major societal goal. Being well means being physically and mentally healthy. Additionally, feeling empowered is also a component of well-being. Recently, self-tracking has been proposed as means to achieve increased awareness, thus, giving the opportunity to identify and decrease undesired behaviours. However, inappropriately communicated self-tracking results might cause the opposite effect. To address this, a subtle self-tracking feedback by mirroring the self's state into an embodied artificial agent has been proposed. By eliciting empathy towards the artificial agent and fostering helping behaviours, users would help themselves as well. We searched the literature to find supporting or opposing evidence for the robot mirroring framework. The results showed an increasing interest in self-tracking technologies for well-being management. Current discussions disseminate what can be achieved with different levels of automation; the type and relevance of feedback; and the role that artificial agents, such as chatbots and robots, might play to support people's therapies. These findings support further development of the robot mirroring framework to improve medical, hedonic, and eudaemonic well-being.**

***Index Terms*—Affective computing, chatbots, well-being, self-quantification, empathy, robotics.**

I. Introduction

Well-being is one of the most basic human concerns. It refers to being well in medical, hedonic and eudaemonic senses. Medical well-being refers to being healthy physically and psychologically. Hedonic well-being refers to the subjective feeling of happiness. Finally, eudaemonic well-being is a state of being that demonstrates our inner potential and makes us feel fulfilled, significant and leads to an active life in connection to our surroundings and other people [1].

Mindfulness and self-compassion have been related to high levels of well-being [2]. A common strategy to improve our well-being is to become aware of the issues undermining it. Self-tracking has been proposed to identify areas of opportunity to improve our well-being. Bodily states are often measured with the goal of monitoring our health and detect anomalies to treat them before they become problematic. Models of Personal Informatics (PI) propose that self-tracking leads to self-insight, and eventually to self-improvement or a positive behaviour change (Self-Improvement Hypothesis) [3], [4].

Other theories have also been proposed to explain and understand the processes that make behaviour change more likely. According to self-attention theory [5], there are three fundamental requirements for any behaviour to occur. Self-focused attention, a salient behavioural standard, and a sufficiently good outcome expectancy to warrant continued efforts. The cybernetic model of self-attention processes proposed that behaviour change is achieved within a test–operate–test–exit (TOTE) loop [6]. To enter the TOTE loop, a person must establish a goal. In the *test* phase, the person compares current and goal state. If there is a mismatch between current and desired state, the person performs the *operate* phase in which actions are performed to reduce the mismatch. The person **tests** again and, depending on whether the person achieves the goal to be met or whether more work is needed, the person either returns to the *operate* phase or *exits* the loop. In this model, emotions play a central role: positive emotions indicate that the person is moving towards the goal and negative emotions indicate a movement away from the goal. Further, the Nudge theory argued that, to alter people's behaviour without limiting freedom of choice, it is more effective to encourage positive choices rather than restricting unwanted behaviour with sanctions [7]. All these theories have in common the notion of knowing what is a positive choice, how far is a person from their goal or the desired standard, and some degree of self-attention, self-awareness or self-insight.

In this work, we challenge the assumption that awareness through self-tracking is necessary for self-improvement. We argue that, sometimes, a lack of awareness might be more beneficial. Overwhelming the user with information might boost unintended effects of self-tracking such as burden of data collection or annoyance with technology [8]. The number of tracking possibilities is vast, and their interpretation is complex. Interventions might not be effective for all personalities [9]. To overcome these challenges, machines could be useful by processing high dimensional raw data. Recent machine learning techniques can be used to automatically match raw information to a human-understandable, more abstract concept. Nevertheless, data interpretation by artificial intelligence (AI) algorithms might not be perfect and situations of distrust or over-trust might arise [10]. Hence, high levels of automation in data interpretation and user feedback should be considered carefully. Furthermore, the automation challenge when providing affective feedback is complex at both the state

978-1-6654-0021-3/21/$31.00 ©2021 IEEE

identification by the machine, and at the social-signal interpretation by the human user. For example, a high-level automated agent might tell the user, "you are depressed" after analysing the raw data. However, an accurate interpretation of the signals by the machine might change according to several factors such as context. Also, the interpretation of the feedback by the users might not be the same across individuals. The interoceptive body perception and the subjective affective experience might be linked trough different words for different people with different contextual experiences [11]. If not treated carefully, using high-level concepts could also lead to misinterpretations.

Therefore, the self-tracking feedback can be provided at multiple levels ranging from raw data to high-level metacognitive verbal statements. Numeric feedback is closer to a raw physical measurement of the user's bodily state. Agents conversing with the user would be at the opposite extreme. Embodied agents (i.e., robots) have the flexibility of being able to represent both high-level cognitive states and low-level embodied states. A physical body would also enable them to convey information through movement [12]; and interoceptive cues through vibration [13], and heat [14].

