International Journal of Social Robotics
https://doi.org/10.1007/s12369-021-00787-5

Can I Feel Your Pain? The Biological and Socio-Cognitive Factors Shaping People's Empathy with Social Robots

Joanna K. Malinowska[1,2]

Accepted: 15 April 2021

Abstract

This paper discuss the phenomenon of empathy in social robotics and is divided into three main parts. Initially, I analyse whether it is correct to use this concept to study and describe people's reactions to robots. I present arguments in favour of the position that people actually do empathise with robots. I also consider what circumstances shape human empathy with these entities. I propose that two basic classes of such factors be distinguished: biological and socio-cognitive. In my opinion, one of the most important among them is a sense of group membership with robots, as it modulates the empathic responses to representatives of our- and other- groups. The sense of group membership with robots may be co-shaped by socio-cognitive factors such as one's experience, familiarity with the robot and its history, motivation, accepted ontology, stereotypes or language. Finally, I argue in favour of the formulation of a pragmatic and normative framework for manipulations in the level of empathy in human–robot interactions.

Keywords Empathy · Social robots · Human–robot interactions · Anthropomorphism

1 Introduction

Social robots are becoming increasingly common elements of our reality. Such robots (also referred to as *companion robots* or *artificial companions*) are defined[1] as "a physically embodied, autonomous agent[s] that communicates and interacts with humans on a social level" [37, p. 2]. They already accompany people in a variety of ways—as sexual partners, caregivers, therapists, personal trainers, priests or servants.[2]

Similarly, as in the case of interpersonal relations, many factors shape human interaction with robots. In this paper, I focus on one of them—empathy. The concept of empathy is one that frequently appears in research from psychology and neuroscience that deals with human–robot interactions [38, 39, 102, 131, 144]. The main aim of this paper is to analyse factors influencing the process of empathising with robots and the consequences that result from this phenomenon. I defend three main theses:

[1] There are many different definitions of social robots in the literature. More on this topic can be found in the text Understanding Social Robots [62].

[2] The diverse group of social robots includes, among others: the sex robots Samantha and Roxxxy [25] the policewoman, KP-Bot, the Buddhist monk, Kannon, the seal, Paro (used e.g. in nursing homes for the elderly [150]), the dinosaur, Pleo [48] or the dog, Aibo [161], robots helping in weight loss [71], robotic hostesses [95, 148] or even those moving on wheels, like the box-shaped therapeutic robot IROMEC [49,72]. Artificial companions are also robots that help scientists develop and study artificial intelligence, human emotional responses to cooperation with robots, etc. These include for example iCat [160] or KISMET – a robotic bust whose construction is focused on expressing emotions through appropriate "mimicry".

✉ Joanna K. Malinowska
malinowska@amu.edu.pl

1 Faculty of Philosophy, Adam Mickiewicz University, Poznań, Poland

2 Department of Philosophy, Jagiellonian University, Kraków, Poland

Published online: 20 May 2021

(1) Using the term "empathy" to study and describe the empathy of people towards robots, understood as psychological phenomena analogous to the one occurring between humans, is accurate;
(2) besides biological factors, the level of people's empathy with robots and the behavioural results of this process are dependent on a number of social and cognitive factors that regulate human intergroup relations. The key role is played here by a sense of group membership with the robot;
(3) the careful manipulation of factors affecting the level of empathising with a robot will play an important role in the process of shaping human–robot interactions. The interplay between affective and cognitive factors that determines the behavioural outcome of empathising is crucial in this situation. Moreover, due to the social importance of the outcome of these manipulations, it is necessary to regulate them with pragmatic and normative principles.

2 Empathy as an Important Element in Building and Maintaining Social Interactions between People and Robots

In this section, I will elaborate on whether the concept of empathy between people can be used in a similar way with robots or not. To accomplish this task, in Sect. 2.1. I will expand definition of empathy developed by de Vignemont and Singer [162, p. 435] and discuss it in the context of different positions on this problem within the research community. In Sect. 2.2. I will use the proposed conceptualization of empathy to answer the question of whether it can be applied to study human–robot interactions.

2.1 Empathy: Recent Conceptualizations of the Phenomena

Empathy is an important element of human social interactions, one which has been studied and analysed for years from various perspectives—psychological [55, 156], anthropological [64], ethological [10, 14, 105], neuroscientific [40, 137, 168], etc. It allows us to understand other people, adopt their perspectives and take actions needed to build and strengthen relationships [108, 140]. In the case of social robots, whose task is, inter alia, to build emotional ties with the user (e.g. in the case of accompanying robots caring for elderly or sick people) this seems to be a key element [84], although until now it has not been clearly established whether it is necessary in every situation. Therefore, it is not surprising that empathy is one of the most frequently studied phenomena in the area of HRI [87, 88].

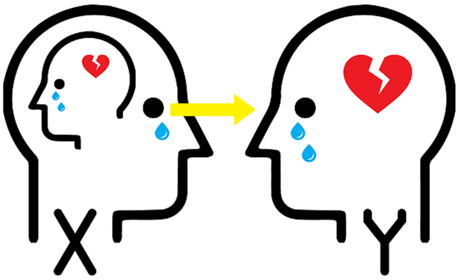

Fig. 1 The diagram visualises the concept of empathy (discussed in detail above). X is in an affective, somatic or cognitive state analogous to state of Y, X's perspective is aimed at understanding the state of Y and X knows that the state of Y is a source of its own state, as the state of X is triggered by observation, imagination or inference about the state of Y

There are hundreds of different definitions of empathy which sometimes coincide with phenomena such as sympathy, emotional contagion, empathic distress or altruism [30, p. 4, 142]. In this paper, I propose to consider empathy when the following conditions are met[3]:

(1) X is in a certain affective, somatic and cognitive state,
(2) the affective or somatic state of X is analogous to the state of Y (these states may vary in intensity of emotions/sensations, but should be similar in type),
(3) the state of X (affective/somatic and cognitive) is triggered by observation, imagination or inference about the affective/somatic state of Y,
(4) the cognitive perspective of X is aimed at understanding the state of Y,
(5) X knows that Y is the source of their affective/somatic and cognitive state.

The above definition tries to capture the results of recent discussions and research on empathising [30, 57, 66]. It is basically a variation of the definition of empathy[4] proposed by De Vignemont and Singer [162, p. 435]. But, as noted by Goldman [57, pp. 31–32], Vignemont and

[3] I presented other positions on this subject in article titled "What does it mean to empathise with a robot?" [87]. I analysed functional, substantial and relational conceptualisations of this phenomena and argued in the favour of the relational one developed by Damiano and Dumouchel [36–35]. However, in order to achieve the aims of this article, the adopted definition of empathising (thanks to the extensive literature on the subject) allows for a better analysis of what elements influence its occurrence. Basically, this approach does not conflict with the relational approach and can be reconciled with it.

[4] The Vignemont and Singer definition of empathy: "One is in an affective state; this state is isomorphic to another person's affective state:, this state is elicited by the observation or imagination of another person's affective state; one knows that the other person is the source of one's own affective state" [162, p. 435].

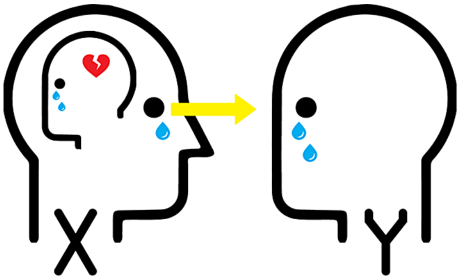

Fig. 2 Empathising with a robot—X (human) is in an affective, somatic or cognitive state. The state of X is analogous to behavioural expressions (simulations) of the state of Y (robot), X's perspective is aimed at understanding the state of Y and X knows that the state of Y is a source of its own state, as the state of X is triggered by observation, imagination or inference about the state of Y

Singer's definition, although widely recognized and used, is too restrictive in some respects, and too vague and broad in others. For example, it limits empathy to the process of sharing emotional states, while, as Goldman argues, it is also possible to empathise with someone's pain and touch [66], although both touch and pain cannot be described as emotions (Figs. 1, 2).

But just like the original definition of empathy coined by Vignemont and Singer, the interpretation of empathising that I propose in this paper distinguishes between the affectiv[5] and cognitive aspects of this process. Many studies have shown that the two types of empathy can sometimes occur and function independently in individuals. Psychopaths, for example, frequently recognize the emotional states of another person accurately but, due to a lack of affective empathy, use this knowledge to achieve their own goals [5, 67]. On the other hand, research conducted by Rogers and Dziobek indicated that people with Asperger's Syndrome, so far considered unable to empathise, show a similar level of affective empathy to other people but they are disturbed by the inability to recognize, understand and respond to the emotions of the other person (and therefore they lack the cognitive aspect of empathising, associated with having a theory of the mind) [120, p. 714]. There are also arguments that other information processing routes lead to different aspects of empathy—affective and somatic empathy prevails in one of them and cognitive empathy in the other [57].

