



Imitating the Human. New Human–Machine Interactions in Social Robots

Johanna Seifert · Orsolya Friedrich · Sebastian Schleidgen

Received: 3 August 2021 / Accepted: 5 June 2022 / Published online: 17 August 2022
© The Author(s) 2022

Abstract Social robots are designed to perform intelligent, emotional, and autonomous behavior in order to establish intimate relationships with humans, for instance, in the context of elderly care. However, the imitation of qualities usually assumed to be necessary for human reciprocal interaction may impact our understanding of social interaction. Against this background, we compare the technical operations based on which social robots imitate human-like behavior with the concepts of emotionality, intelligence, and autonomy as usually attached to humans. In doing so, we aim at making a conceptual contribution to the intuition that the technical imitation of emotional, intelligent, and autonomous behavior differs qualitatively from human forms of emotionality, intelligence, and autonomy. We conclude that the increasing tendency of social robots to imitate human characteristics and the thus increased human tendency of anthropomorphizing technical systems put forth new forms of human–machine interactions (HMI) that call for a critical examination. Thus, the emerging forms of interaction may lead, for instance, to false, reinforced, and perplexing assumptions regarding reciprocity. Therefore, studies are needed that critically examine the potential consequences and psychological effects

that may be associated with the use of social robots, especially with regard to vulnerable groups such as children or the elderly.

Keywords Social robots · Human–machine interaction · Anthropomorphism · Healthcare

Introduction

As a consequence of recent developments in Artificial Intelligence (AI), Machine Learning (ML), and advances in sensor and micro technologies, robots have become part of numerous areas of life and work. In addition to their primary use in industry and the military, they today serve as assistance systems to help with everyday tasks,¹ as companions in geriatric care,² or simply as technical substitute for pets.³ While robots formerly were mainly used in industry where they performed risky and repetitive operations and thus functioned as useful tools, they are now

¹ See, e.g., robotic vacuum cleaners like "Roomba" (iRobot), floor-washing robots or ironing robots.

² In geriatric care, robots mainly perform physical tasks such as washing, transferring and carrying patients (see the model "Robear," for instance). Additionally, they are also used for therapeutic purposes, for example to increase a patients' ability to interact or to alleviate loneliness, anxiety and depression (see, among others, "Paro" and "Pepper").

³ See, e.g., toy robots like "Lovot" (Groove X), or the robotic dog "AIBO" (Sony).

J. Seifert (✉) · O. Friedrich · S. Schleidgen
Junior Professorship of Medical Ethics, Institute
of Philosophy, FernUniversität in Hagen,
Universitätsstraße 33, 58097 Hagen, Germany
e-mail: johanna.seifert@fernuni-hagen.de

– according to the rhetoric of technology developers and media reporting – more and more becoming intimate interaction partners. So-called social robots are central to this context, as they seem to possess characteristics that were previously only attributed to humans: due to their appearance and functionality, they appear intelligent, emotional, and autonomous.⁴ This gives rise to narratives that assert a similarity or even equality between humans and machines (cf. here, for instance, transhumanism [4, 5]) and has considerable consequences for the interaction between humans and machines: People no longer use social robots exclusively as artifacts and instrumental tools, but rather encounter them as a social counterpart.⁵ Crucial to this is their anthropomorphic design and their tendency to imitate human modes of interaction and to simulate socio-emotional behavior. The phenomenon of imitation, which is discussed in this article, thus refers to the complex of anthropomorphic technology design, as it is crucial especially in the field of social robotics (cf. here, among others, [8, 9]).

Against this background, the article examines how social robots perform intelligent, emotional, and autonomous behavior and how these features differ from human forms of emotionality, intelligence, and autonomy, and thus from the conditions we usually assume necessary for human reciprocal interaction. The paper thus takes its starting point from the assumption of a difference between human behavior and technical or robotic modes of operation – an assumption that seems intuitively conclusive, but will be also justified on a conceptual level in the present article. The question of how human and robotic modes of behavior and operation differ is thereby motivated by the following reasons: a) Even if the difference between human and robotic forms of interaction seems intuitively conclusive, it needs to be conceptualized in order to understand what we are dealing with in the case of social robots, on the one hand, and to better understand our relationship to robots, on the other hand. b) The question posed here helps refute hasty analogizations and narratives that

attribute capabilities such as reasoning, emotionality and sociality to robotic systems – as can be observed, for example, in media coverage, in the advertising of technological products, but also in technophilic discourses such as transhumanism [4, 5]. Here, robots are ascribed abilities that they do not (yet) possess; on the other hand, the narrative of an equality of humans and technology is constructed, which must be critically countered on a conceptual level. c) Forms of interaction can be identified in concrete human–robot interaction (HRI) in which the difference between humans and technology seems to be disappearing. Robots are increasingly taking on the role of equivalent or human-like interaction partners, making the question of how humans and robots differ all the more urgent.