Quantification feedback can be delivered subtly, without creating awareness. The user does not necessarily need to be aware of each measure, but rather a reflection of their health is represented in an embodied agent. This is similar to looking at oneself in a mirror, and it allows users to better interpret their state intuitively. Moreover, by provoking empathy with the robot, self-compassion is fostered indirectly. Self-compassion has been connected with the adaptation of healthier behaviours [15]–[18]. As human beings, we are largely influenced and inspired by others. A naïve desire to take care of, and give a good example to our children might motivate us to live a healthier lifestyle [19]. Therefore, we hypothesise that behaviour change can be fostered without necessarily passing through the awareness bridge stated by the Self-Improvement Hypothesis.

In the following sections we provide evidence from the literature in favour and against different self-tracking feedback approaches. First, we explore in detail the assumptions of regular quantification feedback. Next, we present the concept elements of the robot mirroring framework, the assumptions made, and the theoretical support. Finally, we discuss how both options can be used to improve well-being at different levels.

II. Quantification approaches for improved well-being

Self-tracking technologies for self-driven psychological well-being have grown in recent years. Most of these technologies are based on the detection and measurement of different physiological signals such as hearth rate variability and Electroencephalography (EEG) alpha power [20]. In addition, the use of smartphones has made data capture and self-tracking available to everyone. Several authors have proposed mobile apps and wearable devices for self-monitoring of emotional well-being [21], [22] and stress management [20], [23]. The results suggest that engagement with mobile apps is related to improvements in mental health and well-being. There are several assumptions regarding quantification approaches for self-tracking. First, awareness is necessary to produce a behaviour change in the user. Second, increased awareness is useful for everybody. Finally, AI algorithms can be accurate enough and the reasoning behind each result can be communicated to the user in an intuitive manner.

A. Assumption 1: awareness is necessary for behaviour change

Behaviour change models such as the Trans-Theoretical Model of Behaviour Change (TTM) [24] imply that increased awareness includes two elements: awareness of a previously unknown fact, and awareness of reasons why behaviour change is advantageous. In other words, both awareness of the problem and the motivations to solve it are necessary elements for behaviour change. Therefore, for PI systems to induce behavioural changes, the motivations given for behaviour change should outweigh reasons in favour of staying the same. Additionally, for long lasting behaviour change, the perception of self-efficacy in achieving the desired goal is necessary, as well as feedback temporally close to the relevant behaviour [25].

Recent sensing technology and mobile applications have enabled us to monitor and log our behaviour to immediately obtain feedback on the monitored state. For example, heart rate (HR) could be directly displayed to a screen, for the user to interpret the abstract numerical value contextually according to the activity he is realising. This would help users to identify co-occurring events that cause HR variations. Other cues that can add to the interpretability of HR can be tracked simultaneously. For example, HR and respiration rates are known to depend on each other, and respiration exercises are recommended to control one's heart [26], [27]. Even though technology has enabled us to increase the number of tracked features, its interpretation might be complex and can be misunderstood by a naïve user [9]. Furthermore, a modality or channel is required to provide feedback for each type of information. However, it is difficult for a human to handle multiple channels, and more technically complex to implement them. Therefore, appropriate data presentation is necessary to deliver relevant and concise information. Nevertheless, the challenge of avoiding feedback misinterpretation remains. For example, people might only believe information that confirms or suits their beliefs [28]. These confirmation biases may skew the interpretation of raw data feedback. We usually do not think about how many beats per second our heart pumps. Instead, we think about how that makes us feel, and we name those feelings to integrate multiple cues in a concept often described with a single word. However intuitive, word emotion descriptors might be mapped to different bodily cues according to the context [11]. Therefore, in comparison to having a complex conversation about an abstract emotion, there might be an advantage of delivering feedback subtly, in the form of enhanced interoceptive cues mapped into the body of an artificial agent.

B. Assumption 2: increased awareness is useful for everybody

People with different personalities might utilise the provided self-tracking feedback differently. While physiology feedback may be beneficial for individuals high in neuroticism, it may be detrimental for those with high anxiety sensitivity [29]. Furthermore, strategies to deal with symptom tracking differ between introverts and extroverts. A week-long test with an emotion-aware chatbot designed for behaviour change, showed that extroverts preferred the emotion-aware chatbot significantly more than introverts [30].