In sum, I have argued that empathy occurs when a few conditions are met i.e.: the empathiser should be in the affective, cognitive or somatic state analogous to the person they observe, imagine or infer about. The empathiser should also be aware that this person's state is the source of its own state and be interested in understanding it (the state of the other).

At this point, however, we should return to the most important question—are the criteria presented in the definition above met when we talk about people's empathy with social robots?

2.2 Empathising with Robots

At first glance, the issue of people's empathy with robots raises a few doubts, i.e. with whom do they empathise if robots do not have any feelings? Some researchers even postulate treating empathising with robots as a form of illusion [109, 110] or imagining [92]. In opposition to those positions, in this subsection I present arguments in favour of the thesis that using the term "empathy" (understood as psychological phenomena analogous to the one occurring between humans) in the field of HRI is accurate.

Let me start with what seems to be certain—observing robots, especially social robots with whom an emotional bond has been established, creates the impression in many people that they are empathising with these agents. People's empathy with robots can be observed on three levels—the level of declared beliefs, the behavioural level, and at the level of neural activity [87, 88]. This is confirmed by numerous empirical studies conducted in the field of psychology and neuroscience [37, 38, 38, 39, 39, 128, 129]. For example, the analysis of research using fMRI, during which the participants of the experiment watched robots "experiencing" violence—e.g. being kicked—showed that they reacted similarly when they watched people being hurt. In both cases, brain activity associated with affective empathising was observed [53, 121, 122] and it also included the activation of mirror neurons.[6] Although their functions are still the subject of heated discussions, numerous studies indicate that the activity of mirror neurons is an important element of the empathising process [80, 104], 106]. It is primarily about three functions in which these neurons probably participate—anticipating the actions of others, empathising with pain or disgust, and empathising with touch [57], pp. 34–36]. However, based on the analysis of neuronal activation, can we infer anything about the occurrence of a given phenomenon? What's more, if we study human–robot interactions, can we draw conclusions analogous to interactions between

[5] In psychology, emotions and affections (moods) are usually clearly distinguished. However, due to the fact that emotional empathy is usually referred to as affective empathy in the academic literature, in this paper I use the concept of emotion and affect interchangeably.

[6] It was found that when people observed others performing movements, suffering or reacting to touch, their neuronal activation largely coincided with the brain activation when they themselves experienced such states. However, this is not one specific group of neurons but rather various systems involved in given behaviours and their mapping, understanding and sharing with other people.

people? Finally, given that robots cannot currently feel any pain or emotions, who are we empathising with?

At the outset, let us consider the neuroscience research in a slightly broader context. We have just said that many experiments point to the fact that people react by activating parts of the brain associated with empathising at the sight of damaged robots. In such situations, they also report feelings related to empathising. However, doesn't this happen every time we see scenes that we could interpret as cruel or hurtful? In studies conducted by Avenanti et al. [7, 8], it was established that the participants of the experiments who observed people being hurt (e.g. seeing needles being stuck in them) had mirror neurons activated (which is usually interpreted as affective empathising). At the same time, such activity was not observed when the same people watched the needles being stuck into tomatoes. Considering this example, together with the studies on empathising with robots mentioned in the paragraph above [53, 121, 122], it could be an important argument indicating that mirror systems respond to robots in a similar manner to which they respond to people, in contrast to "ordinary" objects [87, 88]. This is probably due to the fact that, by incorporating robots into social relations, we begin to automatically treat them as members of our own group, as individuals [155, p. 5].

Moreover, social robots usually take anthropomorphic forms and the way they form their communications (the verbal messages, their facial expressions and gestures, emulation of pain-like behaviour) are designed in such a way that it is easy for people to understand their activity [87, 88]. For instance, when based on certain computational processes they "conclude" that they should express some emotion, e.g. ask for help and curl up when they are beaten. Furthermore, robots constantly develop and learn from people. We are the closest social environment to them, their knowledge of social behaviour is largely based on information about inter-human and animal interactions. By the observation of robot behaviour, we can draw numerous more or less justified conclusions about their condition. Actually, we continually do the same in our interactions with others as well as with animals. In daily life, the behaviour of other beings is very often the only criterion based on which we assess someone's condition, especially when it comes to pain sensitivity [2]. Thus, it is accurate to say that people's affective response to robot behaviour is adequate to the state that it is communicating. In other words, although the robot itself does not feel pain, it can clearly express a certain affective state with all its behaviour. For example, there are robots which are programmed and designed to signal pain or discomfort to trainee dentists as they learn to drill teeth [1]. Other models, mainly due to artificial emotions, adapt their behaviour to the context in which they are currently located. There is, for example, a robotic rat equipped with 32 million silicon neurons and 13 trillion artificial synapses, which, feeling discomfort and pain, begins to move, e.g. when it falls into the water, it tries to get out of it as soon as possible [2], 5]. The argument that empathy in such a situation does not occur only because the robot "pretends to have feelings" is too simplistic. For instance, during scientific experiments regarding the recognition of emotions or empathy, it is easier in many cases to prepare and control the posed photos. Thus, it transpires that researchers use pictures of actors (usually trained in microexpressions pretending to experience given affective states [45, 90]. Therefore, if we were to accept that robots by now only mark experiencing affective states, we cannot talk about the occurrence of empathy towards them in humans, then we would probably also have to discard or revise all of our knowledge about empathising which is founded on posed photos and videos activity [87, 88].

Given the above arguments, in the process of empathising with robots the focus is on the side of affective and somatic empathy. However, we can also distinguish in it some elements of cognitive empathy.

When interacting with robots, it happens that our affective and cognitive state (or some bodily sensations, such as the impression of pain or touch) is triggered by the interference or observation of the state of the robot (condition (1) and (3)). This state is analogous to the one communicated by the robot (which we recognize based on its speech, behaviour, facial expressions) (condition (2)). Our cognitive perspective is to some extent aimed at understanding the situation of the robot and we know that it is the source of our affective and cognitive states (conditions (4) And (5)) (cf. Section 2.1, see also [87]).

Thus, I argue that when we consider human–robot interaction, although many doubts and controversies remain to be analysed and clarified, that it is correct to use concept of empathy in a similar way with robots as in human interaction. Let me now consider another problem—what factors enable and shape the process of empathising with robots?

3 Factors that Enable and Shape Human Empathy with Robots

In Sect. 2. I expanded the development of the definition of empathy previously proposed by de Vignemont and Singer [162]. Subsequently I used it to analyse whether it is correct to apply the concept of empathy to study people's reactions to robots. I presented various arguments in favour of the affirmative answer to this issue. In this section I will propose a list of factors that can manipulate the process of empathising with robots. I will start with analysing those of them that are biological in nature (in Sect. 3.1). Next (in Sect. 3.2), I turn to characterising the social and cognitive factors that affect our empathic response to robots.

In recent years, many biological foundations for empathising have been established. Thanks to neuroscientific research, empathy is associated with the activation of specific brain areas [66, 133], and genetic studies have identified specific genes that affect a person's predisposition to empathy [21, 151]. It is also known that disorders or the inability to empathise are associated with developmental problems, diseases, brain damage, etc. These biological grounds for empathy are, however, very strongly conditioned by the human cultural environment, experiences, expectations and motivations, as well as by the context in which empathising occurs. On the one hand, factors such as traumatic experiences, especially those experienced in the first years of life, leave their mark on the formation of brain structures and gene expression, which can lead to the inability to feel emotions and empathise as well as other personality disorders [11, 31]. On the other hand, personal values such as e.g. a high level of motivation to not display any racist and xenophobic behaviours raises the level of empathising towards people from another social group, aligning it with the level of empathy we feel towards representatives of our own group [157, 164].

Empathy is therefore a phenomenon co-shaped by several factors. I propose to divide them into two basic categories: biological/evolutionary and socio-cognitive factors. I argue that the biological conditions of human empathising with robots include: (1) individual genetic predispositions, (2a) a tendency to anthropomorphise embodied actors moving on their own, especially if their movement is similar to a biological movement [44, 79], (2b) a tendency to anthropomorphise and empathise with objects resembling human or animal bodies [44, 115, 116], and (3) a tendency to treat representatives of other social groups with less empathy and attention to their individual features [127, 164, 172]. The social and cognitive conditions include: (I) worldview (e.g. stereotypes and beliefs about robots, shared values and ontology etc.), (II) knowledge (e.g. facts about the individual robot and its actual situation).

