



Three Responses to Anthropomorphism in Social Robotics: Towards a Critical, Relational, and Hermeneutic Approach

Mark Coeckelbergh¹

Accepted: 23 February 2021
© The Author(s) 2021

Abstract

Both designers and users of social robots tend to anthropomorphize robots. Focusing on the question how to conceptualize the relation between robots and humans, this paper first outlines two opposite philosophical views regarding this relation, which are connected to various normative responses to anthropomorphism and anthropomorphization. Then it argues for a third view: navigating between what it calls “naïve instrumentalism” and “uncritical posthumanism”, it develops a hermeneutic, relational, and critical approach. Paradoxically, by unpacking the human dimension of robotics in its use and development, this view enables a critical discussion of anthropomorphizing robots. At the same time, and again somewhat paradoxically, it avoids a naïve instrumentalist position by taking robots’ role as an instrument in a larger con-technology seriously. As such, the third view questions the dualism assumed in the debate. The paper then explores what this means for the field of social robotics and the education of computer scientists and engineers. It proposes a reform based on a relational understanding of the field itself and offers suggestions for the role of users-citizens.

Keywords Social robotics · Human–robot relations · Anthropomorphism · Anthropomorphization · Instrumentalism · Posthumanism · Hermeneutics · Relational approach · Power

1 Introduction: Anthropomorphism and Anthropomorphization in Social Robotics

Social roboticists develop robots that are meant to function more naturally in social situations than the machines from the past. For this purpose, they are modelled on human beings and other higher primates, which in turn tend to be understood by means of the metaphor of a kind of social machine—albeit one that develops and has evolved. With the help of literature from psychology, cognitive science, neuroscience, and other fields, researchers in social robotics and human–robot interaction (HRI) identify features that are deemed necessary for social interaction and these are then implemented in the robot. For example, Dautenhahn refers to work in HRI (in particular [1]) and to the social brain hypothesis, which says that primate intelligence evolved because of a need to deal with increasingly complex social

dynamics, to define a number of social interaction characteristics that are implemented in social robots. She claims that ‘socially interactive robots exhibit the following characteristics: express and/or perceive emotions; communicate with high-level dialogue; learn models of or recognize other agents; establish and/or maintain social relationships; use natural cues (gaze, gestures, etc.); exhibit distinctive personality and character; and may learn and/or develop social competencies.’ [2:684; 1:145] Implementing such social features often leads to the robot exhibiting anthropomorphic appearance and behavior. Yet the purpose is not to build human-like robots as such (or primate-like robots for that matter). Anthropomorphization in this context has a more pragmatic aim. In social robotics, it is a strategy to integrate robots in human social environments. It is a means to give robots the ‘capacity to be able to engage in meaningful social interaction with people’ [3:178]. Both designers and users then tend to anthropomorphize such robots as they interact with them, ascribing to them anthropomorphic features such as personality, aliveness, and so on. In the next pages I will refer to ‘anthropomorphism’, ‘anthropomorphizing’ and ‘anthropomorphization’ as respectively the phenomenon and activity of ascribing human-like characteristics to robots

✉ Mark Coeckelbergh
mark.coeckelbergh@univie.ac.at

¹ Universitat Wien, Vienna, Austria

(by users as they interact with the robots) and the creation of such characteristics by robot designers and developers.

Sometimes anthropomorphism is not intended. Social roboticists are aware of this and hence try to control this phenomenon. With their design, they may try to encourage or discourage anthropomorphizing by the user. For example, some roboticists deliberately try to create robots that are as human-like as possible: Ishiguro [4] has argued that humanoids are ideal interfaces. Others build less human-like robots, for example robots that look like an animal, or try to avoid anthropomorphism altogether. In theory, anthropomorphism does not seem to be *necessary* for social robotics; it may well suffice that the robot displays the characteristics mentioned above, without being human-like. But in practice, many social roboticists welcome some degree or form of anthropomorphism, since, so it is argued, it facilitates interactions between humans and robots. Today anthropomorphism is one of the factors that are measured for shaping human robot interaction (HRI) [5]. It is recognized that anthropomorphism presents challenges, but ultimately the aim is ‘facilitating the integration of human-like machines in the real world’ [6] anthropomorphism helps with reaching that goal.

