# Understanding the neural mechanisms of empathy toward robots to shape future applications

Jenna H. Chin[1], Kerstin S. Haring[2]* and Pilyoung Kim[1,3]

[1]Family and Child Neuroscience Lab, Department of Psychology, Brain, Artificial Intelligence, and Child (BAIC) Center, University of Denver, Denver, CO, United States, [2]Humane Robot Technology (HuRoT) Laboratory, Department of Computer Science, Ritchie School of Engineering and Computer Science, University of Denver, Denver, CO, United States, [3]Department of Psychology, Ewha Womans University, Seoul, Republic of Korea

This article provides an overview on how modern neuroscience evaluations link to robot empathy. It evaluates the brain correlates of empathy and caregiving, and how they may be related to the higher functions with an emphasis on women. We discuss that the understanding of the brain correlates can inform the development of social robots with enhanced empathy and caregiving abilities. We propose that the availability of these robots will benefit many aspects of the society including transition to parenthood and parenting, in which women are deeply involved in real life and scientific research. We conclude with some of the barriers for women in the field and how robotics and robot empathy research benefits from a broad representation of researchers.

KEYWORDS

robotics, fNIRS, neuroscience, social robots, empathy

## 1. Introduction

Neurorobotics is an emergent interdisciplinary field including neuroscience, cognitive science, psychology, robotics, and artificial intelligence. For the purpose of this article, we take the following bidirectional perspective on neurorobotics: Robot capabilities can be informed by neuroscience (e.g., brain-inspired algorithms), and the human brain can be evaluated through robots. We are considering neurorobotics from the perspective of what happens in the human brain when people interact with robots. Variations in robot design, including external appearance, movement, and behavior, in addition to robot reactions to external inputs, elicit distinctive and measurable activations in the human brain. Within the context of human–robot interactions, neuroscientific approaches have been applied to developing and understanding empathy toward robots.

This is a rapidly growing field as robots become increasingly capable of advanced social skills. Consequently, robots are more prevalent in domains such as caregiving, teaching, and companionship. There is emerging literature about why and how humans empathize with robots. In this article, we review the current literature on human empathy toward robots and discuss the neural mechanisms that underlie this cognitive process. Furthermore, we propose that increased representation of female scientists is critically needed for advancements in empathetic human–robot interactions. Recent literature suggests that women have particularly sensitive neural mechanisms for empathy and caregiving, which may lead to a greater tendency to empathize with robots. Next, we discuss important considerations for researchers in the development of robot's advanced empathetic abilities. We propose

that the availability of robots with advanced empathetic capabilities will assist parents in the transition to parenthood, benefit women in the society more broadly, and highlight the participation of female researchers in the field. Finally, we highlight the necessity for the use of inclusive language and the importance of representation of diverse groups that are historically underrepresented in the Neurorobotic research.

For the purpose of this study, we are focusing one particular group that is underrepresented in the Science, Technology, Engineering, and Mathematics (STEM) field, women, who make up only 28% of the STEM workforce, and even less in the robotics-related fields of Computer Occupations (25.2%) and Engineering (16.5%) (AAAU, 2020). These data refer broadly to all who self-identify as a woman. The terms *woman* and *female* are used interchangeably in this article. However, we acknowledge that 1) not all individuals who are assigned female at birth identify as women and 2) gender and sex include a broader spectrum than the terms used here.

## 2. Human empathy toward robots

Human empathy toward robots refers to the cognitive capacity of humans to perceive robots as if they have mental and emotional states, similar to those of another human. Understanding human empathy toward robots could be a crucial factor in recognizing or even predicting how people respond to robots and how robots can effectively respond to people. While this research area is a novel field, there are emerging studies that attempt to assess empathy toward robots. In the following, we review studies that sought to investigate this novel field in robotics. Because there is currently no established methodology and robots may vastly differ across research, the results of these studies may not necessarily be directly compared with each other. If methods are a concern, the exact methodology for each study should be reviewed in the cited literature.

Empathy is defined as an "other-oriented emotion elicited by and congruent with the perceived welfare of someone in need" (Batson, 2018). Empathy is a basis of emotions and behavioral responses including sympathy, empathic concerns, and compassion that are more other-oriented (Batson and Shaw, 1991; Nussbaum, 1996). In response to another's suffering, empathy is a vicarious experience of the person's feeling and empathic concerns are the concerns for the person which further motivate compassion. Compassion includes the behavioral responses to others to reduce the person's suffering (Goetz et al., 2010). In this paper, we will use empathy as the term that most broadly includes the other-oriented feelings that can lead to the motivation to help others.

Prior studies indicate that a robot showing empathetic behaviors through facial expressions and utterances was perceived differently (i.e., more friendly), supporting the idea that empathy plays a role in human–robot interactions. It seems that people empathize with robots. For example, it was stated that empathy is often seen as the basis of social cooperation and prosocial behavior and that robots capable of eliciting empathy in humans are a more successful technology (Leite et al., 2013).

How people empathize with robots depends on the human-likeness of the robot. Human-likeness in robotics describes the degree to which robots have the same features as humans [e.g., facial features like eyes, body manipulators like legs or arms, and surface looks like skin (Phillips et al., 2018)]. People empathize less with mechanical looking robots and more with human-like looking robots (Riek et al., 2009). A study examining punishment of a robot showed that not only do a robot's human-like features influence empathy but also its perceived agency (Kwak et al., 2013). Empathy toward a robot also seems to increase when the robot expresses emotions and mirrors expressions in comparison to the same robot not expressing emotions (Gonsior et al., 2012). More accurately expressed, empathy also seems to lead to robots (iCat) being perceived as more dependable and trustworthy (Cramer et al., 2010). Another study found that people had a more empathic response for a robot than a robot on a screen in a scenario where a robot first showed some intelligence and then expressed a fear (i.e., loosing its memory due to a functional problem; Seo et al., 2015).