What is more, all of the above factors, together with the broader situational context in which human–robot interaction takes place and the awareness of the potential consequences of the taken actions, co-modulate the behavioural outcome of empathy [75].

3.1 Biological Factors Shaping People's Empathy with Robots

I start by discussing the biological conditions affecting the process of empathising and focus on second and third of them. In this paper, for pragmatic reasons, I basically omit the analysis of factors that cannot be controlled in any way when designing social robots and social situations with their participation (and thus, for example, the individual genetic predispositions of users).

Empathising with robots is strongly associated with their anthropomorphisation, i.e. giving them human features and properties [170],[7] or interpreting the behaviour of non-human entities through human emotions and mental states [3]. Anthropomorphism can be seen as an important function of the mind, which consists in recognizing that we enter into a relationship with a being who has agency and goals other than we do (in this case, anthropomorphisation is accompanied by a transition from the relationship of using certain objects to interacting with them, which requires the coordination of the perspectives of two agents with separate goals) [34, p. 7]. This phenomenon in the area of HRI is most often interpreted as a common feature of *Homo Sapiens*, an adaptation that strengthened intergroup socialization processes and helped to avoid danger [51, 91, 132, 146. According to this approach, people have tended to anthropomorphise inanimate objects for centuries, seeing faces in the clouds, treetops or stones. This has a deep evolutionary justification—many scientists note that a rapid response to human or animal-like shapes is an adaptation that allowed our ancestors to quickly recognize threats (e.g. various predators) as well as helping them to distinguish enemies from members of their own social group, etc. [60, 91, 153. Anthropomorphisation is associated with the development of religion and magical thinking [21–19], and recent studies even indicate that it may affect the predisposition of some people to develop a pathological attachment to objects and their obsessive collection [146].

Many researchers argue that it is the tendency to anthropomorphise robots that enables us to establish emotional relationships and empathy with them [83, 84]. Nowadays, when everyday objects fail or break, people often interpret it as malice on their part, while when they serve without problems, they become almost close companions over time and even get names [4]. It also happens that children animate their toys, favourite blankets or pillows. Such anthropomorphised objects frequently begin to form the basis for empathy with it. People with a high tendency to anthropomorphise can become sensitive to the view of electrical sockets with "cutefaces" or become sad over the fate of cardboard boxes thrown in the trash (their handles often form the shape similar to a face).

Anthropomorphism intensifies for objects with specific properties. The first of these is the autonomous movement [44, 79]. With regard to robots, this phenomenon can be analysed, among others, thanks to studies on the Roomba vacuum

[7] Although, by definition, anthropomorphisation means giving objects human characteristics or properties, etc. in the literature on relations with robots, the term usually also includes cases of their "equipment", for example a Roomba vacuum cleaner is not treated as a human, but rather like a domestic pet.

cleaner [115, 116], robotic 'bug' HEXBUG [38, 39] or military robots deployed to defuse bombs [24]. If the robot's movement resembles the so-called biological movement, in many cases it causes even more intense reactions in people. Even small gestures performed by the robot can affect the extent to which we will anthropomorphise a given model [124].

Another factor significantly affecting the process of anthropomorphising objects, and then empathising with them, is their appearance, e.g. having a structure resembling a human / animal face or silhouette [44, 47, 50, 115, 116]. Although a human appearance is unnecessary to build an emotional relationship between the robot and the human (and perhaps even in some cases it makes this process difficult, contributing to the phenomenon called *uncanny valley*),[8] it most often provokes reactions related to anthropomorphisation and empathy. Robots with a humanoid shape automatically seem more human to us. Those that look like us are more easily treated as members of our own social group, as is the case with our interactions with people [47]. Depending on the morphological features they possess, robots trigger other feelings in the user—from sympathy, through dislike, to fear, which is why the detailed analysis of how their shape affects their reception by people is one of the key issues analysed in the area of HRI [9, 50, 170.

Even stronger emotional reactions are caused by robots that appear to be vulnerable. The essential biological movement and the shape of the robot also play an important role here. Darling [37, pp. 12–13] describes a "dramatic" experiment with the dinosaur Pleo, who tries to break free and cries out when held upside down. It has also a charming, "childlike" body structure and gives the impression of being fragile. Darling gathered several groups of people and gave each of them a dinosaur. The participants played with him for an hour, after which they were asked to attack him with an axe. No one took up this task, so Darling changed her tactics—she said that the robot of another group should be destroyed to protect their Pleo from a miserable fate. Also, in this variant, nobody decided to reach for the axe. Only when Darling said that if no one were to hit any of the dinosaurs then all of them would be broken, one person volunteered to do so. After a long moment of hesitation, the person "killed" one Pleo, and then all participants of the experiment fell silent for a moment. Coeckelbergh [28] notes that building robots that mimic human sensitivity and fragility can lead to greater ease in establishing social bonds and emotional relations with them, especially empathy. He believes that robots should in this way make it easier for users to empower empathy, and even more so—that the ability of social robots to be recipients of human empathy is essential to create social relationship between them [28, p. 4].

In contrast to the human tendency to anthropomorphize certain objects, which strengthens (and perhaps even enables) empathising with robots, some mechanisms related to learning may weaken and hinder this process. I am convinced that one of the relevant mechanisms in this context is perceptual narrowing [70, 101]. In brief, perceptual narrowing consists of the fact that, during their ontogenetic development, people become experts in recognizing objects and sounds that often appear in their surroundings. On the other hand, the information / objects or sounds we encounter less frequently or which we don't need are often overlooked and we are unable to identify them so well. For example, when a child is born, it has a very broad and universal ability to recognize human faces. But over time, when the baby begins to learn and develop, it becomes a specialist in recognizing the faces of the representatives of its own social group, whom it sees relatively more often and regularly than others [69]. Those elements of reality (people, objects or sounds) that were not frequent enough in their environment are less well recognized and usually treated as representations of a foreign group. This mechanism is to a large extent responsible for the occurrence of the *unfamiliarity homogeneity effect* [89]. In the context of intergroup relations, the unfamiliarity homogeneity effect consists of the fact that we encounter difficulties in differentiating and recognizing representatives of social groups other than our own.

The *unfamiliarity homogeneity effect* (UHE), depending on the criterion of "otherness" and the entity to which it relates to, can take different forms. In intergroup relations, it can take the form of the effect of homogeneity of other ethnic groups (the so-called *other race effect, cross race effect* etc. [54, 99, 139]) or social classes [134]. In the case of sounds—the effect of homogeneity of other languages [16, 78, 103] or types of music, etc. UHE always takes place when something unfamiliar seems to be too difficult to recognise.

The crucial factor here is that UHE is also associated with a lower level of empathising with people outside our own groups [127, 164, 172]. It can be supported by the fact that—if we stop treating other people as individuals, we dehumanise them and they became only representatives of a foreign group for us (resembling a symbol or an object), and our level of empathising decreases. Importantly, UHE, as well as other-group bias in empathy, are modulated by social and cognitive factors.

Above, I have reconstructed and discussed the list of biological factors influencing the process of human empathy. These include: (1) individual genetic predispositions, (2a) a tendency to anthropomorphise embodied actors moving on their own, especially if their movement is similar to a biological movement (2b) a tendency to anthropomorphise

[8] *Uncanny valley* is a term coined by the Japanese engineer Masahiro Mori, which refers to the scientific hypothesis according to which robots and other anthropomorphic performances that look very similar to man, but behave differently from him, cause in people unpleasant sensations on the verge of fear and disgust [94].

and empathise with objects resembling human or animal bodies and (3) a tendency to treat representatives of other social groups with less empathy and attention to their individual features. In the next subsection I turn to the issue of the social and cognitive factors that may affect people's empathising with robots.

3.2 Social and Cognitive Factors Shaping People's Empathy with Robots

UHE, together with other-group bias in empathy, is modulated, and can be considerably removed, inter alia, by the social context, education and motivation of the individual [42, 154, 164, 166, 167]. It decreases when we care about treating a representative of other groups as an individuals, when we know their names, history etc. Inversely, it intensifies when we treat the other person as a representative of a larger community, especially when we are hostile towards this community.

These effects are very susceptible to manipulation and, to a large extent, their intensity depends on who at a given moment we treat as "our" and who as the "other". They may lead to the dehumanization of "others" and further aggression towards them. They are therefore one of the factors creating and strengthening xenophobic and racist behaviour [134, 157, 164]. Sometimes they can be minimalised or removed by employing manipulation tactics of intergroup relation, for example by introducing others as in-group members [65, 127, 159, 164, 166].