The article thus starts from an anthropological-ethical perspective: on the one hand, conclusions can be drawn about the human self-relation and world-relation from the interaction with robotic systems. On the other hand, following the question investigated here, it can be better assessed whether the use of robots in social contexts – such as the care sector, school education or the therapy of people with impairments – is recommendable. Only marginal reference is made to empirical studies, since the topic of human–robot interaction has already been extensively researched on an empirical level and since the article is intended to provide a conceptual contribution to this discussion. With this in mind, we start by analyzing the technical operations based on which social robots imitate human-like behavior. Subsequently, we describe how emotionality, intelligence and autonomy are usually conceptualized in humans. Thus, we combine an analysis of the technical processes realized in social robotics with a discussion of the philosophical and psychological concepts important in discussing human interaction. Emotionality, intelligence, and autonomy are not treated here as sole or sufficient but as necessary features of human interaction, which play a central role in the context of HRI. Imitation, on the other hand, refers to the complex of anthropomorphic technology design with its imperceptible human–machine interfaces through which the acceptance of technical systems is to be increased and the use of technologies is to become as unobtrusive and intuitive as possible. Our aim is thus to make a conceptual contribution to the intuition that the technical imitation of emotional, intelligent, and autonomous

⁴ On social robots see, among others [1–3].

⁵ On the evaluation and quantification of perceived trust in human–robot interaction, see [6]. Cf. here also the psychological requirements for the design of human-like robots [7].

behavior differs qualitatively from human forms of emotionality, intelligence, and autonomy. Therefore, we conclude that caution is advised, as anthropomorphizing robots can lead to false assumptions of reciprocity.⁶ With the notion of reciprocity, we refer to sociological and psychological theories in which reciprocity describes a necessary condition of social relationships. Thus, reciprocity can be explained epistemologically, for instance if, as in Mead [13], the "reciprocity of perspective" is regarded as essential for mutual understanding. On the other hand, reciprocity can be understood as an essential condition of moral relations of recognition [14]. Reciprocity here refers to the fact that interaction partners recognize each other as persons who depend on mutual recognition. The meaning of reciprocity may thus vary depending on the underlying theory, but in most cases refers to a necessary condition of social interactions. However, before we turn more specifically to the complex of imitation, reciprocity, and human conditions for interaction, we first take a look at the field of healthcare, since healthcare is one of the main areas of application in which the use of social robots raises controversial questions about the role and relationship between humans and machines.

Social Robots in Healthcare

In contrast to their industrial predecessors, social robots are designed for applications beyond industry and are characterized by more human-like features and better interactive skills. They are designed to read and interpret human emotions, to display situational awareness, and to express signs of emotional and

cognitive involvement. This is realized by a behavior-based approach that has characterized robotics since the 1980s: robots were no longer considered purely useful tools, but were rather designed as flexible, dynamic, intelligent, and adaptive systems, which were intended to interact with their environment and to adapt to their user's needs. Against this backdrop, social robots were created to develop cognitive, social, and emotional skills, which would enable them to display adaptive behavior and respond to the needs of their users (See among others [1, 2]). The goal of this development is thus to foster increasingly intimate, intuitive, and frictionless human–machine interactions. The use of robotic systems should be as easy and unobtrusive as possible so that even older or ill people are able to use robots for their purposes and in accordance with their needs.

In health care, social robots are mostly used for therapeutic purposes (both physical therapy and psychotherapy [15]), as well as in care facilities such as, for instance, geriatric care [16]. Social care robots are designed to both assist caregivers in everyday tasks such as washing and transferring patients, and to entertain patients and encourage their social interaction. For example, some social care robots, such as the humanoid robot "Myon," are programmed to motivate patients to sing and dance, to entertain them by playing music, or to stimulate their cognitive abilities through memory games.⁷ Other models, such as the robotic seal "Paro," perform the function of an artificial and sterile pet that, through touch and acoustic responses, is intended to relieve loneliness and provide missing social stimuli.⁸ Social robots in healthcare are capable of processing information ("thinking") as well as of adapting their behavior ("learning"). They possess mimetic, gestural, and linguistic abilities, by means of which they are supposed to establish and, at best, maintain a relationship with their human counterpart, thus creating the impression of being emotional, social, intelligent, and autonomous beings ([16], p. 202).

⁶ Cf. here the debate on the role of anthropomorphism for human–robot interaction, as well as on the requirements for robots to be considered cooperators, partners, or companions (cf., e.g., [10, 11]). However, these debates mostly focus on the question of how to design robots to give the *appearance* of possessing certain human competencies and characteristics so that they are *accepted* as partners or companions in accomplishing certain tasks. In contrast, we assume that interaction is a complex social phenomenon that both presupposes several conditions in interaction partners and influences human (understanding of) social practices and sociality as such (cf. [12]). Hence, even if the former conditions are only imitated, this would have implications for social practices, if not for sociality as such (which is open to normative evaluation), in particular if *perceived* as interaction in the narrow sense.

⁷ For further information on the Myon model, see the homepage of the Neurobotics Research Laboratory of the Berliner Hochschule für Technik (BHT): http://www.neurorobotik.de/robots/myon_en.php.

⁸ For more information on Paro, see: <http://www.parorobots.com/index.asp> as well as [17].

A further distinction can be made between humanoid and non-humanoid social robots. While humanoid social robots have human-like features, non-humanoid social robots mostly appear in the form of animals, whereby the interaction with their users is confined to non-verbal communication (see for example the robot seal “Paro,” the robot dog “AIBO,” and the nursing bear “Robear”). With non-humanoid social robots, interaction is often limited to the tactile dimension and encompasses mostly touch, as well as physical and auditory responses from the robot. Humanoid social robots, on the other hand, not only have a more human-like appearance, they are also capable of communicating verbally by using, among others, natural language processing (NLP) and speech dialogue modules [18]. Hence, the technical requirements for humanoid social care robots are high: they must – in addition to the usual requirements for robots such as a stable stand, fluid motion, and finely tuned motor skills – first of all exhibit an authentic, empathic, and consistent behavior. Secondly, they need to be able to remember previous interactions ([19], p. 68), and, thirdly, be capable of adapting to changing circumstances, as well as to the needs of their human counterparts ([19], p. 65). Thus, in addition to the generally required actuators, sensors, and motors, humanoid social care robots operate on the basis of image processing algorithms, NLP, methods to model and express emotions, as well as ML. In order to provide a better understanding of the imitation of human characteristics in social robots, we will in the following analyze a robot model that has been available on the market for several years, that has already been subject to extended empirical research and can thus be used to better illustrate the field of humanoid social robots: the robotic system “Pepper.”