Empathetic responses toward robot pain is a specific variety of empathy toward robots that involves recognizing a robot as if they were able to experience pain similar to how a human would. It seems that people readily perceive robots as being able to experience pain. Using fMRI, it has been shown that when observing another human in pain, the human brain processes empathy in a similar way than it processes pain (we discuss this in more detail in the next section).

## 2.1. How empathy for robots affect human behaviors

Humans have the ability to empathize with robot pain (Suzuki et al., 2015) and express concern and pity for robots that are tortured (Rosenthal-von der Pütten et al., 2013). The empathetic concern for robots leads to more hesitation to strike a robot (Darling et al., 2015). It also has been shown that inducing empathy triggers prosocial human behavior (i.e., increased helpfulness) toward a robot (Kühnlenz et al., 2013), that people seem to be inclined to help a robot find its way (Weiss et al., 2010), and that they indeed empathize with a robot when something bad happens to it (Seo et al., 2015).

However, because humans empathize with robot pain, some argue that the robots do not necessarily need to show social capabilities (Hoffman et al., 2015; Mattiassi et al., 2019). This has been reflected in the language that is used by soldiers feeling for their bomb disposal robots when they got damaged ("injured") or destroyed ("died"). Furthermore, the notion that showing robot pain elicits human empathy should not be interpreted as a blanket application of robots. For example, it has been shown that the empathetic response to pain stimuli is modulated by emotional context (Han et al., 2009). When a robot is highly human-like (i.e., a nearly human-identical android robot), it seems that the opposite effect takes place, such that the robot is perceived less empathetic than a more machine-like robot (Złotowski et al., 2016).

Incidents of bullying robots or robot abuse have also emerged. This vocabulary is used as if we indeed feel a level of empathy

toward robots. Specifically, we use the same terms we use for other agents that can experience pain (e.g., people that are being bullied and animals that are abused) and not the terms we would use for objects (e.g, *it* got destroyed and the car was damaged). This observation of human behaviors and how language is used can be tied to a phenomenological perspective on empathy (Zahavi, 2014), which is a basic class of empathy that indicates an understanding of the other, human or robot, as capable of experience (Quick, 2022). Even if robots do not truly experience the same sensations as humans or animals (e.g., pain), our language does describe what we perceive to be happening (Coeckelbergh, 2018). An early example is the "death" of hitchBOT, a hitchhiking robot that made its way across several countries for over a year, before it was subsequently destroyed by unknown assailants (Smith and Zeller, 2017). The demise of hitchBOT elicited a public reaction (e.g., social media like Twitter) and an analysis of the reactions showed a significant negative emotional reaction to its destruction, suggesting that people had formed an emotional connection with hitchBOT and perceived its destruction as morally wrong (Fraser et al., 2019).

Further research has shown that children sometimes abuse social robots in instances where they insulted the robot, obstructed its path, and even punched or kicked the robot (Nomura et al., 2015). In the same year this study was published, an intoxicated man was arrested in Japan for kicking a Pepper robot in a store (KYODO, 2015). A study on children aged 5–9 years showed that children seemed to abuse the robot, curious as to how the robot would react or whether it would enjoy it (Nomura et al., 2015). In addition, more than half of the children stated that the robot is able to perceive their abuse. It seems that children in this age group exhibit some empathy for the robot and ascribe it the ability to experience (their abuse), but choose to engage in their behaviors anyways. When comparing the bullying of humans and robots, it was found that both instances are connoted with being wrong and immoral; however, different cognitive mechanisms for moral disengagement are used (Sanoubari et al., 2021). The study also showed that, while perceived as wrong, robot mistreatment was associated with bullying less strongly and that people are less likely to intervene when a robot is being bullied.

This does raise research questions on how similarly people feel empathy for robots when compared to humans, what the differences are, and what potential significant factors could decrease or increase empathy for a robot. It has been suggested that empathy as a psychological phenomenon rooted in biology is both agile and fragile and can be broken by social change and enhanced and redirected by novel experiences (Heyes, 2018). It seems to be possible that even adults can re-learn and unlearn to empathize more or less intensively, and it could be the case that more ubiquitous robots, increased robot exposures, and new robot designs shape empathy toward robots.

Here, we propose that neuroscientific measures for empathy toward robots could lead to significantly new insights on the field of human–robot interaction as well as the field of neuroscience as robots are a novel entity in human's lives and with that have little preconceptions.

# 3. Neural mechanisms of human empathy toward robots

## 3.1. Empathy and caregiving motivation in the human brain

Activations in the brain regions that are involved in empathy support humans' empathetic thoughts and behaviors toward robots. In comparison to other species, the literature suggests that humans demonstrate an advanced capacity for empathy. This is partly due to the greater maturation of brain regions implicated in these higher-order cognitive and emotional processes. There have been several philosophical and psychological perspectives of empathy that are closely related to the neuroscientific approach. Goldman (2009) proposed the concept of mindreading as an extended form of empathy. In mindreading, there are two levels (Goldman, 2009). Low-level mindreading is based on a mirroring process whereas high-level mindreading is based on enactment imagination, perspective-taking, and mentalizing. Neuroscientific studies have identified brain regions that support a mirroring process, called the mirror neuron system. The brain regions of the mirror neuron system are the inferior frontal gyrus, the primary motor cortex, the superior temporal sulcus, and the inferior parietal lobule (Iacoboni and Dapretto, 2006). The connected activations in this system are critically involved in mirroring or imitating another person's action. The high-level mindreading such as perspective taking, enactment imagination, or mentalizing is further supported by the brain system called the theory of mind. The brain regions of the theory of mind include the medial prefrontal cortex, posterior cingulate cortex, temporal pole, superior temporal gyrus, and temporoparietal junction. This view of two levels of mindreading is also similar to Stueber (2012)'s basic and reenactive empathy theory. Stuber argues that basic empathy is a broader conception of mirror neurons while reenactive empathy is the process of how one understands the other's action by putting oneself in the other person's shoes, and trying to reenact the other person's thoughts in one's mind (Stueber, 2012).

