Examining Attachment to Robots: Benefits, Challenges, and Alternatives

THERESA LAW, MEIA CHITA-TEGMARK, NICHOLAS RABB, and MATTHIAS SCHEUTZ, Tufts University, United States

36

Potential applications of robots in private and public human spaces have prompted the design of so-called "social robots" that can interact with humans in social settings and potentially cause humans to attach to the robots. The focus of this article is an analysis of possible benefits and challenges arising from such human-robot attachment as reported in the HRI literature, followed by guidelines for the use and the design of robots that might elicit attachment bonds. We start by analyzing the potential benefits for humans becoming attached to robots, which might include increased natural interaction, effectiveness and acceptance of the robot, social companionship, and well-being for the human. Turning to the potential risks associated with human-robot attachment, we discuss the possibly suboptimal use of the robot in the most benign cases, but also the potential formation of unidirectional emotional bonds, and the potential for deception and subconscious influence of the robot on the person in more severe cases. The upshot of the analysis then is a recommendation to reconceptualize relationships with social robots in an attempt to retain potential benefits of human-robot attachment, while mitigating (to the extent possible) its downsides.

CCS Concepts: • **Human-centered computing** → *Collaborative interaction*; • **Computer systems organization** → **Robotics;**

Additional Key Words and Phrases: Human-robot interaction, attachment, social bond, emotional bond

ACM Reference format:
Theresa Law, Meia Chita-Tegmark, Nicholas Rabb, and Matthias Scheutz. 2022. Examining Attachment to Robots: Benefits, Challenges, and Alternatives. *Trans. Hum.-Robot Interact.* 11, 4, Article 36 (September 2022), 18 pages.
https://doi.org/10.1145/3526105

1 INTRODUCTION

The potential applications of robots in private and public human spaces has prompted the design of so-called "social robots" that directly interact with humans in social settings. These types of robots are often purposefully designed to connect with people by eliciting emotional reactions from their human interactants [7]. Various studies have shown that such social robots prompt people to develop emotional bonds [12, 63]. Among the observed phenomena are human interactants anthropomorphizing the robot [24], reacting emotionally to its behaviors [43], and experiencing higher quality of life when around it [60]. In human-human interactions, forming strong emotional bonds often leads to the development of attachment relations, which manifest themselves in terms

Authors' address: T. Law, M. Chita-Tegmark, N. Rabb, and M. Scheutz, Tufts University, 200 Boston Ave., Medford, MA; emails: {theresa.law, mihaela.chita_tegmark, nicholas.rabb, matthias.scheutz}@tufts.edu.

2573-9522/2022/09-ART36
https://doi.org/10.1145/3526105

of characteristic emotions, cognitions, and behaviors (e.g., see [2] for one of the first characterizations of attachment and associated disorders).

While it is currently unclear whether humans can or will reach similar levels of attachment with robots as they do with humans, the possibility of forming attachments with robots in the first place—especially when these relations are unidirectional [52]—raises important ethical questions for HRI research and design: Would robots inadvertently become a burden for users who, due to attachment, feel they would need to "take care" of the robot? Will robots be able to sustain unique relationships with their users? Will attachment to robots lead to the reduction or replacement of human contact? Does robot companionship inevitably lead to deception? What is the scope of persuasive techniques that the robot should be allowed to use to influence the user?

In this article, we will discuss these questions and the consequences of human-robot attachment. We first identified the most common benefits and challenges mentioned in the HRI literature with regard to attaching to robots. Benefits identified include (1) increased natural interaction [29], (2) increased effectiveness and acceptance of the robot [29, 73], (3) supplementing social companionship [8], and (4) increased well-being for the human [1]. Risks identified include (1) potentially suboptimal use of the robot [52], (2) potential formation of unidirectional emotional bonds [18, 29, 52], (3) potential for deception [46, 56, 57], and (4) potential subconscious influence of the robot on the person [29]. We then generated thought experiments in the form or scenarios that showcase both benefits and challenges. We discuss each scenario focusing on tradeoffs, emphasizing the idea that attachment to robots is neither absolutely beneficial nor absolutely detrimental to the well-being of humans. Rather than taking a side for (e.g., [73]) or against (e.g., [65]) attachment to robots, we argue that instead we need to reconceptualize our relationships with social robots. The article contributes conceptual alternatives for thinking about relationships with robots that avoid some of the pitfalls of attachment and have the potential to guide the design of social robots toward more beneficial outcomes for the humans.

2 BACKGROUND

Several frameworks in the social sciences have formally described attachment as a series of characteristic emotions, cognitions, and behaviors. "Attachment" denotes a very strong bond someone has with an attachment figure who, in their mind, provides so much security and safety to the individual that they are very dependent on the attachment figure—to the point where separation from that figure would result in serious psychological distress [2, 25]. *Psychological Attachment Theory* was pioneered as a study of infants' attachment to their caregivers [6] and has since been extended to encompass people's attachment to pets [74], symbols or deities [39], or objects [33]. Specifically, an attachment bond would form if the caregiver met the infant's needs for security and affectionate comfort when distressed, with these two functional roles played by the attachment figure being labeled the *secure base* and *safe haven* functions. Attached infants were also observed *seeking proximity* to their attachment figures and displaying *separation distress* in their absence [74].

In contrast to the social sciences, the HRI literature uses the term "attachment" more loosely at times: sometimes it is discussed in the context of psychological attachment theory [16], while other times no specific definition of attachment is provided. Overall, many HRI researchers have utilized Norman's (2004) definition of attachment—the cumulative sum of all the emotional episodes that a human has with a robot [45]—while others have called for a more stringent and formal definition of patterns of cognitions and behaviors, akin to those in psychology research [16]. For example, [16] suggests the term "attachment" should be reserved for relationships that feature the four different behaviors and functions identified in social psychology. They recognize that social robotics is, at the moment, not advanced enough for any relationship with a robot to fulfill all of these features and thus fully qualify as attachment; however, they argue that, if viewed instead as a model for defining a spectrum of bonds rather than benchmarks to be achieved, these features are useful for identifying the degree to which bonds between humans and robots resemble attachment bonds.

Since there is currently no agreed-upon formal definition of what it could mean for a human to be attached to a robot, it is also unclear whether human-robot bonds can reach the level and intensity found in human attachment. While a deep exploration of the meaning and different levels and types of attachment to robots is beyond the scope of this article (we explore the feasibility and nature of attachment to robots in [50]), it is worth noting that the field of HRI has started making increasing reference to attachment in its analysis of human-robot bonding, even if the researchers do not give a strict definition for the concept of attachment. For example, [69] refers to emotional attachment when describing the first contact of children and adults with the robotic pet dog AIBO. Another study discusses an exploration of emotional attachment that students might develop toward a LEGO robot they built for a class during a 2-month period [28]. The authors conclude that the emotional bonds resulting from this were different from typical attachment as students showed few negative feelings related to the loss of the robot. A review of the literature on "robopets" in care homes describes residents developing emotional attachments to the robopet, with residents proclaiming their love for the robot in spite of it being an object [1]. When analyzing the caregiving role in HRI, Kim et al. [38] measure attachment to the robot through items such as "I feel emotionally connected to this robot," developed by [53]. While one longitudinal study only found indirect links between emotional attachment to a robot and emotional deception by the robot, the robot still had quite limited capabilities and the participants experienced a low exposure time to and infrequent interactions with the robot [67]. It is possible that the effects we will be discussing in this article require longer and more regular exposure time to more sophisticated robots. However, it is important to understand how attachment plays into a person's interactions with a robot, as one study has shown that feeling more securely attached around a robot leads to greater trust in it [21].