Preliminary research indicates that similar techniques not only shape relations among humans but also the level of people's empathy with robots. If robots are introduced to a person as members of their social group (e.g. if they are introduced to the Germans with names which are popular in Germany), they are treated more positively and recognized as more human than robots with foreign-sounding names [47].

I take the position that it is UHE, other-group bias in empathy and other psychological effects correlated with them that are co-responsible for cases of lower empathising with robots, even leading to attacks on them [13]. Social robots form an increasingly large group of non-human actors in our social environment [158]. The more we anthropomorphize them, the more we transfer our prejudices, stereotypes and expectations acquired in our contacts with other people to them. Robots are similar to us, but, at the same time, they are also in some way unfamiliar and different, which may cause fear and resentment. While playing with Pleo, we start to treat it like a certain individual, it becomes a part of our group like a pet, while the robot in the shopping centre is just a representative of an other-group. And like all representatives of other-social groups, it is exposed to distrust, aggression, objectification and decreased level of empathy.

Another way to reduce UHE and other-group bias in empathy is to make "the others" more recognizable. This phenomenon was analysed in the context of human–robot interactions by Darling and her team [38, 39]. She conducted a study in which an insect-like robot called HEXBUG was presented to a group of people. The participants had to hit it with a hammer. In some versions of the experiment, its participants learned the robot's history in advance (e.g. they heard that its name was Frank, it was very friendly, its favourite colour was red and that it had lived in the laboratory for several months, recently it had the opportunity to play with other robots and has been excited ever since [38, 39, p. 772]) and that affected their reactions—they hesitated for a long time before they destroyed the HEXBUG.[9] This is in line with previously conducted research on intergroup relations, in which it has been proven that an appropriate narrative leading to individualisation reduces or eliminates the occurrence of the UHE and other-group bias in empathy.

Studies on human–robot empathy should also take into account such factors as cultural experience, values and beliefs. Distrust, prejudice or contempt for the representatives of other-group and certain stereotypes could reduce the level of empathising and increase the level of aggression toward others. This is the basis for racist (in the case of robots, the term "speciesism" might be more appropriate), xenophobic or sexist behaviours. On the other hand, an individual's strong equitable worldview and motivation not to display racist and xenophobic behaviour modulates UHE and bias in empathy towards other-group [159]. Beliefs likely affect the occurrence of these biases concerning other inter-group divisions too. One of them is probably a gender division.

As in the case of relations between people, gender stereotypes also influence the course of human–robot interactions. Studies show that, depending on which gender they ascribe to the robot (e.g. based on its appearance or voice), people react differently to it. They are more likely to donate money to robots with a male voice [135]. Given the impact of gender stereotypes,[10] the male robot probably seems more convincing and trustworthy. However, the role played in this case by gender bias in empathy is worth analysing. The precise impact of gender stereotypes on the level of empathising has not yet been thoroughly studied. Nevertheless, recent research on the phenomenon of victim-blaming by men indicates that this effect

[9] As Darling herself notes, a number of factors can influence the course of similar experiments, even the fact that some people will refuse to destroy expensive equipment [38, 39, p. 774].

[10] Most gender stereotypes are reproduced in relationships between humans and robots. E.g. people consider robots with more female characteristics to be more suitable for stereotypically female tasks [46]. These gender stereotypes can be strengthened e.g. by giving robots both male and female names that fit into traditional gender roles. "Male" robots are usually robots with "representative" functions and strong names, referring to mythology (like Hermes) while "female" robots usually perform service functions (they are cleaners, hostesses, sex robots) and they are given infantile names such as Candi [37, 38, 39, 102, 119, 131, 144].

might be associated with other-group bias in empathy (in this case—gender bias in empathy) [15]. Perhaps due to, inter alia, group (gender) belonging, men more often empathise with the male perpetrator of violence and blame the victim for it. Importantly, attempts to redirect attention to the perspective of the victim or perpetrator affect the level of empathising with both parties.

Finally, empathy and its determinants are co-shaped, among others, by the ontology we share. Research conducted by 107, p. 295] showed that ontological assumptions reflected in language can be one of the important factors conditioning the human approach to robots, including the tendency to anthropomorphise it. Rakison noticed that, in comparison to children raised in North American culture, Japanese children tend to consider objects as more animated than inanimate, including cases of objects that are difficult to clearly classify as not belonging to either of these two categories. This is partly an expression of the ontology assumed by the oldest religion in Japan, Shinto, which has adopted a universal animism, i.e. that all beings and objects (including human creations) have a certain life energy [56, 107, p. 300, 116, p. 337, 149, p. 78]. Among the inhabitants of Japan, the tendency to anthropomorphise robots and empathise with them is also influenced by the fact that this country developed after World War II mainly as a result of automation, not by the import of labour, and that due to other cultural conditions (e.g. the popularity of manga and anime, in which robots are often depicted as positive characters fighting evil people or other evil robots) the view of robots in the public space is relatively common there and evokes positive associations [56, pp. 76–78, 98, 130].

As can be clearly discerned, the process of empathising with robots (similarly to empathising with people) is extremely complicated and influenced by a number of factors—from evolutionarily shaped mechanisms leading to the anthropomorphisation of inanimate objects, through individual predispositions of a given person and situational context, to common opinions and stereotypes about robots and their development in a given community. Among the social and cognitive factors influencing people's empathy to robots, the most important are the following: worldview (e.g. stereotypes and beliefs about robots, shared values and ontology etc.), and knowledge (e.g. facts about the individual robot and its actual situation). Finally, (in line with the second thesis of this article) one of the most important of those factors is a sense of group membership with the robot, which actively co-shapes the output of psychological reactions to its presence. In the next section, I will answer the question of why it is so important to know how to manipulate people's empathy with robots.

4 Discussion: The Relevance of the Pragmatic and Normative Framework for Manipulating the Level of Empathy in Human–Robot Interactions

Knowledge of the factors modulating the occurrence of empathy gives many opportunities to exploit it. In this section, I will consider the possible ways to use empathy between humans and robots. I will start with the problems that may arise from the occurrence of this phenomenon (in Sect. 4.1) and then analyse how it can be used profitably, e.g. in social education (Sect. 4.2). These examples provide arguments for the last thesis of this article, i.e. that due to the social importance of the outcome of manipulations with people's empathy to robots, it is necessary to regulate this issue with some framework of pragmatic and normative principles.

4.1 The problems with people's empathy toward robots

Let me start with the problematic aspects of people's empathy with robots. Some of the experiments described above could serve as an argument that knowledge about the subject of empathy not only affects the level of empathising with others, but also modulates the behavioural outcome of this process [74, 75, 85, 173]. The level of empathy essentially affects people's decisions, especially those related to altruism. Empathy, or the lack of it, can make the difference in our decision on e.g. whether to help a person in need. It is not exactly known if this is the result of the manipulation of the empathising level, which in itself affects behaviour, or whether these changes are associated with certain differences in the way information is processed and a behaviour decision is made.

However, this situation has to some extent been clarified by recent research conducted by Kossowska et al. [75]. They argue that, although altruistic behaviour is mainly dependent on the individual's desire to help (resulting from empathy, sympathy etc.), it is also modulated by cognitive factors such as the expected effectiveness of the help. Kossowska et al. examined people's responses to the situation of donating money to a person in need. When their affective response was very high and they felt a high desire to help (this is usually associated with a high level of empathising with the person in need), then the additional information about the situation did not play a major role and the participants in the study decided to donate (even if they were informed that their help would not be effective). However, if the affective level was lower and the participants did not show a high degree of willingness to help, then the cognitive factors became significant and co-shaped their behaviour [75, p. 19]. Of course, the above research concerns a very specific situation, but it indicates the very subtle interplay between affective and cognitive factors in shaping empathy and its behavioural outcome. Moreover, it highlights the fact that the affective state in this case (especially affective empathy) may play a major role in the decision-making process and only its lowered level allows people to include rational arguments in it. This is consonant with most scientific texts on the subject, according to which emotions fundamentally influence the decision-making process and rational thinking [163, 169, 171]. But why does this seem important to human–robot interactions?

