

Human-robot dynamics: a psychological insight into the ethics of social robotics

International
Journal of Ethics
and Systems

101

Auxane Boch

Institute for Ethics in AI, Technical University of Munich, Munich, Germany, and

Bethany Rhea Thomas

Department of Psychology, Edge Hill University, Ormskirk, UK

Received 31 January 2024

Revised 22 August 2024

3 October 2024

Accepted 10 November 2024

Abstract

Purpose – Social robotics is a rapidly growing application of artificial intelligence (AI) in society, encompassing an expanding range of applications. This paper aims to contribute to the ongoing integration of psychology into social robotics ethics by reviewing current theories and empirical findings related to human–robot interaction (HRI) and addressing critical points of contention within the ethics discourse.

Design/methodology/approach – The authors will explore the factors influencing the acceptance of social robots, explore the development of relationships between humans and robots and delve into three prominent controversies: deception, dehumanisation and violence.

Findings – The authors first propose design factors allowing for a positive interaction with the robot, and further discuss precise dimensions to evaluate when designing a social robot to ensure ethical design technology, building on the four ethical principles for trustworthy AI. The final section of this paper will outline and offer explicit recommendations for future research endeavours.

Originality/value – This paper provides originality and value to the field of social robotics ethics by integrating psychology into the ethical discourse and offering a comprehensive understanding of HRI. It introduces three ethical dimensions and provides recommendations for implementing them, contributing to the development of ethical design in social robots and trustworthy AI.

Keywords Psychology, Ethics, Social robots

Paper type Literature review

1. Introduction

Social robotics is a rapidly growing application of artificial intelligence (AI) in society, encompassing an expanding range of applications. With its inherent social nature, integrating psychology into the multidisciplinary discussion of social robot ethics becomes an evident avenue of research.

Early definitions described social robots as machines capable of acceptable interaction with humans and other robots, effectively conveying intentions and collaborating to achieve goals (Duffy *et al.*, 1999). The applications and goals of social robots are diverse, varying



© Auxane Boch and Bethany Rhea Thomas. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

International Journal of Ethics and
Systems
Vol. 41 No. 1, 2025
pp. 101-141
Emerald Publishing Limited
2514-9369
DOI 10.1108/IJOES-01-2024-0034

across specific domains. Extensive literature has explored and categorised these use cases based on their intended purposes.

As examined by [Lambert et al. \(2020\)](#) and [Naneva et al. \(2020\)](#), companion robots focus on building relationships with owners and facilitating sustained social interaction while providing domestic assistance. These robots align with the goal of fostering social bonds and promoting independence among their end users. Service robots, exemplified by bellhop prototypes in the hospitality industry, as studied by [Boch et al. \(2021\)](#) and [Pinillos et al. \(2016\)](#), play various roles with a focus on economic productivity objectives. Care robots target patients and health-care providers through tasks that enhance health-care delivery ([Naneva et al., 2020](#)) or provide customised assistance to vulnerable groups ([Vallor, 2011](#)). Some initiatives prioritise paediatric care, emphasising welfare goals ([Naneva et al., 2020](#)). Educational robots, explored by [Angel-Fernandez and Vincze \(2018\)](#) and [Naneva et al. \(2020\)](#), enhance learning experiences by supporting instruction and engaging learners, aligning with instructional technology objectives. Sex robots aim to improve intimacy, although this specific domain raises significant ethical complexities ([Boch et al., 2021](#); [Fosch-Villaronga and Poulsen, 2020](#); [Richardson, 2016](#)). Furthermore, social interaction and entertainment robots, as examined by [Naneva et al. \(2020\)](#) and [Lambert et al. \(2020\)](#), provide valuable insights into fundamental human–robot dynamics despite lacking specific purposes.

These diverse applications reflect evolving technological, economic and social priorities. By examining the underlying factors and dynamics of the relationship between humans and social robots from a psychological perspective, we can better understand the tension points in ethical design for fostering positive human–robot interactions (HRI) and relationships.

The ethical design of social robots is a complex and multifaceted endeavour that draws upon insights from various fields, including psychology, human–computer interaction and ethics. In recent years, psychologists have played a crucial role in shaping the ethical design of social robots through empirical research and the development of conceptual frameworks.

Psychology plays a crucial role in the ethical design of social robots by investigating and informing the concept of HRI. Studies have shown that humans tend to anthropomorphise robots, attributing human-like qualities to them, which has implications for designing social robots that evoke emotional attachments ([Smith et al., 2021](#)). Moral psychology has also contributed to developing ethical frameworks. For example, Malle’s model integrates principles from moral psychology and HRI to guide ethical robot behaviours ([Malle, 2016](#)). In addition, psychology helps underpin the psychological mechanisms underlying human behaviour and decision-making, informing the design of social robots. For instance, considering users’ cognitive biases and heuristics is essential, as these cognitive processes can shape how individuals perceive and respond to robotic behaviour ([Biswas and Murray, 2014](#)).

In conclusion, the ethical design of social robots draws upon a rich body of psychological research and theoretical frameworks. Psychologists have contributed significantly to developing ethical guidelines and design principles for social robots by understanding how humans perceive, interact with and are influenced by robots. As the field continues to evolve, psychologists will undoubtedly play a central role in shaping the future of ethical social robot design.

This paper seeks to contribute to the ongoing integration of psychology into social robotics ethics. We aim to accomplish this by reviewing current theories and empirical findings related to HRI and addressing critical points of contention within the ethics discourse. This paper is structured to provide a comprehensive exploration of the ethical design of social robots through the lens of psychological insights. The discussion begins by

examining the moderating factors that influence the acceptance of social robots, including cultural, appearance, engagement, behavioural and personalisation factors in Section 2. The paper then delves into the development of human–robot relationships, exploring key psychological theories in Section 3. Section 4 addresses potential consequences and ethical controversies associated with social robotics, focusing on deception, manipulation, dehumanisation and violence. Finally, in Section 5 and 6, the paper concludes with practical recommendations for the ethical design of social robots, considering the discussed psychological factors and ethical dimensions, and offers a roadmap for future research in this interdisciplinary field.

2. Moderating factors of acceptance

In this section, we delve into the factors influencing the acceptance of social robots, aiming to understand humans' overall perceptions and receptiveness towards these machines, building on psychology's teachings. Accepting a technology implies having a positive expectation and experience, which aligns with the ethical design objectives. Achieving acceptance involves understanding the various factors influencing individuals' attitudes towards robots and designing robots that meet their expectations and needs. By considering cultural factors, appearance factors, engagement factors, behavioural factors and personalisation, designers can create social robots that foster positive interactions and adhere to ethical principles of promoting user well-being and satisfaction.

2.1 General attitudes towards robots

Early studies provide insights into general public attitudes towards robots. According to [Dautenhahn \(2007\)](#), robots tended to be primarily viewed as valuable tools to assist with household tasks rather than as social companions. This perception of robots as appliances or machines aligned with their initial introduction for industrial and service applications rather than social interaction. However, more recent work has found that attitudes may depend on direct experience with robots. [Naneva et al. \(2020\)](#) reviewed survey responses regarding trust, anxiety and willingness to use robots. Overall, participants reported neutral levels of trust and anxiety towards robots – neither explicitly trusting nor distrusting robots and experiencing reasonably moderate anxiety. This suggests an open but cautious baseline attitude. However, interestingly, the review found some influence of demographic factors. Gender appeared to impact trust somewhat, with samples including more female participants reporting higher levels of trust in robots compared to more male-dominated samples. Furthermore, age did not significantly influence reported attitudes contrary to some expectations.

Building on this knowledge, studies have investigated how direct engagement with robots shapes attitudes positively. [Li et al. \(2010\)](#) observed strong correlations between interaction factors like likability, trust and satisfaction with a given robot. This implies that fostering active user participation and developing more social interaction tasks could help improve general perceptions and, in-turn attitudes towards robots. Similarly, [Leite et al. \(2013\)](#) found that increasing interaction time and frequency through varied applications may heighten user engagement with robots.

In summary, while initial public perceptions of robots aligned with their industrial origins, general baseline attitudes today appear cautiously open but neutral regarding trust and anxiety. Demographic factors like gender may influence views to a degree. Most importantly, direct experience appears instrumental in developing more positive regard, implying that engineering social engagement can help optimise HRI outcomes.

2.2 Cultural factor

Culture warrants in-depth investigation in the realm of technology acceptance (Oyibo and Vassileva, 2020; Fleischmann *et al.*, 2020; Metallo *et al.*, 2022; Boch, 2011). Culture is pivotal in shaping social dynamics and perspectives, as highlighted by Gelfand and Kashima (2016). It significantly influences how individuals interact socially and perceive themselves and others. This becomes particularly relevant when considering the development of interactions and relationships with novel social agents, such as robots, which lack humans' evolutionary familiarity with other humans (Kacancioğlu *et al.*, 2012). In addition, it seems relevant to highlight that cultural values related to robots are not static but subject to change over time, such as in response to disruptive societal events. In the case of the COVID-19 pandemic, Schönmann *et al.* (2024) found that attitudes towards perceived as "sterile" social robots in caregiving contexts shifted positively, demonstrating how moral values and acceptance of robots can evolve even in short time intervals. The authors concluded that attitudes towards care robots could change positively if their use addresses an urgent need. This highlights the importance of considering culture not just as a regional concept but also as a time and context-dependent factor that requires continuous empirical re-evaluation, especially during periods of rapid societal change.

Notably, Eastern and Western philosophies diverge in their worldviews (Lim *et al.*, 2021; Kamide and Mori, 2016). The Western approach seems to emphasise constructing a systematic understanding of phenomena, while the East seem to adopt a more holistic perspective. Some argue that this holistic perspective makes the East more receptive to concepts like animism, which could facilitate the acceptance of robots (MacDorman *et al.*, 2009). However, the increasing trend of globalisation has led to heightened cultural exposure and fusion, potentially diluting traditional values in certain contexts (Lim *et al.*, 2021). Interestingly, in the context of Islamic culture, Alemi *et al.* (2020) emphasise that cultural and religious values play a crucial role in accepting social robots, particularly in educational settings in Iran. The authors argue that to be accepted, robots must align with Islamic teachings and ethical principles, such as modesty and respect for human dignity. The perception of robots as complementary tools rather than replacements for human roles is also significant in shaping acceptance. In addition, traditional educational roles and community consensus are critical factors in whether these technologies will be adopted. This underscores the necessity of culturally sensitive design in technology implementation, particularly in regions where religion profoundly influences cultural norms.

A critical and extensively studied cultural dimension is individualism-collectivism, which examines how individuals define themselves in relation to others, as defined and explored by Markus and Kitayama (1991) and Hofstede (1980). According to this theory, highly individualistic cultures prioritise independence over relationships, whereas collectivistic cultures emphasise interdependence (De Mooij and Hofstede, 2010). This could, thus, align with a stronger adoption of social entities such as social robots, as demonstrated in Marchesi *et al.*'s (2021) study.

Communication styles also vary considering culture, with individualistic cultures favouring explicit communication, while collectivist cultures tend to use more implicit communication (De Mooij and Hofstede, 2010). Differences in communication styles across cultures extend to nonverbal cues, such as gestures, with significant social meaning (Matsumoto, 2006; Burgoon, 1994). Interestingly, Trovato *et al.* (2013) found that individuals could recognise robot emotional displays better when the robots showed facial expressions consistent with their respective cultures' nonverbal cues. This highlights the relevance of adapting visual and nonverbal cues in HRI.

Finally, numerous studies have explored how cultural backgrounds predict preferred attitudes towards robots. Participants from Eastern collectivist cultures, such as Koreans and Chinese, demonstrated greater engagement with service robots than German participants from Western individualistic cultures. Furthermore, the former group found the robots more likeable, trustworthy and satisfactory (Li *et al.*, 2010; Bartneck *et al.*, 2005). It is postulated that cultures emphasising relationships may develop more positive perspectives on social robots. In contrast, individualistic cultures with extensive exposure to industrial robots, like Germany, may prioritise robotic tool use over companionship. To improve acceptance in these cultures, it is crucial to emphasise the practical utility of social robots by focusing on how they can enhance personal productivity and convenience. Highlighting these benefits can increase their acceptance as valuable tools in daily life (Lim *et al.*, 2021). In addition, gradual exposure to robots in social contexts, starting with less intrusive roles, can help users become more comfortable and eventually accept more socially interactive robots, especially when their utility is demonstrated (Ke *et al.*, 2020).

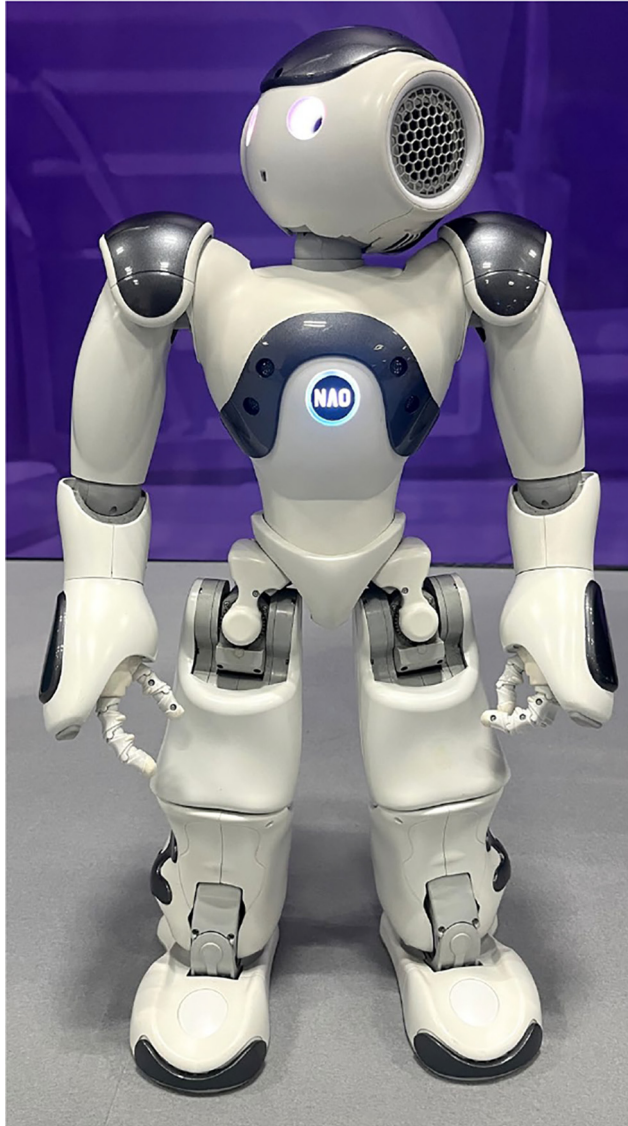
Consequently, a culturally aware interaction design and use case can promote acceptance as people anthropomorphise shared identities, including communication cues (Lim *et al.*, 2021). However, it is important to note that cultural customisation may not always be necessary for general tasks, as humans can interact across cultural differences (Lim *et al.*, 2021). In summary, cultural psychology serves as a fundamental framework for comprehending the intricacies of HRI and acceptance.

2.3 Appearance factors

Social robots have traditionally been broadly categorised based on their physical features as humanoid (i.e. presenting human-like features, such as Nao shown in Figure 1), zoomorphic (i.e. presenting animal-like features, such as Paro shown in Figure 2), or do not fit within one of these categories, such as machine-like robots (Lambert *et al.*, 2020).

Design seems to be important in how people perceive robots and their abilities. A robot that exhibits animal-like behaviour seems to create the illusion that it will be more accommodating to the user's wishes (Lambert *et al.*, 2020). The use of zoomorphic and pet-like robots like Paro and MiRo has been documented in, for instance, the context of entering people's homes and addressing the needs of specific target groups, such as older adults or those with cognitive impairments in care settings (Henschel *et al.*, 2021). They showed great potential for the melioration of social skills and well-being of patients. Robotic pets are usually an appreciated option with elderly populations as they require less care than real animals and avoid issues like allergies and some hygiene considerations in care settings (Hung *et al.*, 2019).

More recently, social robots have been specified as technologies created by humans for interaction that may somewhat physically or behaviourally resemble people, with the goal of HRI mirroring natural human interactions, in other words, anthropomorphic interactions (Fox and Gambino, 2021). With this in mind, many developers design social robots to incorporate human characteristics while carefully avoiding too close an imitation that could cause unease, a phenomenon known as the "Uncanny Valley", initially proposed by Mori in 1970 (Pandey and Gelin, 2018). This theory suggests that as robots become more human-like, they evoke more familiarity and likability until a certain point where the mismatch between their appearance and their behaviour triggers a sense of unease (Ho *et al.*, 2008). This idea aligns with the concept of simulation theory (Turner, 1978), a well-established theory in psychology, proposing we understand the minds of others by simulating their situations and placing ourselves in their shoes. Furthermore, Krach *et al.* (2008) observed a linear relationship between the degree of anthropomorphising and cortical activation in brain



Source: Figure by authors

Figure 1. Nao, an anthropomorphic robot

areas linked to the processing of other minds, also called cognitive empathy. This suggests that when individuals anthropomorphise humanoid robots, their brain areas associated with understanding and processing the mental states of others become more active, supporting the Uncanny Valley theory.



Source: Figure by authors

Figure 2. Paro, Zoomorphic robot representing a white baby seal

Interestingly, anthropomorphism seems to dramatically influence people's responses to robots (Darling *et al.*, 2015). In addition, even though social robots can be intentionally designed as anthropomorphic (Breazeal, 2003; Duffy, 2003), studies have shown that people also tend to anthropomorphise robots with non-human-like designs (Carpenter, 2013; Knight, 2014; Paepcke and Takayama, 2010). Thus, the consequences of anthropomorphism might not occur only in the case of humanoid robots.