Robots are very well suited for this purpose [7]; because of their physical appearance and presence, they have a high ‘anthropomorphizability’ [6]. While already in the 1990s Reeves and Nass found that people treat computers as if they were real people [8], with robots this works even better, especially humanoid robots [9]. But anthropomorphism is not only affected by appearance: robots may offer all kinds of cues during the interaction [10] and there are other factors such as the robot’s autonomy, predictability, etc. Interestingly for the purpose of this paper, however, anthropomorphism is not only the result of how the robot looks or what it does, but also depends on the observer’s characteristics such as social background and gender [6]. For example, people apply social categories such as group membership to robots. [11] This implies that a social angle can help to better understand, and perhaps also better evaluate, anthropomorphism in human–robot interaction. It also suggests that, next to the objective characteristics of the robot, the subjectivity of the user matters.

Psychologists help roboticists with this project of using anthropomorphizing for the purpose of developing social robots by doing empirical studies on the perception of robots (depending on their features)—for example designing questionnaires and analyzing the results in order to study acceptability [12, 13]. Roboticists also import psychological theory about human sociality and do research in an interdisciplinary context: they use key paradigms from psychology and use and replicate findings from human–human interaction; often experimental approaches meet [14]. And next to shaping the design features of the robot, framing is also a method to

have humans anthropomorphize robotic technology: robots may be given a name or a backstory, which effects people’s responses to the robots [15]. Again, anthropomorphism with regard to technology is not new; but social robotics, developing anthropomorphic robots on purpose, seems to increase the effects.

Anthropomorphizing of robots raises all kinds of concerns, and some of these have been identified, recognized, and studied within the social robotics and HRI communities themselves. For instance, Turkle [16] and Scheutz [17] have voiced worries about emotional relationships people develop with such robots. There is also the famous “uncanny valley” problem. Since Japanese roboticist Masahiro Mori [18] introduced the (Freudian) concept of the “uncanny” in robotics, there has been discussion about how robots can evoke an uncanny feeling and how this can be avoided, for example when building androids [19]. For Freud, uncanniness is the psychological experience of strangeness in the familiar. In his 1919 essay on the concept [20], he interprets a story by E.T.A. Hoffmann that features a lifelike doll and other ‘uncanny’ elements: people feel strange when an object that is supposed to be lifeless appears lifelike and are threatened by their own unconscious, hidden impulses. Uncanniness in the context of robotics means that when a robot appears human-like but moves in strange ways, for instance, this may frighten users. One can try to provide psychological explanations for this—Freudian or other—such as fear of death. And from an anthropological view, the phenomenon of anthropomorphizing can be conceptualized as a form of technological animism: animism, the attribution of a soul or spirit to inanimate objects and natural beings, is then not understood as something that belongs to a ‘primitive’ stage of our social evolution or is limited to ‘natural’ phenomena, but as a something that underlies the construction of agency and personhood [21] and that also happens in relation to technological objects such as robots. However, here I will focus on anthropomorphism and anthropomorphizing specifically.

Given the problems raised by anthropomorphism, some roboticists have argued for not aiming at creating anthropomorphic robots as such. Goetz and colleagues, for example, argued that robots should be designed to match their task, rather than taking human-likeness as the main goal. [22] Nevertheless, since building anthropomorphic robots facilitates human–robot interaction and helps to achieve its goals, for example in health care and elderly care, it is often pursued by developers of social robots.

This paper asks: How can one respond to the issue of anthropomorphism from a philosophy of technology point of view? There are interesting possibilities to approach this topic, some of which may connect to the mentioned discussions in social robotics. In this paper, I approach the issue by asking the question how to philosophically frame

the relation between robots and humans. I first set up two opposing philosophical views on this question, which I shall call “naïve instrumentalism” and “uncritical posthumanism”. As they are formulated, they may come across as straw man views, but as I will show there are researchers who hold at least milder versions of them. I will show how each of these views can be connected to different normative positions on anthropomorphism and anthropomorphization: positions that do not concern the description and interpretation of these phenomena, but the question whether we *should* anthropomorphize robots and create anthropomorph robots. Sketching these extreme opposites then helps me to develop a third, alternative view, fueled by a hermeneutic, relational, and critical approach.

While I am aware that this discussion also links to ethics of robotics and discussions about the moral status of robots, I will not directly address these particular normative questions here and instead focus on the topic of anthropomorphism and the related human-technology relations problematic—as relevant to social robotics and HRI. This does not mean that there are no normatively relevant conclusions for social robotics to be drawn from the positions sketched, which I *will* do, but the focus of this paper is not on moral status. At the end of the paper, I will then explore the implications of this approach for doing social robotics and the education of computer scientists and engineers.