If empathy actually plays a key role in shaping people's motivation for altruistic behaviours and rational arguments for taking or rejecting these behaviours only become influential when it is lowered, then it should be studied with particular insight. By the careful manipulation of empathy, we can regulate the interplay between affective and cognitive responses shaping human altruistic behaviour. This problem is very acute in the case of the interaction between humans and robots. Empathy for robots can be beneficial, but it can also be harmful to humans, depending on the circumstances and purposes for which the robot is used. Incorrectly directed empathising with a robot can not only be inconvenient for the user but even dangerous for them. There have already been cases when soldiers empathising with the robots they used on the frontline refused to do exercises with them or saved robots, risking their own lives [24, 38, 39, 52, pp. 4–5]. This is an example of a situation when a robot is meant to perform a specific task, and all the features that lead to its anthropomorphisation and arouse empathy in the user can be an obstacle to its implementation. Furthermore, knowing that the level of empathising frequently depends on a sense of shared group membership, its behavioural outcome can also cause a lot of other dilemmas in human–robot interactions. For example, how will a person who has the choice to help (or risk exposure) a robotic "friend" or a stranger behave? Towards whom will such a person feel greater group membership and empathy? Thorough knowledge of mechanisms building, shaping and supporting the phenomenon of empathising with robots and its behavioural outcome will allow us to avoid, minimize or solve the number of such problems in the future. Thus, this issue requires in-depth empirical research as well as careful normative regulation.

What is more, the human tendency to empathise with robots can be a field of manipulation and abuse of their producers and users [37, 59, 126]. As I noted in the first part of the text, the cognitive aspect of empathising without an affective element is the hallmark of psychopaths, who are famous for their ability to manipulate. Robots, although they do not experience such feelings themselves, are increasingly specialised in recognizing human emotions and needs. As a result, they learn how to do so and adapt their behaviour to achieve a set goal—to enter into a social relationship with a human being with specific tasks like e.g. keeping company, education, therapy or weight loss. It should be emphasized that robots usually do not choose their goals themselves—they differ from people with a psychopathic personality in this respect. However, they can easily become manipulative tools in the hands of people. It is not difficult to imagine that in a capitalist system, social robots will be used to persuade humans to shop in particular online stores, for example, and thus increase company profits.

Finally, many researchers point out the impact that empathising with robots can have on human society, leading them to be treated as partners in a social situation, and then—entering into long-term social relationships (including intimate relationships) with people. This is an extremely complex issue, one which includes problems such as the reality and one-sidedness of these relationships (hypothetical (self) deception of people entering into relationships with robots), or limiting, objectifying and reducing interpersonal relations as a result of people entering into relationships with robots [26, 96, 111, 113, 117, 147]. There is also considerable concern about the impact that sex robots will have on the social status of children [141], women and minorities [22, 36, 43, 77, 113, 117, 152]. As I noted earlier, gender bias in empathy, among other factors, is responsible for the phenomenon of victim-blaming in the case of victims of violence against women, especially sexual violence. Considering the subject of this article, it is worth asking the following question: if, in fact, our moral sense and empathy develop in social interactions, how it will be affected by interactions with sex robots? Can allowing violence against sex robots result in lower levels of empathising with real people, especially woman (or, in case of sex robots looking like kids—children)? More generally—how does objectification and aggression towards robots affect our ability to empathise with representatives of other-groups or animals? The above questions require precise and careful answers.

4.2 The Educational and Therapeutic Potential of Empathising with Robots

Aside from the problems posed by people's empathy to robots, there are a host of potential benefits to this phenomenon. Strengthening interactions between people and robots by building a relationship based on empathy enables robots to perform some specific functions, especially accompanying robots. Darling points out that, for example, children can learn, among other things, to respect life (of people and animals) by taking care of robots and empathising with them

[38, 39], pp. 17–19]. It is also known that building an emotional relationship with a robot helps the elderly, improves their cognitive abilities, and reduces their level of loneliness [20, 93, 150]. By building relationships with the robot and its educational functions, we can shape human cognitive and social competences and implement the values that we consider important for our society. There is a need to carefully analyse the relations between humans and robots to see where the occurrence of empathy is a positive phenomenon, namely the one which is worth supporting (or even strengthening), and in which it should be weakened. The formulation of the pragmatic and normative framework for manipulations in the level of empathy in human–robot interactions is therefore urgently needed.

5 Conclusions

In this article, I argued in favour of the thesis that using the concept of empathy to study, analyse and explain interactions between people and social robots is a correct one. I cited a number of studies showing that empathy with robots is a real process, one discernible at the level of people's beliefs, behaviours and physiological (neuronal) responses.

I have analysed if this phenomenon meets the definitional criteria for empathysing. I have determined that people's affective, somatic and cognitive state can be triggered by the observation or inference about the state of the robot (condition 1 and 3). People's responses in such situations are analogous to the state communicated by the robot (which is recognized based robots verbal communicates, behaviour, facial expressions) (condition 2). Finally, people are aware that robots are the source of their affective and cognitive states as well as their cognitive perspective is to some extent aimed at understanding the situation of these robots (conditions 4 and 5).

The use of the concept of empathy to study human and robot interactions can be valuable both cognitively and scientifically, not only in understanding these interactions, but also in the context of subsequent attempts to capture, describe, and explain the empathising phenomenon itself. For this to happen, researchers of this subject must carefully, precisely and consistently analyse the conclusions that result from using the term *empathy* on grounds other than the relations between biological beings.

Moreover, I argued that people's empathy toward robots is co-shaped by a number of factors. I proposed to divide them into two basic categories: biological/evolutionary and socio-cognitive. The factors from both these categories may facilitate or hinder the robot's anthropomorphisation directly related to its potential to incite empathy in people. Among them, particularly crucial are the appearance and behaviour of the robot itself as well as one's worldview (e.g. accepted ontology, stereotypes) or language.

It is also important that universal psychological and cognitive mechanisms (of evolutionary origin) related to the dehumanization of others and the reduced level of empathising with them can be modulated by socio-cognitive factors. Our reactions to others (including robots) differ depending on whether we treat them as members of our own social group or not. They can be modulated by one's motivation, familiarity with the robot and its history, the robot's accent and others. However, more research on the factors shaping people's empathy with robots needs to be carried out. Not only can it affect the quality of our contacts with robots and their functionality, but also our understanding and shaping human intergroup relation. Just as knowledge from the field of social psychology can say a lot about our interactions with robots, our relationships with robots can teach us a lot about how and why we treat representatives of other-social groups, as well as animals.

Finally, given the importance of empathy between people and robots in human society, it is necessary to regulate the factors that shape it by means of pragmatic and normative principles. They should take into account, inter alia, the functions of the robot (e.g. empathising with robot may facilitate the functionality of care-bots, while it can impede it in the case of military robots), abuse possibilities (using empathising with robots as a means of manipulating their user) and the impact of empathising or lack of empathy with robots on society.

Robotics is developing so rapidly that we need these principles today. I hope that by showing its relevance this paper will start a broader discussion on the topic.

Acknowledgments This research was funded by the Ministry of Science and Higher Education in Poland, National Programme for the Development of Humanities, No. 0068/NPRH4/H2b/83/2016. Fragments of the text in Sect. 2.1 are based on the articles: "The growing need for reliable conceptual analysis in HRI studies: the example of the term 'empathy'" presented at *Robophilosophy* conference in 2020) and "What does it mean to empathise with a robot" [87]. I am grateful to my colleagues Maciej Musiał and Tomasz Żuradzki for all of their comments and guidance. I would also like to thank the Reviewers for their great help in improving this paper.