Pepper was launched in 2014 and has been available on the market since 2015. Pepper is used in care facilities and hospitals, but is also applied, among others, in sales, hotels, banks, and education. In health-care, Pepper works as a health assistant, communicator, data generator, and receptionist.⁹ It consists of a variety of sensors and motors and can be connected to the Internet via WiFi or Ethernet. Pepper has a human-like appearance and can communicate with

its human counterpart not only through touch but also through natural language communication, as well as through a tablet installed in its chest.

Pepper seems sentient and emotional by different operations. In addition to body movements and sound, Pepper transmits information about its “emotional” state via natural language communication, a touch screen in its chest and LEDs behind its eyes. With these LEDs, Pepper is able to demonstrate whether it is listening (blue), processing information (green), or not “being there” at all (white),¹⁰ thereby displaying situation awareness and imitating attentiveness and emotional involvement. Above all, however, Pepper displays emotional and social behavior through its advanced abilities to remember and communicate. Thus, the robot can remember previous interactions and linguistic content. The impression of emotionality and social behavior is further strengthened by Pepper’s ability to interpret or decode facial expressions, gestures, and the tonality of its human counterpart. Pepper uses emotion detection and facial recognition and responds to the detected information with corresponding gestures, sentences, and movements.

Moreover, Pepper appears as a rudimentarily intelligent and autonomous device that imitates the understanding of its human counterpart. While Pepper appears to be sentient due to a variety of sensor technologies, the impression of intelligence is signaled by NLP and emotion detection, and manifests itself in the capability to remember previous interactions and to perform AI-based communication. Besides speech recognition that enables Pepper to communicate via dialogue, it is able to focus its attention and has a certain degree of awareness of its situation and environment [20]. In addition, Pepper also appears to be autonomous to a certain degree. In contrast to humans, however, “autonomy” here refers mainly to the ability to move independently, to detect objects in the immediate surroundings, to sidestep them and thus avoid collision. Autonomy is thus performed through body posture identification, self-configuration recognition and self-body awareness, among others [21]. In other words, robots like Pepper have modules that allow them to confront their human users as apparently independent beings. In addition, Pepper conveys the impression of

⁹ Cf. <https://www.softbankrobotics.com/emea/en/industries/healthcare>.

¹⁰ Cf. http://doc.aldebaran.com/2-4/family/pepper_technical/index_dev_pepper.html.

being intelligent and autonomous by its ability to learn. By means of learning algorithms and machine learning methods, Pepper is capable of adapting to its user's preferences and needs and is able to imitate the user's behavior. Every interactive setting thus provides the robot with new data, based on which it can improve its interactive, social, and linguistic skills.

Differences in Humans and Robots regarding Emotionality, Autonomy, and Intelligence

In the following sections, we analyze how autonomy, emotionality, and intelligence are used differently in the context of humans and robots by elaborating how these qualities may be conceptualized in humans. In doing so, we do not understand autonomy, emotionality, and intelligence as exhaustive properties of human interaction, but rather as necessary features of human interaction that are closely related to epistemological and ethical concepts of reciprocity. Moreover, they are central in the context of an anthropomorphic technology, insofar as they are crucial to the "human likeness" of social robots. Further, the selection of these three features allows to pursue the question of how human and robotic modes of interaction differ on the basis of three different criteria and thus offers impulses for further analysis. With this in mind, we refer to selected philosophical and psychological debates in which emotions, intelligence, and autonomy are analyzed as prominent human characteristics. Subsequently, we examine to what extent these concepts can be applied to social robots.

Emotionality

For a long time, emotions were understood in opposition to rationality and reason. In current debates in the "philosophy of emotions," however, the relationship between rationality and emotions is being reconceptualized. Two lines of thinking can be distinguished here: cognitivist and phenomenological theories of emotion. Cognitivist theories define emotions as cognitions or mental states such as thoughts, beliefs, or evaluations [22–26].¹¹ Hence, emotions are said

to have a propositional structure and, on this basis, are asserted to be structurally analogous to rationality. Phenomenological theories of emotion, however, object this. They criticize the reduction of emotions to mental states and emphasize the bodily relation of emotions ([29], p. 21, 28), as well as their quality of experience. Hence, according to phenomenological theory, it is primarily the (bodily) *experience* that distinguishes emotions from other mental phenomena.¹²

Thus, the realm of emotions is not clearly defined [31]. Further, in the "philosophy of emotions," a distinction is often made between *emotions* and *feelings* [32, 33]. In this sense, at least five characteristics of emotions can be defined. Emotions are – in contrast to feelings or mere bodily and sensory sensations – intentional (1), have a representational and evaluative character (2), a motivational function (emotions motivate to act) (3), possess a bodily component (4) and an experiential quality (5). A decisive criterion of differentiation between emotions and feelings is therefore their intentionality. Emotions differ from feelings or bodily sensations such as pain, pleasure, cold, or warmth first of all by their intentionality, i.e., by their directedness to an object (1) [30]. This results in another property of emotions: their representational and evaluative character. The object to which emotions refer is, qua this reference, an "evaluated" object. Accordingly, emotions represent their intentional content and confer a value on the object to which they are directed. For example, the dog I am afraid of becomes a frightening animal (for me) through my emotion of fear [29, 32]. Furthermore, emotions have a motivating function (for instance, emotions motivate us to act) (3) and a physical component (4): they manifest themselves in facial expressions and gestures and are often accompanied by physical reactions such as sweating, blushing, and others. And lastly, emotions have an experiential quality (5). Thus, emotions not only evoke physical reactions, but are also experienced in terms of a lived-body or a body-subject, i.e., for example, as feelings of constriction, tension, widening or opening ([34], p. 22, 222).