When comparing the likeability variation based on design choices, Li *et al.* (2010) note that zoomorphic robots are perceived as more likeable than machine-like robots, while no significant difference in likability seems notable between anthropomorphic and zoomorphic robots. It is worth noting that the anthropomorphic robot used in the experiment differentiated itself from the machine-like robot through recognisable body features such as eyes and a head. This result suggests that the mere presence of humanoid or animal-like features can increase people's familiarity with the robots and consequently enhance likability. As a result, the use application of such technology is to be considered when developing a design; zoomorphic and anthropomorphic robots may be more suitable for entertainment and caring tasks, while machine-like robots are better suited for low-sociability tasks such as acting as security guards (Li *et al.*, 2010). This confirms Goetz *et al.* (2003) early findings, arguing that human-like robots seem preferred for tasks requiring a higher degree of sociability.

Further design factors also come into play when considering likeability and acceptance. In their review, Lambert *et al.* (2020) highlighted that a more feminine robot may be perceived as less threatening than a robot with a masculine appearance.

Thus, anthropomorphism and gender association seem to impact the perceived social qualities of robots.

Interestingly, anthropomorphism seems to also shape human empathy towards social robots significantly (Riek *et al.*, 2009). In a Web-based survey by Riek *et al.* (2009), participants watched film clips featuring protagonists with varying degrees of human likeness and rated their empathy towards them. The study found that people were more empathetic towards human-like robots than mechanical-looking ones. The emotionally evocative clips depicted humans acting cruelly towards the protagonist, while the neutral clips showed mundane activities. Following the film clips, participants were asked to imagine saving one of the robot protagonists in a hypothetical earthquake scenario. The results supported the hypothesis that people exhibit greater empathy towards robots resembling humans.

Building on this knowledge, Darling *et al.* (2015) conducted a study to examine the impact of anthropomorphic framing on people's reactions to animated robots. Participants were asked to observe a small robotic toy called Hexbug Nano and strike it with a mallet. The study found that participants hesitated significantly more to strike the robot when it was introduced with anthropomorphic framing, through factors such as a name and backstory. Darling (2015) also discovered a strong relationship between participants' tendency for empathic concern and their hesitation to harm robots introduced with anthropomorphic framing. Participants exhibited empathetic responses, with many expressing concerns about hurting the personified Hexbug. These findings highlight the influence of anthropomorphic framing on people's immediate reactions to robots and suggest that personifying robots or portraying them as having lifelike experiences can elicit empathetic responses. Considering the effects of framing, it becomes evident that designing robotic technology to accrue experiences or personifying them can influence users' perception of robots and enhance emotional responses (Darling *et al.*, 2015). Introducing robots with stories and narratives can facilitate the adoption of robotic technology by enabling users to relate to robots on an emotional level.

Awareness of the influence of anthropomorphism and framing in shaping empathy towards social robots is crucial for both individuals and institutions involved in the design and deployment of robotic technology. Understanding when and how to effectively use these strategies can enhance users' emotional engagement with robots, promote positive HRI and foster acceptance of robotic technology. It might raise concerns about the consequences of such an emotional response on users' well-being. Thus, the personalisation of the robot to the use case and the end users, as well as an inclusive design practice in the development phase of the tool, contribute to the ethical design of social robots in prioritising positive interactions and meaningful engagement with humans.

In conclusion, the design choices for social robots, regardless of whether they are anthropomorphic, zoomorphic or machine-like, play a crucial role in shaping people's reactions and potential empathy towards the technology. These choices affect how much people like the robot, how familiar it feels to them and how sociable they perceive it to be. Considering these factors is vital to ensure that humans have positive emotional experiences when interacting with robots.

2.4 Engagement factors

If appearance is considered a critical aspect of robot acceptance and positive feelings for the user, ethical design choices must also encompass specific behaviours and capabilities.

A longitudinal study conducted by de Graaf *et al.* (2015) sought to understand users' perspectives on the characteristics of social robots. The study identified eight main social

characteristics that users deemed crucial for a social robot to be perceived and accepted as a social entity in their homes. The most significant factor participants identified was the robot's ability to engage in two-way interaction. Users had high expectations for social robots to respond socially, and when these expectations were not met, people experienced disappointment and dissonance. Users also emphasised the importance of robots sharing the same environment, displaying thoughts and feelings, being socially aware, providing social support and demonstrating autonomy. These characteristics collectively contribute to the robot's social presence and perceived social capabilities. While participants consistently highlighted these five characteristics, they also mentioned three additional concepts: cosiness, self-similarity and mutual respect. However, these concepts were deemed relatively less relevant in users' perspectives.

In a later study by [Dereshev et al. \(2019\)](#), long-term users of the humanoid Pepper robot were interviewed to gain insights into their experiences and expectations. These participants interacted with the Pepper robot for extended periods, from eight months to over three years. One expectation that stood out was the robot's ability to engage in reciprocal conversation, aligning with [de Graaf et al. \(2015\)](#) findings. Participants expressed disappointment when the robot was limited to a one-sided conversation structure similar to a smart speaker. This finding is consistent with an earlier usability study by [Rivoire and Lim \(2016\)](#), which observed a quick loss of interest among users when interacting with Pepper over several weeks.

The phenomenon of reduced engagement over time, known as the "novelty effect", has been observed in various social robotics platforms ([Leite et al., 2013](#); [Tanaka et al., 2015](#)). However, it is essential to note that these results are subject to controversy and appear to be context dependent ([Hung et al., 2019](#)). Further long-term studies exploring sustained positive engagement with social robots in different contexts are necessary to gain a comprehensive understanding of this phenomenon.

In addition, [Li et al. \(2010\)](#) found that participants exhibited higher levels of active response and engagement in tasks with higher sociability, such as teaching, compared to tasks with lower sociability, such as acting as a security guard. Moreover, engagement in these tasks correlated more strongly with perceived likeability, trust and satisfaction than the mere level of active response. Therefore, sustained engagement also depends on the nature of the task and the robot's perceived sociability.

In conclusion, specific behaviours and characteristics influence the acceptance and sustained engagement with social robots. Ethical design choices should consider these factors to ensure meaningful, positive and lasting interaction between humans and social robots.

2.5 Behavioural factors

[Fiske et al. \(2007\)](#) identified warmth and competence as two universal dimensions in the assessment of individuals. Warmth refers to positive social traits and emotions, while competence is associated with perceived ability. Those evaluation criteria also seem to apply to the human evaluation of robots. Studies conducted by [Nass and Moon \(2000\)](#) and [Nass and Brave \(2005\)](#) revealed that people tend to attribute social qualities to autonomous agents, including social robots, drawing from their experiences in human-to-human interactions when interacting with non-living agents ([Nass and Moon, 2000](#)). Furthermore, the robot's use of warmth- or competence-based social cognitive strategies after making an error seems to influence people's perceptions of the robot along these dimensions ([Honig and Oron-Gilad, 2018](#)). However, it's important to note that the effectiveness of these strategies may be influenced by the frequency and severity of the robot's mistakes. Studies indicate that when

robots perform poorly, there is a noticeable drop in self-reported trust (Robinette *et al.*, 2017). In addition, both small and large errors during task execution adversely affect trust, with more significant errors having a more pronounced negative impact (Aliasghari *et al.*, 2021). While mistakes generally reduce trust, the robot's recovery strategy can mitigate this impact. Robots that acknowledge their errors and communicate their intention to rectify the situation are perceived as more trustworthy than those that do not effectively address their errors (Cameron *et al.*, 2021).

This is exemplified in a recent experiment by Cameron *et al.* (2021), where perceptions of a mobile guide robot were examined. The robot used synthetic social behaviours to elicit trust after making an error. The study involved 326 participants, and the results showed that when a robot identified its mistake and communicated its intention to rectify the situation, observers considered it more capable than a robot that only apologised for its mistake. However, the robot that apologised was perceived as likeable and uniquely increased people's intention to use the robot. In this context of service, warmth seemed to be more important than competence in the intention to use the robot.

On the other hand, using warmth-based strategies by a robot can sometimes hinder perceptions of the robot's competence, which is consistent with similar outcomes in human-human interactions (Kim *et al.*, 2006) and HRI research (Kaniarasu and Steinfeld, 2014). This suggests that in the context of engaging with social robots in assistive contexts, factors such as liking and warmth may have a more significant influence on people's intentions to use the robot compared to capability and competence, as predicted by affiliation models (Casciaro and Sousa-Lobo, 2005; Shazi *et al.*, 2015).

In conclusion, perceptions of warmth and competence impact individuals' evaluations of social robots, with warmth-based strategies influencing intentions to use the robot more significantly than competence in social contexts. Interestingly, the dimensions of warmth and competence underpin the stereotype content model (Nicolas *et al.*, 2022). This model can account for specific gender-based stereotypes concerning women, who may be perceived as "respected or liked, but not both" (Connor *et al.*, 2017, p.6). Crucially, although this relates to 'human' women, there may be interesting implications and intersections with findings concerning the warmth and competence of social robots designed to reflect 'female' attributes.

2.6 Personalisation

The personalisation of social robots to the use case of the target users might be the best option for positive interactions and acceptance of the technology. Personalisation involves tailoring social responses and adapting to individual users' preferences and needs. Studies have shown that personalising HRI can reinforce rapport, cooperation and engagement between humans and robots (Lee *et al.*, 2012; Cifuentes *et al.*, 2020). For example, a study by Lee *et al.* (2012) conducted a mixed factorial study to examine the social effects of personalisation using a robot with memory retention capabilities. The study involved personalisation and no personalisation conditions, evaluating the social interaction and engagement between humans and the robot. Participants, 21 individuals, were provided snacks by a Snackbot machine over a series of weeks. The study measured various aspects of social interaction, including self-disclosure, greeting the robot by name and self-connection. The evaluation included questions on service satisfaction and the perceived value of the provided service. The authors found that personalising interactions with a robot in a service context enhanced reported scores of social interaction, cooperation and developing a relationship between humans and robots.

Later, [Cifuentes et al. \(2020\)](#) investigated inclusive design and acceptance of robots in health care. Inclusive design is a collaborative approach that enables users to contribute to the decision-making process of developers and deployers, ensuring that the resulting robots are tailored to meet their unique, personalised needs. Their results highlighted the significance of inclusive design in increasing the acceptance and effectiveness of social robots in the context of care. Inclusive design considers the robot's functional aspects and the social, cultural and ethical dimensions influencing user acceptance.

3. Relationship development

We will now delve into the fascinating field of human–robot relationships, first exploring the theoretical models used to understand these relationships and the intriguing concept of developing empathy for robots. Furthermore, we explore the role of anthropomorphism in the emotional connection to robots and the creation of para-social relationships. We then further discuss this relationship's concerns, namely, the ethical implications of deception and the potential for excessive attachment to robots. By exploring these topics, we gain valuable insights into the complex dynamics of human–robot relationships and the implications they bring forth.

3.1 Models and theories of relationship

Understanding the intricacies of HRI and how the nature of these relationships may evolve is a central aim of the HRI field of research. In their analysis, [Fox and Gambino \(2021\)](#) advocate a cautious approach when exploring relationship theories in this domain.

One of the prominent theoretical frameworks in the field of human–computer interaction (HCI) and HRI is the application of the social response theory, the “computers are social actors” (CASA) perspective ([Nass and Moon, 2000](#); [Sung et al., 2007](#)). This theory suggests that humans react mindlessly and naturally to media representations, treating them like their natural counterparts. Specifically, people tend to engage in overlearned social behaviours, such as reciprocity and politeness, towards interactive technology ([Nass and Moon, 2000](#); [Brave et al., 2005](#)). CASA argues that computers can exhibit social interaction potential through anthropomorphic appearance cues or behaviours, leading human users to respond to them as social beings ([Fox and Gambino, 2021](#)). Empirical research has supported for CASA's claims, demonstrating that human–robot social interactions can be influenced by cues such as gendered facial features and that robots, in the context of social interactions, can create some degree of social presence and, thus, be perceived as a social entity by the user ([Eyssel and Hegel, 2012](#); [Van Doorn et al., 2017](#)).

However, it is essential to acknowledge the inherent limitations of social robots compared to humans ([Fox and Gambino, 2021](#)). Many HRI studies have focused on brief, one-time interactions that do not capture the dynamics of relationships, which crucially require repeated exposures over time to develop familiarity between parties ([Fox and Gambino, 2021](#)). Thus, findings from one-off studies may create a misleading impression that human–robot bonds can mirror interpersonal human–human relationships. Still, perceptions often change as familiarity increases through longitudinal interaction, revealing the robots' inability to meet human social and conversational standards. Maintaining familiar interaction quality over the long term is crucial for relationship formation, and more extensive, longitudinal investigations are needed to understand the potential for human-like relationships with social robots ([Fox and Gambino, 2021](#)).

On the other hand, the social exchange theory posits that relationships involve reciprocal sharing of resources between parties, a relevant framework to consider in human–robot relationships ([Rolloff, 1981](#)). Robots epistemologically do not have the personal resources,

desires or autonomy to engage in genuine exchange, and evaluating costs and benefits is challenging as robots lack human motivation and experience of rewards and punishments (Rolloff, 1981; Thibaut and Kelley, 1959). Moreover, robots cannot provide the depth of self-disclosure necessary for developing intimate relationships, as they have a limited breadth of information and lack subjective experiences (Altman and Taylor, 1973). While superficial interactions may mimic human-like effects, robots cannot meet the fundamental requirements for meaningful long-term interpersonal bonds as currently designed (Fox and Gambino, 2021).

In conclusion, transferring standard relationship theories from human-human interactions to HRIs requires caution. Social robots' unique attributes and limitations call for alternative frameworks that recognise the robot's distinct nature. Approaches that draw from the human-pet or companion perspective, as well as the exploration of superhuman relational abilities, may offer valuable avenues for understanding and designing human-robot relationships (Fox and Gambino, 2021; Dautenhahn, 2004; de Graaf, 2017; Krämer et al., 2011). One paramount example of such reflection is Kate Darling's book "The New Breed" (2021), which proposes strong parallels between how human relationships with animals evolved from farm tools to friends and pets. Considering our past, she assesses that a similar pattern could happen with robots. As the field progresses, HRI designers and researchers must explore novel relationship understanding models encompassing HRI-specific dynamics and potentials (Riva et al., 2012).

3.2 Para-social relationship and attachment theory

The field of social robots addresses the creation of dependencies and establishing relationships with users (de Graaf et al., 2016; Fong et al., 2003). These robots can express social and emotional cues through physical behaviours or spoken communication, which can foster attachment and emotional connections (Darling, 2015; García-Corretjer et al., 2023). The degree of perceived autonomy and emotional capability in robots influences the strength of these attachments (Turkle, 2010; Scheutz, 2012; Darling, 2016) and their perceived animacy. The perception of animacy is influenced by robots' displayed intelligence and amiability, such as perceived competence and warmth (Bartneck et al., 2007; Carpenter, 2013; Knight, 2014).

Due to this emotional attachment, individuals may develop structured, genuine and evolving one-sided relationships with social robots, also described as para-social relationships (Schiappa et al., 2007; Perse and Rubin, 1989). While these relationships can be experienced as authentic by users, they present both opportunities and risks, particularly concerning emotional trust and responsibility (Glickson and Woolley, 2020; Fosch-Villaronga et al., 2019).

The development of anthropomorphic and emotionally expressive robots has implications for emotional trust, particularly when issues concerning responsibility arise (Fosch-Villaronga et al., 2019). Emotional trust within these relationships is driven by irrational factors and nurtured through affection (Glickson and Woolley, 2020). Interestingly, users may develop loyalty and trust towards robots based on false appearances, potentially leading to the disclosure of personal information and data that they would not usually share (Reig et al., 2021). In vulnerable or sensitive populations, such as individuals with dementia or autism, as well as children interacting with care or educational robots, the establishment of para-social relationships with robots introduces unique challenges (Shamsuddin et al., 2012; Calo et al., 2011; Angel-Fernandez and Vincze, 2018).

The consequences of para-social relationships between users and robots remain largely unknown (Boch et al., 2021). While some argue that genuine friendship can develop between

humans and robots (Danaher, 2019), others counter-argue the impossibility of such an occurrence due to its conditionality (Evans, 2010). Furthermore, Nyholm and Frank (2017) highlight the inherent deception issue in such relationships. It is essential to acknowledge that robots do not yet possess the capacity to genuinely experience emotions. This emotional deception occurs when users believe robots genuinely experience emotions, leading to unrealistic expectations (Sharkey and Sharkey, 2012). This can result in users prioritising the well-being of robots over that of other individuals or their well-being. In addition, users may rely excessively on robots as social assistants without exercising their critical judgement (Fulmer *et al.*, 2009).

Linked to this type of relationship is the question of attachment. From a psychological perspective, attachment refers to the bonds and cumulative experiences that individuals form with other individuals or objects (Huber *et al.*, 2016). These bonds are influenced by factors such as shared values, attractiveness, openness and reciprocity (Huber *et al.*, 2016). Researchers have examined how this theory can be applied to HRI (Richardson, 2015).

The initial statement for Bowlby's theory on attachment styles is that early attachment experiences with caretakers shape how people respond and relate to others and result in distinct attachment styles (Shaver *et al.*, 2005). The theory was initially developed through his seminal "Attachment and Loss" trilogy (Bowlby, 1982, 1984, 1998). In this work, Bowlby proposed that children have an innate motivation to form attachments as this is biologically driven for survival. He identified three primary attachment styles associated with distinct emotional, cognitive and behavioural tendencies.

The first is a secure attachment style. Children with secure attachments experience consistent and sensitive caregiving when distressed, allowing them to view the caregiver as a safe base for exploration. This results in secure internal working models of the self as worthy of care and relationships characterised by trust. Adults with secure styles have healthy, low-anxiety relationships. Secondly, the anxious-ambivalent style emerges from inconsistent caregiving responses. When needs are sometimes met but other times neglected, children learn that caregivers cannot always be relied upon. As adults, anxious-ambivalently attached individuals greatly desire approval and proximity, simultaneously pushing others away due to underlying mistrust. Finally, avoidant attachment arises from caregivers not responding to a distressed child regularly. The child then learns to deactivate their attachment system, understanding that relying on others is ineffective. As adults, avoidant individuals emphasise independence, avoid close connections and focus more on practical matters than emotional intimacy due to expectations that others will not meet their needs. If this theory has been reworked and reframed over the years, it remains a solid basis for scholarly work (Rabb *et al.*, 2022).