While the paper refers to work in social robotics and HRI, given the main audience of this journal it is worth noting explicitly that this paper presents a *philosophical* discussion of various views about human–robot relations and normative positions towards anthropomorphism and anthropomorphizing; it is neither a literature review nor an empirical study. Nevertheless, I hope the paper may stimulate a fruitful dialogue between social roboticists and philosophers. Furthermore, the discussion of views regarding the relation between robots and humans is also relevant to thinking about the relation between technology and humans in general, and may thus appeal to philosophers of technology.

2 Two Opposite, Problematic Views: Naïve Instrumentalism and Uncritical Posthumanism

There are at least two opposite views about the relation between robots and humans, which each can be connected to various normative responses to anthropomorphism and anthropomorphization:

2.1 Naïve Instrumentalism

One is to insist that robots are machines and mere instruments to human purposes. According to this view, which I

call “naïve instrumentalism”, anthropomorphizing robots is a kind of psychological bias. We (scientists) know that the robot is just a tool, but nevertheless when we interact with the robot our psychology (the psychology of users) leads us to perceive the robot as a kind of person. Users think that the robot is human-like, whereas actually it is not. (We scientists know that) robots are just machines. The philosophical basis of this view is twofold. First, it assumes a dualist view of the world in terms of reality versus appearance (a view that is sometimes ascribed to Plato): it is assumed that in reality the robot is an instrument, whereas in appearance it is more like a person. Naïve instrumentalism also assumes a dualist view of humans versus non-humans: humans and non-humans are mutually exclusive categories and there is a deep ontological divide between them. For thinking about technology, this means that technology is to be found on the side of non-humans (e.g., the things, the objects) whereas humans are entirely different (persons, subjects). Robots and humans are part of entirely different ontological categories. Second, it assumes a version of metaphysical and epistemological realism: objects exist independently of our concepts and perceptions, and we can describe them in an objective way. Users are misled by appearances: in reality, the robot is just a machine. Scientists can describe this reality in objective terms and study the way our minds are misled by the anthropomorphic features of the robot, the form of the interaction, and our social and cultural background.

This view may lead to at least two normative positions with regard to anthropomorphism and the task of social robotics. One is that researchers in social robotics can make use of this bias by designing human-like robot in order to improve human–robot interaction and better achieve its goals. Not only the robot is a tool; anthropomorphization itself is also an instrument that can be used and exploited. This is the stance I sketched in the beginning of this paper. It involves an uneven distribution of knowledge about the real state of affairs: the robot designer knows the reality, whereas the user is—at least temporarily—misled to believe that the robot is a person. One could say that the robot designer works as a kind of magician [23] who creates the illusion that the robot is human. As design philosopher Flusser [24] has noted, the very terms design and machine are etymologically related to cunning and deception. Robots are designed in order to produce anthropomorphization by the users, who are tricked into believing that the robot is a human or (taking into account the current state of the art in robotics) are at least willing to temporarily suspend disbelief and embrace the illusion. According to the instrumentalist view, this magic and trickery of anthropomorphization is not problematic if the goal is achieved: if anthropomorphism of the robot and anthropomorphization by the users leads to a better human–robot interaction, and if this in turn achieves the goals humans wanted to achieve (e.g. a specific health

care task), then this trickery is allowed and even recommended. This is again the instrumentalism at work.

Note that, by the same reasoning, if the goals are *not* achieved by anthropomorphism, the instrumentalist will not use it. If a robot, for whatever reason, does not fulfil its function, then instrumentalist reasoning will advise against it.

Another normative position is to use instrumentalism to argue that, regardless of the functioning and effectivity of the robot, it is highly problematic to anthropomorphize machines and to develop such machines *in general and in principle* because, so it is believed and asserted, robots are mere tools. Therefore, it is argued, social roboticists should stop designing them or at least make sure that users are aware that they are mere tools. Consider for example Bryson's position: robots are tools we use to achieve our own goals [25]. To talk about them in a way that suggests that they are people is misleading.

Moreover, there may be independent reasons why anthropomorphism is problematic. For example, one could argue that building such machines is also not desirable because of the potential ethical problems mentioned before: people might get emotionally attached to the machines and, as Sparrow and Sparrow have argued, such robots might disengage them from reality [26]. For example, a social robot in health care may mislead users into thinking that it is really alive or can really be a friend. While the current state of the art does not create this illusion for all users and for a long time, developments in social robotics seem to aim for this, for example with robots such as the baby seal robot Paro which are being used in hospitals and nursing homes. Turkle [16] has argued that robots like Paro provide only the illusion of a relationship.