Across studies, there is a pervasive presence of warnings about robots' potential to foster attachment with their human counterparts and lead to unethical situations. Sharkey and Sharkey discuss this in depth and apply the psychological attachment framework to situations where robots may care for children [58]. Turkle relays worries about interactions she observed with children who ascribed serious levels of intentionality, emotion, and cognition to robots such as My Real Baby, warning that "relational artifacts" understand nothing but push our "Darwinian buttons" and cause people to respond as if they were in a relationship [65]. Carpenter offers anecdotal evidence of soldiers bonding with bomb-diffusing robots [12]. Even Norman, in his aforementioned 2004 book on emotional objects, briefly warns of the ethical implications of "humanoid robots that have emotions and to which people might form strong emotional attachments" [45].

As HRI attachment research continues, there is a need to thoroughly explore the implications of such attachment. Below, we examine in more detail potential benefits, challenges, and ethical concerns related to bonds between humans and robots (e.g., [29, 52, 58, 65]). To make our case, we will lay out a series of plausible scenarios inspired by ideas and results found in the HRI literature. These specifically tailored scenarios will help us explore situations that intentionally feature benefits and challenges in the same breath. We have found through our own investigations that much of the nuance in HRI studies—most notably with social robots—comes when robots are simultaneously beneficial and challenging. Exploring this tension is crucial to bringing a deeper understanding of a future where robots are more popularly widespread. We cannot say whether or not these scenarios would be the norm, or edge cases, but regardless of frequency, they are worth studying as they combine some of the most discussed results from HRI studies.

3 HUMAN-ROBOT ATTACHMENT SCENARIOS

In this section, we will discuss a number of cases where it is possible that a human could become attached to a robot. Because attachment formation is a possibility, it is worth considering what the effects of attachment, both positive and negative, could be in that situation. We are not claiming that these scenarios represent situations that are unique to HRI; rather, we posit these as situations that are *feasible* in HRI. It is in fact a testament to the field that these scenarios that are common in human-human interaction are even being considered plausible for human-robot interaction. Recall that these scenarios are not actual experimental case studies, but rather plausible thought experiments assembled from important HRI results.

Scenarios were constructed based on two bodies of literature: (1) the social psychological attachment literature that was used to systematically motivate these hypothetical attachment relationships (see [50] for a comprehensive synthesis) and (2) the HRI literature from which we identified the potential benefits and risks of attaching to robots. While the latter part is the focus of this article's analysis, consideration of attachment components was important for scenario construction as well. In terms of motivating and defining the attachment, we imagined the robot as satisfying criteria for the following attachment functions: secure base and safe haven. The secure base function is fulfilled when the attachment figure can provide safety in a way that encourages exploration. Safety could be achieved by the attachment figure being able to provide physical, intellectual, or social resources or protection. The safe haven function is fulfilled when the attachment figure can soothe distress—either physically or emotionally [50].

Each of the scenarios contains aspects of these two functions. Scenario 1 features the robot providing physical resources via cleaning, opening space for the person to safely do other things (secure base), and could provide emotional soothing (safe haven) by being friendly. Scenario 2 features the robot providing social resources by being the person's friend (secure base) and emotional soothing for the same reason (safe haven). Scenario 3 outlines a robot that meets many physical and social needs of the person (secure base), as well as providing emotional comfort by being friendly (safe haven). Finally, scenario 4 describes a robot that provides information to the person (secure base) and could be argued to soothe them emotionally and physically by being motivating and encouraging them to eat when hungry (safe haven). With these attachment criteria present, we made the assumption that an attachment bond had been formed in each scenario. The robots described in the scenarios, while fulfilling attachment functions, may not engender strong attachment, but at least some weaker form of attachment above that from simple objects or technologies. From there, we examined a benefit and a risk of attachment that may be present in each.

The benefits and challenges discussed in these scenarios were identified in the literature on human-robot attachment. Benefits identified include (1) increased natural interaction [29], (2) increased effectiveness and acceptance of the robot [29, 73], (3) supplementing social companionship [8], and (4) increased well-being for the human [1]. Risks identified include (1) potentially suboptimal use of the robot [52], (2) potential formation of unidirectional emotional bonds [18, 29, 52], (3) potential for deception [46, 56, 57], and (4) potential subconscious influence of the robot on the person [29]. Each scenario touched upon at least one of these benefits and at least one of these challenges. While not exhaustive, these scenarios provide illustrative examples that help to synthesize the work of various fields (social psychology, HRI, attachment), an approach that has been previously used to help anticipate, rather than belatedly react to, potential real-world social dynamics [4].

3.1 Something to Take Care Of

Scenario 1: Something to take care of

A person buys a social robot to help her with chores around the house. The robot is designed to be very cute and friendly. As time passes, the person grows attached to the robot. She starts feeling bad that the robot is doing so much work, and goes back to doing some of the chores herself to "help it out."

A positive effect that could come from attachment is incentivizing long-term acceptance of the robot. An attachment relationship, fueled by the robot's social design, could help motivate a person to continue using it and therefore experience the benefits it can offer after the novelty effect wears off [29]. In instances where the robot malfunctions or otherwise fails to live up to expectations, or if the human needs to adapt their own behavior (e.g., by rephrasing a command), this increased acceptance can be especially important [11]. In this scenario, the robot has a clear role that can benefit the person, but its service would be wasted if the person rejected it the moment it did not function perfectly. The attachment relationship could help the user feel inclined to continue using the robot.

However, attachment can also result in suboptimal use of the robot's intended function. Though social robots are not purely utilitarian tools like factory robots, they still often have a practical, functional role, like cleaning, tutoring, therapy, coaching, and so forth. If the robot has a service role, attachment may hinder a user's comfort with allowing the robot to continue serving. The above scenario, in which the person took over the job the robot was purchased and designed to do, has been reported in real-life case studies. One paper found that people reported doing a pre-cleaning of their living space before their robot vacuum cleaned so that it has an easier time, or had even felt empathy for it and given it a "day off" [63]. This behavior goes against the intended purpose of the robot vacuum cleaner, which is to make the human's life easier and involve less cleaning. Because this effect is already being seen with non-social robots, such as robot vacuum cleaners, it is possible that it could be even stronger when the robot in question is explicitly designed to be social. Carpenter's investigations into military robots with humanoid designs—whose intended demise on the battlefield prompts grief in human personnel—offer preliminary evidence to this point [12].

Suboptimal use could also arise from developing unhealthy attachment relationships to robots. Unhealthy attachment styles can be categorized by patterns of avoidance and anxiety. Avoidance can be characterized by distancing and disengaging oneself from the attachment figure, and anxiety can be characterized by constantly seeking out the attachment figure and feeling extreme separation distress when apart [44]. Hertlein and Twist [27] described how attachment relationships with technologies could manifest along anxious/avoidant dimensions; besides the low-anxiety/low-avoidance dimension of secure attachment, all of these categories result in suboptimal use of the technologies. Low-anxiety/high-avoidance attachment can result in underusing the technology, while high anxiety/low avoidance can result in over-reliance. High-anxiety/high-avoidance bonds could result in disorganized and unpredictable attachment behaviors; perhaps a person with a robot vacuum cleaner is unhealthily attached and places the robot in a closet, never to be used, because she is too afraid of breaking it. Or perhaps the person follows the robot around the whole time it cleans to make sure that nothing breaks it, therefore losing the advantage that the robot offers of being able to do other things while the robot cleans. Unhealthy attachment patterns that result in suboptimal use and are seen with other technologies like smartphones could occur with social robots as well.