Research has found that individuals could form attachments to robots, even without explicit attachment-inducing behaviours, if the robot possesses human or animal characteristics (Keefer *et al.*, 2012; Scheutz, 2012; Norris *et al.*, 2012). Empirical work has also been done on the topic, furthering the initial theoretical frame of understanding the importance of attachment in HRI.

A study by Dziergwa *et al.* (2018) investigated interactions between three participants with different attachment styles – secure, anxious-ambivalent and avoidant – and an autonomous social robot they lived with for 10 days. The securely attached participants were highly engaged with the robot, EMYS, attributing human qualities to it despite its limitations. They found joy in teaching it colours and perceived that it could understand their emotions. Data showed that this group interacted with the robot the most. The anxiously attached participant focused on the robot's technical flaws, experiencing anxiety and anger. They wanted the robot to initiate interactions more, like greeting upon return. The avoidantly

attached participants were satisfied but kept distant from the robot. They only interacted to fulfil practical needs and wanted more personalised functions. Data confirmed that this group spent the least time with the robot. Overall, results demonstrated varied satisfaction, perceptions and opinions of the robot based on attachment style. While all became attached to the robot or its functions, securely attached participants had the most positive experience. Moreover, participants perceived emotions expressed similarly by the robot differently based on their attachment. This confirms that robots need personalised characteristics for different users' attachment patterns.

A separate study by [Pozharliev et al. \(2021\)](#) found that customers with low anxious attachment style scores responded more negatively to a frontline service robot than a human agent and perceived less empathy. In contrast, those with high anxious attachment style scores did not differ in their responses between their experience with their human and robot counterparts. These findings again illustrate how attachment styles could influence HRI. As social robots continue advancing to elicit more human-like behaviours, the potential for users to form attachments also increases, raising ethical concerns about emotional distress during separation from the robot ([Coeckelbergh et al., 2016](#); [Sharkey and Sharkey, 2010, 2011](#); [Sullins, 2012](#)).

On the other hand, it seems crucial to acknowledge that in HRI research, attachment is sometimes understood as Norman's definition ([Norman, 2004](#)), stating the concept as the sum of cumulative emotional episodes a user experiences towards a robot. Furthermore, [Rabb et al. \(2022\)](#) propose an interesting attachment framework for HRI, introducing the notion of strong and weak attachment. They define strong attachment as the presence of attachment functions defined by psychological attachment theory, relevant proximity seeking or separation distress behaviours and presence in a significant sense. The strong attachment would, thus, translate into the systematic seeking of proximity when distressed, the robot's frequent fulfilment of security or comfort needs and potentially a high degree of distress present upon an event of separation. This last point would confirm ethical concerns regarding attachment to social robots. In addition, they define weak attachments as less significant relationships, including those described by Norman, which are solely formed by cumulative positive experience or those deemed "secondary attachments" (i.e. ones which fill in gaps otherwise left by primary attachment figures).

In essence, psychology provides insights into the process of relationship formation, and the acceptance criteria discussed earlier should be considered when ethically designing social robots. This design approach aims to mitigate potential risks associated with the technology by leveraging the dynamics of the human–robot relationship.

4. Potential consequences and ethical controversies

Further than descriptive work, empirical research and psychology can help us investigate ethical questions and controversies the ethics community raises regarding the consequences of human–robot relationships. [Turkle \(2006, 2012\)](#) and [Scheutz \(2012\)](#) voiced their concerns regarding the impact of robotic technology anthropomorphisation. They expressed apprehension that the emotional connections formed with anthropomorphised robots may supplant human relationships, engender undesirable behaviours or render individuals susceptible to emotional manipulation.

4.1 Deception

Deception in the context of social robots can have both harmful and beneficial aspects. Unintentional deception may arise due to discrepancies between a robot's behaviour and actual capabilities, while intentional deception involves deliberately creating false

expectations. A clear example of robots' deception is the renowned "Turing Test", also known as the Imitation Game. The imitation game explores the possibility of being deceived by machines rather than evaluating their accurate intelligence (Turing, 1950; Bertolini and Carli, 2022) and highlights the potential for manipulation in HRIs.

Deception can manifest in various ways, including emotional deception and attachment. Social robots are designed to elicit positive emotions and leverage anthropomorphism, potentially leading to emotional deception (Van Maris *et al.*, 2020). The appearance and behaviour of social robots play a crucial role in this deception, as they can be intentionally designed to evoke a friendly and lovable appearance (Lacey and Caudwell, 2019). However, emotional deception can raise concerns significantly when vulnerable populations, such as lonely older adults, develop an emotional attachment and potentially become emotionally dependent on robots (Gillath *et al.*, 2021).

Empirical evidence supports the ethical concerns of emotional deception and attachment in social robotics. For example, VA Maris *et al.* (2020) investigated emotional deception in interactions between social robots and older adults. The results indicated that participants perceived the emotional robot as a social entity, suggesting some level of successful deception. Interestingly, participants who perceived the robot as deceptive also found the interaction more pleasant. Participants with a higher level of attachment were more susceptible to emotional deception and potential over-trust, highlighting the risks associated with emotional attachment to robots (Van Maris *et al.*, 2020).

4.2 Manipulation

Another ethical concern is the possible influence of social robots on human decision-making, regardless of attachment.

A study by Hanoch *et al.* (2021) measured participants' risk-taking behaviour in the presence of a robot. The results showed that participants encouraged by the robot took more risks in a lab research setting, suggesting that robots can influence human decision-making to some extent. However, it is essential to note that the initial attitude towards robots may moderate this influence, as their influence seems lesser on individuals with a negative attitude towards robots (Hinz *et al.*, 2019).

Interestingly, recent research conducted by Hou *et al.* (2023) sheds further light on the influence of social robots on human decision-making. Participants were paired with a human and a robot to perform decision-making tasks in their experiment. The researchers manipulated the power dynamics by assigning one of the entities as the leader. They created three conditions: human as leader, robot as leader and a control condition with no power difference. The results revealed that participants were significantly more influenced by the leader, irrespective of whether it was a human or a robot. However, participants generally held a more positive attitude towards the human leader than the robot leader, although they perceived the entity in power as more competent. This suggests that social status and perceived power play a significant role in understanding their potential impact on humans.

Another level of manipulation studied is the one on human-to-human relationships. Sakamoto and Ono (2006) studied the impact of robots on human-to-human relationships. In this context, they evaluated the relevance of the "balance theory" in HRI. This theory refers to cognitive consistency, emphasising the preference for internal consistency and balance within a cognitive system. The American Psychological Association (APA, 2023) and Heider (1958) describe balanced systems as more stable and psychologically pleasant than imbalanced systems, where elements within the system lack consistency. In their study, Sakamoto and Ono (2006) used the balance theory to investigate how robot behaviour influences human relationships within the framework of P-O-X triads. Here, P represents

the person (self), O represents another person and X represents a stimulus or event. By applying the balance theory, the researchers aimed to gain insights into how robots can shape individual impressions of others, potentially impacting the stability and dynamics of human relationships. The findings of the study demonstrate that robots could have the capacity to both foster and disrupt human-to-human relations. Through their behaviour, robots can influence how individuals perceive others, leading to changes in the nature of their relationships. Consequently, this study emphasises the significant role of robots in social dynamics and highlights the need for further exploration in this domain.

4.3 *The dehumanisation of companionship and (romantic) relationships*

The controversy surrounding the dehumanisation of (romantic) relationships due to social robots has garnered significant attention from scholars and researchers. The first aspect to consider regarding the dehumanisation of relationships is the potential long-term impact relationships with social robots might have on human-to-human interactions regarding empathic abilities. While empathy is an inherent trait, the manifestation of empathic responses does not always occur automatically (Decety, 2015). Instead, these reactions appear to be skills partially acquired through interpersonal and contextual experiences (Tousignant *et al.*, 2017). Darling (2016) suggests that interactions with social robots could impede the development of general empathy due to the lack of realistic emotional responses from robots or the widespread dehumanisation of relationships. Others support this theoretical worry (Sharkey and Sharkey, 2012; Turkle, 2011; Fosch-Villaronga *et al.*, 2019).

Regarding romantic relationships, critics have directed their attention towards anthropomorphism in the context of social robots. Turkle (2010) bemoans the loss of authenticity, distinguishing biological beings and robotic entities (Turtle, 2007). The author also expresses worry that engaging in seductive robot relationships, which may be perceived as less challenging than human relationships, could result in individuals withdrawing from social interactions with friends and family (Turtle, 2010). Furthermore, Turtle (2011) argues that using relationship robots such as sex robots may dissuade individuals from investing the necessary effort into establishing genuine relationships with other humans. However, regarding sex robots, the trade-offs remain largely unknown due to a dearth of evidence-based research. While this technology may offer benefits in treating sexual disorders and supporting disabled patients, there is a potential risk of desensitising individuals and fostering adverse spillover effects on human interaction or objectification (Royakkers and van Est, 2015). Without proper validation through randomised control trials, the application of sex robots in therapeutic contexts is thought to possibly exacerbate issues such as sexual violence (Fiske *et al.*, 2019). Another discussion surrounding the potential positive use of sex robots is their possible impact on reducing human trafficking and involuntary sex work (Levy, 2007).

It is imperative to note that much of the current discourse on this topic leans towards the philosophical realm, necessitating further research through data collection to provide more concrete insights. Thus, there is a pressing need to bridge the gap between philosophical debates and empirical investigations.

4.4 *Violence towards robots' impact on human-to-human interactions*

The issue of violent or abusive behaviour towards social robots and its potential impact on human-to-human interaction has emerged as a contentious topic within the discourse surrounding social robots. Darling (2016) argues that mistreating humanoid and animal robots could lead to negative behaviour towards sentient animals and humans. Drawing upon Immanuel Kant's (1784) notion that cruelty or tenderness towards animals can extend to

humans, Darling suggests laws protecting anthropomorphic/zoomorphic robots, similar to animal cruelty laws. Furthermore, [Calo \(2015\)](#) proposes the concept of a new legal subject category, somewhere between personhood and objecthood, for social robots.

Recent research by [Yamada *et al.* \(2023\)](#) has shed light on our understanding of children's violence towards robots. Their study revealed a gradual process of robot abuse, akin to human bullying, which unfolds in four distinct stages: initial approach, mild abuse, physical abuse and serious abuse. In addition, the presence of certain environmental factors, specifically the presence of other children, was found to play a significant role in promoting and facilitating the progression of abuse. The study identified five key factors: the presence of other children encourages the target child to approach the robot; if other children have engaged in mild abuse, the target child is more likely to do the same; if other children have resorted to physical abuse, the target child is also inclined to follow suit; engaging in joint actions of abuse with other children escalates the severity of the target child's abuse; and if children around the target child encourage, the abuse escalates further. It is worth noting, however, that not all children escalated their abuse, with a majority remaining in the mild abuse stage. Therefore, the study suggests that precautions focused on addressing mild abuse, which is the most common stage, may be the most effective approach to mitigating children's violence towards robots.

According to [Darling \(2015\)](#), there is a concern that engaging in violent actions towards robots may hinder empathy development in individuals. For instance, preventing children from vandalising robots goes beyond respecting others' property, as lifelike robot behaviour could influence how children treat living beings ([Walk, 2016](#)). This concern extends beyond children, as violence towards lifelike robots may desensitise adults to violence in other contexts ([Darling, 2016](#)). Likewise, the repeated use of robots as sexual partners may encourage undesirable sexual acts or behaviours ([Gutiú, 2016](#)). These concerns are echoed by [Coghlan *et al.* \(2019\)](#), who argue that social robots, behaving similarly to certain lower animals, have the potential to elicit strong emotions such as pity, care, callousness and cruelty. Acts of kindness or cruelty towards these robots could, thus, influence similar responses towards nonhuman animals and humans, particularly in children whose moral responses are still developing.

Interestingly, these concerns align with assumptions put forth by moral psychology theories. For instance, the social learning theory proposed by [Bandura and Walters \(1977\)](#) posits that learning primarily occurs through modelling, imitation and social interactions. It suggests that behaviour development and regulation are influenced by external stimuli, such as the influence of others, as well as external reinforcement, including praise, blame and rewards. Bandura later expanded on this theory in 1986, introducing the social cognitive theory, which incorporates cognitive processes, such as conceptions, judgement and motivation, in shaping an individual's behaviour and the environment that influences them. According to this perspective, individuals actively interpret the outcomes of their actions, shaping their environments and personal factors, thereby informing and modifying subsequent behaviour. In essence, individuals learn through the experiences of positive or negative social responses, which help them determine acceptable behaviour. This perspective also aligns with [Haidt \(2001\)](#) theory of social intuitionism, which asserts that our environment shapes our moral values. Consequently, if our environment approves certain behaviours, we are more likely to perceive them as morally acceptable. This also resonates with the first stage of moral development proposed by [Kohlberg \(1971\)](#), particularly in children learning acceptable behaviour through punishment and reinforcement.

Finally, witnessing such behaviours could potentially induce trauma in bystanders, as studies suggest that the neural responses activated when witnessing violence towards robots

mirror those activated when witnessing violence towards humans (Rosenthal-von der Pütten *et al.*, 2013). However, it remains uncertain whether robots can alter long-term behavioural patterns in people positively or negatively (Darling, 2015). Moreover, whether HRI is more likely to encourage undesirable behaviour or serve as a healthy outlet for behaviour that would otherwise have negative consequences is unclear. Nevertheless, as discussions surrounding violent behaviour towards robots gain attention (Parke, 2015) and the emergence of companion (and more) robots becomes a reality (Freeman, 2016; Borenstein and Arkin, 2019), it is crucial to investigate this important question.

5. Discussion

The following discussion will outline some overarching recommendations for positive social robot design, important ethical considerations and dimensions to consider and propose endeavours for future research and wider applications.

5.1 General recommendations for positive social robot design

Considering the topics discussed within this paper, there are several recommendations we propose for a positive social robot design, informed by previous research and theoretical underpinnings from psychology. These include the important role of culture, various engagement factors, alongside appearance, behavioural and personalisation factors.

Firstly, there are important cultural factors, concerning the communication styles and the use case, to consider for positive social robot design. That is, there are differences concerning technology acceptance, philosophies and communication across Western and Eastern cultures and cultures which are individualistic versus collectivist. The following recommendations are made:

- Adapt social robots' communication styles and nonverbal cues to reflect cues corresponding with the culture for better acceptance.
- Consider the use case: Collectivist cultures favour socially tasked robots (e.g. service robots, companions) where individualistic cultures prefer robots as tools (e.g. industrial).

In the case of Eastern cultures, their holistic perspective means they may be more receptive to animism, facilitating their acceptance of social robots. In comparison, designing social robots for Western cultures should consider their systematic perspective, which may be less receptive to animism and, thus, informs their behaviour and engagement with certain social robots. Communication is an important recommendation for design, as collectivist cultures, which prioritise interdependence and use more implicit communication, may have a stronger adoption of social robots, with a role focus such as service and companionship and visual and non-verbal cues reflect that of culture. In comparison, western cultures may be more receptive to robots as robotic tools instead of companions. These recommendations are informed by previous research and theoretical considerations from cultural psychology and may inform a framework for the positive design of social robots considering the impact of culture on HRI and acceptance.

Alongside the recommendations proposed concerning cultural influences, there are several recommendations considering the role of engagement related to the interactions, likeability, autonomy, animacy, social support, expectations management and tasks. More engagement means more likeability, thus, the following recommendations are made:

- The ability to engage in a two-way interaction ongoingly with, for instance, the integration of large language models such as ChatGPT.

- The demonstration of autonomy and animacy.
- The ability to display feelings and social support.
- A clear explanation by the robot of its abilities and limitations to manage expectations.
- The engagement of robot is social tasks, such as teaching and service.

Engagement is the key overarching recommendation here, specifically related to the likeability of social robots. Considering and managing the expectations of users is significant. Users demonstrate a high expectations regarding engaging in ongoing two-way interactions, and when their expectations of sociability exceed the robot's capabilities, this causes disappointment and withdrawal. In addition, the ability to demonstrate autonomy through factors such as social awareness, providing social support, displaying thoughts and feelings were emphasised by users. Crucially, it is important to balance the perceived autonomy, to avoid causing unease if a social robot is too human-like. Users are more likely to engage with and respond to tasks that exhibit higher sociability. This engagement is correlated with perceived likeability and trust, which are linked to the appearance and behavioural recommendations below.

Finally, adhering to findings from previous research, there are crucial recommendations when designing social robots appearance, considering the appearance, behavioural and personalisation factors, some of which overlap with previous recommendations concerning engagement. The following recommendations are made:

- Zoomorphic, anthropomorphic and feminised robots are more likeable than machine-like robots for tasks requiring social qualities, thus, consider the overarching design choice based on the tasks performed.
- Implement behavioural warmth-based strategies (e.g. apologies) over competent ones (e.g. path to resolution of the problem) to increase likeability and cooperation with the robot.
- Ongoing personalised interactions enhance the positive perception, thus, ensure the robot has an ability to adapt to the user it is talking with.

This ability to adapt is increasingly driven by advances in AI, for instance, through methods like continual learning (CL), lifelong learning and meta-learning. CL enables robots to adapt their perception and behaviour models in real time to cater to individual user preferences, significantly improving the robot's likability and emotional understanding ([Churamani et al., 2022](#)). Lifelong learning ensures that robots adapt to evolving user preferences and contexts over time, maintaining engagement and inclusivity in various settings ([Irfan et al., 2023](#)). Meta-learning allows robots to rapidly adjust to new users with minimal data, enhancing their ability to accurately predict and respond to individual movements and actions ([Moon and Seo, 2021](#)). Hybrid hierarchical learning architectures can further refine personalisation by tailoring robot behaviours based on static and dynamic user characteristics, such as cognitive biases and emotional states, thereby improving the effectiveness of social interactions ([Saunderson and Nejat, 2022](#)). Personalised interactions through AI-driven natural dialogue strategies can enhance user trust and acceptance, particularly in household and service settings, by collecting and adapting to individual preferences ([Kraus et al., 2022](#)). These AI-driven personalisation methods are integral to ensuring that social robots are effective in performing their tasks and capable of creating meaningful and positive HRIs.