From an instrumentalist perspective, one could support such criticisms and add that social roboticists should rather develop robots that avoid this illusion and just work as a tool. One should not build robots that “pretend” to be more than a tool. If implemented, this might well be the end of the very project of social robotics in the sense outlined above, since if the robot no longer has any features that give rise to anthropomorphizing (or if the user is constantly reminded that it is a mere machine), it is unclear how it can create “natural” social interaction with humans—or indeed “social” interaction at all. On this view, then, we should stop developing and using social robots that invite anthropomorphization. If that means that human–robot interaction then becomes less “natural” or less “social”, then so be it.

But whatever normative position towards anthropomorphization is taken (whether or not we *should* anthropomorphize), both positions assume an instrumentalist view of the relation between robots and humans. They both think that robots are just tools; they only differ when it comes to evaluating anthropomorphism and anthropomorphization. Furthermore, calling this instrumentalism “naïve” by no means

refers to ignorance about anthropomorphism, let alone about robotics. Both normative positions are very well aware of what anthropomorphism is and does (its nature and its effects). The term “naïve” in “naïve instrumentalism” only refers to ignorance concerning the non-instrumental dimension of technology. What this means will become clear in the next sections, especially when I unpack the third position.

2.2 Uncritical Posthumanism

Another view, at the other end of the spectrum and tentatively called “uncritical posthumanism”, is to totally embrace social robots as quasi-persons and “others”. Let me unpack this. Posthumanism can have several meanings, but here I mean theory that criticizes traditional humanist world views that put humans at the center of the world (anthropocentrism) and instead expands the circle of ontological and moral concern to non-humans. This could be non-human animals, but also for example (some) robots. In contrast to the instrumentalist position, here social robots are welcomed as part of a posthumanist ecology or network of humans and non-humans. Instead of a dualist worldview that opposes humans to things, according to this view humans and non-humans are part of the same network or ecology and are entangled in various ways. Posthumanists encourage us to cross the borders and question dualisms and binaries. This includes the humans-technology binary: according to this view, humans are technological beings and technologies are human: humans have always used technologies, humans create technologies, and technologies are part of our world in ways that are not just instrumental. For example, we may talk to technology as if it was a thing, we may expect more from technology than it can offer (something that often happens in robotics and also for example in AI), we can worship technology, and so on. For thinking about robots, this means that according to this view robots should not be seen as mere instruments and things if that means that the only relation we have to them is instrumental; instead, we can acknowledge them as part of our human and cultural world and even “meet” them as others: not as human others but also as social entities, entities that share our world. Moreover, instead of a realist view of the world and of how we know the world, posthumanists tend to be non-realist: they believe that we cannot have an objective view of “reality” (as if we could know a reality independent from our ways of knowing) and that that scientific beliefs (e.g., about robots) are a social construction. What robots “are”, then, is not exhaustively described by science and cannot be known independently of human subjectivity and culture. Consider for example the very terms “robot” and “machine”: the meaning of these terms is underdetermined by scientific and engineering definitions, since they have their own cultural history

and are dependent on contemporary usage. What “robot” means can depend, for instance, on the social context and the related language use. For example, we generally don’t call self-driving cars robots (in most contexts we call them cars), whereas according to most technical definitions they certainly are.

Such a posthumanist and social constructivist approach could be inspired by Haraway, who in her ‘A Cyborg Manifesto’ [27] and subsequent work has argued for crossing boundaries by including machines and animals into the political, or by Latour, whose non-modern approach includes non-humans in the social [28]. Haraway’s point is not only about the creation of literal cyborgs (mergers of humans and machines); she aims to change our *thinking* about (the importance of) boundaries between humans and machines, metaphysically but also politically. One could say, for instance, that robots should be included in society. Similarly, Latour’s thinking would see robots as non-humans that need to be included in the social and political collective; to exclude them would be to maintain a strict boundary between nature and culture, and between humans and non-humans. A non-modern approach goes beyond such binaries. Or one could interpret Gunkel’s investigations that use Levinas to explore whether robots can be considered as ‘Other’ [29, 30] as not identical to (his view is more critical), but potentially leading to such a view. In traditional phenomenology, the other (sometimes written: “Other”) is opposed to the self and usually refers to other human beings. The philosopher Emmanuel Levinas added to this a specific use of the term: the ‘Other’ is the radical counterpart of the self, which is absolutely other and we are called to ethically respond to that Other. Yet Levinas was still concerned with human beings. From a posthumanist point of view, one can then ask: could this other or Other also be a non-human animal or a robot?