Against this backdrop, the question arises to what extent social robots can perform or imitate emotional behavior. For example, understanding emotions as

¹¹ Other forms of cognitivism define emotions as perceptions or as states similar to perceptions [27, 28].

¹² See on emotional experience further [30].

physiological or biochemical processes and conceptualizing them within the framework of stimulus–response schemes could lead to the assumption that machines can, in principle, acquire emotions. If in the future, for instance, Pepper was able to read and interpret the facial expressions, gestures and moods of its human counterparts as stimuli, and respond with corresponding emotional reactions, it could be assumed that Pepper has a form of emotionality comparable to that of humans. However, this assertion appears problematic if one considers the premises of this assumption. Thus, the understanding of emotions that defines them as cognitive states or bodily sensations presupposes the existence of mental states, i.e., of consciousness. The question of whether social robots or AI-based technologies in general will one day be able to develop consciousness-like states is, however, doubtful and highly controversial [35]. Moreover, the aforementioned assumption becomes even more critical if one considers the phenomenological understanding of emotions, which situates emotions within the relationship of the self and the world, and thus links emotions to the concept of corporeality. For even if social robots already show a certain degree of situational awareness, and even though research is currently being conducted to implement self-awareness or bodily awareness in robots [21, 36], it still seems that there is a long way to go before felt corporeality and an existentially experienced self-relationship can be achieved, if at all.¹³

Autonomy

Autonomy, on the other hand, is one of the central concepts of ethics and political theory and covers a wide range of meanings. It can refer to self-rule, self-determination, self-knowledge, absence of external causation, and free will ([37], p. 267). At the same time, autonomy can be understood as a capacity, a competence, a constitution, an ideal, or a right [38]. Etymologically, autonomy refers to the capacity for self-legislation (autós: self, nómos: law). In this meaning, the term also found its way into Kant's philosophy, where autonomy refers to the human

capacity for self-legislation, from which the demand is derived to make actions dependent on the generalizability of their maxims. Moreover, it is often emphasized that autonomy means freedom from external coercion or domination by others. A person who is autonomous in accordance with this understanding is able to make choices and decisions based on their own preferences without any major interference [37]. A third approach defines autonomy as personal autonomy and thus as the ability to act and live according to one's own reasons [39]. Here, the focus is on the ability to critically evaluate one's own beliefs and identify with higher-order volitions. In this context, authenticity or the coherence of a set of beliefs, long-term values, or life plans is considered a condition of autonomy [39–42]. Another approach conceives autonomy as relational autonomy. Here, the traditional concepts of autonomy, which are based on the notions of self and identity, are criticized; instead, the importance of relationships and interdependencies for the constitution of the individual self is stressed [43–45]. The autonomous self is thus not one that precedes interaction and relations, but rather one that is always influenced and shaped by other individuals and entities, be they human or non-human. Relational theories of autonomy thus criticize the idea of an "atomistic" and self-sufficient self, and argue that identity is relational, i.e., pluralistic, and exposed to others [46].

What does this mean for social robots? In what ways can we speak of social robots as autonomous? Although the term "autonomous devices" is often used in connection with new technologies (cf. the notion of "autonomous" vehicles, weapons systems, or robots), this terminology is revealed to be misleading when we take a closer look at its underlying concept. Nevertheless, the concept of autonomy harbors a certain ambivalence regarding AI-based technologies, insofar as these technologies – especially in the field of robotics – seem to exhibit a certain degree of self-activity. If, for example, Pepper performs tasks independently and optimizes them through repeated execution, it appears to act free of external coercion and thus "autonomously." However, insofar as Pepper's scope of action depends on its programming, Pepper is ultimately still subject to external – i.e., human – legislation.

If, however, one considers autonomy as the ability to act and live according to one's own reason, it is

¹³ Again, this amounts to the question of whether technology will one day be able to develop human-like states of consciousness.

conceivable that social robots could form a hierarchy of values and follow higher-order values (such as, for instance, the value of politeness), even while making individual decisions. Or, at least, social robots could aim to achieve a certain degree of coherence in individual decisions and the pre-programmed higher-level values. However, this would still not be comparable to those forms of reflection, evaluation, and identification that human individuals undertake in terms of their personal autonomy, concerning their beliefs and value judgements. The latter, in fact, presupposes a lively interest in one's own desires, value attitudes, and life plans.