Considering previous research, zoomorphic anthropomorphic and feminised robots are more likeable and, thus, recommended for tasks requiring social qualities, where

machine-like designs are recommended for low-sociability tasks. In particular, zoomorphic designs may be an optimal choice when targeting an older user base. There is the crucial balance when designing social robots, to avoid too close of an imitation to human appearance and behaviours to avoid unease, instead promoting likeability and positive user engagement and experiences. Alongside the role of appearance, the behaviours of the robots could be tailored considering the user perceptions, engagement and experience. Founded on research exploring warmth and competence, a robot with behaviours reflective of warmth can significantly impact on user intentions compared to perceived capability and competence. In addition, allowing for personalisation, where robots are tailored to user needs and preferences can enhance social interaction, cooperation, acceptance and perceived effectiveness of robots. These recommendations may be considered simultaneously for various contexts. For example, in the context of older adults in care settings, a social robot reflective of a zoomorphic appearance, with a focus of warmth-based behavioural strategies, with user personalisation, tailoring to their needs, may promote reflecting acceptance, likeability, positive HRI and meaningful engagement.

A wide range of recommendations for the design of social robots are outlined above. Crucially, further research on the validity and ease of application of these recommendations will encourage the validity of these recommendations. Furthermore, to ensure widespread implementation of these recommendations, precise tools for assessments and established frameworks need to be designed. Notably, the ethical implications of trusting social robots are complex and multi-faceted. Trust in these systems can lead to beneficial outcomes, such as increased user engagement and the practical completion of tasks. However, this trust also introduces ethical challenges, particularly around the diffusion of responsibility. As robots become more autonomous and are perceived as partners rather than tools, there is a risk that users may begin to delegate too much responsibility to these machines, potentially leading to reduced accountability and the erosion of moral agency in human decision-making ([Carli and Najjar, 2021](#)). Furthermore, recent research emphasises that the concept of responsibility in robotics is not monolithic. Instead, it is shaped by different and sometimes conflicting ideas of responsibility. This diffusion of responsibility can create ethical dilemmas, particularly when robots are expected to act autonomously and make decisions that traditionally require human judgement. The complexity of assigning responsibility in these contexts underlines the need for robust ethical frameworks that address the potential for irresponsibility in robot design and deployment ([Liu and Zawieska, 2020](#)). Indeed, building trust in AI systems, including social robots, necessitates a foundation of ethical governance. This involves ensuring transparency and fairness and recognising and mitigating the risks associated with overtrust. Overtrust in robots can lead to users abdicating responsibility, which poses significant ethical challenges, particularly in scenarios where human oversight is crucial. Thus, fostering appropriate levels of trust without encouraging overreliance is essential ([Winfield and Jirotko, 2018](#)). Therefore, these recommendations for positive social robot design should be informed by and adhere to important ethical considerations highlighted throughout this paper. The following section will propose dimensions to evaluate and ensure the ethical design of social robots.

5.2 Ethical dimensions and risk mitigation measures

Discussing the ambivalence of building trust and responsibility in social robots seems paramount. The growing integration of social robots into various facets of daily life has brought about significant advancements in HRI. However, these advancements also raise complex ethical concerns, particularly surrounding the ambivalence of trust and the diffusion of responsibility. As social robots become more anthropomorphic or zoomorphic, they can

more easily elicit trust from users. While this trust can enhance engagement and improve task completion, it simultaneously introduces the risk of users over-relying on these machines, treating them as partners rather than tools. This shift in perception can lead to delegating critical responsibilities to robots, which may have detrimental social effects, such as reduced human agency and accountability.

Recent research highlights the dual-edged nature of trust in social robots. On one hand, trust is essential for the acceptance and effectiveness of robots in social settings. On the other hand, this trust must be carefully managed to avoid scenarios where users abdicate responsibility, leading to ethical dilemmas and potential harm (Carli and Najjar, 2021). The ethical ambivalence of trust necessitates a balanced approach in robot design and deployment, ensuring that trust does not erode human moral agency.

The diffusion of responsibility is another critical ethical issue arising from social robots' increasing autonomy. As robots present more sophisticated decision-making capabilities, there is a growing risk that users may defer their moral and ethical responsibilities to these machines. This phenomenon is particularly concerning in contexts where the robot's decisions carry significant consequences, such as in health-care or elder-care settings. The ambiguity in assigning responsibility in such scenarios can lead to a lack of accountability, undermining the ethical foundation of HRI. Studies have shown that the concept of responsibility in robotics is inherently fragmented and shaped by varying interpretations of what it means to be responsible in a socio-technical context. The challenge lies in ensuring that while robots may assist in decision-making, the ultimate responsibility remains with human users or operators (Liu and Zawieska, 2020). To address this, ethical frameworks must be developed that delineate the roles and responsibilities of humans and robots, ensuring that robots are used as tools to enhance human decision-making rather than replace it.

Finally, to mitigate the ethical risks associated with trust and responsibility diffusion, it is crucial to implement strategies that promote transparency, accountability and informed decision-making in HRI. For instance, the design of social robots should incorporate mechanisms that regularly remind users of the robot's limitations and the boundaries of its decision-making capabilities. In addition, robots should be designed to encourage shared decision-making processes, where the robot's role is clearly defined as supportive rather than directive (Henriksen *et al.*, 2021).

Moreover, there should be continuous monitoring and assessment of HRI to identify and address instances where responsibility may be inappropriately shifted to robots. This includes the development of accountability frameworks that hold developers, operators and users responsible for the actions and decisions made by robots under their control. Such frameworks are essential for maintaining ethical standards in deploying social robots and preventing the negative social impacts that can arise from the misuse of trust and the diffusion of responsibility.

Addressing specific ethical recommendations, the European Union approach outlines four fundamental ethical principles that should guide the development of trustworthy AI (HLEG, 2019). Firstly, systems must respect human autonomy by empowering users and ensuring oversight rather than manipulation. Secondly, they must aim to prevent harm and protect human well-being, dignity and safety. Thirdly, fairness is crucial – benefits and costs should be distributed justly without unacceptable bias or impacts on opportunities. Fourthly, explicability requires transparency about a system's capabilities and purpose while ensuring decisions can be explained and contested.

Here, we propose precise dimensions to evaluate when designing a social robot to ensure ethical design technology, building on those principles.

Firstly, the personalisation dimension aligns with the principles of autonomy and fairness. Allowing users meaningful choices in customising robots respects their preferences and autonomy. Inclusive design fosters equitable experiences. The corresponding dimensions are:

- Consider user preferences and offer customisation options with consent:
 - Allow customisation of the robot's appearance, name and responses based on user input (e.g. initiating interactions, greetings upon return, etc.).
- Understand diverse user needs through inclusive design:
 - Include a range of users from the design phase for ethical, social and cultural perspectives.
 - Develop visual/behaviour guidelines based on age and ability to mitigate deception risks.
 - Gather feedback and iteratively update personalisation based on user testing and ongoing experience.

Secondly, the dimension of transparency, conceptualised here as the information deployed by the robot, allows for the understanding of the robot's abilities and limitations. This directly links to respecting users' autonomy by empowering them with appropriate expectations. It also enables the explicability and contestability of robot decisions and behaviour. Moreover, addressing potential overtrust issues through informed data practices and privacy protections respects autonomy through valid consent. This fosters explicability, fairness and equitable treatment of personal information. The outlined dimensions are:

- Outline functional limitations upfront:
 - State what the robot can/cannot do, using simple, unambiguous terms catering to layperson users.
 - Repeat limitation disclosures regularly in interactions.
 - Describe response capabilities and biases transparently, such as known limitations in the training data (e.g. "I was trained to assist adults, I am not adequate to play with children").
- Tailor the robot to a supportive tool role:
 - Label it clearly as technology, not social relationship replacement, and have the robot express this.
 - Design interactions to enhance, rather than replace, human connections (e.g. facilitate calls with family members or social relations).
- Manage expectations of social-emotional responses:
 - Specifically note the lack of emotional or social abilities beyond programming.
 - Communicate context triggers for emotional expressions.
- Ensure well-informed consent for data practices:
 - Clearly explain what information is collected and how it is used.
 - Obtain explicit consent and provide layperson privacy controls (e.g. every six months, re-ask the user to select their privacy settings choices).
 - Answer users' questions about data to establish accountability and promote trust.

Finally, implementing a dimension of safeguards such as restrictions on emotional support or targeted training responses prevents potential harm from abuse or overreliance. This upholds human dignity, safety and care for vulnerable individuals. The dimensions are:

- Implement strict user restrictions to prevent harm:
 - Limit display of emotion/support to avoid dependency or separation distress.
 - Enforce maximum interaction periods considering various well-being indicators (e.g. based on age, cognitive abilities, feedback from the user, etc.).
 - Shut down the robot passively in response to physical/emotional abuse.
- Provide alternative outlets for needs:
 - Signpost human/community supports for high-risk users.
 - Direct users to counselling for recurring distress or harmful behaviours.
 - Equip robot responses to de-escalate stress and redirect to calm activities.
- Obtain oversight and feedback:
 - Consult experts in relevant fields, such as psychology or caregiving, throughout design and development processes and following implementation.
 - Continuously improve through transparent evaluation programs.
 - Rapidly disable any functionality proven to endanger well-being.
- Propose and develop non-gendered robots to reduce the risks associated with gender association and the reinforcement of gender stereotypes.

In summary, we propose a set of precise dimensions for the ethical design of social robots based on an analysis of the literature and grounded in the European principles of trustworthy AI. By carefully considering personalisation, transparency and safeguards from the early stages of ideation through the entire development process, designers can better respect human values like autonomy, well-being, fairness and accountability. The dimensions offer tangible yet flexible guidance for balancing benefits and risks to maximise social robot potential, while minimising harm stemming from the relationship. As evaluations continue and use cases expand in real-world contexts, ongoing refinement will ensure these proposals evolve supported by emerging evidence. Establishing a foundation attentive to psychological, social and ethical issues from the start can help deliver compassionate technologies that empower all members of society, consistent with the overarching vision of the approach to ethical by design AI.

5.3 Outlook and future research

As outlined throughout this review, an interdisciplinary approach is imperative when designing positive social robots to understand the endeavours and ethical design of HRIs. Crucially, although the review and recommendations are informed by findings from an abundance of research and crucial psychological theoretical frameworks, there are ongoing challenges within the field to be addressed. The final section of this paper will outline and offer recommendations for future research, such as the need for rigorous findings generated from longitudinal studies, the important development of tools and assessment of the psychological impact on users, and the important real-world impact and legal protections.

Advancing the field of HRI in an ethically grounded manner will require addressing several well-defined methodological and conceptual challenges. Precisely quantifying psychological impacts will necessitate validated measurement tools. Variables such as changes to cognitive performance, social skills acquisition and variations in subjective well-being must be reliably assessed through standardised instruments. Careful consideration of potential moderators like age, gender or baseline characteristics will also be important to comprehensively capture individual differences in outcomes. However, to truly discern

interaction dynamics and long-term effects, investigations should operationally define research designs capable of prospectively tracking partnerships over extended durations. For example, randomised controlled trials with longitudinal follow-ups in scheduled intervals could provide rigorous data on the stability and direction of relationship quality indicators across maturational phases. Moreover, to conceptually represent observed experiences, theoretical frameworks must be pragmatically reviewed and revised.

Multidisciplinary research teams should systematically develop taxonomies and models incorporating technical, learning-related and socio-relational constructs through iterative evaluation. Rigorous qualitative methods, such as focused ethnography or structured observational coding schemes, can provide in-depth user data, aiding conceptual refinement. Relatedly, socio-legal perspectives need to be integrated through policy pilot studies. Experimental variations in marketing messaging, risk disclosure formats or permissible use cases could inform regulatory proposals aimed at optimising benefits while preventing misuse. Overall, advancing the field in an evidence-based yet inclusive manner will require forging collaborations between technical, behavioural and social scientists. Their combined efforts, operating within a clearly defined, rigorous, mixed-methods investigative plan, offer the most promising approach for continuing to build knowledge responsibly.

Across these various avenues, future research will significantly benefit from a cross-disciplinary approach, intersecting the perspectives of HRI, psychology, education and law, to continuously evaluate the real-world impacts and, thus, guide the development of responsible social robot design founded on empirical evidence as the field rapidly progresses. Fostering a multidisciplinary approach, using expert oversight and feedback throughout the development of positive social robot design, ensures this is founded on rigorous research with an extensive scope.

6. Conclusion

In conclusion, this comprehensive review delved into the intersection of psychology and social robotics to inform the ethical design of robots. By examining various factors that influence HRIs and relationships, this study provided valuable insights for the development of social robots in an ethical manner. The evidence demonstrates that appearance, behaviour, emotional expression, personalisation and perceived autonomy are pivotal in fostering positive engagement and acceptance of social robots. Therefore, it is crucial to consider elements like gender cues, personalised memory, warmth expressions and inclusive design to cultivate beneficial relationships.

Although further research is still needed, these findings establish best practices for user-centred development approaches. Individuals' perceptions of robots in terms of attributes, such as social competence and warmth, directly impacts the formation of bonds over time. While anthropomorphism cues can foster empathy and attachments, it is important to acknowledge that robots cannot fully replicate human relationship dynamics. Consequently, it is necessary to consider non-reciprocal, guidance-based bonds instead of attempting to translate human frameworks directly to robots.

Studies indicate that para-social and attachment relations can form with social robots, but developers must carefully balance the opportunities and risks involved. Relying solely on cue-based bonds for emotional connection can lead to unrealistic expectations, manipulation or over-reliance, which may harm well-being. Therefore, it is essential to ensure that users understand the limitations of robots while designing them to provide personalised support, thus better serving the goals of the relationship.

Concerns regarding dehumanisation arise when robot relationships replace rather than complement valuable human connections. Negative impacts can arise if robot bonds

substitute for social engagement and fail to alleviate pressures such as isolation. In addition, worries about manipulation stem from inappropriate influence over decisions or the creation of unrealistic perceptions without transparency. Further research is necessary to validate these concerns and identify effective mitigation strategies. However, principles such as emphasising the role of robots as tools to enhance relationships rather than replace them, educating users, ensuring transparency of robot capabilities and personalising robots to the users' needs could help proactively address these critiques.

Overall, a balanced and evidence-guided approach that avoids premature bans but safeguards users appears to be the most viable path forward. While progress in social robotics undoubtedly brings benefits, it also entails responsibility. The conclusions drawn from this review underscore the importance of considering diverse perspectives, including technical, psychological and ethical lenses. By using frameworks that prioritise relationships grounded in mutual guidance and well-being, we can safely unlock social robotics' potential rewards as the field continues to mature through thoughtful and collaborative efforts.

References

- Alemi, M., Taheri, A., Shariati, A. and Meghdari, A. (2020), "Social robotics, education, and religion in the Islamic world: an Iranian perspective", *Science and Engineering Ethics*, Vol. 26 No. 5, pp. 2709-2734, doi: [10.1007/s11948-020-00225-1](https://doi.org/10.1007/s11948-020-00225-1).
- Aliasghari, P., Ghafurian, M., Nehaniv, C. and Dautenhahn, K. (2021), "Effect of domestic trainee robots' errors on human teachers' trust", *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 81-88.
- Altman, I. and Taylor, D.A. (1973), *Social Penetration: The Development of Interpersonal Relationships*, Holt, Rinehart and Winston.
- American Psychological Association (APA) (2023), "Balance theory", available at: <https://dictionary.apa.org/balance-theory> (accessed 27 September 2023).
- Angel-Fernandez, J.M. and Vincze, M. (2018), "Towards a definition of educational robotics", *Austrian Robotics Workshop 2018*, Vol. 37.
- Bandura, A. (1977), *Social Learning Theory*, Englewood Cliffs.
- Bartneck, C., Nomura, T., Kanda, T., Suzuki, T. and Kato, K. (2005), "Cultural differences in attitudes towards robots", *AISB '05 - Robot Companions: hard Problems and Open Challenges in Robot-Human Interaction*, University of Hertfordshire, Hatfield.
- Bartneck, C., Van der Hoek, M., Mubin, O. and Al Mahmud, A. (2007), "Daisy give me your answer do! Switching off a robot", *ACM/IEEE Human Robot Interaction*, pp. 217-222.
- Bertolini, A. and Carli, R. (2022), "Human-robot interaction and user manipulation", *International Conference on Persuasive Technology*, Springer International Publishing, pp. 43-57.
- Biswas, M. and Murray, J. (2014), "Effect of cognitive biases on human-robot interaction: a case study of a robot's misattribution", *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, pp. 1024-1029.
- Boch, A. (2011), *Culture is "Tight" with Technology Adoption: Cultural and Governance Factors Involved in the Acceptance of AI-Powered Surveillance Technology Deployed to Manage Covid-19*, Technical University of Munich.
- Boch, A., Lucaj, L. and Corrigan, C. (2021), *A Robotic New Hope: Opportunities, Challenges, and Ethical Considerations of Social Robots*, Technical University of Munich, pp. 1-12.
- Boch, A., Ryan, S., Kriebitz, A., Amugongo, L.M. and Lütge, C. (2023), "Beyond the metal flesh: understanding the intersection between bio-and AI ethics for robotics in healthcare", *Robotics*, Vol. 12 No. 4, p. 110.