According to a posthumanist view, the project of social robotics is not necessarily problematic and can even be embraced, since we are invited to include non-humans into the sphere of the social, and these anthropomorphizations may help with that. We should move beyond the anthropocentrism of the instrumentalist position: the human should not be the center. Social robotics may help with the project of de-centering the human. This position would also endorse anthropomorphism in social robots: it is fine and perhaps even desirable to build robots that look like humans (anthropomorphism) and are built and perceived as such (anthropomorphization), not because they are instruments to our purposes but since such robots offer us an “other” with which we can interact and which we can include in the social. Posthumanism rejects anthropocentrism: it rejects the idea that humans should be the center of the world and of moral concern. Human

purposes, therefore, are not the only purposes that count, and we may also want to create social robots that are not anthropomorph at all.

The latter position needs not reject anthropomorphism as such; it merely points out that there are many other possibilities. However, based on Gunkel’s view [29, 30] and taking a postmodern approach, one could also take another, less favorable normative stance towards anthropomorphization: one could argue that, viewed from a difference- or otherness-oriented posthumanist approach, anthropomorphizing in the sense of perceiving robots as humans *is* problematic to the extent that it does not respect the difference or otherness of the robot, understood as artificial other. Instead of projecting our humanness on the robots (how we are as humans and how we see ourselves as humans) and thus (ab)use them as mere project screens of our self-image, we should treat them as others in their own right and respect their difference. If we are not able to do this, we should not build anthropomorphic robots (but we can build other ones). This normative stance thus also relates anthropomorphism to anthropocentrism, but rejects both.

2.3 Problems

Both naïve instrumentalism and uncritical posthumanism are problematic. The first is naïve if it fails to fully understand that robots are not mere tools but have also unintended consequences and are bound up with humans through experience, language, social relations, narratives, and so on. Its dualist and realist view of the world, which configures the relation between humans and technology as external, prevents us from seeing that robots are not mere things (that are the result of technical construction and writing) but are also at the same time humanly and socially constructed. The second is uncritical if it forgets that robots are human-made machines, which might well confront us as ‘quasi-other’ [31] but are never totally other or completely non-human since they are created, interpreted and given meaning by humans. While they may confront us as an external other, they are never entirely external to us. The posthumanist’s focus on the otherness and social-cultural construction of robots leads to ignoring their origin in human and material practices. Robots are *also* machines. And these machines are made by humans. In other words, *both* views do not sufficiently recognize how social robots are intrinsically and (as I will put it below) internally related to humans in various ways.

I put these criticisms in terms of a lack of *full* understanding and *forgetting* since both might well have some knowledge of respectively unintended consequences and of the origin of the technology, or could in principle acquire this knowledge. To be more precise, therefore: the point is not that naïve instrumentalists do not know in principle that their technology can have unintended consequences or that

uncritical posthumanists do not know in theory that technology is created. They know that there are these relations or can acquire that knowledge. Rather, the problem is that they usually fail to understand the deeper implications of this knowledge for (thinking about) the relation between humans and robots, and hence for the problem of anthropomorphism. In particular, both views fail to understand how strong the entanglement of humans and robots really is, i.e., that it is an internal relation (see below).

Yet there is an alternative, “third” position, which remedies this problem but still enables us to be critical of anthropomorphism and anthropomorphizing—and indeed of the project of social robotics. How can we arrive at such a view? One way to go would be to try to find a “middle” position between the two extremes: social robots are neither entirely instrumental nor entirely other. But it is not clear what this means. What is the “non-instrumental” aspect of the robot? And what does it mean that a robot can be “other” to some extent? To find a real alternative, therefore, I propose that we question and change the assumptions that support both positions. In particular, I propose to drop the dualist assumption that the relationship between humans and robots is an external one and to show how human social robots and social robotics are. This creates a new, entirely different position that is not a middle position but instead escapes the field defined by the opposite views outlined here.