C. Assumption 3: AI can deliver an accurate and explainable interpretation of the data

Artificial Intelligence algorithms excel at the task of processing high amounts of information. Recent machine learning techniques can be used to automatically match raw information to a more abstract, human-understandable concept. However, data interpretation by artificial intelligence algorithms might not be perfect and situations of distrust or over-trust might arise [28]. Therefore, high levels of automation in data interpretation and user feedback should be considered carefully. It is important to define up to which level of automation the interpretations are reliable. Previously, four levels of automation have been proposed: (1) information acquisition; (2) information analysis; (3) decision and action selection; and (4) action implementation [31]. Current self-tracking systems deal only with information acquisition and analysis. High cognitive tools like chatbots, on the other hand, tend to have level 3 automation [32]. The control over or burden to make decisions and implement behaviour change remains with the user. Nevertheless, even results of information analysis should be treated carefully. For example, incorrectly detecting that someone is angry and stating so might lead to two outcomes. Either the user's trust in the feedback decreases because it was incorrect, or a self-fulfilling prophecy is created where the user will get angry after being told that they are angry. Usually, the degree to which a machine learning system can tolerate false alarms can be controlled a priori and strongly depends on the system application. It might be better to falsely diagnose a disease than to miss the symptoms and fail to provide proper treatment to a sick person. Hence, the feedback provided to a user who wants to lose weight because of aesthetic reasons should be different from the feedback provided to a diabetic patient at risk of complications. In non-medical conditions, users might benefit from more lenient feedback thresholds.

III. The Robot mirroring framework

Previous work has proposed the Robot Mirroring framework [13]. This framework aims to improve human well-being by encouraging well-being directed actions towards an embodied agent or robot that mirrors the user's emotional state. There are two core elements to this concept. First, empathy towards an artificial agent is fostered through the mirroring of behaviour and physiological signals. In this manner, the information acquired through self-tracking is delivered intuitively, and with a minimum of information analysis. As a strategy to amplify the perception of the user's own body, the pet reflects the current user state (e.g., user's heart rate) through the user's interoception augmentation. Empathy can also be fostered by allowing the users to personalise their agent (e.g., a pet), and own them. Second, the agent should appear weaker and be embodied, as this would trigger more caring behaviours from the users (Fig. 1).

By using this framework, it is possible to provide therapeutic support without going into more complex language-dependent applications. Currently there are applications in the market and extensive research to push the limits of AI chatbots as effective therapists [32], [33]. In these, empathetic behaviours have been demonstrated in a variety of contexts. An empathetic chatbot can serve as a buffer against the adverse effects of social ostracism [34]; and bots speaking in conversation topics building from general to specific can help to manage stress [35]. Nevertheless, these approaches require higher levels of automation [36] and the automation level needs to be appropriately communicated to the user [37]. Moreover, chat-bots lack a physical body. Previous studies have shown that emotional interaction between humans and artificial agents is increased whey they have a strong physical presence [38]. Therefore, the robot mirroring concept places a strong emphasis on embodied representations of affect. In this concept, the knowledge is derived from research on chatbots, but the focus relies less on high level cognitive conversations and more on the intuitive care of a physical body that can break.

For Robot Mirroring approaches to work, there are different assumptions. First, the robot design is sufficient to represent the user's emotional state. Second, the robot is able to clearly communicate his state to evoke empathetic understanding with the user. At the same time, feedback can be provided subtly to avoid unintended effects of self-tracking such as excessive awareness and over-trust. Third, users develop a stronger feeling of responsibility for agents of weak appearance, or through personalising the agent (pet) itself. Lastly, by attempting to mirror user emotional state to the agent, the user is encouraged to inadvertently practice self-compassion.

A. Assumption 1: a robot is sufficient to represent the user's emotional state

A robot is an embodied agent whose actions to cope with the environment depend largely on its body sensing capabilities. Because of this, software, behaviour, and interactions with the environment and other agents strongly depend on its hardware design. A large amount of research in the Human-Robot Interaction community uses out-of-the-shelf robots that were designed with a general purpose. The number of robots designed completely for a specific purpose is more limited. For example, one work proposed to design robots considering expected movements to interact with people even before the hardware design [40]. This is of utmost importance when