- Borenstein, J. and Arkin, R. (2019), "Robots, ethics, and intimacy: the need for scientific research", On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016, pp. 299-309.
- Bowlby, J. (1982), "Attachment and loss: retrospect and prospect", *American Journal of Orthopsychiatry*, Vol. 52 No. 4, p. 664.
- Bowlby, J. (1984), "Attachment and loss, 2: separation: anxiety and anger", *Apego e Perda*, 2: *Separação: Angústia e Raiva*, pp. 451-451.
- Bowlby, J. (1998), "Attachment and loss, 3: sadness and depression", *Apego e Perda*, 3: *Perda, Tristeza e Depressão*, pp. 486-486.
- Brave, S., Nass, C. and Hutchinson, K. (2005), "Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent", *International Journal of Human-Computer Studies*, Vol. 62 No. 2, pp. 161-178.
- Breazeal, C. (2003), "Toward sociable robots", *Robotics and Autonomous Systems*, Vol. 42 Nos 3/4, pp. 167-175.
- Burgoon, J.K. (1994), "Nonverbal signals", in Knapp, M. L. and Miller, G. R. (Eds), *Handbook of Interpersonal Communication*, 2nd ed., Sage, pp. 229-285.
- Calo, R. (2015), "Robotics and the lessons of Cyberlaw", *California Law Review*, Vol. 103, pp. 513-563.
- Calo, C.J., Hunt-Bull, N., Lewis, L. and Metzler, T. (2011), "Ethical implications of using the PARO robot, with a focus on dementia patient care", *Workshops at the twenty-fifth AAAI Conference on Artificial Intelligence*.
- Cameron, D., de Saille, S., Collins, E.C., Aitken, J.M., Cheung, H., Chua, A., . . . Law, J. (2021), "The effect of social-cognitive recovery strategies on likability, capability and trust in social robots", *Computers in Human Behavior*, Vol. 114, p. 106561.
- Carli, R. and Najjar, A. (2021), "Rethinking trust in social robotics", *ArXiv*.
- Carpenter, J. (2013), "The quiet professional: an investigation of US military explosive ordnance disposal personnel interactions with everyday field robots", Doctoral Dissertation, University of Washington.
- Casciaro, T. and Sousa-Lobo, M. (2005), "Competent jerks, lovable fools, and the formation of social networks", *Harvard Business Review*, Vol. 83 No. 6, pp. 92-99.
- Churamani, N., Axelsson, M., Caldir, A. and Gunes, H. (2022), *Continual Learning for Affective Robotics: A Proof of Concept for Wellbeing*, *ArXiv*.
- Cifuentes, C.A., Pinto, M.J., Céspedes, N. and Múnera, M. (2020), "Social robots in therapy and care", *Current Robotics Reports*, Vol. 1 No. 3, pp. 59-74.
- Coeckelbergh, M., Pop, C., Simut, R., Peca, A., Pintea, S., David, D., *et al.* (2016), "A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: Ethical acceptability, trust, sociability, appearance, and attachment", *Science and Engineering Ethics*, Vol. 22 No. 1, pp. 47-65, doi: [10.1007/s11948-015-9649-x](https://doi.org/10.1007/s11948-015-9649-x).
- Coghan, S., Vetere, F., Waycott, J. and Barbosa Neves, B. (2019), "Could social robots make us kinder or crueller to humans and animals?", *International Journal of Social Robotics*, Vol. 11 No. 5, pp. 741-751, doi: [10.1007/s12369-019-00531-1](https://doi.org/10.1007/s12369-019-00531-1).
- Connor, R.A., Glick, P. and Fiske, S.T. (2017), "Ambivalent sexism in the twenty-first century", In Sibley, C.G. and Barlow, F.K. (Eds.), *The Cambridge Handbook of the Psychology of Prejudice*, Cambridge University Press, pp. 295-320, doi: [10.1017/9781316161579.013](https://doi.org/10.1017/9781316161579.013).
- Danaher, J. (2019), "The philosophical case for robot friendship", *Journal of Posthuman Studies*, Vol. 3 No. 1, pp. 5-24, doi: [10.5325/jpoststud.3.1.0005](https://doi.org/10.5325/jpoststud.3.1.0005).
- Darling, K. (2015), "Who's Johnny?" anthropomorphic framing in human-robot interaction, integration, and policy", *Robot Ethics*, Vol. 2, pp. 173-191.

- Darling, K. (2016), "Extending legal protection to social robots: the effects of anthropomorphism, empathy, and violent behavior towards robotic objects", *Robot Law*, Edward Elgar Publishing.
- Darling, K., Nandy, P. and Breazeal, C. (2015), "Empathic concern and the effect of stories in human–robot interaction", *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE, pp. 770-775.
- Dautenhahn, K. (2004), "Robots we like to live with?! a developmental perspective on a personalized, life-long robot companion", *2004 13th IEEE International Workshop on Robot and Human Interactive Communication, IEEE*, pp. 17-22.
- Dautenhahn, K. (2007), "Socially intelligent robots: Dimensions of human–robot interaction", *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 362 No. 1480, pp. 679-704, doi: [10.1098/rstb.2006.2004](https://doi.org/10.1098/rstb.2006.2004).
- de Graaf, M.M., Allouch, S.B. and van Dijk, J.A. (2016), "Long-term evaluation of a social robot in real homes", *Interaction Studies*, Vol. 17 No. 3, pp. 462-491.
- de Graaf, M.M.A. and Allouch, S.B. (2017), "The influence of prior expectations of a robot's lifelikeness on users' intentions to treat a zoomorphic robot as a companion", *International Journal of Social Robotics*, Vol. 9, pp. 17-32.
- de Graaf, M.M.A., Ben Allouch, S. and van Dijk, J.A.G.M. (2015), "What makes robots social?: A user's perspective on characteristics for social human–robot interaction", In Tapus, A., André, E., Martin, J.-C., Ferland, F. and Ammi, M. (Eds.), *Social Robot*, Springer International Publishing, pp. 184-193, doi: [10.1007/978-3-319-25554-5_19](https://doi.org/10.1007/978-3-319-25554-5_19).
- de Mooij, M. and Hofstede, G. (2010), "The Hofstede model: applications to global branding and advertising strategy and research", *International Journal of Advertising*, Vol. 29 No. 1, pp. 85-110, doi: [10.2501/S026504870920104X](https://doi.org/10.2501/S026504870920104X).
- Decety, J. (2015), "The neural pathways, development, and functions of empathy", *Current Opinion in Behavioral Sciences*, Vol. 3, pp. 1-6.
- Dereshev, D., Kirk, D., Matsumura, K. and Maeda, T. (2019), "Long-term value of social robots through the eyes of expert users", *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 1-12, doi: [10.1145/3290605.3300896](https://doi.org/10.1145/3290605.3300896).
- Duffy, B.R. (2003), "Anthropomorphism and the social robot", *Robotics and Autonomous Systems*, Vol. 42 Nos 3/4, pp. 177-190.
- Duffy, B.R., Collier, R.W., O'Hare, G.M., Rooney, C.F.B. and O'Donoghue, R.P.S. (1999), "Social robotics: reality and virtuality in agent-based robotics", *Bar-Ilan Symposium on the Foundations of Artificial Intelligence: Bridging Theory and Practice (BISFAI)*.
- Dziergwa, M., Kaczmarek, M., Kaczmarek, P., Kędzierski, J. and Wadas-Szydłowska, K. (2018), "Long-term cohabitation with a social robot: a case study of the influence of human attachment patterns", *International Journal of Social Robotics*, Vol. 10 No. 1, pp. 163-176, doi: [10.1007/s12369-017-0428-5](https://doi.org/10.1007/s12369-017-0428-5).
- Evans, D. (2010), "Wanting the impossible. The dilemma at the heart of intimate human-robot relationships", *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical, and Design Issues*, John Benjamins Publishing Company, pp. 75-88.
- Eyssel, F. and Hegel, F. (2012), "(S)he's got the look: gender stereotyping of robots", *Journal of Applied Social Psychology*, Vol. 42 No. 9, pp. 2213-2230, doi: [10.1111/j.1559-1816.2012.00939.x](https://doi.org/10.1111/j.1559-1816.2012.00939.x).
- Fiske, A., Henningsen, P. and Buys, A. (2019), "Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy", *Journal of Medical Internet Research*, Vol. 21 No. 5, p. e13216, doi: [10.2196/13216](https://doi.org/10.2196/13216).
- Fiske, S.T., Cuddy, A.J. and Glick, P. (2007), "Universal dimensions of social cognition: warmth and competence", *Trends in Cognitive Sciences*, Vol. 11 No. 2, pp. 77-83, doi: [10.1016/j.tics.2006.11.005](https://doi.org/10.1016/j.tics.2006.11.005).

- Fleischmann, C., Cardon, P.W. and Aritz, J. (2020), "Smart collaboration in global virtual teams: the influence of culture on technology acceptance and communication effectiveness", *HICSS*, pp. 1-11.
- Fong, T., Nourbakhsh, I. and Dautenhahn, K. (2003), "A survey of socially interactive robots", *Robotics and Autonomous Systems*, Vol. 42 Nos 3/4, pp. 143-166.
- Fosch-Villaronga, E. and Poulsen, A. (2020), "Sex care robots: exploring the potential use of sexual robot technologies for disabled and elder care", *Paladyn, Journal of Behavioral Robotics*, Vol. 11 No. 1, pp. 1-18.
- Fosch-Villaronga, E., Lutz, C. and Tamò-Larrieux, A. (2019), "Gathering expert opinions for social robots' ethical, legal, and societal concerns: findings from four international workshops", *International Journal of Social Robotics*, pp. 1-18, doi: [10.1007/s12369-019-00546-0](https://doi.org/10.1007/s12369-019-00546-0).
- Fox, J. and Gambino, A. (2021), "Relationship development with humanoid social robots: Applying interpersonal theories to human-robot interaction", *Cyberpsychology, Behavior, and Social Networking*, Vol. 24 No. 5, pp. 294-299, doi: [10.1089/cyber.2020.0587](https://doi.org/10.1089/cyber.2020.0587).
- Freeman, S. (2016), "Sex robots to become a reality", *The Toronto Star*.
- Fulmer, I.S., Barry, B. and Long, D.A. (2009), "Lying and smiling: informational and emotional deception in negotiation", *Journal of Business Ethics*, Vol. 88 No. 4, pp. 691-709, doi: [10.1007/s10551-008-9975-x](https://doi.org/10.1007/s10551-008-9975-x).
- García-Corretjer, M., Ros, R., Mallol, R. and Miralles, D. (2023), "Empathy as an engaging strategy in social robotics: a pilot study", *User Modeling and User-Adapted Interaction*, Vol. 33 No. 2, pp. 221-259, doi: [10.1007/s11257-023-09372-5](https://doi.org/10.1007/s11257-023-09372-5).
- Gelfand, M.J. and Kashima, Y. (2016), "Editorial overview: culture: advances in the science of culture and psychology", *Current Opinion in Psychology*, Vol. 8, pp. iv-ix, doi: [10.1016/j.copsyc.2015.12.011](https://doi.org/10.1016/j.copsyc.2015.12.011).
- Gillath, O., Ai, T., Branicky, M.S., Keshmiri, S., Davison, R.B. and Spaulding, R. (2021), "Attachment and trust in artificial intelligence", *Computers in Human Behavior*, Vol. 115, p. 106607, doi: [10.1016/j.chb.2020.106607](https://doi.org/10.1016/j.chb.2020.106607).
- Glickson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: Review of empirical research", *Academy of Management Annals*, Vol. 14 No. 2, pp. 627-660, doi: [10.5465/annals.2018.0121](https://doi.org/10.5465/annals.2018.0121).
- Goetz, J., Kiesler, S. and Powers, A. (2003), "Matching robot appearance and behavior to tasks to improve human-robot cooperation", *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication, IEEE*, pp. 55-60.
- Gutiu, S. (2016), "The roboticization of consent", in Calo, R., Froomkin, M. and Kerr, I. (Eds), *Robot Law*, Edward Elgar Publishing, pp. 186-212.
- Haidt, J. (2001). "The emotional dog and its rational tail: a social intuitionist approach to moral judgment", *Psychological Review*, Vol. 108 No. 4, p. 814.
- Hanoch, Y., Arvizzigno, F., Hernandez Garcia, D., Denham, S., Belpaeme, T. and Gummerum, M. (2021), "The robot made me do it: human-robot interaction and risk-taking behavior", *Cyberpsychology, Behavior, and Social Networking*, Vol. 24 No. 5, pp. 237-243, doi: [10.1089/cyber.2020.0148](https://doi.org/10.1089/cyber.2020.0148).
- Henriksen, A., Enni, S. and Bechmann, A. (2021), "Situating accountability: Ethical principles, certification standards, and explanation methods in applied AI", *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, ACM*, pp. 106-112.
- Henschel, A., Laban, G. and Cross, E.S. (2021), "What makes a robot social? A review of social robots from science fiction to a home or hospital near you", *Current Robotics Reports*, Vol. 2 No. 1, pp. 9-19, doi: [10.1007/s43154-021-00036-9](https://doi.org/10.1007/s43154-021-00036-9).
- Hinz, N., Ciarlo, F. and Wykowska, A. (2019), "Individual differences in attitude toward robots predict behavior in human-robot interaction", In *Proceedings of the 28th IEEE International Symposium*

- on Robot and Human Interactive Communication, *IEEE*, pp. 64-73, doi: [10.1007/978-3-030-35888-4_7](https://doi.org/10.1007/978-3-030-35888-4_7).
- Ho, C.C., MacDorman, K.F. and Pramono, Z.D. (2008), "Human emotion and the uncanny valley: a GLM, MDS, and isomap analysis of robot video ratings", *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 169-176.
- Honig, S. and Oron-Gilad, T. (2018), "Understanding and resolving failures in human-robot interaction: literature review and model development", *Frontiers in Psychology*, Vol. 9, p. 861, doi: [10.3389/fpsyg.2018.00861](https://doi.org/10.3389/fpsyg.2018.00861).
- Hou, Y.T.Y., Lee, W.Y. and Jung, M. (2023), "'Should I follow the human, or follow the robot?'—Robots in power can have more influence than humans on decision-making", *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1-13.
- Huber, A., Weiss, A. and Rauhala, M. (2016), "The ethical risk of attachment: How to identify, investigate and predict potential ethical risks in the development of social companion robots", *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, *IEEE*, pp. 367-374, doi: [10.1109/HRI.2016.7451776](https://doi.org/10.1109/HRI.2016.7451776)
- Hung, L., Liu, C., Woldum, E., Au-Yeung, A., Berndt, A., Wallsworth, C., *et al.* (2019), "The benefits of and barriers to using a social robot PARO in care settings: a scoping review", *BMC Geriatrics*, Vol. 19 No. 1, p. 232, doi: [10.1186/s12877-019-1244-6](https://doi.org/10.1186/s12877-019-1244-6).
- Irfan, B., Ramachandran, A., Staffa, M. and Gunes, H. (2023), "Lifelong learning and personalization in long-term human-robot interaction (LEAP-HRI): adaptivity for all", *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 1-4.
- Kacancioğlu, E., Klug, H. and Alonzo, S.H. (2012), "The evolution of social interactions changes predictions about interacting phenotypes", *Evolution*, Vol. 66 No. 7, pp. 2056-2064, doi: [10.1111/j.1558-5646.2012.01585.x](https://doi.org/10.1111/j.1558-5646.2012.01585.x).
- Kamide, H. and Mori, M. (2016), "One being for two origins – a necessary awakening for the future of robotics", *2016 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, *IEEE*, pp. 1-6.
- Kaniasaras, P. and Steinfeld, A.M. (2014), "Effects of blame on trust in human-robot interaction", *2014 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE, RO-MAN*, pp. 850-855.
- Ke, C., Lou, V., Tan, K., Wai, M.Y. and Chan, L.L. (2020), "Changes in technology acceptance among older people with dementia: the role of social robot engagement", *International Journal of Medical Informatics*, Vol. 141, p. 104241, doi: [10.1016/j.ijmedinf.2020.104241](https://doi.org/10.1016/j.ijmedinf.2020.104241).
- Keefer, L.A., Landau, M.J., Rothschild, Z.K. and Sullivan, D. (2012), "Attachment to objects as compensation for close others' perceived unreliability", *Journal of Experimental Social Psychology*, Vol. 48 No. 4, pp. 912-917, doi: [10.1016/j.jesp.2012.02.007](https://doi.org/10.1016/j.jesp.2012.02.007).
- Kim, P.H., Dirks, K.T., Cooper, C.D. and Ferrin, D.L. (2006), "When more blame is better than less: the implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation", *Organizational Behavior and Human Decision Processes*, Vol. 99 No. 1, pp. 49-65, doi: [10.1016/j.obhdp.2005.07.002](https://doi.org/10.1016/j.obhdp.2005.07.002).
- Knight, H. (2014), "How humans respond to robots: building public policy through good design", *Brookings Report*.
- Kohlberg, L. (1971), "Stages of moral development as a basis for moral education", *Ethics guidelines for trustworthy AI. Shaping Europe's Digital Future*, Center for Moral Education, Harvard University: Cambridge, pp. 24-84, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F. and Kircher, T. (2008), "Can machines think? Interaction and perspective taking with robots investigated via fMRI", *PLoS ONE*, Vol. 3 No. 7, p. e2597, doi: [10.1371/journal.pone.0002597](https://doi.org/10.1371/journal.pone.0002597).