3 Towards a Third View: Critical, Relational, and Hermeneutic

Let me outline such a position, which is applicable to robots in general but is formulated in reply to the opposite positions that emerged in response to anthropomorphism and anthropomorphizing in social robotics. In order to constitute a real alternative as defined above, it should consist of at least the following elements:

3.1 Robots are Human, But Not Humans: They are Created and used, and They Shape Our Goals

First, social robots do not just appear from the wilderness or out of nowhere (as if they are part of “non-human” nature or as if they appear as aliens from outer space) but are designed and made by human beings. This implies that the “tool” or the “other” brought forward in the two positions we considered are not only non-human but also human at the same time. Robots are not only related to our goals and intentions, as if they were external things that merely confront us from the outside so to speak; they are also *created* by us in concrete techno-scientific practices and they are *used* by us. In this sense, they are human. And this also means that they can never be totally “other”. There is also

sameness. Even robots that are not anthropomorphic have a human aspect, since we created them and use them. They don’t need to be welcomed and brought into “our” sphere; they are already part of it since and as we use them. And since we created them, they are never mere tools but are infused with our ends, our meanings, and our values. They are instruments, but they are *our* instruments. And with their unintended effects, they also shape our goals. Instrument and goal are thus interdependent.

For example, a robot may be given anthropomorphic features in order to function in a particular health care context. The goal has to do with health care, for example providing a way to communicate with patients. But by having robots communicate with patients, the content of this goal is changed: what communication to patients means is changed from one meaning (where humans are the ones who communicate as care givers) to another (care giver-patient communication is something that can also be done by a machine). Furthermore, if we (as patients, as care givers) forget that the robots are created by humans and that we give meaning to them, this can lead to uncritical adoption of the robot and closing possibilities for change. The robot and the interaction should not be taken as given but as changeable. Let me say more about this meaning-giving in interaction.

3.2 The Linguistic and Social Construction of Robots

Second, even if we take the robot as given, that is, as it appears to us and as we interact with it as users (and designers are also a kind of users, albeit a special category of users), we should question again the assumption that there is an external relation between humans and robots. This can be done by considering what I propose to call *epistemological* anthropocentrism. It means that what we say a robot “is”, what we know of the robot, and how we relate to the robot will always depend on human subjectivity, meaning-making, narratives, language, metaphors, etc. Whether or not we can move away from anthropocentrism in a moral or metaphysical sense, taking a critical philosophical approach (after Kant, phenomenology, and much of twentieth century philosophy) means recognizing that we always have to “go through the human” when we talk about, think about, and interact with social robots. What the robot “is”, is not independent of what we say about it and how we respond to the robot in interaction. Humans do not only materially create robots but also (during development, use and interaction) “construct” them by means of language and in social relations, which must be presupposed when we think about these robots and interact with them. This can be revealed by critical philosophical and social-scientific efforts that include a temporal and historical perspective and indeed the role of language and metaphor [32–34]. Consider the role of language: if I give a personal name to a social robot, then

I have already constructed the robot as a companion, other, as having a particular gender, etc. [35]. The robot does not necessarily stop functioning an instrument and may be seen as an instrument at other times or from a different perspective. But the name giving, next to other factors such as the appearance of the robot, co-shapes the meaning of the robot. Furthermore, this meaning is connected to the socio-cultural environment in which the robot is embedded. By giving it a particular name, users may also tap into an entire culture of naming and gendering. For example, giving a female name to a social robot that serves in the household (in response to features of the robot that invites this social behavior) does not only “make” that robot into a kind of person with “whom” I will interact in ways that are accepted in my community and society; it also mirrors, confirms, and proliferates a culture in which women are supposed to take such a role. In this sense, social robots are neither “mere machines” nor totally “other”; they are linked up with human society and human culture in various ways and are made possible by that society and that culture. In this sense, robots are not “external” to humans and their goals but there is an internal relation between humans and robots.

3.3 Relationality and Meaning Making: Robots are Embedded in Cultural Wholes

Third, and related to the previous points: an alternative view should be relational all the way. “Relational” must be understood here in the following senses. An alternative position should acknowledge the relationality of robots as not only embedded in larger technical systems but also larger sociotechnical systems: like other technologies, robots are ‘intertwined with the social practices and systems of meaning of human beings’ [36:195]. They are thus “related” to such practices and systems of meaning. One could also say that they are linked to social relations [32] or that they are embedded in cultural wholes.