These views are important in understanding particularly the cognitive aspect of empathy. In addition, emotional aspects of empathy have been discussed from different views including emotional contagion and sympathy. Emotional contagions refer to automatic responses to feel and mimic other person's facial expressions or actions. Hume's and Smith's views on sympathy argue that one can feel another person's feelings without being mentally engaged in imagination or reenactment (Sayre-McCord, 2013).

Another view of empathy proposes neurobiological mechanisms of these two sub-components of human empathy— emotional empathy and cognitive empathy. Emotional empathy refers to the ability to detect others' emotions and share feelings that others feel. Emotional empathy regions primarily include brain regions that are involved in emotional information processing such as the inferior frontal gyrus, anterior insula, and anterior cingular gyrus. Cognitive empathy refers to the ability to understand what others feel and think by taking the perspective of others (Decety and Jackson, 2004). Cognitive empathy regions include brain regions that are referred to earlier as being involved in theory

of mind (Shamay-Tsoory et al., 2009; Yu and Chou, 2018). The human brain is also particularly sensitive to others' experiences of pain. In one study examining this sensitivity, participants received a shock or their romantic partner received a shock. Here, both conditions similarly activated the anterior insula and the anterior cingulate cortex (Singer et al., 2004).

Humans seemingly engage some of these same neural circuitries when experiencing empathy in interactions with non-human agents. How people empathize with robots can be measured by their behaviors toward robots (Spatola and Wudarczyk, 2020), through explicit measures like surveys (Nomura et al., 2006; Carpinella et al., 2017), and more recently through neuroscientific measures (e.g., EEG, MRI, and fNIRS). The latter measures are believed to be less subject to biases than explicit measures (Ogunyale et al., 2018; Spatola et al., 2021) and allow for more temporal measures as they can be administered during the stimuli which is of interest to robotics research and could allow to identify the effect of robot behaviors better as opposed to measures administered after a stimuli. We are also currently conducting a neuroimaging studying using fNIRS (Pittman et al., 2022) in conjunction with robot facial expressions in embodied robots to gain new insight into the empathy and theory of mind humans form of robots and combine them with current, explicit measures.

Indeed, research utilizing multi-modal neuroimaging has begun to identify the neural signatures of empathy and broader sociocognitive processes during human interactions with robots. In an early neurocognitive examination of human–robot interactions utilizing neuroimaging methodology, researchers found early evidence of prefrontal and superior frontal cortex activations (i.e., regions implicated in theory of mind tasks) during an interpersonal behavior paradigm performed with both human and robot opponents (Hegel et al., 2008). There is also a particular interest in examining human empathy toward robot pain. In one study utilizing fMRI, researchers found similar neural activation patterns in the amygdala, insula, and inferior frontal gyrus in participants while viewing videos of both human and robot scenarios, further supporting the idea that humans employ empathy and theory of mind circuitry in response to robot pain (Rosenthal-von der Pütten et al., 2013). EEG has also been used to assess temporal aspects of neural responses to human and robot pain scenarios. In a study conducted by Suzuki et al., researchers found similar neural responses in the P3 component in both the robot and human conditions, suggesting that humans use similar top-down processing to empathize with robots (Suzuki et al., 2015). More recent studies of human–robot interactions aim to explore empathetic relationships between humans and robots beyond laboratory settings, though early findings remain mixed (Henschel et al., 2020).

While the evidence on the development of close bonds with robots is currently limited, robots with more advanced and human-like social skills may activate the brain circuits for forming intimate relationships in humans. This is due to humans having highly developed neural circuitry related to caregiving and affiliative behaviors. The literature suggests that human caregiving involves several brain networks, primarily those involved with reward and motivation and social and emotional information processing (Kim et al., 2016; Feldman, 2017). First, caregiving relies on the perception of a wide range infant stimuli including cries, emotional

faces, and gestures. Responding to these infant cues, humans form a model motivational network, recruiting dopamine and oxytocin pathways. These regions include the ventral tegmental area, substantia nigra, the striatum, and the medial prefrontal cortex. When responding to these infant cues, caregivers must also effectively regulate their own emotions, particularly when the cues may be distressing (e.g., infant cries). Key regions involved in this process include the anterior cingulate cortex and prefrontal cortex. The amygdala, which also contains an abundance of oxytocin receptors, interacts both the reward circuitry as well as emotion regulation circuitry to further modulate motivated responses and emotional salience to infant cues. Activation in these regions are closely related to the brain circuits of empathy, and are also important in supporting affiliative behaviors in close human–human relationships beyond parenting, including romantic relationships and friendships (Eslinger et al., 2021).

## 3.2. Sex differences in empathy and caregiving motivation

Individual differences may further influence how humans perceive, interact, or even develop close bonds with robots. For instance, research has found evidence for biological sex differences in social cognitive processes. Notably, differences in emotional sensitivity have been observed, such that women are more sensitive to fearful and sad stimuli whereas men are more sensitive to anger-provoking stimuli (He et al., 2018; Li et al., 2020). Previous work has also found differences in emotion identification between men and women. One meta-analysis found that while women may demonstrate a slight advantage in emotion recognition, this effect is moderated by factors such as the specific emotion and whether the emotion is positive or negative (Thompson and Voyer, 2014).