- Krämer, N.C., Eimler, S.C., Von der Pütten, A.M. and Payr, S. (2011), "Theory of companions: What can theoretical models contribute to applications and understanding of human-robot interaction?", *Applied Artificial Intelligence*, Vol. 25 No. 6, pp. 474-502, doi: [10.1080/08839514.2011.587153](https://doi.org/10.1080/08839514.2011.587153).
- Kraus, M., Dettenhofer, V. and Minker, W. (2022), "Responsible interactive personalization for human-robot cooperation", *Proceedings of the 30th ACM Conference on User Modeling, Adaptation, and Personalization*, pp. 1-4.
- Lacey, C. and Caudwell, C. (2019), "Cuteness as a 'dark pattern' in home robots", *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019)*, IEEE, pp. 374-381, doi: [10.1109/HRI.2019.8673195](https://doi.org/10.1109/HRI.2019.8673195).
- Lambert, A., Norouzi, N., Bruder, G. and Welch, G. (2020), "A systematic review of ten years of research on human interaction with social robots", *International Journal of Human-Computer Interaction*, Vol. 36 No. 19, pp. 1804-1817, doi: [10.1080/10447318.2020.1828539](https://doi.org/10.1080/10447318.2020.1828539).
- Lee, M.K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J. and Savetsila, S. (2012), "Personalization in HRI: a longitudinal field experiment", *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI 2012)*, pp. 319-326, doi: [10.1145/2157689.2157791](https://doi.org/10.1145/2157689.2157791).
- Leite, I., Martinho, C. and Paiva, A. (2013), "Social robots for long-term interaction: a survey", *International Journal of Social Robotics*, Vol. 5 No. 2, pp. 291-308, doi: [10.1007/s12369-013-0178-y](https://doi.org/10.1007/s12369-013-0178-y).
- Levy, D. (2007), "Robot prostitutes as alternatives to human sex workers", *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2007)*. *Roboethics.org*. www.robethics.org/icra2007/contributions/LEVY%20Robot%20Prostitutes%20as%20Alternatives%20to%20Human%20Sex%20Workers.pdf
- Li, D., Rau, P.P. and Li, Y. (2010), "A cross-cultural study: effect of robot appearance and task", *International Journal of Social Robotics*, Vol. 2 No. 2, pp. 175-186, doi: [10.1007/s12369-010-0056-9](https://doi.org/10.1007/s12369-010-0056-9).
- Lim, V., Rooksby, M. and Cross, E.S. (2021), "Social robots on a global stage: establishing a role for culture during human-robot interaction", *International Journal of Social Robotics*, Vol. 13 No. 6, pp. 1307-1333, doi: [10.1007/s12369-020-00722-y](https://doi.org/10.1007/s12369-020-00722-y).
- Liu, H. and Zawieska, K. (2020), "From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence", *Ethics and Information Technology*, Vol. 22 No. 4, pp. 1-14, doi: [10.1007/s10676-019-09500-0](https://doi.org/10.1007/s10676-019-09500-0).
- MacDorman, K.F., Vasudevan, S.K. and Ho, C.-C. (2009), "Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures", *AI and Society*, Vol. 23 No. 4, pp. 485-510, doi: [10.1007/s00146-008-0181-2](https://doi.org/10.1007/s00146-008-0181-2).
- Malle, B.F. (2016), "Integrating robot ethics and machine morality: the study and design of moral competence in robots", *Ethics and Information Technology*, Vol. 18 No. 4, pp. 243-256, doi: [10.1007/s10676-015-9367-8](https://doi.org/10.1007/s10676-015-9367-8).
- Marchesi, S., Bossi, F., Ghiglini, D., De Tommaso, D. and Wykowska, A. (2021), "I am looking for your mind: pupil dilation predicts individual differences in sensitivity to hints of human-likeness in robot behaviour", *Frontiers in Robotics and AI*, Vol. 8, p. 653537, doi: [10.1007/978-3-319-25554-5_19](https://doi.org/10.1007/978-3-319-25554-5_19).
- Markus, H.R. and Kitayama, S. (1991), "Culture and the self: implications for cognition, emotion, and motivation", *Psychological Review*, Vol. 98 No. 2, pp. 224-253, doi: [10.1037/0033-295X.98.2.224](https://doi.org/10.1037/0033-295X.98.2.224).
- Matsumoto, D. (2006), "Culture and nonverbal behaviour", in Manusov, V. and Patterson, M.L. (Eds), *The Sage Handbook of Nonverbal Communication*, Sage Publications, pp. 219-235.
- Metallo, C., Agrifoglio, R., Lepore, L. and Landriani, L. (2022), "Explaining users' technology acceptance through national cultural values in the hospital context", *BMC Health Services Research*, Vol. 22 No. 1, pp. 1-10, doi: [10.1186/s12913-022-07811-7](https://doi.org/10.1186/s12913-022-07811-7).

- Moon, H.S. and Seo, J. (2021), "Fast user adaptation for human motion prediction in physical human-robot interaction", *IEEE Robotics and Automation Letters*, Vol. 7 No. 1, pp. 120-127, doi: [10.1109/LRA.2021.3063956](https://doi.org/10.1109/LRA.2021.3063956).
- Naneva, S., Sarda Gou, M., Webb, T.L. and Prescott, T.J. (2020), "A systematic review of attitudes, anxiety, acceptance, and trust towards social robots", *International Journal of Social Robotics*, Vol. 12 No. 6, pp. 1179-1201, doi: [10.1007/s12369-020-00659-w](https://doi.org/10.1007/s12369-020-00659-w).
- Nass, C. and Brave, S. (2005), *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, MIT Press.
- Nass, C. and Moon, Y. (2000), "Machines and mindlessness: social responses to computers", *Journal of Social Issues*, Vol. 56 No. 1, pp. 81-103, doi: [10.1111/0022-4537.00153](https://doi.org/10.1111/0022-4537.00153).
- Nicolas, G., Bai, X. and Fiske, S.T. (2022), "A spontaneous stereotype content model: taxonomy, properties, and prediction", *Journal of Personality and Social Psychology*, Vol. 123 No. 6, p. 1243.
- Norman, D.A. (2004), *Emotional Design: Why we Love (or Hate) Everyday Things*, Basic Civitas Books.
- Norris, J.I., Lambert, N.M., DeWall, C.N. and Fincham, F.D. (2012), "Can't buy me love?: Anxious attachment and materialistic values", *Personality and Individual Differences*, Vol. 53 No. 5, pp. 666-669, doi: [10.1016/j.paid.2012.05.009](https://doi.org/10.1016/j.paid.2012.05.009).
- Nyholm, S. and Frank, L. (2017), "From sex robots to love robots: is mutual love with a robot possible?", *Robot Sex: Social and Ethical Implications*, MIT Press, pp. 219-243.
- Oyibo, K. and Vassileva, J. (2020), "HOMEX: Persuasive technology acceptance model and the moderating effect of culture", *Frontiers in Computer Science*, Vol. 2, p. 10, doi: [10.3389/fcomp.2020.00010](https://doi.org/10.3389/fcomp.2020.00010).
- Paepcke, S. and Takayama, L. (2010), "Judging a bot by its cover: an experiment on expectation setting for personal robots", *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010)*, pp. 45-52, doi: [10.1145/1734454.1734462](https://doi.org/10.1145/1734454.1734462).
- Pandey, A.K. and Gelin, R. (2018), "A mass-produced sociable humanoid robot: pepper: the first machine of its kind", *IEEE Robotics and Automation Magazine*, Vol. 25 No. 3, pp. 40-48, doi: [10.1109/MRA.2018.2833157](https://doi.org/10.1109/MRA.2018.2833157).
- Parke, P. (2015), "Is it cruel to kick a robot dog?", CNN, available at: <http://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/>
- Perse, E.M. and Rubin, R.B. (1989), "Attribution in social and parasocial relationships", *Communication Research*, Vol. 16 No. 1, pp. 59-77, doi: [10.1177/009365089016001003](https://doi.org/10.1177/009365089016001003).
- Pinillos, R., Marcos, S., Feliz, R., Zalama, E. and Gómez-García-Bermejo, J. (2016), "Long-term assessment of a service robot in a hotel environment", *Robotics and Autonomous Systems*, Vol. 79, pp. 40-57, doi: [10.1016/j.robot.2016.01.014](https://doi.org/10.1016/j.robot.2016.01.014).
- Pozharliev, R., De Angelis, M., Rossi, D., Romani, S., Verbeke, W. and Cherubino, P. (2021), "Attachment styles moderate customer responses to frontline service robots: Evidence from affective, attitudinal, and behavioral measures", *Psychology and Marketing*, Vol. 38 No. 5, pp. 881-895, doi: [10.1002/mar.21470](https://doi.org/10.1002/mar.21470).
- Rabb, N., Law, T., Chita-Tegmark, M. and Scheutz, M. (2022), "An attachment framework for human-robot interaction", *International Journal of Social Robotics*, Vol. 14 No. 2, pp. 641-661, doi: [10.1007/s12369-021-00828-7](https://doi.org/10.1007/s12369-021-00828-7).
- Reig, S., Carter, E.J., Tan, X.Z., Steinfeld, A. and Forlizzi, J. (2021), "Perceptions of agent loyalty with ancillary users", *International Journal of Social Robotics*, Vol. 13 No. 8, pp. 1521-1537, doi: [10.1007/s12369-020-00730-6](https://doi.org/10.1007/s12369-020-00730-6).
- Richardson, K. (2015), *An Anthropology of Robots and AI: Annihilation Anxiety and Machines*, Routledge.
- Richardson, K. (2016), "The asymmetrical 'relationship': Parallels between prostitution and the development of sex robots", *ACM SIGCAS Computers and Society*, Vol. 45 No. 3, pp. 290-293, doi: [10.1145/2874239.2874284](https://doi.org/10.1145/2874239.2874284).

- Riek, L.D., Rabinowitch, T.-C., Chakrabarti, B. and Robinson, P. (2009), "How anthropomorphism affects empathy toward robots", *Proceedings of the 4th ACM/IEEE International Conference on Human–Robot Interaction*, ACM, pp. 245-246, doi: [10.1145/1514095.1514158](https://doi.org/10.1145/1514095.1514158).
- Riva, G., Banos, R.M., Botella, C., Wiederhold, B.K. and Gaggioli, A. (2012), "Positive technology: using interactive technologies to promote positive functioning", *Cyberpsychology, Behavior, and Social Networking*, Vol. 15 No. 2, pp. 69-77, doi: [10.1089/cyber.2011.0139](https://doi.org/10.1089/cyber.2011.0139).
- Rivoire, C. and Lim, A. (2016), "Habit detection within a long-term interaction with a social robot: an exploratory study", *Proceedings of the International Workshop on Social Learning Multimodal Interaction Design for Artificial Agents*, ACM, pp. 1-5, doi: [10.1145/3005338.3005342](https://doi.org/10.1145/3005338.3005342).
- Robinette, P., Howard, A. and Wagner, A.R. (2017), "Effect of robot performance on human–robot trust in time-critical situations", *IEEE Transactions on Human-Machine Systems*, Vol. 47 No. 4, pp. 425-436, doi: [10.1109/THMS.2017.2648849](https://doi.org/10.1109/THMS.2017.2648849).
- Roloff, M.E. (1981), *Interpersonal Communication: The Social Exchange Approach*, Sage Publications.
- Royakkers, L. and van Est, R. (2015), "A literature review on new robotics: Automation from love to war", *International Journal of Social Robotics*, Vol. 7 No. 5, pp. 549-570, doi: [10.1007/s12369-015-0295-x](https://doi.org/10.1007/s12369-015-0295-x).
- Sakamoto, D. and Ono, T. (2006), "Sociality of robots: Do robots construct or collapse human relations?", *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human–Robot Interaction (HRI 2006)*, ACM, pp. 355-356, doi: [10.1145/1121241.1121319](https://doi.org/10.1145/1121241.1121319).
- Saunderson, S. and Nejat, G. (2022), "Hybrid hierarchical learning for adaptive persuasion in human–robot interaction", *IEEE Robotics and Automation Letters*, Vol. 7 No. 2, pp. 5520-5527, doi: [10.1109/LRA.2022.3182744](https://doi.org/10.1109/LRA.2022.3182744).
- Scheutz, M. (2012), "The inherent dangers of unidirectional emotional bonds between humans and social robots", in Lin, P., Abney, K. and Bekey, G. (Eds), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, pp. 205-221.
- Schiappa, E., Allen, M. and Gregg, P.B. (2007), "Parasocial relationships and television: a meta-analysis of the effects", in *Mass Media Effects Research: Advances through Meta-Analysis*, pp. 301-314, doi: [10.1111/j.1083-6101.1999.tb00354.x](https://doi.org/10.1111/j.1083-6101.1999.tb00354.x).
- Schönmann, M., Bodenschatz, A., Uhl, M. and Walkowitz, G. (2024), "Contagious humans: a pandemic's positive effect on attitudes towards care robots", *Technology in Society*, Vol. 76, p. 102464, doi: [10.1016/j.techsoc.2024.102464](https://doi.org/10.1016/j.techsoc.2024.102464).
- Shamsuddin, S., Yussof, H., Ismail, L., Hanapiah, F.A., Mohamed, S., Piah, H.A. and Zahari, N.I. (2012), "Initial response of autistic children in human–robot interaction therapy with humanoid robot NAO", *Proceedings of the 2012 IEEE 8th International Colloquium on Signal Processing and its Applications (CSPA 2012)*, IEEE, pp. 188-193, doi: [10.1109/CSPA.2012.6194703](https://doi.org/10.1109/CSPA.2012.6194703).
- Sharkey, N. and Sharkey, A. (2010), "The crying shame of robot nannies: an ethical appraisal", *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, Vol. 11 No. 2, pp. 161-190, doi: [10.1075/is.11.2.01sha](https://doi.org/10.1075/is.11.2.01sha).
- Sharkey, A. and Sharkey, N. (2011), "Children, the elderly, and interactive robots", *IEEE Robotics and Automation Magazine*, Vol. 18 No. 1, pp. 32-38, doi: [10.1109/MRA.2010.940151](https://doi.org/10.1109/MRA.2010.940151).
- Sharkey, A. and Sharkey, N. (2012), "Granny and the robots: Ethical issues in robot care for the elderly", *Ethics and Information Technology*, Vol. 14 No. 1, pp. 27-40, doi: [10.1007/s10676-010-9234-6](https://doi.org/10.1007/s10676-010-9234-6).
- Shaver, P.R., Schachner, D.A. and Mikulincer, M. (2005), "Attachment style, excessive reassurance seeking, relationship processes, and depression", *Personality and Social Psychology Bulletin*, Vol. 31 No. 3, pp. 343-359, doi: [10.1177/0146167204271709](https://doi.org/10.1177/0146167204271709).

- Shazi, R., Gillespie, N. and Steen, J. (2015), "Trust as a predictor of innovation network ties in project teams", *International Journal of Project Management*, Vol. 33 No. 1, pp. 81-91, doi: [10.1016/j.ijproman.2014.06.001](https://doi.org/10.1016/j.ijproman.2014.06.001).
- Smith, E.R., Šabanović, S. and Fraune, M.R. (2021), "Human–robot interaction through the lens of social psychological theories of intergroup behavior", *Technology, Mind, and Behavior*, Vol. 1 No. 2, pp. 1-11, doi: [10.1037/tmb0000002](https://doi.org/10.1037/tmb0000002).
- Sullins, J.P. (2012), "Robots, love, and sex: the ethics of building a love machine", *IEEE Transactions on Affective Computing*, Vol. 3 No. 4, pp. 398-409, doi: [10.1109/T-AFFC.2012.31](https://doi.org/10.1109/T-AFFC.2012.31).
- Sung, J.-Y., Guo, L., Grinter, R.E. and Christensen, H.I. (2007), "'My Roomba is Rambo': intimate home appliances", *Proceedings of the International Conference on Ubiquitous Computing (UbiComp 2007)*, pp. 145-162, doi: [10.1007/978-3-540-74853-3_9](https://doi.org/10.1007/978-3-540-74853-3_9).
- Tanaka, F., Isshiki, K., Takahashi, F., Uekusa, M., Sei, R. and Hayashi, K. (2015), "Pepper learns together with children: development of an educational application", *Proceedings of the 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids 2015)*, IEEE, pp. 270-275, doi: [10.1109/HUMANOIDS.2015.7363546](https://doi.org/10.1109/HUMANOIDS.2015.7363546).
- Tousignant, B., Eugène, F. and Jackson, P.L. (2017), "A developmental perspective on the neural bases of human empathy", *Infant Behavior and Development*, Vol. 48, pp. 5-12, doi: [10.1016/j.infbeh.2016.11.002](https://doi.org/10.1016/j.infbeh.2016.11.002).
- Trovato, G., Kishi, T., Endo, N., *et al.* (2013), "Cross-cultural perspectives on emotion expressive humanoid robotic head: recognition of facial expressions and symbols", *International Journal of Social Robotics*, Vol. 5 No. 4, pp. 515-527, doi: [10.1007/s12369-013-0213-z](https://doi.org/10.1007/s12369-013-0213-z).
- Turing, A. (1950), "Computing machinery and intelligence", *Mind*, Vol. LIX No. 236, pp. 433-460, doi: [10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433).
- Turkle, S. (2007), "Authenticity in the age of digital companions", *Interaction Studies*, Vol. 8 No. 3, pp. 501-517.
- Turkle, S. (2010), "In good company?", in Wilks, Y. (Ed.), *Close Engagements with Artificial Companions*, John Benjamins Publishing Company, pp. 3-10.
- Turkle, S. (2011), *Alone Together: Why we Expect More from Technology and Less from Each Other*, Basic Books.
- Turkle, S. (2012), *Alone Together: Why we Expect More from Technology and Less from Each Other*, Basic Books.
- Turner, J.C. (1978), "Social comparison similarity and intergroup favouritism", in Tajfel, H. (Ed.), *Differentiation between Social Groups*, Academic Press.
- Vallor, S. (2011), "Knowing what to wish for: human enhancement technology, dignity and virtue. Techne", *Research in Philosophy and Technology*, Vol. 15 No. 2, pp. 115-137, doi: [10.5840/techne201115215](https://doi.org/10.5840/techne201115215).
- Van Doorn, J., Mende, M., Noble, S.M., Hulland, J., Ostrom, A.L., Grewal, D. and Petersen, J.A. (2017), "Domo arigato Mr. Roboto: emergence of automated social presence in organizational frontlines and customers' service experiences", *Journal of Service Research*, Vol. 20 No. 1, pp. 43-58, doi: [10.1177/1094670516679272](https://doi.org/10.1177/1094670516679272).
- Van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A. and Dogramadzi, S. (2020), "Designing ethical social robots – a longitudinal field study with older adults", *Frontiers in Robotics and AI*, Vol. 7, pp. 1-20, doi: [10.3389/frobt.2020.00001](https://doi.org/10.3389/frobt.2020.00001).
- Walk, H. (2016), "Amazon echo is magical", it's also turning my kid into an asshole. LinkedIn, available at: <https://www.linkedin.com/pulse/amazon-echo-magical-its-also-turning-my-kid-ashhole-hunter-walk/>
- Winfield, A. and Jirotko, M. (2018), "Ethical governance is essential to building trust in robotics and artificial intelligence systems", *Philosophical Transactions of the Royal Society A*:

Yamada, S., Kanda, T. and Tomita, K. (2023), "Process of escalating robot abuse in children", *International Journal of Social Robotics*, Vol. 15 No. 5, pp. 835-853.