The concept of relational autonomy would ultimately be most applicable to social robots, as it emphasizes the role of interactions in the formation of the self and identity and, thus, for autonomy. Thus, the transferability of the concept of autonomy to social robots seems plausible, but still only to a limited extent, since autonomy is inextricably related to the concept of vulnerability [47]. Vulnerability, however, refers to a moment of existential exposure and social interdependency at least to a certain degree that is inherent only to corporeal beings.¹⁴ Even though the concepts of embodiment and situatedness determine the construction of robotic systems, social robots are still technically constructed bodies that lack the experience of contingency and interdependency. Unlike the living human body, social robots are replaceable, exchangeable, and not subject to finitude [48].

Intelligence

Intelligence is one of the main characteristics of human existence, and it plays a decisive role in situations of interpersonal interaction. In the history of philosophy, intelligence has been elevated to the defining characteristic of human beings, distinguishing them from other entities such as animals, plants, and machines (cf., among others, the definition of "human beings" as *animal rationale* or *zoon logon echon*). However, this line of thinking had been

challenged at least from two directions by the twentieth century. Firstly, ethology and comparative behavioral research proved that animals also demonstrate intelligent behavior. Secondly, research on artificial intelligence from the 1950s onwards considered the possibility that "machines can also think" [49]. But what is intelligence and to what extent are human and machine intelligence similar or different?

If one looks for a definition of "intelligence," one encounters a multitude of different approaches. Intelligence is, for instance, defined as the ability to *adapt to new situations*, as the competence to *efficiently solve problems*, to *learn from experience* or as *symbol manipulation*.¹⁵ Some theories assume a *general intelligence factor* that underlies all intellectual performances and represents the cognitive performance of an individual (two-factor theory) (cf., for example [51]). Against this backdrop, Raymond B. Cattell distinguished between two different forms of intelligence: a fluid intelligence, and a crystallized (general) intelligence. While fluid intelligence includes all those abilities that are innate, rather than socially or culturally conditioned (such as general comprehension), crystallized intelligence refers to experience-based abilities and thus encompasses explicit knowledge (such as factual knowledge), as well as implicit knowledge (such as body techniques).

In general, *factorial*, *multidimensional*, and *process-oriented theories* of intelligence can be distinguished. Multidimensional intelligence models, as formulated by Joy P. Guilford or within the framework of the Berlin Intelligence Structure Model, understand intelligence as a collection of different abilities that are differentiated into complex systems.¹⁶ In doing so, they provide a counter-model to the hierarchical approach of older factorial intelligence models. Thus, Robert Sternberg abandons the assumption of a general intelligence and shifts the focus from a hierarchical concept of intelligence to the study of information processing [53]. In contrast to the hierarchical model of factorial theories of intelligence, Sternberg distinguishes between three

¹⁴ Cf. e.g., "[...] in its surface and its depth, the body is a social phenomenon: it is exposed to others, vulnerable by definition. Its very persistence depends upon social conditions and institutions, which means that in order to 'be' in the sense of 'persist' it must rely on what is outside itself" ([46], p. 33).

¹⁵ Cf. on the latter definition the "Physical Symbol System Hypothesis (PSSH)" by Allen Newell and Herbert A. Simon, which links intelligence to the manipulation of symbols and thus defines it as symbol processing [50].

¹⁶ See, for example, Guilford's Structure of Intellect Theory [52].

different components of intelligence (analytical, creative, and practical intelligence) and at the same time adds two further aspects to the concept of intelligence: “the ability to adapt to a continuously changing environment” and “the motivation to acquire knowledge” [54]. Against this backdrop, Howard Gardner, in his “Theory of Multiple Intelligences” claims seven different “intelligences” which are inter-related but still relatively independent of each other [55]: verbal, logical-mathematical, spatial, musical, bodily kinesthetic, intrapersonal (self-reflection), and interpersonal (social skills) intelligences.¹⁷

How can this be transferred to the field of social robotics? In view of the various models of intelligence, as they have been elaborated on since the beginning of the twentieth century, a transfer of the concept of intelligence to technical systems seems possible. In particular, information processing theories, which attribute intelligent behavior to information processing, seem to suggest a proximity of human and machine intelligence (see, e.g., Sternberg). Similarly, the concept by Allen Newell and Herbert A. Simon, which explains intelligence as the manipulation of symbols or symbol processing, seems applicable to both human and machine entities. One reason for this is the functionalist perspective, according to which intelligence is based on information processing, processing speed, and numerical abilities.

Despite the conceptual proximity that the field of machine intelligence shares with its natural model, it is striking that “natural” forms of intelligence encompass a much broader scope than their technical simulations. While technical intelligence is primarily specialized in a specific area, human intelligence expresses itself in more diverse ways: for example, as social, intrapersonal, bodily kinesthetic, or even musical intelligence [55]. Human intelligence therefore does not consist exclusively of logical reasoning and information processing but can also be expressed through visual, emotional, or social abilities. Moreover, human intelligence is more error-prone, unspecialized, and thus in some ways more irrational than its technical counterpart. It is by comparison

versatile, open, and unspecialized [57]. This may, however, change in the field of social robotics, as social robots could in principle develop certain forms of interpersonal and embodied intelligence through their interactive skills and embodiment. However, intrapersonal intelligence, i.e., the ability of self-reflection, still seems a long way off for social robots. In addition, human intelligence is always bound to a biological body, which is inherently temporal, and has, moreover, a close proximity to consciousness and thought. Thus, human intelligence and technical intelligence ultimately refer to very different phenomena, while the question of technical intelligence points to that of artificial consciousness. Just as artificial consciousness, if it were technically realizable, would in all likelihood be fundamentally different from human states of consciousness, technical or artificial forms of intelligence also describe an entirely new phenomenon.