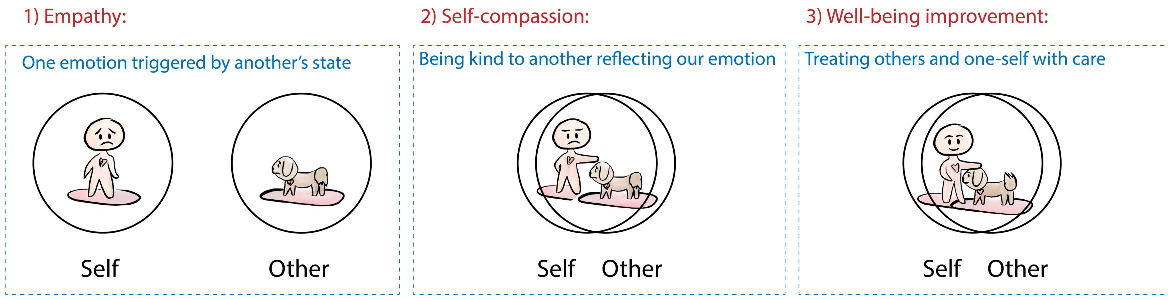

Fig. 1. The care-receiving pet mirrors and magnifies the user's state. Thus, by fostering empathy with the agent, we also indirectly foster empathy with one-self. In addition, by making the agent appear weaker, we also trigger compassionate action towards it. Therefore, since the agent is a mirror of the user's state, the user's actions can be interpreted as self-compassionate actions aimed to improve one-self's well-being. The circles represent the self and other in a similar manner as the Inclusion of Other in the Self (IOS) scale [39].

designing robots with a specific application. A robot representing user's emotional state requires specialised embodied characteristics if it possesses a larger body for example body language. Robotic-supported positive psychology interventions have already been demonstrated. Social robots have been shown to build the rapport and working alliance necessary for successful therapies. A social robot coach led to significant student well-being improvement after the intervention [9]. Moreover, interactions with empathetic chat-bots have been shown to lead to a higher percentage of positive mood reports [30].

B. Assumption 2: the robot is able to communicate his state to evoke empathetic understanding with the user

According to Strayer [41] and Davis [42], empathy definitions fall under two categories [43]: an affective or cognitive approach. In the cognitive approach, empathy is usually defined as *"the ability to cognitively understand anothers' feelings and thoughts through perspective-taking"*. In contrast, in the affective approach, empathy is defined as *"an affective response more appropiate to another's situation than one's own"*.

To show caring and empathetic behaviour, a robot must be attuned to the affective state of the user [44]. Several authors have proposed empathetic virtual agents to promote social and healthy behaviours in users. For example, an application of virtual drama to help children address bullying situations [45] or an Embodied Conversational Agent (ECA) that emulates the behaviour of a human therapist with a Wizard of Oz setup, to promote interest in healthy eating to the users [46]. More recently, a virtual reality game was proposed in which users controlled a superhero character to fulfil altruistic tasks [47], and a virtual human that acts as an interviewer designed to address problems of depression, anxiety, or post-traumatic stress disorder in patients [48]. However, a problem of virtual agents is the limited physical interaction resulting from the use of a computer screen and predefined dialogues to inform the virtual agent the user's emotional state [49]. The interaction with robots tends to be more natural by allowing to adapt the robot's behaviour to the user's emotional state during the interaction. Robots that accurately signal emotions have already been developed. For example, a robot with the form of a chimpanzee that mimics the facial expressions and head movements of users [50]; and an empathetic anthropomorphic robot (torso) that shows happiness, fear and neutral facial expressions as response to the affective speech signal of the user [51]. The results suggest that users perceived the robot to react more adequately to emotional aspects of a situation ("situation fit"). Furthermore, an expressive robot head was used to confirm that helping behaviours towards a robot can be increased by proactively adapting the robot's behaviour to the mood of the user [52]. Moreover, the role of empathy in long-term interaction between children and social robots was explored with a social robot, iCat, that plays chess with children [53]. The proposed social robot generated supportive behaviours according to the detected affective state of the children. In a more recent study, empathy towards a wearable robot was encouraged by providing visual and haptic feedback. The visual feedback communicated multiple valence and arousal levels. The haptic modality was used to enhance heart beat interoception [13].

Although previous work in artificial agent design has actively encouraged empathy to deal with own distress or to promote social behaviours, none of them explored what would be the effect of mirroring the user's health state in the artificial agent to promote self-helping behaviours.

C. Assumption 3: Customisation of the agent leads to a higher empathy and effectiveness of the intervention

Customisation is one of the strategies that can be used to boost empathy with an agent. People tend to assign more value to things they have customised (IKEA effect [54]), and to things they own (Endowment effect [55]). Moreover, customisation of artificial agents such as chatbots is deemed desirable when the required user effort does not exceed users' mental and motivational capacity and when ample resources are available [56]. In practice, customisation of physical robots is more challenging as it requires a certain level of technical knowledge and/or the interest to learn. Because of this, not the effects of robot customisation on empathy have been underexplored.