Prior research has shown that women, in comparison to men, demonstrate a greater neural response to social information (e.g., faces and people) opposed to inanimate stimuli (Proverbio et al., 2008). Relatedly, studies have indicated that women are more likely to see faces in objects (Zhou and Meng, 2020). Together, these enhancements in broad social domains set the stage for differences in later lower order processes. For instance, there may be functional differences in empathy as well as theory of mind associated with biological sex. Behavioral measures of empathy also suggest that women score higher on self-report questionnaire measures of empathy (Eisenberg and Lennon, 1983). There is also EEG evidence showing differential neural responses to pain between male and female participants such that women are more reactive to vicarious experiences of pain (Han et al., 2008). In addition, other studies have demonstrated differences in the sub-components of these processes. For instance, one study showed that women performed better than men on an affective Theory of Mind task (Baron-Cohen et al., 2003). Overall, prior research demonstrates that women may have differential brain activations and enhanced performance in cognitive domains that underlie human–robot interactions.

Even more, women may also have greater affiliative bonds with robots due to biological sex differences that underlie caregiving behaviors. Indeed, caregiving is a primary domain in which prior literature suggests differences between male and female

behaviors. Evolutionary perspectives of psychobiology suggest that these differences are crucial to reproductive processes such as in sexual mate selection and subsequent caregiving. Furthermore, biophysiological changes during pregnancy also increase positive caregiving behaviors in female animals and human birthing parents. One such evolutionarily advantageous adaptive behavior increased responses to infant cues such as faces and cries. Cognitive neuroscience literature has shown that infant faces are more perceptually salient for women than men (Hahn et al., 2013), providing evidence for the "Baby Schema Effect" in which women show enhanced affective valence and orienting toward infant or child faces in comparison to adult faces (Brosch et al., 2007). Taken together, these findings suggest that those assigned female gender at birth may be more likely to engage with robots in a more prosocial manner and further develop emotional connections with robots due to being more attuned to robot social cues and enhanced sensitivity to affiliative motivation.

In discussing these findings, it is also important to note that interactions between biological sex and social cognition are complex, and interpretations of sex-based differences in social cognition require careful consideration. Specifically, sex (e.g., female and male) is determined by biological, anatomical, and chromosomal sex characteristics whereas gender (e.g., woman, man, etc.) refers to the socially constructed roles that are associated with masculinity and femininity. Previous research, such as the studies discussed in this section, oftentimes combine definitions of biological sex and gender into a single binary categorization when describing participants. Researchers in future studies should extend upon these findings by utilizing a more precise and inclusive approach to understanding the interplay among, biological sex, social constructions of gender, and social cognitive functioning.

## 4. Empathetic abilities in robots

As humans have empathy toward robots, humans respond positively to robots with empathetic abilities. In recent years, robots are providing caregiving roles in many areas including elderly care, child education, and mental health support (Broadbent, 2017; Belpaeme et al., 2018). As a robot's role in caregiving grows, evidence suggests that the empathetic abilities of the robots increase their effectiveness in caregiving roles for humans. For example, their empathetic abilities lead to more positive relationships between humans and robots (Broadbent, 2017). It was also shown that children's learning was more effectively improved when interacting with robots that are more empathetic than not empathetic (Leite et al., 2014; Marchetti et al., 2018; Kory-Westlund and Breazeal, 2019).

As discussed earlier, empathy in humans is defined by two primary components—emotional empathy and cognitive empathy. Both components—emotional empathy and cognitive empathy—are critical for appropriate and natural empathetic responses in human social interactions. Therefore, a robot's empathetic abilities in interactions with humans may require both emotional empathy and cognitive empathy. Previous studies suggest that for a robot's empathetic capacity, a robot needs to (1) model a human's emotion and (2) adjust its social behaviors based on the human's emotion (Leite et al., 2014). The first aspect is closely aligned with emotional

empathy while the second aspect requires cognitive empathy. There have been studies that robots mirrored humans' emotions or provide empathetic responses to humans in specific contexts such as children playing games together with a robot (Leite et al., 2013). For more complex interactions with a wide range of contexts and individuals, robots not only need to mirror humans' emotions but also should be able to provide cognitive empathy by taking humans' internal perspectives (Asada, 2015). This is a challenging task for robots because, for effective and accurate perspective-taking, a robot needs to understand the background information of the human such as demographic factors, personality, previous experiences, and the context of the situation. Therefore, more work in this area is needed. Mirroring how human empathy is processed and generated in the brain, it will be critical for robots to be able to integrate the two components of emotional empathy and cognitive empathy.

## 4.1. An area that robots with empathetic abilities benefit women

The advanced robot's empathetic abilities are important in caregiving roles in a wide range of areas. We would like to propose one area where empathetic robots could significantly contribute; however, this area is currently receiving less attention than other applications: The robot's potential role is to support those who are in the transition to becoming new parents. This is also the area where women would particularly benefit more from the availability of the robot and the role of female scientists may also be important in developing robots with empathetic abilities. We used the gendered terms "mothers" and "fathers" in this section to match the findings of previous studies; however, we acknowledge that not all birthing parents identify themselves as "mothers" and "women."

Having a new baby is an exciting event for new mothers, but it is also an event that often causes high levels of stress to mothers. Due to the financial, physical, and emotional demands of caring for young infants, many mothers experience a significant decrease in their wellbeing (Kim, 2021). For example, nearly 20% of new mothers experience depression or anxiety during the perinatal period (Gavin et al., 2005; Sidebottom et al., 2021). A leading cause of postpartum death is drug-related death and suicide and their risks are associated with the postpartum psychopathology (Goldman-Mellor and Margerison, 2019). What is even more concerning is that postpartum depression is one of the most untreated disorders (Marcus and Heringhausen, 2009; Vigod et al., 2016). Postpartum psychopathology is concerning not only because of the mother's own wellbeing but also because of its negative impacts on the infant's brain, cognitive, and socioemotional development (Brand and Brennan, 2009; Kingston et al., 2012; Monk et al., 2012; Waters et al., 2014; Goodman, 2019).