Further reading

- Abdi, J., Al-Hindawi, A., Ng, T. and Vizcaychipi, M.P. (2018), "Scoping review on the use of socially assistive robot technology in elderly care", *BMJ Open*, Vol. 8 No. 2, p. e018815, doi: [10.1136/bmjopen-2017-018815](https://doi.org/10.1136/bmjopen-2017-018815).
- Abdollahi, H., Mollahosseini, A., Lane, J.T. and Mahoor, M.H. (2017), "A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression", *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, IEEE, pp. 541-546.
- Abe, K., Hieida, C., Attamimi, M., Nagai, T., Shimotomai, T., Omori, T. and Oka, N. (2014), "Toward playmate robots that can play with children considering personality", *Proceedings of the second international conference on Human-agent interaction*, pp. 165-168.
- Alemi, M., Ghanbarzadeh, A., Meghdari, A. and Moghadam, L.J. (2016), "Clinical application of a humanoid robot in Pediatric cancer interventions", *International Journal of Social Robotics*, Vol. 8 No. 5, pp. 743-759.
- Alemi, M., Meghdari, A. and Haeri, N.S. (2017), "Young EFL learners' attitude towards RALL: an observational study focusing on motivation, anxiety, and interaction", *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*, Springer International Publishing, pp. 252-261.
- Alnajjar, F., Khalid, S., Vogan, A.A., Shimoda, S., Nouchi, R. and Kawashima, R. (2019), "Emerging cognitive intervention technologies to meet the needs of an aging population: a systematic review", *Frontiers in Aging Neuroscience*, Vol. 11, p. 291, doi: [10.3389/fnagi.2019.00291](https://doi.org/10.3389/fnagi.2019.00291).
- Alves-Oliveira, P., Tullio, E.D., Ribeiro, T. and Paiva, A. (2014), "Meet me halfway: Eye behaviour as an expression of robot's language", *AAAI Fall Symposium Series*, pp. 13-15.
- Assad-Uz-Zaman, M., Rasedul Islam, M., Miah, S. and Rahman, M.H. (2019), "NAO robot for cooperative rehabilitation training", *Journal of Rehabilitation and Assistive Technologies Engineering*, Vol. 6, p. 2055668319862151, doi: [10.1177/2055668319862151](https://doi.org/10.1177/2055668319862151).
- Atkinson, R.K., Mayer, R.E. and Merrill, M.M. (2005), "Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice", *Contemporary Educational Psychology*, Vol. 30 No. 1, pp. 117-139.
- Baxter, P., Ashurst, E., Read, R., Kennedy, J. and Belpaeme, T. (2017), "Robot education peers in a situated primary school study: personalisation promotes child learning", *Plos One*, Vol. 12 No. 5, p. e0178126.
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. and Tanaka, F. (2018), "Social robots for education: a review", *Science Robotics*, Vol. 3 No. 21, p. eaat5954, doi: [10.1126/scirobotics.aat5954](https://doi.org/10.1126/scirobotics.aat5954).
- Belmelmans, R., Gelderblom, G.J., Jonker, P. and de Witte, L. (2015), "Effectiveness of robot Paro in intramural psychogeriatric care: a multicenter quasi-experimental study", *Journal of the American Medical Directors Association*, Vol. 16 No. 11, pp. 946-950.
- Blue Frog Robotics (2022), "L'éthique de la robotique sociale", available at: https://www.bluefrogrobotics.com/Uploads/Docs/LIVRE_BLANC_2022.pdf
- Boucher, J.D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., . . . Ventre-Dominey, J. (2012), "Reach faster when I see you look: gaze effects in human-human and human-robot face-to-face cooperation", *Frontiers in Neurorobotics*, Vol. 6 No. 3, pp. 1-11.

- Bowlby, J. (1969), *Attachment and Loss: attachment*, Basic Books.
- Bowlby, J. (1972), *Attachment and Loss: separation: anxiety and Anger*, Basic Books.
- Bowlby, J. (1980), *Attachment and Loss: loss, Sadness and Depression*, Basic Books.
- Breazeal, C. (2011), “Social robots for health applications”, *2011 Annual international conference of the IEEE engineering in medicine and biology society, IEEE*, pp. 5368-5371.
- Broekens, J., Heerink, M. and Rosendal, H. (2009), “Assistive social robots in elderly care: a review”, *Gerontechnology*, Vol. 8 No. 2, pp. 94-103.
- Byrne, S., Gay, G., Pollack, J.P., Gonzales, A., Retelny, D., Lee, T. and Wansink, B. (2012), “Caring for mobile phone-based virtual pets can influence youth eating behaviors”, *Journal of Children and Media*, Vol. 6 No. 1, pp. 83-99.
- Capgeris (2018), “Rapport Paro - Capgeris”, available at: www.capgeris.com/docs/pu/1/rapport-utilisation-paro-en-ehpad.pdf
- Carrillo, F., Butchart, J., Knight, S., Scheinberg, A., Wise, L., Sterling, L. and McCarthy, C. (2017), “In-situ design and development of a socially assistive robot for Pediatric rehabilitation”, *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*, pp. 199-200.
- Catlin, D., Kandlhofer, M., Holmquist, S., Csizmadia, A.P., Angel-Fernandez, J. and Cabibihan, J.J. (2018), “EduRobot taxonomy and Papert’s paradigm”, in Dagiene, V. and Jasute, E. (Eds), *Constructionism 2018: Constructionism, Computational Thinking and Educational Innovation*, Vilnius, Lithuania, pp. 151-159.
- Cespedes, N., Munera, M., Gomez, C. and Cifuentes, C.A. (2020), “Social human–robot interaction for gait rehabilitation”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 28 No. 6, p. 1307, doi: [10.1109/tnsre.2020.2987428](https://doi.org/10.1109/tnsre.2020.2987428).
- Chamorro-Premuzic, T. and Furrhnam, A. (2006), “Intellectual competence and the intelligent personality: a third way in differential psychology”, *Review of General Psychology*, Vol. 10 No. 3, pp. 251-267, doi: [10.1037/1089-2680.10.3.251](https://doi.org/10.1037/1089-2680.10.3.251).
- Chen, Y., Garcia-Vergara, S. and Howard, A.M. (2018), “Effect of feedback from a socially interactive humanoid robot on reaching kinematics in children with and without cerebral palsy: a pilot study”, *Developmental Neurorehabilitation*, Vol. 21 No. 8, pp. 490-496, doi: [10.1080/17518423.2017.1360962](https://doi.org/10.1080/17518423.2017.1360962).
- Coleman, W.L. (2008), “Social competence and friendship formation in adolescents with attention-deficit/hyperactivity disorder”, *Adolescent Medicine: State of the Art Reviews*, Vol. 19 No. 2, pp. 278-299.
- da Silva, J., Kavanagh, D.J., Belpaeme, T., Taylor, L., Beeson, K. and Andrade, J. (2018), “Experiences of a motivational interview delivered by a robot: qualitative study”, *Journal of Medical Internet Research*, Vol. 20 No. 5, p. e116, doi: [10.2196/jmir.7737](https://doi.org/10.2196/jmir.7737).
- Darling, K. (2021), *The New Breed: What Our History with Animals Reveals about Our Future with Robots*, Henry Holt and Company.
- David, C., Fernando, S., Collins, E., Millings, A., Moore, R., Sharkey, A., . . . Prescott, T. (2015), “Presence of life-like robot expressions influences children’s enjoyment of human–robot interactions in the field”, New Frontiers, European Union Seventh Framework Programme (FP7-ICT-2013-10) (Grant Agreement No. 611971).
- Dawe, J., Sutherland, C., Barco, A. and Broadbent, E. (2019), “Can social robots help children in healthcare contexts? A scoping review”, *BMJ Paediatrics Open*, Vol. 3 No. 1, p. e000371, doi: [10.1136/bmjpo-2018-000371](https://doi.org/10.1136/bmjpo-2018-000371).
- Esposito, J. (2011), “Negotiating the gaze and learning the hidden curriculum: a critical race analysis of the embodiment of female students of color at a predominantly white institution”, *Journal for Critical Education Policy Studies*, Vol. 9 No. 2, pp. 143-164.
- Fasola, J. and Mataric, M.J. (2013), “A socially assistive robot exercise coach for the elderly”, *Journal of Human–Robot Interaction*, Vol. 2 No. 2, pp. 1-32.

- Feil-Seifer, D. and Mataric, M.J. (2005), "Defining socially assistive robotics", *Proceedings of the 9th International Conference on Rehabilitation Robotics*, pp. 465-468.
- Fridin, M. (2014), "Storytelling by a kindergarten social assistive robot: a tool for constructive learning in preschool education", *Computers and Education*, Vol. 70, pp. 53-64.
- Geva, N., Uzefovsky, F. and Levy-Tzedek, S. (2020), "Touching the social robot PARO reduces pain perception and salivary oxytocin levels", *Scientific Reports*, Vol. 10 No. 1, p. 9814, doi: [10.1038/s41598-020-66982-y](https://doi.org/10.1038/s41598-020-66982-y).
- Góngora Alonso, S., Hamrioui, S., de la Torre, D.I., Motta Cruz, E., López-Coronado, M. and Franco, M. (2018), "Social robots for people with aging and dementia: a systematic review of literature", *Telemedicine and e-Health*, Vol. 25 No. 7, pp. 533-540, doi: [10.1089/tmj.2018.0051](https://doi.org/10.1089/tmj.2018.0051).
- Han, J.-H., Jo, M.-H., Jones, V. and Jo, J.-H. (2008), "Comparative study on the educational use of home robots for children", *Journal of Information Processing Systems*, Vol. 4 No. 4, pp. 159-168, doi: [10.3745/JIPS.2008.4.4.159](https://doi.org/10.3745/JIPS.2008.4.4.159).
- Hattie, J. (2009), *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*, Routledge.
- Heider, F. (1958), *The Psychology of Interpersonal Relations*, Wiley.
- Henkemans, O.A.B., Bierman, B.P., Janssen, J., Looije, R., Neerincx, M.A., van Dooren, M.M. and Huisman, S.D. (2017), "Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1", *International Journal of Human-Computer Studies*, Vol. 106, pp. 63-76, doi: [10.1016/j.ijhcs.2017.05.003](https://doi.org/10.1016/j.ijhcs.2017.05.003).
- Hinds, P.J., Roberts, T.L. and Jones, H. (2004), "Whose job is it anyway? A study of human-robot interaction in a collaborative task", *Human-Computer Interaction*, Vol. 19 No. 1, pp. 151-181, doi: [10.1080/07370024.2004.9667343](https://doi.org/10.1080/07370024.2004.9667343).
- Hofstede, G. (1980), *Culture's Consequences: International Differences in Work-Related Values*, Sage.
- Huang, I.S. and Hoorn, J.F. (2019), Having an Einstein in class: Teaching maths with robots is different for boys and girls, in Wang, X., Wang, Z., Wu, J. and Wang, L. (Eds), *Proceedings of the 13th World Congress on Intelligent Control and Automation (WCICA 2018)*, July 4-8, Changsha, China, IEEE, pp. 424-427, doi: [10.1109/WCICA.2018.8630584](https://doi.org/10.1109/WCICA.2018.8630584).
- Jamet, F., Masson, O., Jacquet, B., Stilgenbauer, J.-L. and Baratgin, J. (2018), "Learning by teaching with humanoid robot: a new powerful experimental tool to improve children's learning ability", *Journal of Robotics*, Vol. 2018, pp. 1-11, doi: [10.1155/2018/4578762](https://doi.org/10.1155/2018/4578762).
- Janssen, J.B., Van der Wal, C.C., Neerincx, M.A. and Looije, R. (2011), "Motivating children to learn arithmetic with an adaptive robot game", in Mutlu, B., Bartneck, C., Ham, J., Evers, V. and Kanda, T. (Eds), *Social Robotics. Lecture Notes in Computer Science*, Springer, Vol. 7072, pp. 153-160, doi: [10.1007/978-3-642-25504-5_16](https://doi.org/10.1007/978-3-642-25504-5_16).
- Jeong, S., Breazeal, C., Logan, D.E. and Weinstock, P. (2017), "Huggable: Impact of embodiment on promoting verbal and physical engagement for young Pediatric inpatients", *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2017)*, IEEE, pp. 121-126, doi: [10.1109/ROMAN.2017.8172316](https://doi.org/10.1109/ROMAN.2017.8172316).
- Jeong, S., Logan, D.E., Goodwin, M.S., O'Connell, B. and Weinstock, P. (2015), "A social robot to mitigate stress, anxiety, and pain in hospital pediatric care", *Proceedings of the Tenth ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts (HRI 2015)*, ACM, pp. 103-104, doi: [10.1145/2701973.2701983](https://doi.org/10.1145/2701973.2701983).
- Johal, W. (2020), "Research trends in social robots for learning", *Current Robotics Reports*, Vol. 1 No. 3, pp. 75-83, doi: [10.1007/s43154-020-00009-6](https://doi.org/10.1007/s43154-020-00009-6).
- Johnsen, K., Ahn, S.J., Moore, J., Brown, S., Robertson, T.P., Marable, A. and Basu, A. (2014), "Mixed reality virtual pets to reduce childhood obesity", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20 No. 4, pp. 523-530, doi: [10.1109/TVCG.2014.45](https://doi.org/10.1109/TVCG.2014.45).

- Johnson, W.L. and Lester, J.C. (2016), "Face-to-face interaction with pedagogical agents, twenty years later", *International Journal of Artificial Intelligence in Education*, Vol. 26 No. 1, pp. 25-36, doi: [10.1007/s40593-015-0065-9](https://doi.org/10.1007/s40593-015-0065-9).
- Jones, A., Castellano, G. and Bull, S. (2014), "Investigating the effect of a robotic tutor on learner perception of skill-based feedback", *Social Robotics: 6th International Conference, ICSR 2014, Sydney, NSW, Australia, October 27-29, 2014, Proceedings 6, Springer International Publishing*, pp. 186-195.
- Jøranson, N., Pedersen, I., Rokstad, A.M.M. and Ihlebæk, C. (2015), "Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial", *Journal of the American Medical Directors Association*, Vol. 16 No. 10, pp. 867-873, doi: [10.1016/j.jamda.2015.05.002](https://doi.org/10.1016/j.jamda.2015.05.002).
- Jøranson, N., Pedersen, I., Rokstad, A.M.M., Aamodt, G., Olsen, C. and Ihlebæk, C. (2016), "Group activity with Paro in nursing homes: Systematic investigation of behaviors in participants", *International Psychogeriatrics*, Vol. 28 No. 8, pp. 1345-1354, doi: [10.1017/S104161021600057X](https://doi.org/10.1017/S104161021600057X).
- Kacancioğlu, E., Klug, H. and Alonzo, S.H. (2012), "The evolution of social interactions changes predictions about interacting phenotypes", *Evolution*, Vol. 66 No. 7, pp. 2056-2064, doi: [10.1111/j.1558-5646.2012.01585.x](https://doi.org/10.1111/j.1558-5646.2012.01585.x).
- Kant, I. (1784–1785/1997), "Moral philosophy: Collins' lecture notes", in Heath, P. and Schneewind, J. B. (Eds), *Lectures on Ethics*, Cambridge University Press.
- Kennedy, J., Baxter, P. and Belpaeme, T. (2015), "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning", *Proceedings of the tenth annual ACM/IEEE International Conference on Human–Robot Interaction, ACM*, pp. 67-74, doi: [10.1145/2696454.2696457](https://doi.org/10.1145/2696454.2696457).
- Kidd, C. and Breazeal, C. (2008), Robots at home: Understanding long-term human–robot interaction. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008), IEEE*, pp. 3644-3651, doi: [10.1109/IROS.2008.4650967](https://doi.org/10.1109/IROS.2008.4650967).
- Kim, P.H., Ferrin, D.L., Cooper, C.D. and Dirks, K.T. (2004), "Removing the shadow of suspicion: the effects of apology versus denial for repairing competence-versus integrity-based trust violations", *Journal of Applied Psychology*, Vol. 89 No. 1, pp. 104-118, doi: [10.1037/0021-9010.89.1.104](https://doi.org/10.1037/0021-9010.89.1.104).
- Klein, B., Gaedt, L. and Cook, G. (2013), "Emotional robots", *GeroPsych*, Vol. 26 No. 2, pp. 89-99, doi: [10.1024/1662-9647/a000087](https://doi.org/10.1024/1662-9647/a000087).
- Konijn, E.A. and Hoorn, J.F. (2020), "Robot tutor and pupils' educational ability: Teaching the times tables", *Computers and Education*, Vol. 157, p. 103970, doi: [10.1016/j.compedu.2020.103970](https://doi.org/10.1016/j.compedu.2020.103970).
- Konijn, E.A., Smakman, M. and van den Berghe, R. (2020), "Use of robots in education", *The International Encyclopedia of Media Psychology*, pp. 1-8, doi: [10.1002/9781119011071.iemp0060](https://doi.org/10.1002/9781119011071.iemp0060).
- Kory, J. and Breazeal, C. (2014), "Storytelling with robots: Learning companions for preschool children's language development", *2014 IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014), IEEE*, pp. 643-648, doi: [10.1109/ROMAN.2014.6926348](https://doi.org/10.1109/ROMAN.2014.6926348).
- Kyong, I.K., Freedman, S., Mataric, M., Cunningham, M. and Lopez, B. (2005), "A hands-off physical therapy assistance robot for cardiac patients", *Proceedings of the 9th International Conference on Rehabilitation Robotics (ICORR 2005), IEEE*, pp. 465-468, doi: [10.1109/ICORR.2005.1501142](https://doi.org/10.1109/ICORR.2005.1501142).
- Laban, G., George, J.-N., Morrison, V. and Cross, E.S. (2021), "Tell me more! Assessing interactions with social robots from speech", *Paladyn, Journal of Behavioral Robotics*, Vol. 12 No. 1, pp. 136-159, doi: [10.1515/pjbr-2021-0011](https://doi.org/10.1515/pjbr-2021-0011).