In recent work, I have expressed this social-cultural embeddedness of robots by using Wittgenstein’s concepts ‘language games’ and ‘form of life’: like using words, using things such as robots is also embedded in (technology) games and a form of life [33, 37]. I use the term ‘form of life’ not in a biological sense but interpret (by using Wittgenstein) it as culturally defined games and practices that make up a whole or wholes, which give meaning to our activities and technologies. These wholes enable us to make sense of the robots and interact with them, and the robots in turn contribute to, or are part of, our activities of meaning making. We do not only tell things about technology; technologies also actively contribute to the making of meaning. That web or horizon of meaning in turn enables us to make sense of concrete experiences and things. This point can also be made by referring to the term ‘narrative’.

Using Ricoeur’s narrative theory, Wessel Reijers and I [38] have argued that, like text, technologies co-constitute narratives by configuring characters and events into a meaningful whole. For example, a particular human–robot interaction may be placed within a wider narrative of building friendship relations with robots: this can be done by telling stories *about* the robot, but the point is that the robot itself also actively contributes to shaping this narrative. This can be done, for example, through the anthropomorphic shape of a robot, which creates characters and events (e.g. two friends that meet and ask how things go).

More generally, it can be said that technologies have a hermeneutic function as they contribute to the making of meaning (the term “hermeneutic” is used here as referring to meaning and interpretation). This also happens when we interact with robots and through our interactions with robots. They are not only the object of our meaning making and interpretations (e.g., I interpret the robot as a “mere machine”) but also shape how we make sense of the world. For example, a household robot that interacts with humans as a servant might convey the meaning that the social world is one of masters and servants and that the user is a master. Robots are thus not mere instruments if that means that they would stand outside the realm of human culture; instead, they are included in our hermeneutic activities, in a passive way (we talk about them and make sense of them) and in an active way (they contribute to shaping the meaning of our stories).

For the underlying philosophical basis, this relational and hermeneutic approach implies that ontological “is” language (philosophical claims or questions about what something or someone is, e.g. What is a robot?), used by realists and instrumentalists, becomes problematic, especially if this “is” is not understood in a relational way. There is neither a “thing” or “tool” in itself nor an “other” that is unrelated to humans. All technologies, as created and used by humans, are related to humans. Not only so-called “social” robots but *all* robots are already social and relational in this sense, even without or before they have features that are considered “social”. For example, an industrial robot is embedded within language and technology games that belong to a particular social-industrial context, in which human workers have been replaced by machines under capitalist conditions. But with social robots, the social-relational character of robots is even more apparent; they so to speak put on display our social environment and culture. For example, an AI-enabled social robot with voice interface will “talk” to us in ways (and we will talk with it in ways) that are connected with how “we” (in our family, in our community, our country, our language community, etc.) talk. If it succeeds in being humorous, for example, then that is only possible on the basis of tapping into a shared (human) basis of knowledge, experience, and skill. And if it managed to make a joke by means of machine learning, the

robot would have learned to play “our” game. The success of its performance depends entirely on a relational whole, on a form of life, that must be presupposed for the robot to be funny and, more generally, to *make sense*. Social roboticists may have the robot imitate this sense making and meaning making (I will use sense and meaning interchangeably here) by enabling the robot to learn from material that it finds on the world wide web, where text corpora materialize some of the meaning that is around. In artificial intelligence research there is already research on ethically problematic aspects of using language corpora from the internet; for example, there is gender bias in these corpora and in our ordinary language, which tends to be replicated by AI [39]. When AI technology would be increasingly implemented in robots, such problems may proliferate. But considered at the level of the human–robot interaction, which by definition involves humans and hence also connects to games and wider horizons of meaning, the interaction is always already a (real-time) activity of meaning making. Again, in many ways the robot is more than a mere tool or instrument: it is an element in activities and games that make meaning and depend on larger wholes for their meaning. Moreover, the robot is not totally “other” in sense that, to the extent that it is really social, there is a lot of sameness acquired through its relations to human society and human culture. It can only succeed, that is, perform successfully as a social robot, if there is enough sameness in terms of providing a basis for shared meaning. An alien “social” robot would not be regarded as social by us, humans, since it would lack all these relations to human meaning.