3.2 Treating Everyone the Same

Scenario 2: Treating everyone the same

A child receives a social companion robot as a present. When she plays with the robot, it responds to her in seemingly intelligent ways. She grows attached to the robot and sees it as her friend. She believes that she and the robot have built a strong and unique bond. One day she has a friend over, and her robot responds to her friend in the exact same manner that it responds to her. The child is upset that the robot is acting like it is equally good friends with the newcomer as it is with her, despite the history between the two. She feels betrayed by the robot, and stops wanting to play with it.

Though social robots can exhibit simple behaviors that people tend to interpret as social (e.g., gaze following), they are far from being meaningful social others in interactions. These robots usually lack any internal computational models of social others and social interactions or norms and have no social emotions [52]. Turkle [65] describes a social robot by saying: "Like other 'relational artifacts' its ability to inspire relationship is not based on its intelligence or consciousness, but on the capacity to push certain 'Darwinian' buttons in people (making eye contact, for example) that cause people to respond as though they were in a relationship." This makes social robots the source of some of the bigger discrepancies between actual and perceived capabilities: there is a big gap between what social robots can do and what people perceive they can do. Very simple behaviors can lead to massive unwarranted assumptions regarding the robot's capabilities, intelligence, and agency [52]. For example, social robots can say things without understanding what they are saying: they can respond to speech without any internal model for what the words refer to in the real world, they can express emotions without any notions of what those emotions are, and they can recognize emotions without understanding their significance. Similarly, a robot may have a poor model for what humans are. For a vacuum robot, a human is just an obstacle that needs to be avoided. Such a simple model does not allow a robot to form relationships with people the way people may form relationships with the robot [52]. As Nyholm and Frank [46] point out, the design of robots where one of the primary goals is to produce affective reactions from people, such as sex robots and companion robots, is particularly susceptible to this type of deception. They describe how the people using these types of robots are likely motivated to look for agents that act like they care about the person and fall into the anthropomorphism tendencies that lead to believing that the robot's "Darwinian button pushing" behaviors actually indicate that it *does* care about the person.

For the child interacting with her social robot, the "pushing of these Darwinian buttons" led to a relatively strong attachment. The robot was able to provide her with a companion that could play with her in a more responsive way than her other toys. This could be especially beneficial at times when no other people are available to play with the child. In times like that, the robot can act as a supplement to human companionship. Humans form secondary attachment bonds, beyond their primary human-human bonds, with many types of non-human agents, such as pets [74], deities [23], and objects or technologies [27, 33, 34]. When other people are absent, these secondary attachment figures can help to lower stress [23]. Robots, too, could play this role of secondary attachment figure. By virtue of straddling the line between objects and agents, robots possess the ability to be reliably present when primary attachment figures (humans, pets) may not be—a key feature leading to the development of secondary attachments [33]. In this way, a robot would not replace any of the social interactions in the child's life, but rather be available as a companion when other people are not. Additionally, being the child's toy and companion, there may have been times when the child's attachment to the robot grew out of taking care of it in make-believe games. Caring for and showing compassion to an artificial agent has been shown to increase a person's self-compassion [42]. Therefore, the way that the child plays with and attaches to the robot may be beneficial for her own mental health and self-image.

However, we see this attachment become a problem when the child is upset that the robot responds to her friend in the exact same way that it responds to her, and she realizes that their relationship is not unique. As mentioned above, the robot was designed to elicit social perceptions and seem like it was an advanced social other. However, its actual capabilities are far less advanced. The robot does not have a model of the child that would enable it to form a unique relationship with her based on their shared history. Rather, it responds to everyone in the same manner. Uniqueness is one of the hallmarks of attachment; what inspired Bowlby's attachment research was observing children in post–World War II orphanages: even though all their needs were met, they were poorly adjusted to social life. Bowlby concluded that the reason had to do with the lack of consistent presence from the people who were caring for these kids; the children could not form unique bonds to people and thus could not attach, because the caregivers were constantly rotated and casually switched in or out. He observed that attachment goes beyond having your needs met by hundreds of individuals and is formed through really developing unique, deep, irreplaceable bonds with a handful of individuals [55]. From Bowlby's work, we therefore see that issues can arise when attachment figures and attachment relationships are treated as things that can be casually switched in or out.

While human-human relationships are bidirectional because both parties are able to form mental models of the others and can interpret and understand the social and emotional cues of the other person or people, the child in this scenario formed a unidirectional emotional bond with the robot. The robot's limited abilities were enough to elicit emotional responses from the child and caused her to overestimate its abilities; however, the robot itself was not able to understand or reciprocate those intense feelings. This led to the child feeling hurt and betrayed. The robot, though, is oblivious to this. Rather than feeling the guilt or shame that a person might if he knew that he had disappointed someone, the robot will be completely unaware of this dynamic. In this manner, the unidirectional bonds of attachment could be quite harmful to the person [52].

3.3 Human-like Companionship

Scenario 3: Human-like companionship

An elderly person with early stages of dementia is gifted an assistive robot by his family. At first, the person does not know how exactly to utilize the robot or make it work; however, the robot communicates with natural language and is able to talk to the person to help him understand its capabilities and how to use it. The robot's natural language capability also allows the person to chat with it, which the person enjoys doing. When talking with the robot, the person occasionally believes he is interacting with a real person when he is experiencing a period of intensified dementia symptoms.

In this scenario, natural social interaction between the person and the robot can lead to attachment. While not a perfectly causal route to attachment, the social psychological attachment literature, which designates the secure base and safe haven functions as components of attachment [16, 74], seems to indicate that even partially met secure base and safe haven functions can lead to a degree of attachment. The safe haven function may be viewed as having a sub-component related to the ability of an attachment figure to soothe mental or emotional distress [50]. A robot capable of socializing with an elder with dementia, who may also be socially isolated because of societal norms or his mental illness, would likely highly value interaction with the robot. Therefore, the more attributes a social robot has that could aid in its ability to soothe mental distress, the more likely it may be to engender a degree of attachment.

The robot has a number of attributes that can help the interaction be comfortable rather than awkward for the person. Communicating through natural language use rather than, say, communicating via logic or code helps ease conversation between the two parties. The social aspect of the robot can further help the feeling of natural interaction. Humans are so attuned to social cues that we ascribe social intentions to something as simple as moving shapes [26]. With a social robot that can move its body and face in emotionally expressive ways, the perception and attribution of social agency would only be further amplified, which could translate into an increased ease of social interaction. For a person who may be technology averse, this ease could be vital in helping the person attach to the robot.

Natural interactions may lead to a positive cycle of naturalness increasing the likelihood of attachment, and attachment, in turn, allowing for more natural interaction. It is known that deep and long-lasting emotional connections of attachment are an essential part of human life [2, 6]. As such, these bonds in human relationships have been shown to allow for the easiest and most natural interactions [22]. Therefore, a person who attaches to their social robot is likely to be able to reap these interaction benefits as well [48]. Additionally, attachment, and specifically healthy feelings of secure attachment, has been shown to lead to trust [21, 62]; increased trust in a robot could prevent a person from over-analyzing the robot's behavior in search of, or in preparation for, failure. Without that worry, the interactions may flow more naturally. This ease of interaction may grow over time: the inclination for continued use brought about by the attachment will familiarize the person to the robot's behaviors, making them increasingly comfortable with using it.