Funding This research was funded by the Ministry of Science and Higher Education in Poland, National Programme for the Development of Humanities, No. 0068/NPRH4/H2b/83/2016.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Abe S et al (2018) Educational effects using a robot patient simulation system for development of clinical attitude. Eur J Dent Educ 22(3):327–336
2. Adamo SA (2016) Do insects feel pain? A question at the intersection of animal behaviour, philosophy and robotics. Anim Behav 118:75–79
3. Airenti G (2015) The cognitive bases of anthropomorphism: from relatedness to empathy. Int J Soc Robot 7(1):117–127
4. Aggarwal P, McGill AL (2007) Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. J Consum Res 34(4):468–479
5. Ames H et al (2012) The animat new frontiers in whole brain modeling. IEEE Pulse 3:47–50
6. Ali F, Amorim IS, Chamorro-Premuzic T (2009) Empathy deficits and trait emotional intelligence in psychopathy and Machiavellianism. Personality Indiv Dif 47:758–762
7. Avenanti A et al (2005) Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. Nat Neurosci 8:955–960
8. Avenanti A et al (2006) Stimulus-driven modulation of motor-evoked potentials during observation of others' pain. Neuroimage 32:316–324
9. Bartneck C, Kulić D, Croft E, Zoghbi S (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J Soc Robot 1(1):71–81
10. Bekoff M (2010) The emotional lives of animals: a leading scientist explores animal joy, sorrow, and empathy—and why they matter. New World Library
11. Blair RJR, Peschardt KS, Budhani S, Mitchell DGV, Pine DS (2006) The development of psychopathy. J Child Psychol Psychiatry 47(3–4):262–276
12. Blasco PG, Moreto G (2012) Teaching empathy through movies: reaching Learners' affective domain in medical education. J Edu Learn 1(1):22
13. Bromwich JE (2019) Why do we hurt robots? They are like us, but unlike us, and both fearsome and easy to bully. https://www.nytimes.com/2019/01/19/style/why-do-people-hurt-robots.html. Accessed 19 Apr 1999.
14. Brothers L (1989) A biological perspective on empathy. Am J Psychiatry 146(1):10–19
15. Bongiorno R et al (2019) Why women are blamed for being sexually harassed: the effects of empathy for female victims and male perpetrators. Psychol Women Q 0361684319868730
16. Bosch L, Sebastián-Gallés N (1997) Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. Cognition 65(1):33–69
17. Boyer P (2008) Being human: religion: bound to believe? Nature 455(7216):1038
18. Boyer P (2008) Religion explained. Random House
19. Boyer P (1997) Further distinctions between magic, reality, religion, and fiction. Child Dev 68(6):1012–1014
20. Broekens J, Heerink M, Rosendal H (2009) Assistive social robots in elderly care: a review. Gerontechnology 8(2):94–103
21. Buck R, Ginsburg B (1997) Communicative genes and the evolution of empathy. Ann NY Acad Sci 807(1):481–483
22. Can WSR, Seibt SDJ (2016) Are sex robots as bad as killing robots? What Soc Robots Can Should Do: Proce Robophilos 2016/TRANSOR 2016 290:27
23. Cañamero L (2005) Emotion understanding from the perspective of autonomous robots research. Neural Netw 18(4):445–455
24. Carpenter J (2016) Culture and human–robot interaction in militarized spaces: a war story. Ashgate
25. Cheok ADI, Zhang EY (2019) Sex and a history of sex technologies. In: Human–robot intimate relationships. Human–computer interaction series. Springer, Cham
26. Chesney T, Lawson S (2007) The illusion of love: Does a virtual pet provide the same companionship as a real one? Interaction Studies 8(2):337–342
27. Coeckelbergh M (2014) The moral standing of machines: towards a relational and non-Cartesian moral hermeneutics. Philos Technol 27(1):61–77
28. Coeckelbergh M (2010a) Artificial companions: empathy and vulnerability mirroring in human–robot relations. Stud Eth, Law, Technol 4(3)
29. Coeckelbergh M (2010) Robot rights? Towards a social-relational justification of moral consideration. Eth Inf Technol 12(3):209–221
30. Coplan A (2011) Understanding empathy: its features and effects. Empathy: Philos Psychol Perspect 5–18
31. Craparo G, Schimmenti A, Caretti V (2013) Traumatic experiences in childhood and psychopathy: a study on a sample of violent offenders from Italy. Eur J Psychotraumatol 4(1)
32. Damasio AR (1999) The feeling of what happens: body and emotion in the making of consciousness. Houghton Mifflin Harcourt
33. Damiano L, Dumouchel P (2020) Emotions in relation. Epistemological and ethical scaffolding for mixed human–robot social ecologies. HUMANA MENTE J Philos Stud 13(37):181–206
34. Damiano L, Dumouchel P (2018) Anthropomorphism in human–robot co-evolution. Front Psychol 9:468
35. Damiano L, Dumouchel P, Lehmann H (2014) Towards human-robot affective co-evolution. Int J Soc Robot. https://doi.org/10.1007/s12369-014-0258-7
36. Danaher J (2019) Building better sex robots: lessons from feminist pornography. In: AI love you. Springer, Cham, pp 133–147
37. Darling K. (2016). Extending legal protection to social robots: the effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In: Robot law, Calo, Froomkin, Kerr red. Edward Elgar
38. Darling K (2015) 'Who's Johnny? 'Anthropomorphic framing in human-robot interaction, integration, and policy. In: Anthropomorphic framing in human–robot interaction, integration, and policy, ROBOT ETHICS, vol 2
39. Darling K, Nandy P, 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. https://doi.org/10.1109/ROMAN.2015.7333675
40. Decety JE, Ickes WE (2009) The social neuroscience of empathy. MIT Press
41. Decety J, Hodges SD (2006) The social neuroscience of empathy'. Bridging Social Psychology: Benefits Transdiscipl Approaches 103–109
42. Devine PG et al (2002) The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. Cognition 82(5):835
43. Devlin K (2015) In defence of sex machines: why trying to ban sex robots is wrong. The conversation

44. Duffy BR (2003) Anthropomorphism and the social robot. Robot Auton Syst 42(3–4):177–190. https://doi.org/10.1016/S0921-8890(02)00374-3
45. Ekman P (1992) Are there basic emotions? Psychol Rev 99(3):550–553. https://doi.org/10.1037/0033-295X.99.3.550
46. Eyssel F, Hegel F (2012) (s) he's got the look: gender stereotyping of robots 1. J Appl Soc Psychol 42(9):2213–2230
47. Eyssel F, Kuchenbrandt D (2012) Social categorization of social robots: anthropomorphism as a function of robot group membership. Br J Soc Psychol 51(4):724–731
48. Fernaeus Y, Håkansson M, Jacobsson M, Ljungblad S (2010) How do you play with a robotic toy animal?: a long-term study of pleo. In: Proceedings of the 9th international conference on interaction design and children. ACM, pp 39–48
49. Ferrari E, Robins B, Dautenhahn K (2009) Therapeutic and educational objectives in robot assisted play for children with autism. In: RO-MAN 2009-The 18th ieee international symposium on robot and human interactive communication. IEEE, pp 108–114
50. Fink J (2012) Anthropomorphism and human likeness in the design of robots and human–robot interaction. In: Ge SS, Khatib O, Cabibihan JJ, Simmons R, Williams MA (eds) Social robotics. ICSR 2012. Lecture notes in computer science, vol 7621. Springer, Berlin, Heidelberg
51. Gallup Jr GG, Marino L, Eddy TJ (1997) Anthropomorphism and the evolution of social intelligence: a comparative approach
52. Garreau J (2007) Bots on the ground: In the field of battle (or even above it), robots are a soldier's best friend. Wash Post 6. https://www.pressreader.com/usa/the-washington-post-sunday/20070506/282192236550791
53. Gazzola V, Rizzolatti G, Wicker B, Keysers C (2007) The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. Neuroimage 35(4):1674–1684
54. Ge L, Zhang H, Wang Z, Quinn PC, Pascalis O, Kelly D et al (2009) Two faces of the other race effect: Recognition and categorisation of Caucasian and Chinese faces. Cognition 38:1199–1210
55. Gladstein GA (1983) Understanding empathy: Integrating counseling, developmental, and social psychology perspectives. J Couns Psychol 30(4):467–482. https://doi.org/10.1037/0022-0167.30.4.467
56. Glaskin K (2012) Empathy and the robot: a neuroanthropological analysis. Ann Anthropol Pract 36(1):68–87
57. Goldman A (2011) Two routes to empathy. Empathy: Philos Psychol Perspect 31
58. Goldman AI (2006) Simulating minds: the philosophy, psychology, and neuroscience of mindreading. Oxford University Press, Oxford
59. Govindarajulu NS, Bringsjord S, Ghosh R, Sarathy V (2019) Toward the engineering of virtuous machines. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp 29–35
60. Guthrie SE, Guthrie S (1995) Faces in the clouds: a new theory of religion. Oxford University Press on Demand
61. Guzzi J, Giusti A, Gambardella LM, Di Caro GA (2018) Artificial emotions as dynamic modulators of individual and group behavior in multi-robot system. In: Proceedings of the 17th international conference on autonomous agents and multiagent systems. International Foundation for Autonomous Agents and Multiagent Systems, pp 2189–2191
62. Hegel F, Muhl C, Wrede B, Hielscher-Fastabend M, Sagerer G (2009) Understanding social robots. In: 2009 2nd international conferences on advances in computer-human interactions. IEEE, pp 169–174
63. Hodges SD, Wegner DM (1997) Automatic and controlled empathy, Empathic accuracy. The Guilford Press, New York, pp 311–339
64. Hollan DW, Throop CJ (eds) (2011) The anthropology of empathy: experiencing the lives of others in Pacific societies, vol 1. Berghahn Books, New York
65. Ito TA, Bartholow BD (2009) The neural correlates of race. Science 13(12):524–531
66. Jackson PL, Meltzoff AN, Decety J (2005) How do we perceive the pain of others? A window into the neural processes involved in empathy. Neuroimage 24(3):771–779
67. Jones AP, Happé FG, Gilbert F, Burnett S, Viding E (2010) Feeling, caring, knowing: different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. J Child Psychol Psychiatry 51(11):1188–1197
68. Keay A (2012) The Naming of Robots: biomorphism, gender and identity. Master thesis in digital cultures, University of Sydney
69. Kelly D, J. et al (2009) Development of the other-race effect during infancy: evidence toward universality? J Exp Child Psychol 104(1):105–114
70. Kelly DJ et al (2007) The other-race effect develops during infancy: Evidence of perceptual narrowing. Psychol Sci 18(12):1084–1089
71. Kidd CD, Breazeal C (2007) A robotic weight loss coach. In: Proceedings of the national conference on artificial intelligence, vol 22, no 2. AAAI Press, MIT Press, London, p 1985
72. Klein T, Gelderblom GJ, de Witte L, Vanstipelen S (2011) Evaluation of short term effects of the IROMEC robotic toy for children with developmental disabilities. In: 2011 IEEE international conference on rehabilitation robotics. IEEE, pp 1–5
73. Khusumadewi A, Juliantika YT (2018) The effectiveness of cinema therapy to improve student empathy. In: 2nd international conference on education innovation (ICEI 2018). Atlantis Press
74. Kogut T, Ritov I (2015) Target dependent ethics: discrepancies between ethical decisions toward specific and general targets. Current Opin Psychol 6:145–149
75. Kossowska M, Szumowska E, Szwed P et al. (2020) Helping when the desire is low: Expectancy as a booster. Motiv Emot 44: 819–831. https://doi.org/10.1007/s11031-020-09853-3
76. Ku H, Choi JJ et al (2018) Shelly, a tortoise-like robot for one-to-many interaction with children. In: Companion of the 2018 ACM/IEEE international conference on human–robot interaction. ACM, pp 353–354
77. Kubes T (2019) New materialist perspectives on sex robots. A feminist dystopia/utopia? Soc Sci 8(8):224
78. Kuhl PK (2000) Language, mind, and brain: experience alters perception. Cognition 2:99–115
79. Kupferberg A, Glasauer S, Huber M, Rickert M, Knoll A, Brandt T (2011) Biological movement increases acceptance of humanoid robots as human partners in motor interaction. AI Soc 26(4):339–345
80. Lamm C, Majdandžić J (2015) The role of shared neural activations, mirror neurons, and morality in empathy–a critical comment. Neurosci Res 90:15–24
81. Lamm C, Meltzoff AN, Decety J (2010) How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. J Cogn Neurosci 22(2):362–376
82. Lamm C, Batson CD, Decety J (2007) The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. J Cogn Neurosci 19(1):42–58
83. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2014) Empathic robots for long-term interaction. Int J Soc Robot 6(3):329–341
84. Leite I, Pereira A, Mascarenhas S, Martinho C, Prada R, Paiva A (2013) The influence of empathy in human–robot relations. Int J Hum Comput Stud 71(3):250–260