Mothers are often isolated due to the constant demands of caring for new infants and also feel guilt in expressing negative thoughts and feelings about being a parent. These risk factors are associated with why stress that mothers have can lead to more severe psychopathology and their conditions are undertreated. Here, robots with advanced empathetic abilities can help to detect the symptoms of postpartum psychopathology and provide the

support that mothers need to cope with their stress. Robots can detect symptoms of depression and suicidal thoughts and connect those mothers to the appropriate healthcare providers. With the accurate detection of the symptoms, robots can also provide emotional support that the mothers need to reduce their stress during this stressful time.

Indeed, in the psychology and psychiatry literature, social support from a partner, family members, or friends is one of the best protective factors for preventing and reducing the symptoms of postpartum depression (Sherbourne and Stewart, 1991). Social support includes two different types—one is emotional support and the other is instrumental support. Robots that have advanced empathetic abilities can contribute to providing both types of support. For emotional support, robots can detect the mothers' moods and thoughts and provide empathetic responses to the mothers. The robots can also provide suggestions on adaptive coping strategies and share stories of other mothers who share similar experiences. Instrumental supports are a type of support that can help to care for infants. Robots with empathetic caregiving abilities can help new mothers to navigate challenges in caring for infants, for example, help understand the meaning of baby cry sounds and make suggestions on how to care for infants.

The role of support for the robot will benefit both mothers and fathers. However, it is still in the US, mothers spend significantly more hours caring for children compared to fathers (Aguiar and Hurst, 2007; Zamarro and Prados, 2021). Also, women are more vulnerable to postpartum psychopathology compared to men (Albert, 2015; Kuehner, 2017). Therefore, the availability of a robot is critically needed to support mothers, and women researchers may provide a deeper understanding of what is needed for developing robots with empathetic abilities.

# 5. Discussion and suggestions for future directions

In this perspective study, we reviewed the neural mechanisms of human empathy toward robots and the potentials for robots with empathetic abilities in assisting humans. We proposed that women who have more sensitive neural systems for empathy and caregiving compared to men are likely to exhibit the higher levels of empathy toward robots. We also proposed that robots with greater empathetic abilities may assist women, including their transition to parenthood. In studying both human empathy toward robots and robots' empathetic abilities, more inclusion of female researchers will be critical.

Here, we also propose additional considerations for future studies when discussing the importance of the role of women in neurorobotic research. First, biological sex and socially constructed gender categories are distinct. This understanding of the differences between biological sex and gender prompts additional consideration when interpreting prior findings on sex differences. Even more, biological sex is complex, and binary categorizations of assigned sex at birth do not capture all variations in hormonal, anatomical, and chromosomal differences (e.g., intersex individuals). Relatedly, inclusion efforts within the field of neurorobotics should not be limited to cisgender women. Inclusion and diversity within scientific research require an intersectional lens, taking into account how multiple minoritized identities may result in multiple forms of oppression and exclusion. Future research should aim to elevate the perspectives and work of women and other minoritized gender identities. Ultimately, this literature highlights the necessity for diverse perspectives in the creation of novel technologies that advance neurorobotics.

Neurorobotic research benefits from diverse representations of the researchers. It is shown very quickly that the field of robotics, as well as the field of neurorobotics suffers from under-representation. For example, looking only at women in the field, it shows that only 19% of computer science degrees in 2016 was awarded to women, in the History of Neuroscience in Autobiography and only 12% of the essays are authored by distinguished women neuroscientists, and between 1901 and 2019, the Nobel Prizes were awarded to a mere 21 women (counting Marie Curie two times) out of 615 scientists and many contributions of women have been simply left out or prize went to their male colleagues (Metitieri and Mele, 2022). When considering other underrepresented and minoritized groups in the field, including but not limited to Black, Indigenous, and People of Color (BIPOC), LGBTQ+ individuals, first-generation college graduates, immigrants, and people who identify in more than one or across these categories, the numbers look even more dire (MORAN, 2017). The way robots are created, including the robot technology and the interdisciplinary field surrounding robotics, is not appropriately balanced. As a result, robotics as a field is missing out on the opportunities that come with a broad participation in the research and development, and robot design process. It has been shown that full integration of minoritized scientists, under conditions of equitable and integrated work environments, leads to creativity, innovation, productivity, and positive reputational effects (Smith-Doerr et al., 2017). Gender gaps in STEM pursuits have been in part attributed to the STEM fields being deterring to communally oriented individuals (i.e., women and underrepresented minorities) as the fields impede goals of directly benefitting others, altruism, or collaboration (Diekman et al., 2015). A field like neurorobotics, however, can be highly communally oriented (e.g., empowering or helping others through neurorobotics) in addition to being highly interdisciplinary.

Overall, we argue that bidirectional empathy (i.e., human empathy toward robots and robot empathetic responses toward humans) is a primary factor in robot–human interactions. Prior examinations of the neural correlates of specific processes such as empathy, as well as social cognitive processes and behaviors more broadly, suggest that women and men may interact differently with robots due to differences in neural processes that underlie robot perception. To extend upon these findings, we propose caregiving to be a domain where a focus on bidirectional empathy between humans and robots to be particularly beneficial. Importantly, the design, implementation, and evaluation of such robots require increased diversity and inclusion within the field.

# Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

## Funding

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

AAAU (2020). *The STEM Gap: Women and Girls in Science, Technology, Engineering and Mathematics*. Available online at: https://www.aauw.org/resources/research/the-stem-gap/

Aguiar, M., and Hurst, E. (2007). Measuring trends in leisure: the allocation of time over five decades. *Q. J. Econ.* 122, 969–1006. doi: 10.1162/qjec.122.3.969

Albert, P. R. (2015). Why is depression more prevalent in women? *J. Psychiatry Neurosci.* 40, 219. doi: 10.1503/jpn.150205

Asada, M. (2015). Development of artificial empathy. *Neurosci. Res.* 90:41–50. doi: 10.1016/j.neures.2014.12.002

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., and Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 361–374. doi: 10.1098/rstb.2002.1206

Batson, C. D. (2018). *A Scientific Search for Altruism: Do we Only Care About Ourselves*? Oxford: Oxford University Press.

Batson, C. D., and Shaw, L. L. (1991). Evidence for altruism: toward a pluralism of prosocial motives. *Psychol. Inq.* 2, 107–122. doi: 10.1207/s15327965pli0202_1

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3, eaat5954. doi: 10.1126/scirobotics.aat5954

Brand, S. R., and Brennan, P. A. (2009). Impact of antenatal and postpartum maternal mental illness: how are the children? *Clin. Obstet. Gynecol.* 52, 441–455. doi: 10.1097/GRF.0b013e3181b52930

Broadbent, E. (2017). Interactions with robots: the truths we reveal about ourselves. *Annu. Rev. Psychol.* 68, 627–652. doi: 10.1146/annurev-psych-010416-043958

Brosch, T., Sander, D., and Scherer, K. R. (2007). That baby caught my eye... attention capture by infant faces. *Emotion* 7, 685–689. doi: 10.1037/1528-3542.7.3.685

Carpinella, C. M., Wyman, A. B., Perez, M. A., and Stroessner, S. J. (2017). "The robotic social attributes scale (ROSAS) development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna: IEEE), 254–262.

Coeckelbergh, M. (2018). Why care about robots? empathy, moral standing, and the language of suffering. *Kairos. J. Philosophy Sci.* 20, 141–158. doi: 10.2478/kjps-2018-0007

Cramer, H., Goddijn, J., Wielinga, B., and Evers, V. (2010). "Effects of (in) accurate empathy and situational valence on attitudes towards robots," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Osaka: IEEE), 141–142.

Darling, K., Nandy, P., and Breazeal, C. (2015). "Empathic concern and the effect of stories in human-robot interaction," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Kobe: IEEE), 770–775.

Decety, J., and Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 71–100. doi: 10.1177/1534582304267187

Diekman, A. B., Weisgram, E. S., and Belanger, A. L. (2015). New routes to recruiting and retaining women in stem: policy implications of a communal goal congruity perspective. *Soc. Issues Policy Rev.* 9, 52–88. doi: 10.1111/sipr.12010

Eisenberg, N., and Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychol. Bull.* 94, 100. doi: 10.1037/0033-2909.94.1.100

Eslinger, P. J., Anders, S., Ballarini, T., Boutros, S., Krach, S., Mayer, A. V., et al. (2021). The neuroscience of social feelings: mechanisms of adaptive social functioning. *Neurosci. Biobehav. Rev.* 128, 592–620. doi: 10.1016/j.neubiorev.2021.05.028

Feldman, R. (2017). The neurobiology of human attachments. *Trends Cogn. Sci.* 21, 80–99. doi: 10.1016/j.tics.2016.11.007

Fraser, K. C., Zeller, F., Smith, D. H., Mohammad, S., and Rudzicz, F. (2019). "How do we feel when a robot dies? emotions expressed on twitter before and after hitchbot's destruction," in *Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 62–71.

Gavin, N. I., Gaynes, B. N., Lohr, K. N., Meltzer-Brody, S., Gartlehner, G., and Swinson, T. (2005). Perinatal depression: a systematic review of prevalence and incidence. *Obstetrics Gynecol.* 106(5 Part 1), 1071–1083. doi: 10.1097/01.AOG.0000183597.31630.db

Goetz, J. L., Keltner, D., and Simon-Thomas, E. (2010). Compassion: an evolutionary analysis and empirical review. *Psychol. Bull.* 136, 351. doi: 10.1037/a0018807

Goldman, A. I. (2009). "Mirroring, mindreading, and simulation," in *Mirror Neuron Systems: The Role of Mirroring Processes in Social Cognition*, 311–330.

Goldman-Mellor, S., and Margerison, C. E. (2019). Maternal drug-related death and suicide are leading causes of postpartum death in california. *Am. J. Obstet. Gynecol.* 221, 489-e1. doi: 10.1016/j.ajog.2019.05.045

Gonsior, B., Sosnowski, S., Buß, M., Wollherr, D., and Kühnlenz, K. (2012). "An emotional adaption approach to increase helpfulness towards a robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vilamoura-Algarve: IEEE).

Goodman, J. H. (2019). Perinatal depression and infant mental health. *Arch. Psychiatr. Nurs.* 33, 217–224. doi: 10.1016/j.apnu.2019.01.010

Hahn, A. C., Xiao, D., Sprengelmeyer, R., and Perrett, D. I. (2013). Gender differences in the incentive salience of adult and infant faces. *Q. J. Exp. Psychol.* 66, 200–208. doi: 10.1080/17470218.2012.705860

Han, S., Fan, Y., and Mao, L. (2008). Gender difference in empathy for pain: an electrophysiological investigation. *Brain Res.* 1196:85–93. doi: 10.1016/j.brainres.2007.12.062

Han, S., Fan, Y., Xu, X., Qin, J., Wu, B., Wang, X., et al. (2009). Empathic neural responses to others' pain are modulated by emotional contexts. *Hum Brain Mapp.* 30, 3227–3237. doi: 10.1002/hbm.20742

He, Z., Liu, Z., Wang, J., and Zhang, D. (2018). Gender differences in processing fearful and angry body expressions. *Front. Behav. Neurosci.* 12, 164. doi: 10.3389/fnbeh.2018.00164

Hegel, F., Krach, S., Kircher, T., Wrede, B., and Sagerer, G. (2008). "Understanding social robots: a user study on anthropomorphism," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication* (Munich: IEEE), 574–579.