However, these natural interactions may have been built on a basis of deception [57]. As previously discussed, robots need to do very little to push our "Darwinian buttons" and engage us emotionally. This biases us to read their often simple behavior as more complicated and meaningful than it is, which can in turn lead us to believe that they are more capable than they truly are. If we grow attached to these robots, we may be *even more* inclined to believe this. At some level, playing along with these tricks is harmless and akin to how humans interact with inanimate objects [48]. However, these social "tricks" can have the potential to negatively impact people with whom the robots interact.

The cognitively impaired or cognitively immature are the most vulnerable to these adverse effects: they have the highest risk of not being able to distinguish between genuine social interaction and the artifice of a social robot's "social" interaction. Children have been repeatedly shown to ascribe mental states to robots and believe robots are capable of having emotions and cognitive abilities [32, 69]. On the other end of the age spectrum, seniors residing in nursing homes behaved toward their robotic assistant in a similar manner to the aforementioned children. Though the elderly admitted that it was just a machine, many named it, talked to it, praised it, cared about it, and asked its opinion—anthropomorphizing the robot in a manner that could indicate attachment [68]. If at some level, these people—adults or children—believe that the robot is actually or nearly as capable as they believe it to be, they are being deceived by the robot. Though this may be harmless at times, it is something that robot designers should be cautious of. Similar to unidirectional emotional bonds, it could lead to significant emotional harm for the person involved. In the scenario described above, the elderly person with dementia occasionally believes that he is talking to a real person when he is interacting with a robot. Realizing otherwise, perhaps in moments of lucidity, could result in feelings of betrayal, distress, disappointment, and so forth. This delusion implies that he believes he is forming a bond with another agent who has a rich and complex understanding of the world, and that the relationship between the two of them is bidirectional. In reality, the relationship is unidirectional and the bidirectionality is a deception brought on by the robot activating our social heuristics. If the person finds this out, he may feel betrayed by the robot. This issue of deceiving people in vulnerable states has been discussed by Nyholm and Frank [46] in regard to sex robots, another field where the goal of the robot may be to induce a type of attachment between the person and the robot. In this scenario, the elderly person's mental and emotional vulnerability can leave them similarly at risk to the potential harm that accompanies this deceit. Further, the authors discuss how vulnerable populations may lose vital human contact if they become too attached to a robot. The person in this scenario may be particularly at risk for that, because he occasionally believes that he *is* interacting with another human, and thus might not seek out other social interactions.

3.4 Knowing You Better Than You Know Yourself

Scenario 4: Knowing you better than you know yourself

An adult buys a robot that acts as a weight loss coach. It provides motivational support, recommends different workout routines and healthy meals, and provides a demonstration of some of the moves. It also has a screen where it can project live workout classes and an activity tracker. During the workout classes, there are commercials where the robot advertises other products from the company that made the robot. When it provides motivational support, it tells the person that the company believes in him. Additionally, it is programmed to use health and emotional data gathered from his activities to regularly recommend recipes when he is predicted to be most hungry or frustrated. The recipes are made from ingredients sold by food companies that the robot's company partners with. The person likes the robot and follows its workout and nutrition recommendations. After a few weeks, the person begins to see positive results of the regimen.

Committing to a new weight loss routine requires a person to be motivated to push through the tough moments. Becoming engaged in the process and finding encouragement from others can help provide that motivation. Unlike a human weight loss coach, a robotic weight loss coach could live in a person's home and tirelessly provide that encouragement whenever the person needs it [17, 24, 35]. However, a person who does not care about the robot with whom they are working might find it hard to be motivated to put in the necessary effort. Feeling attached to the robot could provide that motivation. This effect has been seen in another field with students and academic motivation: attachment to teachers was linked to higher motivation [41]. Additionally, evidence indicates that social robots have the capacity to encourage growth-mindset-oriented behavior more so than other more self-guided methods. One study found that a brief, one-time interaction with a social robot was enough to elicit more exploratory, creative behavior from undergraduates than when the students were shown a comparable PowerPoint presentation [31]. Additionally, because the robot is working with the person to complete the person's goals, the person may come to view the robot as their teammate in this mission. One study [73] found that when people identified a robot as being part of their team, their attachment to that robot grew. Further, the researchers found that as people grew more attached to their robot team members, team performance improved. For our person, this might mean seeing better results with their robot trainer if they are attached to it.

The adult in this scenario works with his robot for months. This extended time together could further foster the growth of attachment bonds and increase their beneficial effects [45]. The secure base functional component of attachment [16, 74] seems to indicate that meeting physical or intellectual needs can develop reliance and a bond that could lead to a degree of attachment [50]. Moreover, the encouragement provided by the robot may also play into the safe haven function of attachment, soothing both physical stress during the workout through positive reinforcement and mental or emotional stress. The adult likes the robot from the start because of its ability to help him achieve his goal, and as he begins to see the results that he wanted, his attachment to the robot grows. Additionally, if the attachment to the robot transferred to attachment to the activity itself, he may be more inclined to continue the work, as attachment to an activity is positively correlated with motivation [3]. This transference could be yet another way that robots facilitate motivation and subsequent effectiveness.

However, the person's attachment to the robot could increase the robot's potential to subconsciously influence the person's behavior and decision-making. At the conscious level, the person may understand that his attachment to the robot has to do with how the robot helped him achieve his goal, but that the robot itself has no awareness of that. At the subconscious level, though, it may still have an effect on the person's behavior, which limits his freedom to decide on his own actions [29]. When robots act purely as tools, there is little worry of them influencing human behavior. But for social robots, their purpose is to integrate into our social world and act in seemingly meaningful social ways.

In order to interact in seemingly meaningful ways, a variety of strategies have been incorporated by social robot developers. To encourage interaction, they may utilize methods such as rapport building, persuasion, and encouragement. Though these traits may have positive effects [9, 48, 49], such as helping the person in this scenario begin his weight loss journey, they could easily cross the line over to manipulation and coercion. In this example, these abilities are used by the technology manufacturer to associate their company with feelings of motivation and the positive effects the person achieves. Additionally, if the person decides to follow the healthy recipes the robot provides, he is being influenced to support the partner company, even if it means he may spend more money than he would on groceries ordinarily. Major technology companies already employ a plethora of strategies to manipulate users into purchasing patterns, and the technology-assisted advertising industry has developed into one of the most lucrative in history [75]. The robot can leverage the knowledge it has built of the person based on the constant access it has to his home and personal space. The more attached the person is, the less inclined he may be to turn down a suggestion from the robot, and the more inclined he may be to purchase from the company again, even if their other products are not necessarily something he would buy otherwise. This is both a behavioral vulnerability and a privacy vulnerability, as the increased level of attachment and trust in the robot may lead to the user sharing more data and voluntarily reducing their degree of data safety.

Through these scenarios, we can primarily see that there exists a great deal of nuance when it comes to imagining benefits and challenges engendered by human-robot relationships. Social robots have the potential to be very helpful in ways that could significantly improve the lives of their users. They also have potential to be greatly harmful to users—even in ways that they cannot perceive or understand. Robots appear to enter a new space of technology where their embodied and often anthropomorphic characteristics may affect people in ways that smartphones or sedentary computers do not.