85. Loewenstein G, Lerner JS (2003) The role of affect in decision making. In: Davidson RJ, Scherer KR, Goldsmith HH (eds) Series in affective science. Handbook of affective sciences. Oxford University Press, pp 619–642
86. Lyon C, Nehaniv CL, Saunders J (2012) Interactive language learning by robots: the transition from babbling to word forms. PLoS ONE 7(6):e38236
87. Malinowska JK (2021) What does it mean to empathise with a robot? Mind Mach. https://doi.org/10.1007/s11023-021-09558-7
88. Malinowska JK (2020) The growing need for reliable conceptual analysis in HRI studies: the example of the term 'empathy'. In: Frontiers in artificial intelligence and applications, Volume 335: culturally sustainable social robotics, pp 96–104
89. Malinowska JK (2016) Cultural neuroscience and the category of race: the case of the other-race effect. Synthese 193(12):3865–3887
90. Mafessoni F, Lachmann M (2019) The complexity of understanding others as the evolutionary origin of empathy and emotional contagion. Sci Rep 9(1):5794
91. Mithen S, Boyer P (2019) Anthropomorphism and the evolution of cognition. J R Anthropol Inst 2(4):717+ Academic OneFile
92. Misselhorn C (2009) Empathy with inanimate objects and the uncanny valley. Mind Mach 19(3):345
93. Mordoch E, Osterreicher A, Guse L, Roger K, Thompson G (2013) Use of social commitment robots in the care of elderly people with dementia: A literature review. Maturitas 74(1):14–20
94. Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. IEEE Robotics & Automation Magazine 19(2):98–100
95. Murphy J, HofackerC Gretzel U (2017) Dawning of the age of robots in hospitality and tourism: Challenges for teaching and research. European J Tourism Res 15(2017):104–111
96. Musiał M (2019) Enchanting robots: intimacy, magic, and technology. Springer
97. Musiał M (2018) Loving dolls and robots: from freedom to objectification, from solipsism to autism? In: Exploring erotic encounters. Brill Rodopi, pp 152–168
98. Nakamura M (2007) Marking bodily differences: mechanized bodies in hirabayashi hatsunosuke's "robot" and early showa robot literature. Jpn Forum 19(2):169–190
99. Natu V, Radboy D, O'Toole AJ (2010) Neural correlates of own- and other-race face perception: spatial and temporal responce differences. Proc Natl Acad Sci 54:2547–2555
100. Nehaniv CL, Dautenhahn KE (2007) Imitation and social learning in robots, humans and animals: behavioural, social and communicative dimensions. Cambridge University Press
101. Nelson CA (2001) The development and neural bases of face recognition. Infant Child Dev: Int J Res Pract 10(1–2):3–18
102. Niculescu A, van Dijk B, Nijholt A, Li H, See SL (2013) Making social robots more attractive: the effects of voice pitch, humor and empathy. Int J Soc Robot 5(2):171–191
103. Palmer SB, Fais L, Golinkoff RM, Werker JF (2012) Perceptual narrowing of linguistic sign occurs in the 1st year of life. Proc Natl Acad Sci 83(2):543–553
104. Praszkier R (2016) Empathy, mirror neurons and SYNC. Mind Soc 15(1):1–25
105. Preston SD, Hofelich AJ, Stansfield RB (2013) The ethology of empathy: a taxonomy of real-world targets of need and their effect on observers. Front Hum Neurosci 7:488
106. Preston SD, de Waal FB (2017) Only the PAM explains the personalized nature of empathy. Nat Rev Neurosci 18(12):769
107. Rakison DH (2003) Parts, motion, and the development of the animate-inanimate distinction in infancy. In: Rakison DH, Oakes LM (eds) Early category and concept development: making sense of the blooming, buzzing confusion. Oxford University Press, Oxford, pp 159–192
108. Redmond MV (1989) The functions of empathy (decentering) in human relations. Hum Relat 42(7):593–605
109. Redstone J (2014) Making sense of empathy with social robots. In: Robophilosophy, pp 171–177
110. Redstone J (2017) Making sense of empathywith sociable robots: a new look at the "imaginative Perception of Emotion". Social Robots:Boundaries, Potential, Challenges (Ed. Nørskov M) 19 Routlage: 19-38.
111. Ribeiro T, Paiva A (2012) The illusion of robotic life: principles and practices of animation for robots. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, pp 383–390
112. Richardson K (2018) Challenging sociality: an anthropology of robots, autism, and attachment. Springer
113. Richardson K (2016) The asymmetrical "relationship": parallels between prostitution and the development of sex robots. ACM SIGCAS Comput Soc 45(3):290–293
114. Richardson K (2015) An anthropology of robots and AI: annihilation anxiety and machines. Routledge
115. Riek LD, Rabinowitch TC, Chakrabarti B, Robinson P (2009) Empathizing with robots: fellow feeling along the anthropomorphic spectrum. In: 2009 3rd international conference on affective computing and intelligent interaction and workshops. IEEE, pp 1–6
116. Riek LD, Rabinowitch TC, Chakrabarti B, Robinson P (2009) How anthropomorphism affects empathy toward robots. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction. ACM, pp 245–246
117. Richardson K (2016) Sex robot matters: slavery, the prostituted, and the rights of machines. IEEE Technol Soc Mag 35(2):46–53
118. Robertson J (2017) Robo sapiens Japanicus: robots, gender, family, and the Japanese nation. University of California Press
119. Robertson J (2010) Gendering humanoid robots: robo-sexism in Japan. Body Soc 16(2):1–36
120. Rogers K, Dziobek I, Hassenstab J, Wolf OT, Convit A (2007) Who cares? Revisiting empathy in Asperger syndrome. J Autism Dev Disord 37(4):709–715
121. Rosenthal-Von Der Pütten AM, Schulte FP, Eimler SC, Sobieraj S, Hoffmann L, Maderwald S et al (2014) Investigations on empathy towards humans and robots using fMRI. Comput Hum Behav 33:201–212
122. Rosenthal-Von Der Pütten AM, Krämer NC, Hoffmann L, Sobieraj S, Eimler SC (2013) An experimental study on emotional reactions towards a robot. Int J Soc Robot 5(1):17–34
123. Royzman EB, Cassidy KW, Baron J (2003) "I know, you know": Epistemic egocentrism in children and adults. Rev Gen Psychol 7(1):38–65
124. Salem M, Eyssel F, Rohlfing K, Kopp S, Joublin F (2013) To err is human (-like): effects of robot gesture on perceived anthropomorphism and likability. Int J Soc Robot 5(3):313–323
125. Salichs SA, Malfaz M (2011) A new approach to modeling emotions and their use on a decision-making system for artificial agents. IEEE Trans Affect Comput 3(1):56–68. https://doi.org/10.1109/T-AFFC.2011.32
126. Sarathy V, Arnold T, Scheutz M (2019) When exceptions are the norm: Exploring the role of consent in hri. ACM Transactions on Human-Robot Interaction (THRI) 8(3):1–21
127. Sheng F, Han S (2012) Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. Neuroimage 61(4):786–797
128. Scheutz M, Arnold T (2017) Intimacy, bonding, and sex robots: examining empirical results and exploring ethical ramifications (Unpublished manuscript)