Implications for Human–Robot Interaction

As we have shown, there are major differences between human emotionality, autonomy and intelligence, and the characteristics that social robots exhibit. This has far-reaching implications for human interactions with social robots. On the one hand, the human-like yet different characteristics of social robots lead to increased acceptance and a higher degree of usability of the technical systems. On the other hand, the technical imitation of human characteristics and modes of interaction can have problematic implications for assumptions about interpersonal relationships.

Especially in the healthcare sector discussed earlier, social robots that are autonomous, problem solving, and responsive to the emotional needs of humans can lead to improvements and innovations. In the treatment of dementia patients, for instance, social robots can help to relieve nursing staff by taking over routine tasks such as transferring, carrying and washing patients, administering medication, or taking blood samples. At the same time, they can bring about new forms of therapy and allow for new methods of monitoring and communication [58, 59]. In addition, there have recently been more and more studies investigating the use of social robots in autism therapy. Here, it is observed that children with autism

¹⁷ In his later writings, Gardner speculates of integrating three more “intelligences” in his “Theory of Multiple Intelligences”: a naturalistic, an existential, and a pedagogical intelligence [56].

seem to engage more easily with a robotic system as a social counterpart [60]. While interaction with human subjects is often stressful and overwhelming for the children, they face a lower level of social stimuli and demands with social robots. In this way, one takes advantage of the fact that social robots have so far only limited communication and interaction capabilities, which is why interaction with them is less complex and therefore easier to process and understand. Social robots thus function as a practice tool that children with autism can use to gradually approach the field of complex social interactions.

This is made possible by their human-like, yet different design and interaction mode. Thus, several studies have shown that children in robot-assisted autism therapy were able to improve their communication and interaction skills as well as their understanding of emotional responses [61, 62]. Hence, new interaction scenarios can be observed in the use of social robots in care and therapy that are characterized by increasing closeness, intimacy and trust between humans and technology. Further, social robots can become "devices of trust" especially for patients with cognitive or emotional impairments, compensating for the lack of social interactions and providing patients with missing or adequate social stimuli. On the other hand, however, imitating human behavior also has problematic effects. The problem, however, does not so much lie in the different manifestations of the characteristics discussed here, but rather in the simulation itself. Because robots merely imitate autonomous, emotional and intelligent behavior, i.e., they have no intrinsic social motivation, people can be deceived about the fact that they are interacting with a technical system [63]. In other words, the use of social robots in care involves the risk that users do not realize that the interaction with the robotic systems is algorithmically pre-programmed, i.e., that robots do not respond to their counterparts in a responsive manner [64, 65], but merely follow preset programs. Although this argument has been criticized in some contexts (cf., for example, [66]), it seems to remain relevant in the field of care and therapy, where social robots encounter people with physical, emotional, or social impairments. For if robots do not behave socially of their own accord, but only on the basis of their programming, it means that they ultimately perceive only those social signals that they can measure, calculate, and process qua their

programming, as a result of which both sympathy and attention as well as spontaneity and flexibility in care and therapy are likely to decrease.

In addition to these implications, which are problematic especially for the care sector, the imitation of human behavior and the paradigm of human likeness in technology design could also have problematic effects on a more fundamental and general level. Thus, it could challenge the intuitively shared basis of interpersonal interaction and reinforce tendencies already observed in other technological contexts, such as the phenomena of "automation bias," for instance. *Automation bias* – the tendency of people to trust automatically generated decisions more than human decision-making processes – is particularly relevant in the field of social robotics: first, because social robots appear as quasi-independent, embodied and "understanding" entities through their ability to imitate and mimic human behavior; and second, because social robots are mostly used by vulnerable groups whose critical judgment can be limited. In this context, studies have shown that, especially in the field of social robotics, people are more likely to be influenced in their individual judgments by robots than by humans ([67], p. 41). The increased trust in automated decision-making processes observed here could thus lead to a loss of trust in human decision-making and to new forms of manipulation.

Moreover, the imitation of human characteristics could decisively change the shared basis of human interaction. Thus, it is crucial for human–human relationships to assume that the social counterpart is capable of being emotional, autonomous, and intelligent and that this applies to all interaction partners, i.e., that interaction is characterized by reciprocity. However, we cannot assume that the emotions people demonstrate in situations of interpersonal interaction are necessarily real or similar to our own emotions. Nevertheless, interpersonal interaction is based on the assumption of a fundamental similarity regarding the disposition of being emotional, intelligent, and autonomous. If the basic assumption of fundamental similarity and reciprocity is now transferred to artifacts that merely mimic human properties, the common basis of human interaction could change decisively. For example, dealing with social robots that are always obedient and never show interests, wishes or needs on their own could reduce the willingness to engage in complex and conflictual interactions [68, 69]. To the extent

that social robots exhibit socio-emotional behavior but always adapt to the needs of their users and never appear with divergent interests and demands, interacting with social robots – assuming an all-encompassing use of such systems – might lead to a decrease in the willingness to deal with unruly and non-adaptive behavior, i.e., to acknowledge the conflictual and demanding dimension of social relationships.