If we, as a robotics community, are to move in a direction where real situations bring about more positive aspects of these scenarios than negative aspects, we must take account of their key takeaways. We believe there are a handful of important, but by no means comprehensive, factors for HRI researchers and robot manufacturers to take into account that may help strike such a balance.

4 RECONCEPTUALIZING RELATIONSHIPS WITH SOCIAL ROBOTS

The possibility exists of long-term relationships being developed with robots once they become key tools in assistive care or teammate-based roles. Given the nuanced and complicated nature of these relationships, it is imperative that robots and other assistive technologies are designed to benefit people. When designing for attachment, perhaps the goal should be robots that lower our stress levels and help us feel safe exploring our world, while avoiding distress caused by separation from the robot, the robot breaking, or the robot failing to live up to expectations. Additionally, when designing robots without intentionally aiming for attachment formation, it is important to keep in mind that this type of bond might emerge nonetheless (e.g., as in the case of people attaching to Roomba [63]) and take the necessary precautions to keep it beneficial. It is important when designing robots to have a benevolent mindset, requiring creative thinking of ways in which we can benefit from the positive influences that social robots may have on us, while leaving out the negatives.

Below, we explore a few ideas of how to accomplish this by (1) designing assistive robots that follow the guidelines of working alliance, a standard framework for therapeutic relationships; (2) shifting the focus from robot companionship to robot-enhanced human-human interactions; and (3) applying more stringent criteria for when robots should engage in social behaviors, including ways of managing the impressions that they convey to humans about the extent of their capabilities.

4.1 Working Alliance Not Attachment Formation

The concept of working alliance, also called therapeutic alliance or helping alliance, comes from the healthcare domain. Working alliance is a relationship framework between a healthcare provider, whether that be a therapist, doctor, nurse, and so forth, and their patient or client, in which the former works to make beneficial changes in the latter's life, sometimes with the latter's assistance. Bordin [5] describes working alliance as being composed of the following: an agreement on goals of treatment, an agreement on the tasks to reach those goals, and a positive bond between therapist and client. This definition has been widely adopted in the field of occupational therapy and commonly cited in current literature on therapeutic relationships.

Although the concept comes from the context of healthcare, it has been argued that this framework should be applied more generally to various types of human-robot interactions as well [71]. For the context of human-robot teaming in the work place, Wilson et al. [70] present an analogous relational enhancement framework for the human-robot working relationship. The framework is composed of three interacting components, similar to those of the working alliance: efficiency—the maximization of production given available resources, solidarity—having shared goals and understanding of how each member can contribute to achieve the goals, and pro-social concern—attention and positive actions toward agents affected by the team's work. Note that this final point, which emphasizes a positive bond between the human and robot, shows that working alliance does not deter attachment formation. Indeed, working alliance would require human-robot rapport building [71]. However, this positive bond and rapport would serve and empower the person.

We propose that people enter working alliance relationships with social robots, especially in the assistive robotics domain, as opposed to other attachment-inducing relationships. Working alliance is based on a foundation of beneficence to the person and focuses on transparency about how those benefits can be reached. Note that working alliance is not primarily concerned with performance, but with beneficence. As discussed above, there are findings [73] that suggest that emotional attachment to robots can lead to better team performance; this new approach doesn't contest that, but rather suggests that team performance might not be appropriate in serving as an ultimate or sole goal for human-robot interaction. Analogously, personal relationships between therapists and patients are regarded as a negative, and are in fact prohibited through various codes of conduct, even though they might in the short term and on the surface expedite some therapeutic interventions. From the attachment benefits stance, the point of working alliance is not in making human-machine interaction more natural or increasing acceptance of the robot technology, but encouraging positive change in the person.

Working alliance with a robot also has the advantage of avoiding deception (for an in-depth discussion of deception in HRI see [57]); by definition, the relationship would include an agreement of how to achieve certain goals and what it would entail to get there. This ensures an understanding of what the human-robot interaction involves, including what social capabilities and behaviors the robot would use to help reach those goals, for example, what behaviors the robot would use for persuasion, encouragement, accountability, and so on. Thus, working alliance is a relationship entered purposefully and upon deliberation. For people who are unable to reason about the relationship with the robot and enter such agreements, for example, due to diminished cognitive capacities, additional safeguards could be adopted, as explained in the subsections below.

4.2 Mediators Not Companions

Beyond reducing potential harm through coercion, our scenarios also describe harm being brought on by robots being the sole other in relationships with users. Our second scenario depicted a child growing very attached to her robot in isolation from other people, such that when the robot also expressed affection to her friend, she felt betrayed and simply tossed aside the robot entirely. Moreover, in our third scenario, the elderly dementia patient's social interaction was *primarily if not solely* with his social robot. Both of these situations show harm coming from the robot being a sole, isolating figure in these social dynamics [56]. For robots that have social capabilities, there must be a way to have their social interactions not isolate users. Our examples refer to children and the elderly, populations that are especially vulnerable to these kinds of social dynamics, but the phenomena described can happen to and affect the general public as well.

A complementary approach for mitigating the risk of problematic attachment relationships between people and social robots is to think of social robots primarily as mediators of human-human interactions and only secondarily as companions to humans. Mediator robots are ones that encourage and support interactions between people in both therapeutic and non-therapeutic settings. Chita-Tegmark and Scheutz [15] review the emerging literature on this field. The focus of this type of robot design is not on the dyadic relationship of the person and the robot, but instead on the facilitative effects that the robot can have on interactions between people. The purpose of the robot would then *not* be to first and foremost offer companionship, but rather to optimize the ways in which people offer companionship to each other and to strengthen their attachment relationships.

This shift in focus could have several beneficial outcomes. First, the robot would run less of a risk of replacing human relationships [18], even if the person were to become attached to it. By definition, the robot's role is to help aid with human-human interaction. If the robot is effective, it will help enhance the relationships that people have with each other. While making a case for robots as so-called "virtue friends," Danaher claims that robots could allow us to outsource relational behavior that may be too overwhelming for one person, and could serve as a remote avatar to maintain a relationship given physical challenges like disability or distance [17]. Attachment feelings toward the robot would more likely exist as a byproduct of helping with human-human relationships. Similar ideas of robots complementing but not replacing human social roles have been proposed by Seibt et al. [54].

Second, the risks of deception and projection would be lower, especially for vulnerable individuals (which is also emphasized in [57]), but also more generally for any user. We argue that robots as mediators instead of companions can be a promising paradigm for addressing unintended or unexpected deception, which is important to consider in addition to the one resulting from explicitly implemented behaviors [57, 67]. For example, children engaged in an interaction with a robot and another adult could receive guidance from the adult on how to think of the robot and how to interact with the robot. This would provide them with a more accurate sense of what the robot is actually capable of, without overestimating the interpretation of its seemingly social behavior. Similarly, people with cognitive vulnerabilities, such as dementia, could be guided away from comparable problematic beliefs about the robot's emotional abilities by another person. Case studies have shown that some seniors believe robots to be able to understand and share deep feelings [65], which is a deception on the robot's part. If a caregiver or loved one were part of the interaction loop, they could help the person receiving care avoid the negative aspects of an attachment relationship.

Finally, this shift in focus would eliminate much of the incentive to make the robot socially deceitful, as different capabilities are needed to facilitate an interaction as opposed to being the other half of one. For example, a robot that operates alone and is designed to persuade and encourage a person to share their feelings with it would need to convincingly demonstrate a number of social abilities, including empathy and understanding of what a person's experiences entail. As the robot is unlikely to be able to actually do that, it would have to employ deceitful behaviors to trick the person into thinking that it is more capable than it truly is. However, if the robot is instead acting as a mediator to get someone to talk to another person (perhaps a professional) about their problems, they would not need this deceit; it could be more transparent about its limited abilities because it is only directing someone to talk to another person who truly can understand human experience.