Henschel, A., Hortensius, R., and Cross, E. S. (2020). Social cognition in the age of human-robot interaction. *Trends Neurosci.* 43, 373–384. doi: 10.1016/j.tins.2020.03.013

Heyes, C. (2018). Empathy is not in our genes. *Neurosci. Biobehav. Reviews* 95, 499–507. doi: 10.1016/j.neubiorev.2018.11.001

Hoffman, G., Zuckerman, O., Hirschberger, G., Luria, M., and Shani Sherman, T. (2015). "Design and evaluation of a peripheral robotic conversation companion," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: IEEE), 3–10.

Iacoboni, M., and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* 7, 942–951. doi: 10.1038/nrn 2024

Kim, P. (2021). How stress can influence brain adaptations to motherhood. *Front. Neuroendocrinol.* 60, 100875. doi: 10.1016/j.yfrne.2020.100875

Kim, P., Strathearn, L., and Swain, J. E. (2016). The maternal brain and its plasticity in humans. *Horm. Behav.* 77:113–123. doi: 10.1016/j.yhbeh.2015.08.001

Kingston, D., Tough, S., and Whitfield, H. (2012). Prenatal and postpartum maternal psychological distress and infant development: a systematic review. *Child Psychiatry Hum. Dev.* 43, 683–714. doi: 10.1007/s10578-012-0291-4

Kory-Westlund, J. M., and Breazeal, C. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front. Robot. AI* 6:81. doi: 10.3389/frobt.2019.00081

Kuehner, C. (2017). Why is depression more common among women than among men? *Lancet Psychiatry* 4, 146–158. doi: 10.1016/S2215-0366(16)30263-2

Kühnlenz, B., Sosnowski, S., Buß, M., Wollherr, D., Kühnlenz, K., and Buss, M. (2013). Increasing helpfulness towards a robot by emotional adaption to the user. *Int. J. Soc. Robot.* 5, 457–476. doi: 10.1007/s12369-013-0182-2

Kwak, S. S., Kim, Y., Kim, E., Shin, C., and Cho, K. (2013). "What makes people empathize with an emotional robot?: the impact of agency and physical embodiment on human empathy for a robot," in *2013 IEEE RO-MAN* (Gyeongju: IEEE), 180–185.

KYODO (2015). *Drunken Kanagawa Man Arrested After Kicking SoftBank Robot.*

Leite, I., Castellano, G., Pereira, A., Martinho, C., and Paiva, A. (2014). Empathic robots for long-term interaction. *Int. J. Soc. Robot.* 6, 329–341. doi: 10.1007/s12369-014-0227-1

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., and Paiva, A. (2013). The influence of empathy in human-robot relations. *Int. J. Hum. Comput. Stud.* 71, 250–260. doi: 10.1016/j.ijhcs.2012.09.005

Li, G., Zhang, S., Le, T. M., Tang, X., and Li, C.-S. R. (2020). Neural responses to negative facial emotions: sex differences in the correlates of individual anger and fear traits. *Neuroimage* 221, 117171. doi: 10.1016/j.neuroimage.2020.117171

Marchetti, A., Manzi, F., Itakura, S., and Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie* 226, 98. doi: 10.1027/2151-2604/a000326

Marcus, S. M., and Heringhausen, J. E. (2009). Depression in childbearing women: when depression complicates pregnancy. *Primary Care* 36, 151–165. doi: 10.1016/j.pop.2008.10.011

Mattiassi, A. D. A., Sarrica, M., Cavallo, F., and Fortunati, L. (2019). "Degrees OF Empathy: Humans' empathy toward humans, animals, robots and objects," in *Ambient Assisted Living. ForItAAL 2017. Lecture Notes in Electrical Engineering, Vol 540,* eds N. Casiddu, C. Porfirione, A. Monteriù, and F. Cavallo (Cham: Springer). doi: 10.1007/978-3-030-04672-9_7

Metitieri, T., and Mele, S. (2022). "Women in neuroscience: A short time travel," in *Encyclopedia of Behavioral Neuroscience*, 2nd Edn. p. 71–76. doi: 10.1016/B978-0-12-819641-0.00007-4

Monk, C., Spicer, J., and Champagne, F. A. (2012). Linking prenatal maternal adversity to developmental outcomes in infants: the role of epigenetic pathways. *Dev. Psychopathol.* 24, 1361–1376. doi: 10.1017/S.0954579412000764

MORANB (2017). *LGBTQ+ Issues in Stem Diversity.* Available online at: https://www.bu.edu/articles/2017/lgbt-issues-stem-diversity/

Nomura, T., Suzuki, T., Kanda, T., and Kato, K. (2006). Measurement of negative attitudes toward robots. *Interact. Stud.* 7, 437–454. doi: 10.1075/is.7.3.14nom

Nomura, T., Uratani, T., Kanda, T., Matsumoto, K., Kidokoro, H., Suehiro, Y., et al. (2015). "Why do children abuse robots?," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts (HRI'15 Extended Abstracts)* (New York, NY: Association for Computing Machinery), 63–64. doi: 10.1145/2701973.2701977

Nussbaum, M. (1996). Compassion: The basic social emotion. *Soc. Philos. Po|Licy* 13, 27–58. doi: 10.1017/S0265052500001515

Ogunyale, T., Bryant, D., and Howard, A. (2018). Does removing stereotype priming remove bias? A pilot human-robot interaction study. *arXiv preprint* arXiv:1807.00948. doi: 10.48550/arXiv.1807.00948

Phillips, E., Zhao, X., Ullman, D., and Malle, B. F. (2018). "What is human-like? Decomposing robots' human-like appearance using the anthropomorphic robot (ABOT) database," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL: IEEE), 105–113.