Conclusion

Humans have always interacted with non-human entities, transferring certain human characteristics, such as emotions or motivations, to non-human entities. This ranges from animism to assumptions about divine human likeness, from human-like animals in fairy tales and fables to childhood games with dolls and cuddly toys. What is new in the context of social robots, however, is that the tendency to anthropomorphize non-human entities is brought to a technological level and is thus deliberately reinforced. Anthropomorphism migrates into the devices and is reflected in their design and modes of operation. However, as has been shown, the resulting forms of simulated emotionality, intelligence, and autonomy differ greatly from their human model. This, however, may lead to false, reinforced, and confusing assumptions regarding the reciprocity and basic assumptions of social interactions. Therefore, further studies are needed that critically examine and evaluate the potential risks of false assumptions about reciprocity and social robots as interaction partners. This is particularly important for the use of social robots in healthcare, i.e., among vulnerable groups such as children or the elderly. Thus, educating individuals and society about social robots is a task that will become increasingly important as the development and use of social robots becomes more widespread.

Acknowledgements We would like to thank Ms. Jessica Harreaves for the language corrections.

Funding Open Access funding enabled and organized by Projekt DEAL. The work on this publication was funded by the German Research Foundation (DFG) | project number 418201802.

Declarations

Conflicts of Interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Robinson H, MacDonald B, Broadbent E (2014) The role of healthcare robots for older people at home: A review. *Int J Soc Robot* 6:575–591. <https://doi.org/10.1007/s12369-014-0242-2>
2. Royakkers L, van Est R (2015) A literature review on new robotics: Automation from love to war. *Int J Soc Robot* 7:549–570
3. Jones RA (2017) What makes a robot “social”? *Soc Stud Sci* 47:556–579. <https://doi.org/10.1177/0306312717704722>
4. Kurzweil R (2005) *The singularity is near: When humans transcend biology*. Penguin US
5. Vita-More N (2008) Designing human 2.0—regenerative existence. *Artifact* 2(3–4):145–152
6. Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human-robot interaction. *Hum Factors* 53:517–527. <https://doi.org/10.1177/0018720811417254>
7. Kahn PH, Ishiguro H, Friedman B, Kanda T (2006) What is a human? - Toward psychological benchmarks in the field of human-robot interaction. *ROMAN* 364–371. <https://doi.org/10.1109/ROMAN.2006.314461>
8. Roesler E, Onnasch L (2020) Teammitglied oder Werkzeug – Der Einfluss anthropomorpher Gestaltung in der Mensch-Roboter-Interaktion. In: Buxbaum HJ (ed) *Mensch-Roboter-Kollaboration*. Springer Gabler, Wiesbaden, pp 163–175
9. Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In: *Proceedings, Social Robotics. 4th International Conference, ICSR 2012, Chengdu, China, October 29–31 2012*. Springer, Berlin, Heidelberg, pp 199–208
10. Duffy BR (2003) Anthropomorphism and the social robot. *Robot Auton Syst* 42:177–190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3)