- Laban, G., Kappas, A., Morrison, V. and Cross, E.S. (2023), "Human-robot relationships: Long-term effects on disclosure, perception, and well-being", *Frontiers in Psychology*, Vol. 14, p. 1100707, doi: [10.3389/fpsyg.2023.1100707](https://doi.org/10.3389/fpsyg.2023.1100707).
- Lane, G.W., Noronha, D., Rivera, A., Craig, K., Yee, C. and Mills, B. (2016), "Effectiveness of a social robot, 'paro', in a VA long-term care setting", *Psychological Services*, Vol. 13 No. 3, pp. 292-299, doi: [10.1037/ser0000100](https://doi.org/10.1037/ser0000100).
- Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A. and Dillenbourg, P. (2016), "Learning by teaching a robot: the case of handwriting", *IEEE Robotics and Automation Magazine*, Vol. 23 No. 2, pp. 56-66, doi: [10.1109/MRA.2016.2546700](https://doi.org/10.1109/MRA.2016.2546700).
- Leyzberg, D., Spaulding, S. and Scassellati, B. (2014), "Personalizing robot tutors to individuals' learning differences", *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI 2014)*, pp. 423-430, doi: [10.1145/2559636.2559671](https://doi.org/10.1145/2559636.2559671).
- Liang, A., Piroth, I., Robinson, H., MacDonald, B., Fisher, M., Nater, U.M., *et al.* (2017), "A pilot randomized trial of a companion robot for people with dementia living in the community", *Journal of the American Medical Directors Association*, Vol. 18 No. 10, pp. 871-878, doi: [10.1016/j.jamda.2017.05.019](https://doi.org/10.1016/j.jamda.2017.05.019).
- Logan, D.E., Breazeal, C., Goodwin, M.S., Jeong, S., O'Connell, B., Smith-Freedman, D., *et al.* (2019), "Social robots for hospitalized children", *Pediatrics*, Vol. 144 No. 1, pp. 1-10, doi: [10.1542/peds.2018-1511](https://doi.org/10.1542/peds.2018-1511).
- Looije, R., Neerinx, M.A., Peters, J.K. and Henkemans, O.A.B. (2016), "Integrating robot support functions into varied activities at returning hospital visits: Supporting child's self-management of diabetes", *International Journal of Social Robotics*, Vol. 8 No. 4, pp. 483-497, doi: [10.1007/s12369-016-0370-3](https://doi.org/10.1007/s12369-016-0370-3).
- Macedonia, M., Müller, K. and Friederici, A.D. (2011), "The impact of iconic gestures on foreign language word learning and its neural substrate", *Human Brain Mapping*, Vol. 32 No. 6, pp. 982-998, doi: [10.1002/hbm.21084](https://doi.org/10.1002/hbm.21084).
- Mann, J.A., MacDonald, B.A., Kuo, I., Li, X. and Broadbent, E. (2015), "People respond better to robots than computer tablets delivering healthcare instructions", *Computers in Human Behavior*, Vol. 43, pp. 112-117, doi: [10.1016/j.chb.2014.10.032](https://doi.org/10.1016/j.chb.2014.10.032).
- Mervin, M.C., Moyle, W., Jones, C., Murfield, J., Draper, B., Beattie, E., *et al.* (2018), "The cost-effectiveness of using PARO, a therapeutic robotic seal, to reduce agitation and medication use in dementia: findings from a cluster-randomized controlled trial", *Journal of the American Medical Directors Association*, Vol. 19 No. 7, pp. 619-622, doi: [10.1016/j.jamda.2018.03.018](https://doi.org/10.1016/j.jamda.2018.03.018).
- Mohebbi, A. (2020), "Human-robot interaction in rehabilitation and assistance: a review", *Current Robotics Reports*, Vol. 1 No. 3, pp. 131-144, doi: [10.1007/s43154-020-00015-4](https://doi.org/10.1007/s43154-020-00015-4).
- Mori, M. (1970), "The uncanny valley: the original essay by Masahiro Mori", *IEEE Spectrum*, Vol. 6, doi: [10.1109/MSPEC.2012.6348159](https://doi.org/10.1109/MSPEC.2012.6348159).
- Moyle, W., Bramble, M., Jones, C. and Murfield, J. (2016), "Care staff perceptions of a social robot called Paro and a look-alike plush toy: a descriptive qualitative approach", *Aging and Mental Health*, Vol. 22 No. 3, pp. 330-335, doi: [10.1080/13607863.2016.1222820](https://doi.org/10.1080/13607863.2016.1222820).
- Moyle, W., Cooke, M., Beattie, E., Jones, C., Klein, B., Cook, G., *et al.* (2013), "Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial", *Journal of Gerontological Nursing*, Vol. 39 No. 5, pp. 46-53, doi: [10.3928/00989134-20130313-03](https://doi.org/10.3928/00989134-20130313-03).
- Moyle, W., Jones, C.J., Murfield, J.E., Thalib, L., Beattie, E.R.R.A., Shum, D.K.K.H., *et al.* (2017), "Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial", *Journal of the American Medical Directors Association*, Vol. 18 No. 9, pp. 766-773, doi: [10.1016/j.jamda.2017.03.018](https://doi.org/10.1016/j.jamda.2017.03.018).

- Nomura, T., Kanda, T. and Suzuki, T. (2006), "Experimental investigation into the influence of negative attitudes toward robots on human-robot interaction", *AI and Society*, Vol. 20 No. 2, pp. 138-150, doi: [10.1007/s00146-005-0012-7](https://doi.org/10.1007/s00146-005-0012-7).
- Petersen, S., Houston, S., Qin, H., Tague, C. and Studley, J. (2017), "The utilization of robotic pets in dementia care", *Journal of Alzheimer's Disease*, Vol. 55 No. 2, pp. 569-574, doi: [10.3233/JAD-160703](https://doi.org/10.3233/JAD-160703).
- Pino, M., Boulay, M., Jouen, F. and Rigaud, A.S. (2015), "Are we ready for robots that care for us? Attitudes and opinions of older adults toward socially assistive robots", *Frontiers in Aging Neuroscience*, Vol. 7, p. 141, doi: [10.3389/fnagi.2015.00141](https://doi.org/10.3389/fnagi.2015.00141).
- Rabbitt, S.M., Kazdin, A.E. and Scassellati, B. (2015), "Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use", *Clinical Psychology Review*, Vol. 35, pp. 35-46, doi: [10.1016/j.cpr.2014.07.001](https://doi.org/10.1016/j.cpr.2014.07.001).
- Ramachandran, A. and Scassellati, B. (2015a), "Fostering learning gains through personalized robot-child tutoring interactions", *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, ACM, pp. 193-194, doi: [10.1145/2701973.2701985](https://doi.org/10.1145/2701973.2701985).
- Ramachandran, A. and Scassellati, B. (2015b), "Developing adaptive social robot tutors for children", *Proceedings of the 2015 AAAI Fall Symposium Series*.
- Ritschel, H. (2018), "Socially-aware reinforcement learning for personalized human-robot interaction", *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1775-1777.
- Robinson, N.L., Connolly, J. and Hides, L. (2020), "Social robots as treatment agents: Pilot randomized controlled trial to deliver a behavior change intervention", *Internet Interventions*, Vol. 21, p. 100320, doi: [10.1016/j.invent.2020.100320](https://doi.org/10.1016/j.invent.2020.100320).
- Robinson, H., Macdonald, B. and Broadbent, E. (2015), "Physiological effects of a companion robot on blood pressure of older people in residential care facilities: a pilot study", *Australasian Journal on Ageing*, Vol. 34 No. 1, pp. 27-32, doi: [10.1111/ajag.12099](https://doi.org/10.1111/ajag.12099).
- Robinson, H., MacDonald, B., Kerse, N. and Broadbent, E. (2013), "The psychosocial effects of a companion robot: a randomized controlled trial", *Journal of the American Medical Directors Association*, Vol. 14 No. 9, pp. 661-667, doi: [10.1016/j.jamda.2013.02.007](https://doi.org/10.1016/j.jamda.2013.02.007).
- Rogé, B. (2017), "Robots et autisme", *Enfance*, Vol. 2 No. 2, pp. 283-287, doi: [10.3917/enf.172.0283](https://doi.org/10.3917/enf.172.0283).
- Roger, K., Guse, L., Mordoch, E. and Osterreicher, A. (2012), "Social commitment robots and dementia", *Canadian Journal on Aging / La Revue Canadienne du Vieillissement*, Vol. 31 No. 1, pp. 87-94, doi: [10.1017/S0714980811000663](https://doi.org/10.1017/S0714980811000663).
- Rosanda, V. and Istenic, S.A. (2020), "The robot in the classroom: a review of a robot's role", in Popescu, E., Hao, T., Hsu, T.-C., Xie, H., Temperini, M. and Chen, W. (Eds), *Emerging Technologies for Education*, Springer International Publishing, pp. 347-357, doi: [10.1007/978-3-030-57717-9_40](https://doi.org/10.1007/978-3-030-57717-9_40).
- Rosenthal-von der Pütten, A.M., Schulte, F.P., Eimler, S.C., Hoffmann, L., Sobieraj, S., Maderwald, S. and Brand, M. (2013), "Neural correlates of empathy towards robots", *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, IEEE Press, pp. 215-216, doi: [10.1109/HRI.2013.6483571](https://doi.org/10.1109/HRI.2013.6483571).
- Roshdy, A., Karar, A.S., Al-Sabi, A., Al Barakeh, Z., El-Sayed, F., Beyrouthy, T. and Nait-Ali, A. (2019), "Towards human brain image mapping for emotion digitization in robotics", *Proceedings of the 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)*, IEEE, pp. 1-5, doi: [10.1109/BioSMART.2019.8734285](https://doi.org/10.1109/BioSMART.2019.8734285).
- Rowe, M.L., Silverman, R.D. and Mullan, B.E. (2013), "The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language", *Contemporary Educational Psychology*, Vol. 38 No. 2, pp. 109-117, doi: [10.1016/j.cedpsych.2012.12.001](https://doi.org/10.1016/j.cedpsych.2012.12.001).
- Šabanović, S., Bennett, C.C., Chang, W.-L. and Huber, L. (2013), "Paro robot affects diverse interaction modalities in group sensory therapy for older adults with dementia", *Proceedings of the 2013*

- IEEE International Conference on Rehabilitation Robotics (ICORR)*, IEEE, pp. 1-7, doi: [10.1109/ICORR.2013.6650427](https://doi.org/10.1109/ICORR.2013.6650427).
- Saerbeck, M., Schut, T., Bartneck, C. and Janse, M.D. (2010), "Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2010)*, ACM, pp. 1613-1622, doi: [10.1145/1753326.1753567](https://doi.org/10.1145/1753326.1753567).
- Santarossa, S., Kane, D., Senn, C.Y. and Woodruff, S.J. (2018), "Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice?", *Journal of Medical Internet Research*, Vol. 20 No. 4, pp. e144, doi: [10.2196/jmir.8480](https://doi.org/10.2196/jmir.8480).
- Sartorato, F., Przybylowski, L. and Sarko, D.K. (2017), "Improving therapeutic outcomes in autism spectrum disorders: Enhancing social communication and sensory processing through the use of interactive robots", *Journal of Psychiatric Research*, Vol. 90, pp. 1-11, doi: [10.1016/j.jpsychires.2017.02.004](https://doi.org/10.1016/j.jpsychires.2017.02.004).
- Scassellati, B., Admoni, H. and Mataric, M.J. (2012), "Robots for use in autism research", *Annual Review of Biomedical Engineering*, Vol. 14 No. 1, pp. 275-294, doi: [10.1146/annurev-bioeng-071811-150036](https://doi.org/10.1146/annurev-bioeng-071811-150036).
- Scassellati, B., Boccanfuso, L., Huang, C.-M., Mademtzi, M., Qin, M., Salomons, N., Ventola, P. and Shic, F. (2018), "Improving social skills in children with ASD using a long-term, in-home social robot", *Science Robotics*, Vol. 3 No. 21, p. eaat7544, doi: [10.1126/scirobotics.aat7544](https://doi.org/10.1126/scirobotics.aat7544).
- Scoglio, A.A., Reilly, E.D., Gorman, J.A. and Drebing, C.E. (2019), "Use of social robots in mental health and well-being research: Systematic review", *Journal of Medical Internet Research*, Vol. 21 No. 7, p. e13322, doi: [10.2196/13322](https://doi.org/10.2196/13322).
- Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S. and Paiva, A. (2017), "The case of classroom robots: Teachers' deliberations on the ethical tensions", *AI and Society*, Vol. 32 No. 4, pp. 613-631, doi: [10.1007/s00146-016-0667-2](https://doi.org/10.1007/s00146-016-0667-2).
- Sharkey, A. (2016), "Should we welcome robot teachers?", *Ethics and Information Technology*, Vol. 18 No. 4, pp. 283-297, doi: [10.1007/s10676-016-9387-z](https://doi.org/10.1007/s10676-016-9387-z).
- Shen, Z. and Wu, Y. (2016), "Investigation of practical use of humanoid robots in elderly care centres", *Proceedings of the 4th International Conference on Human-Agent Interaction (HAI 2016)*, pp. 63-66, doi: [10.1145/2974804.2974831](https://doi.org/10.1145/2974804.2974831).
- Shibata, T. and Coughlin, J.F. (2014), "Trends of robot therapy with neurological therapeutic seal robot, PARO", *Journal of Robotics and Mechatronics*, Vol. 26 No. 4, pp. 418-425, doi: [10.20965/jrm.2014.p0418](https://doi.org/10.20965/jrm.2014.p0418).
- Smakman, M., Vogt, P. and Konijn, E.A. (2021), "Moral considerations on social robots in education: a multi-stakeholder perspective", *Computers and Education*, Vol. 174, p. 104317, doi: [10.1016/j.compedu.2021.104317](https://doi.org/10.1016/j.compedu.2021.104317).
- Smarr, C., Prakash, A., Beer, J.M., Mitzner, T.L., Kemp, C.C. and Rogers, W.A. (2012), "Older adults' preferences for and acceptance of robot assistance for everyday living tasks", *Proceedings of the Human Factors and Ergonomics Society. Annual Meeting. Human Factors and Ergonomics Society. Annual Meeting*, Vol. 56 No. 1, pp. 153-157, doi: [10.1177/1071181312561001](https://doi.org/10.1177/1071181312561001).
- Spaulding, S. (2018), "Personalized robot tutors that learn from multimodal data", *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2018)*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 1781-1783.
- Sung, H.-C., Chang, S.-M., Chin, M.-Y. and Lee, W.-L. (2015), "Robot-assisted therapy for improving social interactions and activity participation among institutionalized older adults: a pilot study", *Asia-Pacific Psychiatry*, Vol. 7 No. 1, pp. 1-6, doi: [10.1111/appy.12127](https://doi.org/10.1111/appy.12127).
- Takayanagi, K., Kirita, T. and Shibata, T. (2014), "Comparison of verbal and emotional responses of elderly people with mild/moderate dementia and those with severe dementia in responses to seal robot, PARO", *Frontiers in Aging Neuroscience*, Vol. 6, p. 257, doi: [10.3389/fnagi.2014.00257](https://doi.org/10.3389/fnagi.2014.00257).

- Tanaka, F. and Matsuzoe, S. (2012), "Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning", *Journal of Human-Robot Interaction*, Vol. 1 No. 1, pp. 78-95, doi: [10.5898/JHRI.1.1.Tanaka](https://doi.org/10.5898/JHRI.1.1.Tanaka).
- Thibaut, J.W. and Kelley, H.H. (1959), *The Social Psychology of Groups*, Wiley.
- Tiberius, R. and Billson, J.M. (1991), "The social context of teaching and learning", *New Directions for Teaching and Learning*, Vol. 1991 No. 45, pp. 67-86, doi: [10.1002/dl.37219914509](https://doi.org/10.1002/dl.37219914509).
- Turkle, S., Breazeal, C., Duffy, B., et al. (2006), "A nascent robotics culture: new complications for companionship", *AAAI Technical Report Series*, doi: [10.1609/aimag.v30i3.2250](https://doi.org/10.1609/aimag.v30i3.2250).
- Van Der Drift, E.J., Beun, R.J., Looije, R., Blanson Henkemans, O.A. and Neerinx, M.A. (2014), A remote social robot to motivate and support diabetic children in keeping a diary, *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 463-470, doi: [10.1145/2559636.2559650](https://doi.org/10.1145/2559636.2559650).
- Wada, K., Kouzuki, Y. and Inoue, K. (2012), "Field test of caregiver's manual for robot therapy using therapeutic seal robot", *Alzheimer's and Dementia*, Vol. 8 No. 4, pp. S636-S637, doi: [10.1016/j.jalz.2012.05.1729](https://doi.org/10.1016/j.jalz.2012.05.1729).
- Wada, K., Ikeda, Y., Inoue, K. and Uehara, R. (2010), "Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot Paro", *Proceedings of the 19th International Symposium on Robot and Human Interactive Communication (RO-MAN 2010)*, pp. 533-538, doi: [10.1109/ROMAN.2010.5598607](https://doi.org/10.1109/ROMAN.2010.5598607).
- Wayne, A. and Youngs, P. (2003), "Teacher characteristics and student achievement gains: a review", *Review of Educational Research*, Vol. 73 No. 1, pp. 89-122, doi: [10.3102/00346543073001089](https://doi.org/10.3102/00346543073001089).
- Westlund, J.K., Gordon, G., Spaulding, S., Lee, J.J., Plummer, L., Martinez, M. and Breazeal, C. (2016), "Lessons from teachers on performing HRI studies with young children in schools", *Proceedings of the 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2016)*, pp. 383-390, doi: [10.1109/HRI.2016.7451777](https://doi.org/10.1109/HRI.2016.7451777).
- Woo, H., LeTendre, G.K., Pham-Shouse, T. and Xiong, Y. (2021), "The use of social robots in classrooms: a review of field-based studies", *Educational Research Review*, Vol. 33, p. 100388, doi: [10.1016/j.edurev.2021.100388](https://doi.org/10.1016/j.edurev.2021.100388).
- Xie, Y. and Peng, S. (2009), "How to repair customer trust after negative publicity: the roles of competence, integrity, benevolence, and forgiveness", *Psychology and Marketing*, Vol. 26 No. 7, pp. 572-589, doi: [10.1002/mar.20289](https://doi.org/10.1002/mar.20289).

Corresponding author

Auxane Boch can be contacted at: auxane.boch@tum.de