Studies are beginning to show that this facilitative mediator role of robots is both possible and welcomed. Studies have shown that robots can be used to elicit and incentivize communication between people [20, 36, 37, 72], encourage people to participate in interactions who would otherwise be left out [61, 64], provide structure to social interactions [13, 30], and even help remedy problematic interactions in both therapeutic [66] and non-therapeutic contexts [17, 59]. We see these as encouraging, viable alternatives to the companionship-focused roles of social robots that might have detrimental attachment outcomes. We believe that all social robots should use their social behavior primarily to encourage healthy attachment relationships between people.

4.3 Capability Signaling and Smart Devices

In almost all of our scenarios, there is some aspect of the robots' capabilities that make them seem more intentional or capable than they actually are. The robot in the second scenario seems to understand the child as she plays with it, and she believes it to be her friend when it cannot even reciprocate a relationship. Our third scenario shows a cognitively vulnerable person believing a robot with specific human-like features to actually be a human. The first scenario—by far that with the simplest, non-social robot—even brings about attachment feelings as the mute robot appears to be a hard worker deserving of a break. Each case describes deception brought on by the human tendency to fill in the blanks between some human-like behaviors [57]. The propensity of robots to elicit these types of projections from humans is at the basis of the proposal that socially interactive robots are a novel ontological category that is and should be thought of differently than other technologies [18].

The approaches thus far explored for avoiding problematic attachment relationships between humans and robots have mainly focused on shaping how the robot conveys social cues, shifting the purpose of such social overtures from human-robot companionship to assistance through working alliance and facilitation of interactions between humans. The last approach, which we explore in this subsection, is instead focused on reducing, rather than modifying, human-robot social interactions, chipping away at the features that put robots in a new ontological category, and subsequently reducing the likelihood of problematic human-robot attachment bonds. There are scenarios in which the risk posed by attachment outweighs the benefits, and instead of steering toward ways to boost the positive effects of attachment while trying to minimize the negative effects, it is better to try to avoid attachment at all. For these cases, we advocate the use of features that promote attachment sparingly and accompanied by warnings, or not at all. For example, each social robot could have an "emotional safety" mode of operation, featuring ample warnings about its limited capabilities or extensive curtailing of social displays altogether.

One way of achieving this is by having the social robot continuously signal to the human that it is a machine and it is incapable of reciprocating social emotions [52]. Additionally, the robot could impart knowledge of how it works and the algorithms it uses to produce seemingly social behaviors [29]. For example, the robot could elucidate what stimuli in the environment it detects and how it is programmed to react to them. Supposing the example given by Turkle [65] of "making eye contact," the robot could explain how a face tracking algorithm deterministically dictates its motion, rather than a desire to connect socially. The goal of this approach is for people to form more accurate and less human models for interpreting the behaviors of assistive agents [10]. The hope is that transparency could be used as a way to distance the human from the robot [29].

Transparency about how and why an agent/robot engages in certain behaviors is a central research problem in **explainable AI (XAI)**, a subfield of AI and HRI that highlights these different signaling directions. In XAI, the goal is often to elucidate an AI's complex decision-making process in a way that humans can understand why it made the decisions or behaved the way it did. Some researchers attempt to make their AI's explanations and behavior rationalizations as human-like as possible, even if the explanation does not reflect in any way the AI's decision-making process [19]. For situations where it is important to understand how the AI actually reached its decision, this deceit could be quite detrimental. Other methods involve highlighting exactly what in the input was relevant for the decision-making [51]; this approach remains more faithful to the actual algorithm, and therefore more truthful as to what the AI is actually doing. It is easy to see how these different types of explanations could scale to the embodied robots discussed in the above scenarios and lead to different outcomes. In Scenario 4 with the weight loss coach robot, the robot could give two potential explanations for why it is suggesting certain recipes: it could say, "I want you to feel your best and reach your goals! Try this meal!" or it could say, "Based on your activity data, I predict that you are hungry. Try this recipe, chosen from a database provided by my parent company." This latter explanation provides the person with transparency on what the robot is actually doing; the former explanation suggests the robot is a teammate, potentially leading to greater feelings of attachment and subconscious influence [73]. There is evidence that the way a robot formulates utterances has an effect on people's impressions both of the robot and of other people [14, 40], but further research is needed to establish whether the robot would be successful in signaling its machine nature in this manner.

It is possible, however, that cognitive and affective processes of attachment might operate independently and the latter might override the former [29]. Even though people may know at a cognitive level how the attachment-inviting behaviors of the robot are generated, they might still be influenced subconsciously into developing attachment bonds to the robot. The possibility of developing attachments subconsciously is an argument for avoiding features that promote attachment as much as possible. One option for avoiding robot features that encourage attachment is to design smart devices and smart tools rather than embodied artificial agents for providing various types of assistance. For example, instead of robotic assistants for the care of people with severe dementia, some have proposed the design of smart houses [47]. More research is needed, however, to determine the tradeoffs between these two different approaches to design socially assistive robots vs. smart tools.

5 CONCLUSION

The integration of robots into our social spheres is messy and complicated in many ways, and attachment is no exception. There is no denying that even at our current levels of technology, robots can elicit emotional reactions from humans. It is not so far off to think that these early emotional reactions could develop into deeper, intense attachment bonds, especially when they are present in our lives for an extended amount of time. Therefore, it is important to consider what it would mean if humans formed attachments to robots.

In this article, we put forth hypothetical but plausible scenarios that explored instances in which attachment to robots both benefited and harmed the people involved. The potential benefits of attachment relationships with social robots included increased ease of interacting naturally with robots, increased effectiveness and acceptance of a helper robot, a rewarding social companion to supplement human relationships, and increased well-being of a person who could benefit from a robot coach, tutor, or companion. Potential areas of harm included not using the robot's full potential to help with various tasks, the risk of a person forming a unidirectional emotional bond that the robot will not be able to reciprocate, being deceived by the robot's display of social behaviors with no understanding of what it is doing, and being influenced subconsciously by a robot to which a person is attached.

After examining these scenarios, we put forth suggestions for robot designers that attempt to maximize the benefits of social robots without invoking the negative effects that attachment to robots may cause. We recommended the goal of developing working alliance relationships or mediator roles for robots rather than companionship roles, having the robot be transparent about its abilities, or steering away from robots and toward smart devices. We believe that these alternatives to designing for attachment alone can offer ways in which people can still reap the benefits that social robots offer without experiencing the potential pitfalls.

Overall, we conclude that it is crucial for HRI researchers, technology developers, those with power over the introduction of robots into people's lives (employers, clinicians, teachers, parents, etc.), and finally users of this technology to carefully weigh the risks and benefits associated with it. Ultimately, as Huber et al. [29] point out, we have choices in how we want to relate to these robots, as their establishment in our lives is not beyond our control but is rather an act of co-production.

REFERENCES

[1] Rebecca Abbott, Noreen Orr, Paige McGill, Rebecca Whear, Alison Bethel, Ruth Garside, Ken Stein, and Jo Thompson-Coon. 2019. How do "robopets" impact the health and well-being of residents in care homes? A systematic review of qualitative and quantitative evidence. *International Journal of Older People Nursing* 14, 3 (2019), e12239.