D. Assumption 4: users develop a stronger feeling of responsibility for agents of weak appearance

In a longitudinal study, children who took care of a robot were observed to treat the robot as if it was asleep when suddenly running out of battery [57]. This behaviour is more salient for robots that are somehow perceived as weaker than the user. Because of this, feeble robots can be used as supporting tools in education settings [58]. This concept is often referred as care-receiving robots. A more intuitive example is that people often overestimate their ability to cope with difficult situations [59], [60]. Thus, they do not take preventive strategies. However, when there is a person with whom we empathise, and this person is perceived as unable to cope or is akin to us (i.e., our child), we would tend to take caring strategies to help them adopt healthier behaviours [61]–[63]. Moreover, work on chatbots showed that a care-receiving entity would foster an increased sense of self-compassion [64].

E. Assumption 5: by mirroring the users emotional state to the agent, the user is encouraged to inadvertently practice self-compassion

Understanding emotions of others is important to foster compassionate behaviours to support each other [65]. Caring and supporting others often leads to a sense of satisfaction and fulfilment that is beneficial for both one's own and others well-being [66]. When speaking only about one's own emotions, the terms used are mindfulness for self-understanding, and self-compassion for understanding oneself and treating oneself with kindness, care, and concern in the face of negative life events [16]. If an artificial agent is representing one's affective state, being compassionate towards it would also mean being compassionate towards one-self indirectly. Self-compassion has been related to promoting successful self-regulation of health-related behaviours [15]. Moreover, framing medical problems and their treatment in ways that foster self-compassion may enhance people's ability to manage their health-related behaviour and deal with medical problems [17]. Thus, being self-compassionate often leads to a healthier lifestyle and may affect our sense of self-efficacy to achieve our behaviour change goals.

IV. DISCUSSION AND CHALLENGES

We have summarised the assumptions that are made in general quantification approaches and highlighted some of their unaddressed problems. We acknowledge that awareness is helpful to encourage behaviour change. However, simply feeding back a large number of raw features might overwhelm certain users and is prone to misinterpretation and confirmation biases. The feedback in these approaches is often presented in raw numerical form and can be complex and highly abstract. Artificial intelligence has the potential to provide a human-friendly interpretation. While AI is excelling at processing multiple streams of data in real-time, its interpretations are only as good as the data it was trained with. This may lead to situations in which the interpretation is not accurate enough (or incorrect). In this case, the users might either distrust the validity of the information and not benefit from the feedback at all; or accept the information as absolute, and blindly follow any advice they receive. Hence, if requested by the user, the agent should be able to explain how it came to a decision. Furthermore, humans are diverse. A one-fits-all assumption like "everybody benefits from increased awareness" ignores the challenges that vulnerable groups face. Thus, the robot mirroring approach proposes to consider individual traits of users to provide a personalised, relevant mirror.

To avoid the potential downsides of raw feedback, the robot mirroring framework is focused on subtle feedback without numbers in form of an embodied and empathetic agent that simply mirrors the physiological data of the user. This approach is also different from previous attempts of empathetic agents, as it only tries to reproduce the users state instead of relying on AI to classify the users emotion and initiating a predefined reaction. However, a challenge in this approach is the selection of efficient and appropriate channels to display accurately and subtly the users emotional state. The modality, its time resolution, sensitivity and interpretability have to be considered when implementing the feedback.

Another challenge of the robot mirroring framework is to increase the users interoceptive awareness and encourage interaction with the agent in order to foster empathy toward the agent. Since the agent is mirroring the user's emotional state, helping behaviours towards the agent must be promoted to create the inadvertent practice of self-compassion. Moreover, making a weak and customised pet agent can improve the closeness between the user and the pet, and induce stronger caring behaviours toward this agent. This customisation poses a challenge. The user must have a high range of personalisation options of his/her agent in order to achieve a greater degree of closeness.

V. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we assessed the assumptions related to self-tracking and well-being improvement. First, we assessed self-tracking quantification systems that provide feedback numerically. The robot mirroring framework suggests to create empathy and behaviour change by fostering helping behaviours towards an artificial agent mirroring the self; and indirectly, self-compassion. The assumptions and challenges in this framework were also outlined and discussed according to the literature.

As a future work, we will explore existing technologies and approaches that allow us to deal with the challenges identified in this work. Then, a robot will be designed to evaluate the different assumptions of the robot mirroring framework under different scenarios. Finally, the assumptions and challenges stated open discussions about new design alternatives to improve self-tracking for well-being.