129. Scheutz M, ArnoldT (2016) Are we ready for sex robots? 2016 11th ACM/IEEE International Conference on Human–Robot Interaction (HRI), pp. 351-358. https://doi.org/10.1109/HRI.2016.7451772
130. Schodt FI (2007) The Astro boy essays: Osamu Tezuka, mighty atom, and the manga/anime revolution. Stone Bridge Press, Berkeley
131. Seo SH, Geiskkovitch D, Nakane M, King C, Young JE (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). IEEE, pp 125–132
132. Serpell J (2003) Anthropomorphism and anthropomorphic selection—beyond the" cute response". Soc Anim 11(1):83–100
133. Shamay-Tsoory SG (2011) The neural bases for empathy. Neuroscientist 17(1):18–24
134. Shriver ER, Young SG, Hugenberg K, Bernstein MJ, Lanter JR (2008) Class, race, and the face: social context modulates the cross-race effect in face recognition. Proc Natl Acad Sci 34(2):260–274
135. Siegel M, Breazeal C, Norton MI (2009) Persuasive robotics: the influence of robot gender on human behavior. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, pp 2563–2568
136. Singer T (2017) Plasticity of empathy and prosocial motivation: from outgroup hate to ingroup favouritism. https://www.youtube.com/watch?v=TOa-sPMDNGg. Accessed 10 Aug 2019
137. Singer T, Lamm C (2009) The social neuroscience of empathy. Ann N Y Acad Sci 1156(1):81–96
138. Sparrow R (2017) Robots, rape, and representation. Int J Soc Robot 9(4):465–477
139. Sporer SL (2001) Recognizing faces of other ethnic groups: an integration of theories. Psychol Publ Policy Law 7(1):36
140. Stephan WG, Finlay K (1999) The role of empathy in improving intergroup relations. J Soc Issues 55(4):729–743
141. Strikwerda L (2017) Legal and moral implications of child sex robots. Robot Sex: Soc Eth Implic 133-152
142. Stueber K (2019) Empathy, The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/fall2019/entries/empathy/. Accessed 15 Aug 2019
143. Stueber K (2006) Rediscovering empathy: agency, folk psychology, and the human sciences. MIT Press, Cambridge
144. Tapus A, Mataric MJ (2007) Emulating empathy in socially assistive robotics. In: AAAI spring symposium: multidisciplinary collaboration for socially assistive robotics
145. Tavani H (2018) Can social robots qualify for moral consideration? Reframing the question about robot rights. Information 9(4):73
146. Timpano KR, Shaw AM (2013) Conferring humanness: the role of anthropomorphism in hoarding. Personal Individ Differ 54(3):383–388
147. Turkle S (2007) Authenticity in the age of digital companions. Interaction studies 8(3):501–517
148. Tussyadiah IP, Park S (2018) Consumer evaluation of hotel service robots. In: Information and communication technologies in tourism 2018. Springer, Cham, pp 308–320
149. Ulbrick A (2008) Rodney's robot revolution. Documentary, 53 min. Essential Media and Entertainment, Sydney
150. Wada K, Shibata T (2007) Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. IEEE Trans Rob 23(5):972–980
151. Walter H (2012) Social cognitive neuroscience of empathy: concepts, circuits, and genes. Emot Rev 4(1):9–17
152. Weber J (2005) Helpless machines and true loving care givers: a feminist critique of recent trends in human-robot interaction. J Inf Commun Ethics Soc 3(4):209–218
153. Westh P (2009) Anthropomorphism in god concepts: the role of narrative. Orig Relig, Cogn Cult 396–413
154. Wheeler ME, Fiske ST (2005) Controlling racial prejudice social-cognitive goals affect amygdala and stereotype activation. Neuropsychologia 16(1):56–63
155. Wiese E, Metta G, Wykowska A (2017) Robots as intentional agents: using neuroscientific methods to make robots appear more social. Front Psychol 8:1663
156. Wisp L (1987) History of the concept of empathy. Empathy Dev 17–37
157. Van Bavel JJ, Cunningham WA (2012) A social identity approach to person memory: group membership, collective identification, and social role shape attention and memory. Pers Soc Psychol Bull 38(12):1566–1578
158. Vanman EJ, Kappas A (2019) "Danger, Will Robinson!" The challenges of social robots for intergroup relations. Soc Pers Psychol Compass 13(8):e12489
159. Van Bavel JJ, Packer DJ, Cunningham WA (2008) The neural substrates of in-group bias a functional magnetic resonanceimaging investigation. Science 19(11):1131–1139
160. van Breemen A, Yan X, Meerbeek B (2005) iCat: an animated user-interface robot with personality. In Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems pp 143–144
161. Veloso MM, Rybski PE, Lenser S, Chernova S, Vail D (2006) CMRoboBits: Creating an intelligent AIBO robot. AI magazine 27(1):67–67
162. De Vignemont F, Singer T (2006) The empathetic brain: how, when, and why? Trends Cogn Sci 10:435–441
163. Vohs KD, Baumeister RF, Loewenstein G (eds) (2007) Do emotions help or hurt decision making?: A hedgefoxian perspective. Russell Sage Foundation
164. Xu X, Zuo X, Wang X, Han S (2009) Do you feel my pain? Racial group membership modulates empathic neural responses. J Neurosci 29(26):8525–8529
165. Young JE, Hawkins R, Sharlin E, Igarashi T (2009) Toward acceptable domestic robots: applying insights from social psychology. Int J Soc Robot 1(1):95
166. Young SG, Hugenberg K, Bernstein MJ, Sacco DF (2012) Perception and motivation in face recognition: a critical review of theories of the cross-race effect. Neuropsychologia 16(2):116–142
167. Young ST, Zhou G, Pu X, Tse C (2015) Effects of divided attention and social categorization on the own-race bias in face recognition. Vis Cogn 22(9–10):1296–1310
168. Zaki J, Ochsner KN (2012) The neuroscience of empathy: progress, pitfalls and promise. Nat Neurosci 15(5):675
169. Zeelenberg M, Nelissen RM, Breugelmans SM, Pieters R (2008) On emotion specificity in decision making: why feeling is for doing. Judgm Decis Mak 3(1):18
170. Złotowski J, Proudfoot D, Yogeeswaran K, Bartneck C (2015) Anthropomorphism: opportunities and challenges in human–robot interaction. Int J Soc Robot 7(3):347–360
171. Zhu J, Thagard P (2002) Emotion and action. Philos Psychol 15(1):19–36
172. Zuo X, Han S (2013) Cultural experiences reduce racial bias in neural responses to others' suffering. Cult Brain 1:34–46
173. Żuradzki T (2018) The normative significance of identifiability. Eth Inf Technol 1–11

Joanna K. Malinowska works as an assistant professor at Adam Mickiewicz University in Poznań (Poland) at the Faculty of Philosophy (Epistemology and Cognitive Science Research Unit). She is interested in the philosophy of neuroscience, philosophy of medicine, philosophy of race, evolutionary epistemology, bioethics and social robotics. She published, e.g. in Synthese, The American Journal of Bioethics, AJOB Neuroscience or Minds and Machines. Currently, she is working on the issue of using the category of race in biomedical research and healthcare.