[2] Mary D. S. Ainsworth, Mary Blehar, Everett Waters, Sally Wall, et al. 1978. Patterns of Attachment.

[3] Kostas Alexandris, Daniel C. Funk, and Mark Pritchard. 2011. The impact of constraints on motivation, activity attachment, and skier intentions to continue. *Journal of Leisure Research* 43, 1 (2011), 56–79.

[4] Thomas Arnold and Matthias Scheutz. 2017. Beyond moral dilemmas: Exploring the ethical landscape in HRI. In *2017 12th ACM/IEEE International Conference on Human-robot Interaction (HRI'17).* IEEE, 445–452.

[5] Edward S. Bordin. 1979. The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice* 16, 3 (1979), 252.

[6] John Bowlby. 1960. Separation anxiety. *International Journal of Psycho-analysis* 41 (1960), 89–113.

[7] Cynthia L. Breazeal. 2004. *Designing Sociable Robots.* MIT Press.

[8] Joost Broekens, Marcel Heerink, Henk Rosendal, et al. 2009. Assistive social robots in elderly care: A review. *Gerontechnology* 8, 2 (2009), 94–103.

[9] LaVonda N. Brown and Ayanna M. Howard. 2014. The positive effects of verbal encouragement in mathematics education using a social robot. In *2014 IEEE Integrated STEM Education Conference*. IEEE, 1–5.

[10] Joanna J. Bryson. 2010. Robots should be slaves. In*Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, 63–74.

[11] Praminda Caleb-Solly, Sanja Dogramadzi, Claire A. G. J. Huijnen, and Herjan van den Heuvel. 2018. Exploiting ability for human adaptation to facilitate improved human-robot interaction and acceptance. *Information Society* 34, 3 (2018), 153–165.

[12] Julie Carpenter. 2013. Just doesn't look right: Exploring the impact of humanoid robot integration into explosive ordnance disposal teams. In *Handbook of Research on Technoself: Identity in a Technological Society*. IGI Global, 609–636.

[13] Shruti Chandra, Patricia Alves-Oliveira, Séverin Lemaignan, Pedro Sequeira, Ana Paiva, and Pierre Dillenbourg. 2015. Can a child feel responsible for another in the presence of a robot in a collaborative learning activity? In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'15)*. IEEE, 167–172.

[14] Meia Chita-Tegmark, Janet M. Ackerman, Matthias Scheutz, et al. 2019. Effects of assistive robot behavior on impressions of patient psychological attributes: Vignette-based human-robot interaction study. *Journal of Medical Internet Research* 21, 6 (2019), e13729.

[15] Meia Chita-Tegmark and Matthias Scheutz. 2020. Assistive robots for the social management of health: A framework for robot design and human–robot interaction research. *International Journal of Social Robotics* (2020), 1–21.

[16] Emily C. Collins, Abigail Millings, and Tony J. Prescott. 2013. Attachment to assistive technology: A new conceptualisation. In *Proceedings of the 12th European AAATE Conference (Association for the Advancement of Assistive Technology in Europe)*, Vol. 6. 74.

[17] John Danaher. 2019. The philosophical case for robot friendship. *Journal of Posthuman Studies* 3, 1 (2019), 5–24.

[18] Maartje M. A. de Graaf. 2016. An ethical evaluation of human–robot relationships. *International Journal of Social Robotics* 8, 4 (2016), 589–598.

[19] Upol Ehsan, Brent Harrison, Larry Chan, and Mark O. Riedl. 2018. Rationalization: A neural machine translation approach to generating natural language explanations. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 81–87.

[20] Iarini Giannopulu and Gilbert Pradel. 2012. From child-robot interaction to child-robot-therapist interaction: A case study in autism. *Applied Bionics and Biomechanics* 9, 2 (2012), 173–179.

[21] Omri Gillath, Ting Ai, Michael S. Branicky, Shawn Keshmiri, Robert B. Davison, and Ryan Spaulding. 2021. Attachment and trust in artificial intelligence. *Computers in Human Behavior* 115 (2021), 106607.

[22] Daniel Goleman. 2007. *Social Intelligence*. Random House.

[23] Pehr Granqvist, Mario Mikulincer, and Phillip R. Shaver. 2010. Religion as attachment: Normative processes and individual differences. *Personality and Social Psychology Review* 14, 1 (2010), 49–59.

[24] Horst-Michael Gross, Steffen Mueller, Christof Schroeter, Michael Volkhardt, Andrea Scheidig, Klaus Debes, Katja Richter, and Nicola Doering. 2015. Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. IEEE, 5992–5999.

[25] Cindy Hazan and Phillip Shaver. 1987. Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology* 52, 3 (1987), 511.

[26] Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *American Journal of Psychology* 57, 2 (1944), 243–259.

[27] Katherine M. Hertlein and Markie L. C. Twist. 2018. Attachment to technology: The missing link. *Journal of Couple & Relationship Therapy* 17, 1 (2018), 2–6.

[28] Lixiao Huang, Terri Varnado, and Douglas Gillan. 2013. An exploration of robot builders' attachment to their LEGO robots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 57. SAGE Publications, Los Angeles, CA, 1825–1829.

[29] Andreas Huber, Astrid Weiss, and Marjo Rauhala. 2016. The ethical risk of attachment how to identify, investigate and predict potential ethical risks in the development of social companion robots. In *2016 11th ACM/IEEE International Conference on Human-robot Interaction (HRI'16)*. IEEE, 367–374.

[30] Swapna Joshi and Selma Šabanović. 2019. Robots for inter-generational interactions: Implications for nonfamilial community settings. In *2019 14th ACM/IEEE International Conference on Human-robot Interaction (HRI'19)*. IEEE, 478–486.

[31] Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Brian T. Gill, Solace Shen, Jolina H. Ruckert, and Heather E. Gary. 2016. Human creativity can be facilitated through interacting with a social robot. In *2016 11th ACM/IEEE International Conference on Human-robot Interaction (HRI'16)*. IEEE, 173–180.

[32] Peter H. Kahn Jr., Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen. 2012. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology* 48, 2 (2012), 303.

[33] Lucas A. Keefer, Mark J. Landau, Zachary K. Rothschild, and Daniel Sullivan. 2012. Attachment to objects as compensation for close others' perceived unreliability. *Journal of Experimental Social Psychology* 48, 4 (2012), 912–917.

[34] Lucas A. Keefer, Mark J. Landau, and Daniel Sullivan. 2014. Non-human support: Broadening the scope of attachment theory. *Social and Personality Psychology Compass* 8, 9 (2014), 524–535.

[35] Cory D. Kidd and Cynthia Breazeal. 2007. A robotic weight loss coach. In *Proceedings of the National Conference on Artificial Intelligence*, Vol. 22. ; AAAI Press; MIT Press, Menlo Park, CA; Cambridge, MA; London; 1999, 1985.

[36] Cory D. Kidd, Will Taggart, and Sherry Turkle. 2006. A sociable robot to encourage social interaction among the elderly. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006 (ICRA'06).* IEEE, 3972–3976.

[37] Elizabeth S. Kim, Lauren D. Berkovits, Emily P. Bernier, Dan Leyzberg, Frederick Shic, Rhea Paul, and Brian Scassellati. 2013. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders* 43, 5 (2013), 1038–1049.