11. Damiano L, Dumouchel P (2018) Anthropomorphism in human-robot co-evolution. *Front Psychol* 9:468. <https://doi.org/10.3389/fpsyg.2018.00468>
12. Gallagher S (2020) Action and interaction. Oxford University Press, Oxford
13. Mead GH (1934) Mind, self and society from the standpoint of a social behaviorist. University of Chicago Press, Chicago
14. Honneth A (1992) Kampf um Anerkennung. Zur moralischen Grammatik sozialer Konflikte. Suhrkamp, Frankfurt/M
15. Fiske A, Henningsen P, Buyx A (2019) Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J Med Internet Res* 21
16. Bendel O (ed) (2018) Pflegeroboter. Springer Gabler, Wiesbaden
17. Marti P, Pollini A, Rullo A, Shibata T (2005) Engaging with artificial pets. In Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics, University of Athens, 2005 (EACE '05) 99–106
18. Sato M, Iwasawa T, Sugiyama A, Nishizawa T, Takano Y (2009) A single-chip speech dialogue module and its evaluation on a personal robot, PaPeRo-Mini. *IEEE Int Conf Acoust Speech Signal Process* 3697–3700
19. Janowski K, Ritschel H, Lugin B, André E (2018) Sozial interagierende Roboter in der Pflege. Universität Augsburg; Springer Fachmedien, Augsburg, Wiesbaden
20. Perera V, Pereira T, Connell J, Veloso M (2017) Setting up pepper for autonomous navigation and personalized interaction with users. <https://arxiv.org/pdf/1704.04797>
21. Rodriguez I, Martinez-Otzeta JM, Lazkano E, Ruiz T, Sierra B (2018) On how self-body awareness improves autonomy in social robots. In 2017 IEEE ROBOT, 2017/12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/robio.2017.8324661>
22. Solomon RC (1993) The passions: Emotions and the meaning of life, 3rd edn. Hackett Publishing Company, Indianapolis/Cambridge
23. Solomon RC (2003) Not passion's slave: Emotions and choice. Oxford University Press, New York
24. Roberts RC (2003) Emotions: An essay in aid of moral psychology. Cambridge Univ. Press, Cambridge
25. Stocker M (1996) Valuing emotions. Cambridge Univ. Press, Cambridge
26. Helm BW (2002) Felt evaluations: A theory of pleasure and pain. *Am Philos Q* 39(1):13–30
27. DeSousa R (1987) The rationality of emotion. MIT Press, Cambridge, Mass.
28. Döring SA, Mayer V (eds) (2002) Die Moralität der Gefühle. De Gruyter, Berlin/Boston
29. Perler D (2011) Transformationen der Gefühle: Philosophische Emotionstheorien 1270–1670. Fischer E-Books, Frankfurt am Main
30. Deonna J, Teroni F (2012) The emotions: A philosophical introduction. Routledge, London
31. Prinz JJ (2004) Gut reactions: A perceptual theory of emotion (philosophy of mind series). Oxford University Press, Oxford, New York
32. Döring SA (ed) (2009) Philosophie der Gefühle. Suhrkamp, Frankfurt am Main
33. Weber-Guskar E (2009) Die Klarheit der Gefühle: Was es heißt, Emotionen zu verstehen. De Gruyter, Berlin
34. Landweer H, Demmerling C (2007) Philosophie der Gefühle. Von Achtung bis Zorn. Metzler, Stuttgart
35. Coeckelbergh M (2010) Moral appearances: Emotions, robots, and human morality. *Ethics Inf Technol* 12:235–241. <https://doi.org/10.1007/s10676-010-9221-y>
36. Chella A, Pipitone A, Morin A, Racy F (2020) Developing self-awareness in robots via inner speech. *Front Robot AI*. <https://doi.org/10.3389/frobt.2020.00016>
37. Agich GJ (2009) Key concepts: Autonomy. *Philos Psychiatr Psychol* 1(4):267–269
38. Feinberg J (1986) Harm to self. Oxford Univ. Press, New York
39. Betzler M (2013) Einleitung: Begriff, Konzeptionen und Kontexte der Autonomie. In M Betzler (Ed) Autonomie der Person. Mentis Verlag, Paderborn
40. Betzler M (2009) Authenticity and self-governance. In: Salmela M, Mayer V (eds) Emotions, ethics, and authenticity. Consciousness and emotion book series, Vol 5. John Benjamins, Amsterdam, pp 51–68
41. Frankfurt HG (1971) Freedom of the will and the concept of a person. *J Philos* 68(1):5–20
42. Dworkin G (1988) The theory and practice of autonomy. Cambridge University Press, Cambridge
43. Grimshaw J (1988) Autonomy and identity in feminist thinking. In: Griffiths M, Whitford M (eds) Feminist perspectives in philosophy. Macmillan Press, Basingstoke, pp 90–108
44. Mackenzie C, Stoljar N (eds) (2000) Relational autonomy: Feminist perspectives on autonomy, agency, and the social self. Oxford Univ. Press, New York
45. Abrams K (1999) From autonomy to agency: Feminist perspectives on self-direction direction. *William Mary Law Rev* 40(3):805–846
46. Butler J (2009) Frames of war: When is life grievable? Verso, London
47. Delgado J (2019) Re-thinking relational autonomy: Challenging the triumph of autonomy through vulnerability. *Bioethics Update* 5:50–65. <https://doi.org/10.1016/j.bioet.2018.12.001>
48. Anders G (2002) Die Antiquiertheit des Menschen. C.H. Beck, München
49. Turing AM (1950) Computing machinery and intelligence. In *Mind*, LIX. <https://doi.org/10.1093/mind/lix.236.433>
50. Newell A, Simon HA (1976) Computer science as empirical inquiry. *Commun ACM* 19:113–126. <https://doi.org/10.1145/360018.360022>
51. Spearman C (1904) "General intelligence," objectively determined and measured. *Am J Psychol* 15:201. <https://doi.org/10.2307/1412107>
52. Guilford JP (1967) The nature of human intelligence. McGraw-Hill, New York
53. Sternberg RJ (1977) Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Erlbaum, Hillsdale
54. Sternberg RJ (1985) Beyond IQ: A triarchic theory of human intelligence. Cambridge Univ. Pr, Cambridge
55. Gardner H (1983) Frames of mind: The theory of multiple intelligences. Basic Books, New York

56. Gardner H (1999) *Intelligence reframed: Multiple intelligences for the 21st century*. Basic Books, New York
57. Grunwald A (2019) *Der unterlegene Mensch: Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern*. Riva premium riva verlag, München
58. Wada K, Shibata T (2006) Robot therapy in a care house - its sociopsychological and physiological effects on the residents. In: *Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006*, pp 3966–3971
59. Manzeschke A (2014) Altersgerechte Assistenzsysteme: Ethische Herausforderungen technologischen Wandels. In *Informationsdienst Altersfragen* 41(3):10–18
60. Warren Z et al (2015) Can robotic interaction improve joint attention skills? *J Autism Dev Disord* 45:3726–3734
61. Mavadati S et al (2016) Robot-based therapeutic protocol for training children with Autism. In: *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, New York, NY, USA, 26–31 August 2016, pp 855–860
62. Eichenberg C (2020) Robotik in der Psychotherapie: Anwendungsfelder – Effektivität – Praxisbeispiele. In: Bauer M, Deinzer L (eds) *Bessere Menschen? Technische und ethische Fragen in der transhumanistischen Zukunft*. Springer, Berlin, pp 97–125
63. Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. *Minds Mach* 16(2):141–161
64. Waldenfels B (2006) *Schattenrisse der Moral*. Suhrkamp, Frankfurt/M
65. Waldenfels B (1994) Response und Responsivität in der Psychologie. *J für Psychol* 2(2):71–80
66. Coeckelbergh M (2016) Care robots and the future of ICT-mediated elderly care: A response to doom scenarios. *AI Soc* 4:455–462
67. Ullrich D, Diefenbach S (2017) Truly social robots - understanding human-robot interaction from the perspective of social psychology. In *Proceedings of the 12th International*, pp 39–45
68. Sharkey N, Sharkey A (2010) The crying shame of robot nannies: An ethical appraisal. *Interac Stud* 11(2):161–190
69. Prescott T, Robillard J (2020) Are friends electric? The benefits and risks of human-robot relationships. *iScience* 24(1)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.