REFERENCES

[1] J. Watanabe, Y. Ooishi, S. Kumano, M. Perusquia-Hernandez, T. Sato, A. Murata, and R. Mugitani, "Measuring, Understanding, and Cultivating Wellbeing in the Age of Technology | NTT Technical Review," *NTT Technical Review*, vol. 30,

no. 9, pp. 29–32, 2018. [Online]. Available: https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa6.html

[2] R. A. Baer, E. L. B. Lykins, and J. R. Peters, "Mindfulness and self-compassion as predictors of psychological wellbeing in long-term meditators and matched nonmeditators," *The Journal of Positive Psychology*, vol. 7, no. 3, pp. 230–238, May 2012, zSCC: 0000412 Publisher: Routledge _eprint: https://doi.org/10.1080/17439760.2012.674548. [Online]. Available: https://doi.org/10.1080/17439760.2012.674548

[3] E. T. K. van Dijk, J. H. Westerink, F. Beute, and W. A. IJsselsteijn, "Personal informatics, self-insight, and behavior change: A critical review of current literature," *Human–Computer Interaction*, vol. 32, no. 5-6, pp. 268–296, 2017.

[4] E. Kersten - van Dijk, "Quantified stress : toward data-driven stress awareness," Ph.D. dissertation, Industrial Engineering and Innovation Sciences, Apr. 2018, proefschrift.

[5] B. Mullen, *Self-Attention Theory: The Effects of Group Composition on the Individual*. New York, NY: Springer New York, 1987, pp. 125–146.

[6] C. Carver, "A cybernetic model of self-attention processes," *Journal of Personality and Social Psychology*, vol. 37, no. 8, pp. 1251–1281, Aug. 1979.

[7] R. Thaler and C. Sunstein, *NUDGE: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, 06 2009, vol. 47.

[8] E. Kersten van Dijk, J. Westerink, and W. Ijsselsteijn, "Self-tracking of stress: what are the effects?" in *European Conference on Ambient Intelligence*, 11 2014.

[9] S. Jeong, S. Alghowinem, L. Aymerich-Franch, K. Arias, A. Lapedriza, R. Picard, H. W. Park, and C. Breazeal, "A Robotic Positive Psychology Coach to Improve College Students' Wellbeing," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2020, pp. 187–194, zSCC: 0000008 ISSN: 1944-9437.

[10] E. Kersten van Dijk, F. Beute, J. Westerink, and W. Ijsselsteijn, "Unintended effects of self-tracking," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 04 2015.

[11] L. F. Barrett, "The theory of constructed emotion: an active inference account of interoception and categorization," *Social Cognitive and Affective Neuroscience*, vol. 12, no. 1, pp. 1–23, Jan. 2017, zSCC: 0000798. [Online]. Available: https://doi.org/10.1093/scan/nsw154

[12] H. Striepe, M. Donnermann, M. Lein, and B. Lugrin, "Modeling and evaluating emotion, contextual head movement and voices for a social robot storyteller," *International Journal of Social Robotics*, pp. 1–17, 2019.

[13] M. Perusquía-Hernández, M. C. Balda, D. A. Gómez Jáuregui, D. Paez-Granados, F. Dollack, and J. V. Salazar, "Robot mirroring: Promoting empathy with an artificial agent by reflecting the user's physiological affective states," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2020, pp. 1328–1333.

[14] D. Peña and F. Tanaka, "Human Perception of Social Robot's Emotional States via Facial and Thermal Expressions," *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 4, pp. 26:1–26:19, May 2020, zSCC: 0000223[s0]. [Online]. Available: https://doi.org/10.1145/3388469

[15] K. Raab, "Mindfulness, self-compassion, and empathy among health care professionals: A review of the literature," *Journal of health care chaplaincy*, vol. 20, pp. 95–108, 07 2014.

[16] K. NEFF, "Self-compassion: An alternative conceptualization of a healthy attitude toward oneself," *Self and Identity - SELF IDENTITY*, vol. 2, pp. 85–101, 04 2003.

[17] M. Terry and M. Leary, "Self-compassion, self-regulation, and health," *Self and Identity*, vol. 10, pp. 352–362, 07 2011.

[18] K. Birnie, M. Speca, and L. Carlson, "Exploring self-compassion and empathy in the context of mindfulness-based stress reduction (mbsr)," *Stress and Health*, vol. 26, p. 359, 12 2010.

[19] L. Munasinghe, Y. Yuan, E. Faught, N. Willows, and P. Veugelers, "Parental encouragement of healthy lifestyles for their children and personally caring about healthy lifestyles is positively associated with children using vitamin d supplements," *Nutrients*, vol. 8, p. 596, 09 2016.

[20] A. Millings, J. Morris, A. Rowe, S. Easton, J. Martin, D. Majoe, and C. Mohr, "Can the effectiveness of an online stress management program be augmented by wearable sensor technology?" *Internet Interventions*, vol. 37, 05 2015.

[21] D. Bakker and N. Rickard, "Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: Moodprism," *Journal of Affective Disorders*, vol. 227, 11 2017.