Pittman, D. E., Haring, K. S., Kim, P., Dossett, B., Ehman, G., Gutierrez-Gutierrez, E., et al. (2022). "A novel online robot design research platform to determine robot mind perception," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo: IEEE), 986–990.

Proverbio, A. M., Zani, A., and Adorni, R. (2008). Neural markers of a greater female responsiveness to social stimuli. *BMC Neurosci.* 9, 56. doi: 10.1186/1471-2202-9-56

Quick, O. S. (2022). Empathizing and sympathizing with robots: implications for moral standing. *Front. Robot. AI* 8:412. doi: 10.3389/frobt.2021.791527

Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., and Robinson, P. (2009). "How anthropomorphism affects empathy toward robots," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (IEEE), 245–246.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., and Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *Int. J. Soc. Robot.* 5, 17–34. doi: 10.1007/s12369-012-0173-8

Sanoubari, E., Young, J., Houston, A., and Dautenhahn, K. (2021). "Can robots be bullied? a crowdsourced feasibility study for using social robots in anti-bullying interventions," in *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (Vancouver, BC: IEEE), 931–938.

Sayre-McCord, G. (2013). Hume and Smith on sympathy, approbation, and moral judgment. *Soc. Philos. Policy.* 30, 208–236.

Seo, S. H., Geiskkovitch, D., Nakane, M., King, C., and Young, J. E. (2015). "Poor thing! would you feel sorry for a simulated robot? a comparison of empathy toward a physical and a simulated robot," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Portland, OR: IEEE), 125–132.

Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132, 617–627. doi: 10.1093/brain/awn279

Sherbourne, C. D., and Stewart, A. L. (1991). The mos social support survey. *Soc. Sci. Med.* 32, 705–714. doi: 10.1016/0277-9536(91)90150-B

Sidebottom, A., Vacquier, M., LaRusso, E., Erickson, D., and Hardeman, R. (2021). Perinatal depression screening practices in a large health system: identifying current state and assessing opportunities to provide more equitable care. *Arch. Womens Mental Health* 24, 133–144. doi: 10.1007/s00737-020-01035-x

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Smith, D. H., and Zeller, F. (2017). The death and lives of hitchbot: The design and implementation of a hitchhiking robot. *Leonardo* 50, 77–78. doi: 10.1162/LEON_a_01354

Smith-Doerr, L., Alegria, S. N., and Sacco, T. (2017). How diversity matters in the us science and engineering workforce: a critical review considering integration in teams, fields, and organizational contexts. *Engag. Sci. Technol. Soc.* 3, 139–153. doi: 10.17351/ests2017.142

Spatola, N., Marchesi, S., and Wykowska, A. (2021). The instance task: How to measure the mentalistic bias in human-robot interaction. Preprint.

Spatola, N., and Wudarczyk, O. A. (2020). Implicit attitudes towards robots predict explicit attitudes, semantic distance between robots and humans, anthropomorphism, and prosocial behavior: from attitudes to human-robot interaction. *Int. J. Soc. Robot.* 2020, 1–11. doi: 10.1007/s12369-020-00701-5

Stueber, K. R. (2012). Varieties of empathy, neuroscience and the narrativist challenge to the contemporary theory of mind debate. *Emot. Rev.* 4, 55–63. doi: 10.1177/1754073911421380

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., and Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Sci. Rep.* 5, 1–9. doi: 10.1038/srep15924

Thompson, A. E., and Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: a meta-analysis. *Cogn. Emot.* 28, 1164–1195. doi: 10.1080/02699931.2013.875889

Vigod, S. N., Wilson, C. A., and Howard, L. M. (2016). Depression in pregnancy. *BMJ* 352, bmj.i1547. doi: 10.1136/bmj.i1547

Waters, C. S., Hay, D. F., Simmonds, J. R., and van Goozen, S. H. (2014). Antenatal depression and children's developmental outcomes: potential mechanisms and treatment options. *Eur. Child Adolescent Psychiatry* 23, 957–971. doi: 10.1007/s00787-014-0582-3

Weiss, A., Igelsböck, J., Tscheligi, M., Bauer, A., Kühnlenz, K., Wollherr, D., et al. (2010). "Robots asking for directions–the willingness of passers-by to support robots," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Osaka: IEEE), 23–30.

Yu, C.-L., and Chou, T.-L. (2018). A dual route model of empathy: a neurobiological prospective. *Front. Psychol.* 9, 2212. doi: 10.3389/fpsyg.2018.02212

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame.* Oxford: Oxford University Press.

Zamarro, G., and Prados, M. J. (2021). Gender differences in couples' division of childcare, work and mental health during COVID-19. *Rev. Econ. Househ.* 19, 11–40. doi: 10.1007/s11150-020-09534-7

Zhou, L.-F., and Meng, M. (2020). Do you see the "face"? individual differences in face pareidolia. *J. Pacific Rim Psychol.* 14, e2. doi: 10.1017/prp.2019.27

Złotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., and Ishiguro, H. (2016). Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy. *Paladyn J. Behav. Robot.* 7, 55–66. doi: 10.1515/pjbr-2016-0005