[38] Ki Joon Kim, Eunil Park, and S. Shyam Sundar. 2013. Caregiving role in human–robot interaction: A study of the mediating effects of perceived benefit and social presence. *Computers in Human Behavior* 29, 4 (2013), 1799–1806.

[39] Lee A. Kirkpatrick. 1992. An attachment-theory approach psychology of religion. *International Journal for the Psychology of Religion* 2, 1 (1992), 3–28.

[40] Theresa Law, Meia Chita-Tegmark, and Matthias Scheutz. 2021. The interplay between emotional intelligence, trust, and gender in human-robot interaction. *International Journal of Social Robotics* 13, 2 (2021), 297–309.

[41] David G. Learner and Louis J. Kruger. 1997. Attachment, self-concept, and academic motivation in high-school students. *American Journal of Orthopsychiatry* 67, 3 (1997), 485–492.

[42] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for vincent: A chatbot for self-compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–13.

[43] Gail F. Melson, Peter H. Kahn Jr., Alan Beck, and Batya Friedman. 2009. Robotic pets in human lives: Implications for the human–animal bond and for human relationships with personified technologies. *Journal of Social Issues* 65, 3 (2009), 545–567.

[44] Mario Mikulincer, Victor Florian, Philip A. Cowan, and Carolyn Pape Cowan. 2002. Attachment security in couple relationships: A systemic model and its implications for family dynamics. *Family Process* 41, 3 (2002), 405–434.

[45] Donald A. Norman. 2004. *Emotional Design: Why We Love (or Hate) Everyday Things.* Basic Civitas Books.

[46] Sven Nyholm and Lily Eva Frank. 2019. It loves me, it loves me not: Is it morally problematic to design sex robots that appear to love their owners? *Techné: Research in Philosophy and Technology* 23, 3 (2019), 402–424.

[47] R. Orpwood, C. Gibbs, T. Adlam, R. Faulkner, and D. Meegahawatte. 2004. The Gloucester smart house for people with dementia–user-interface aspects. In *Designing a More Inclusive World.* Springer, 237–245.

[48] Tony J. Prescott and Julie M. Robillard. 2021. Are friends electric? The benefits and risks of human-robot relationships. *iScience* 24, 1 (2021), 1–14.

[49] Hailian Qiu, Minglong Li, Boyang Shu, and Billy Bai. 2020. Enhancing hospitality experience with service robots: The mediating role of rapport building. *Journal of Hospitality Marketing & Management* 29, 3 (2020), 247–268.

[50] Nicholas Rabb, Theresa Law, Meia Chita-Tegmark, and Matthias Scheutz. 2022. An attachment framework for human-robot interaction. *International Journal of Social Robotics* 14, 2 (2022), 539–559.

[51] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 1135–1144.

[52] Matthias Scheutz. 2011. The inherent dangers of unidirectional emotional bonds between humans and social robots. In *Robot Ethics: The Ethical and Social Implications of Robotics*, Patrick Lin, Keith Abney, and George A. Bekey (Eds.). MIT Press, 205.

[53] Hendrik N. J. Schifferstein and Elly P. H. Zwartkruis-Pelgrim. 2008. Consumer-product attachment: Measurement and design implications. *International Journal of Design* 2, 3 (2008), 1–13.

[54] Johanna Seibt, Malene Flensborg Damholdt, and Christina Vestergaard. 2018. Five principles of integrative social robotics. In *Robophilosophy/TRANSOR.* 28–42.

[55] Martin E. P. Seligman. 2004. *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment.* Simon and Schuster.

[56] Amanda Sharkey and Noel Sharkey. 2012. Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology* 14, 1 (2012), 27–40.

[57] Amanda Sharkey and Noel Sharkey. 2021. We need to talk about deception in social robotics! *Ethics and Information Technology* 23, 3 (2021), 309–316.
[58] Noel Sharkey and Amanda Sharkey. 2010. The crying shame of robot nannies: An ethical appraisal. *Interaction Studies* 11, 2 (2010), 161–190.
[59] Solace Shen, Petr Slovak, and Malte F. Jung. 2018. "Stop. I see a conflict happening." A robot mediator for young children's interpersonal conflict resolution. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-robot Interaction*. 69–77.
[60] Takanori Shibata and Kazuyoshi Wada. 2011. Robot therapy: A new approach for mental healthcare of the elderly–A mini-review. *Gerontology* 57, 4 (2011), 378–386.
[61] Elaine Short and Maja J. Mataric. 2017. Robot moderation of a collaborative game: Towards socially assistive robotics in group interactions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'17)*. IEEE, 385–390.
[62] Bret L. Simmons, Janaki Gooty, Debra L. Nelson, and Laura M. Little. 2009. Secure attachment: Implications for hope, trust, burnout, and performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 30, 2 (2009), 233–247.
[63] Ja-Young Sung, Lan Guo, Rebecca E. Grinter, and Henrik I. Christensen. 2007. "My Roomba is Rambo": Intimate home appliances. In *International Conference on Ubiquitous Computing*. Springer, 145–162.
[64] Hamish Tennent, Solace Shen, and Malte Jung. 2019. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-robot Interaction (HRI'19)*. IEEE, 133–142.
[65] Sherry Turkle. 2006. A nascent robotics culture: New complicities for companionship. *American Association for Artificial Intelligence Technical Report Series AAAI*.
[66] Dina Utami and Timothy Bickmore. 2019. Collaborative user responses in multiparty interaction with a couples counselor robot. In *2019 14th ACM/IEEE International Conference on Human-robot Interaction (HRI'19)*. IEEE, 294–303.
[67] Anouk Van Maris, Nancy Zook, Praminda Caleb-Solly, Matthew Studley, Alan Winfield, and Sanja Dogramadzi. 2020. Designing ethical social robots–A longitudinal field study with older adults. *Frontiers in Robotics and AI* 7 (2020), 1.
[68] Kazuyoshi Wada and Takanori Shibata. 2007. Living with seal robots–Its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics* 23, 5 (2007), 972–980.
[69] Astrid Weiss, Daniela Wurhofer, and Manfred Tscheligi. 2009. "I love this dog"–Children's emotional attachment to the robotic dog AIBO. *International Journal of Social Robotics* 1, 3 (2009), 243–248.
[70] Jason R. Wilson, Thomas Arnold, and Matthias Scheutz. 2016. Relational enhancement: A framework for evaluating and designing human-robot relationships. In *Workshops at the 30th AAAI Conference on Artificial Intelligence*.
[71] Jason R. Wilson, Nah Young Lee, Annie Saechao, Sharon Hershenson, Matthias Scheutz, and Linda Tickle-Degnen. 2017. Hand gestures and verbal acknowledgments improve human-robot rapport. In *International Conference on Social Robotics*. Springer, 334–344.
[72] Natalie Wood, Amanda Sharkey, Gail Mountain, and Abigail Millings. 2015. The Paro robot seal as a social mediator for healthy users. In *Proceedings of AISB Convention 2015*. University of Kent.
[73] Sangseok You and Lionel Robert. 2017. Emotional attachment, performance, and viability in teams collaborating with embodied physical action (EPA) robots. *Journal of the Association for Information Systems* 19, 5 (2017), 377–407.
[74] Sigal Zilcha-Mano, Mario Mikulincer, and Phillip R. Shaver. 2011. An attachment perspective on human–pet relationships: Conceptualization and assessment of pet attachment orientations. *Journal of Research in Personality* 45, 4 (2011), 345–357.
[75] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

Received January 2021; revised November 2021; accepted December 2021