[22] B. Pardamean, H. Soeparno, A. Budiarto, B. Mahesworo, and J. Baurley, "Quantified self-using consumer wearable device: Predicting physical and mental health," *Healthcare Informatics Research*, vol. 26, pp. 83 – 92, 2020.

[23] W. Wu, S. Pirbhulal, H. Zhang, and S. Mukhopadhyay, "Quantitative assessment for self-tracking of acute stress based on triangulation principle in a wearable sensor system," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 05 2018.

[24] J. Prochaska and W. Velicer, "The transtheoretical model of health behavior change," *American journal of health promotion : AJHP*, vol. 12, no. 1, p. 38—48, 1997.

[25] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," *Advances in Behaviour Research and Therapy*, vol. 1, no. 4, pp. 139–161, 1978.

[26] C. Gilbert, "Clinical applications of breathing regulation: Beyond anxiety management," *Behavior Modification*, vol. 27, no. 5, pp. 692–709, 2003, pMID: 14531162.

[27] B. Yu, L. Feijs, M. Funk, and J. Hu, "Breathe with touch: A tactile interface for breathing assistance system," in *Human-Computer Interaction – INTERACT 2015*, J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, Eds. Cham: Springer International Publishing, 2015, pp. 45–52.

[28] J. Klayman, "Varieties of confirmation bias," *Psychology of Learning and Motivation*, vol. 32, pp. 385–418, 1995.

[29] E. T. van Dijk, J. H. D. M. Westerink, F. Beute, and W. A. IJsselsteijn, "In Sync: The Effect of Physiology Feedback on the Match between Heart Rate and Self-Reported Stress," *BioMed Research International*, vol. 2015, p. e134606, Jun. 2015, zSCC: 0000016 Publisher: Hindawi. [Online]. Available: https://www.hindawi.com/journals/bmri/2015/134606/

[30] A. Ghandeharioun, D. McDuff, M. Czerwinski, and K. Rowan, "Towards Understanding Emotional Intelligence for Behavior Change Chatbots," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Sep. 2019, pp. 8–14, iSSN: 2156-8111.

[31] R. Parasuraman, T. Sheridan, and C. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, May 2000, zSCC: 0003662 Conference Name: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans.

[32] P. Brandtzaeg, M. Skjuve, K. Dysthe, and A. Følstad, "When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots," in *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI'21, Yokohama, Japan, 2021, p. 13, zSCC: NoCitationData[s0]. [Online]. Available: https://doi.org/10.1145/3411764.3445318

[33] "Wysa - Everyday Mental Health." [Online]. Available: https://www.wysa.io/

[34] 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, 2020, zSCC: 0000012. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.03061/full

[35] L. Medeiros, C. Gerritsen, and T. Bosse, "Towards Humanlike Chatbots Helping Users Cope with Stressful Situations," in *Computational Collective Intelligence*, ser. Lecture Notes in Computer Science, N. T. Nguyen, R. Chbeir, E. Exposito, P. Aniorté, and B. Trawiński, Eds. Cham: Springer International Publishing, 2019, pp. 232–243.

[36] A. Ghandeharioun, D. McDuff, M. Czerwinski, and K. Rowan, "EMMA: An Emotion-Aware Wellbeing Chatbot," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Sep. 2019, pp. 1–7, iSSN: 2156-8111.

[37] X. Yang and M. Aurisicchio, "Designing Conversational Agents: A Self-Determination Theory Approach," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–16, zSCC: 0000000. [Online]. Available: https://doi.org/10.1145/3411764.3445445

[38] S. S. Kwak, Y. Kim, E. Kim, C. Shin, and K. Cho, "What makes people empathize with an emotional robot?: The impact of agency and physical

embodiment on human empathy for a robot," in *2013 IEEE RO-MAN*, Aug. 2013, pp. 180–185, zSCC: 0000082 ISSN: 1944-9437.

[39] A. Aron, E. N. Aron, and D. Smollan, "Inclusion of Other in the Self Scale and the structure of interpersonal closeness." *Journal of Personality and Social Psychology*, vol. 63, no. 4, pp. 596–612, 1992.

[40] G. Hoffman, O. Zuckerman, G. Hirschberger, M. Luria, and T. Shani-Sherman, "Design and evaluation of a peripheral robotic conversation companion," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2015, 03 2015.

[41] J. Strayer, *Affective and cognitive perspectives on empathy*. Cambridge University Press, 1987.

[42] M. Davis, *Empathy: A Social Psychological Approach*. Westview Press, 02 1996.

[43] I. Leite, A. Pereira, S. Mascarenhas, C. Martinho, R. Prada, and A. Paiva, "The influence of empathy in human–robot relations," *International Journal of Human-Computer Studies*, vol. 71, no. 3, pp. 250–260, 2013.

[44] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[45] R. Aylett, S. Louchart, J. Dias, A. Paiva, and M. Vala, "Fearnot! - an experiment in emergent narrative," in *IVA*, 2005.

[46] F. D. Rosis, A. Cavalluzzi, I. Mazzotta, and N. Novelli, "Can embodied conversational agents induce empathy in users ?" in *AISB ' 05 Virtual Social Characters Symposium*, 2005.

[47] R. S. Rosenberg, S. L. Baughman, and J. N. Bailenson, "Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior," *PLOS ONE*, vol. 8, no. 1, pp. 1–9, 01 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0055003

[48] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, G. Lucas, S. Marsella, F. Morbini, A. Nazarian, S. Scherer, G. Stratou, A. Suri, D. Traum, R. Wood, Y. Xu, A. Rizzo, and L.-P. Morency, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, ser. AAMAS '14. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2014, p. 1061–1068.

[49] A. Paiva, I. Leite, H. Boukricha, and I. Wachsmuth, "Empathy in virtual agents and robots: A survey," *ACM Transactions on Interactive Intelligent Systems*, vol. 7, pp. 1–40, 09 2017.

[50] L. Riek, P. C. Paul, and P. Robinson, "When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry," *Journal on Multimodal User Interfaces*, vol. 3, pp. 99–108, 2009.

[51] F. Hegel, T. Spexard, B. Wrede, G. Horstmann, and T. Vogt, "Playing a different imitation game: Interaction with an empathic android robot," in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 56–61.

[52] B. Gonsior, S. Sosnowski, M. Buß, D. Wollherr, and K. Kühnlenz, "An emotional adaption approach to increase helpfulness towards a robot," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 2429–2436.

[53] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic robots for long-term interaction," *International Journal of Social Robotics*, vol. 6, pp. 329–341, 08 2014.

[54] M. I. Norton, D. Mochon, and D. Ariely, "The IKEA effect: When labor leads to love," *Journal of Consumer Psychology*, vol. 22, no. 3, pp. 453–460, Jul. 2012. [Online]. Available: http://doi.wiley.com/10.1016/j.jcps.2011.08.002

[55] J. L. Knetsch, "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *The American Economic Review*, vol. 79, no. 5, pp. 1277–1284, 1989, zSCC: 0001396 Publisher: American Economic Association. [Online]. Available: https://www.jstor.org/stable/1831454

[56] R. Zhang, K. E. Ringland, M. Paan, D. C. Mohr, and M. Reddy, "Designing for Emotional Well-being: Integrating Persuasion and Customization into Mental Health Technologies," in *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI'21, Yokohama, Japan, 2021, p. 13, zSCC: NoCitationData[s0]. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445771

[57] F. Tanaka, A. Cicourel, and J. Movellan, "Socialization between toddlers and robots at an early childhood education center," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 17 954–8, 11 2007.

[58] F. Tanaka and T. Kimura, "Care-receiving robot as a tool of teachers in child education," *Interaction Studies - INTERACT STUD*, vol. 11, 06 2010.

[59] N. Weinstein, "Unrealistic optimism about susceptibility to health problems," *Journal of Behavioral Medicine*, vol. 5, pp. 441–460, 1982.

[60] ——, "Unrealistic optimism about future life events," *Journal of Personality and Social Psychology*, vol. 39, p. 806–820, 1980.

[61] J. Dovidio and L. Penner, *Helping and Altruism*. Wiley Online Library, 2004, p. 247–280.

[62] V. Konecni and E. Ebbesen, "Effects of the presence of children on adults' helping behavior and compliance: Two field studies," *The Journal of Social Psychology*, vol. 97, pp. 181–193, 12 1975.

[63] J. Dovidio, "Helping behavior and altruism: an empirical and conceptual overview," *Advances in Experimental Social Psychology*, vol. 17, pp. 361–427, 1984.

[64] M. Lee, S. Ackermans, N. van As, H. Chang, E. Lucas, and W. IJsselsteijn, "Caring for Vincent: A Chatbot for Self-Compassion," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–13, zSCC: 0000043. [Online]. Available: https://doi.org/10.1145/3290605.3300932

[65] M. Paciello, R. Fida, L. Cerniglia, C. Tramontano, and E. Cole, "High cost helping scenario: the role of empathy, prosocial reasoning and moral disengagement on helping behavior," *Personality and Individual Differences*, vol. 55, pp. 3–7, 07 2013.

[66] C. Saltzman, "Empathy in patient care–antecedents, development, measurement, and outcomes," *Psychodynamic Practice*, vol. 14, pp. 342–345, 08 2008.