3.4 Lack of Hermeneutic Control

Fourth, however, to say that these activities of material making and meaning making are human (are the kind of thing that humans do), is not to say that they are totally under our control. We do not fully control meaning, also not in “purely” human situations (which are seldom “pure” since usually—if not always—mediated by other technologies and things). There is also emergence of meaning when we interact with robots and there are encounters and events. Sometimes meanings emerge which we do not expect; these are unintended emergences of meaning. For example, we may laugh or feel fear when a robot suddenly sounds or looks very human. As research in HRI shows, robots often surprise people. In this sense and in such cases, there *is* an otherness to them—or at least, to the interaction and to the robot-in-relation. Developers and designers of such robots and the users may try to control meaning (for example a designer might try to avoid “uncanny” effects), but hermeneutic control is never absolute. When humans engage with the world and with one another, there is always room for surprises, for meanings we didn’t expect. In that sense, too, robots are never mere machines, if “machines” evokes the meaning of “mechanical” and if that

is interpreted to mean “predictable”. Human–robot interaction, like all interactions, can produce unintended and unexpected meanings. The point is not only that there are unexpected behaviors in HRI, as has been documented (e.g., [40]); humans can and will also make sense of the interaction in unexpected ways. For example, a gesture from the robot that is meant to be friendly might be suddenly interpreted as hostile. Such unexpected interpretations are unavoidable and are part of how we exist in the world. Yet such emergent meanings can also not be conceptualized as constituting *total* otherness, since the emergence of meaning—for example unintended meanings as in the example—is again entirely dependent on human meaning making and meaning experience. We will always be able to compare what the robot does to something that we know from (the rest of) the human world and from our own social environments, to something that is the *same* and that relates to self. It is never entirely different or other.

3.5 Power

Finally, social robotics understood as interaction and as a thoroughly human activity of meaning making, social use, and material construction, is always interwoven with power. Social robots in use and interaction are not just tools or purely “technical” activities but have social and political meanings and effects, and this includes an aspect of power. For example, the language of “slaves” used by Bryson [25] is deeply problematic [41]—as the author has recently recognized. And social robots in a health care context may raise questions about the politics of healthcare without robots: how are things done now and how people treated now, and what is the quality and justification of that way of doing and that way of treating people? Which interests are served by introducing these social robots in that particular context? Power and technology are connected. Naïve instrumentalism misses this point, because it tends to separate technology and the human/the social. By putting social robots safely in the category of “technology”, it unintentionally keeps out “human” questions concerning society and power. (And sometimes instrumentalism is not naïve and does blind us to social relations and power relations *on purpose*.) But social robots are also not “others” or “non-humans” if that means that they are mere partners in pleasant power-neutral postmodern cultural and material play (a direction which, I believe, Gunkel avoids but might be an effect of at least one interpretation of Gunkel or of Haraway-style postmodernism). Taking seriously the political side of Haraway, but also inspired by Marx [42] and Foucault [43], we should point to the danger that posthumanist fantasies about interaction with robots as alterities may mask how the project of social robotics *can* be embedded in, and contributing to, narratives of domination and gender inequality, capitalist socio-economic systems and neoliberal narratives, and micro and meso power games at the

level of human interactions and organizations. For example, robots can be used to collect data which are sold to third parties, their use may take away jobs, they may proliferate problematic ways of treating women, their production may involve exploitation of people, and robots can be used to display the power of a big company or state. If anything, these power-relevant problems show the reality and possibility of a posthumanist dystopia rather than an utopia. Social robotics may well present robots as “others” and “friends”; but behind the curtain (and actually not all that well-hidden), there may be manipulation, exploitation, and disciplining. One could also consider the labor relations which make possible social robots. Celebrations of otherness and alterity and arguments for the inclusion of non-humans in the social may leave out the wider socio-economic context and hide that, as Foucault [44] argues, power is present in all kinds of relations—including relations we have with other people when developing and using social robots. Focusing on the robot as “other” may distract from humans using robots to extract data from humans and to effect other humans’ ‘bodies and souls, thoughts, conduct, and way of being’ [45:18]. For example, social robots may be used for manipulation of people into buying certain goods and for stimulating other behavior, which is in the interest of a party that is not visible in the human robot interaction itself. Behind the anthropomorphic mask of the robot, presented as a playful invitation for power-neutral pleasant and helpful interaction with a quasi-other, hides the serious face of human power and human power relations. Relationality is not just fun and games are not just about play; this perspective on social robotics (and indeed human being) also opens up a Pandora’s box of problems related to power. Social robotics is about games (in a Wittgensteinian sense, see above), and some of them are very serious games indeed: power games. The fact that this is also true for all kinds of other digital tools is not an excuse to look away from the power dimension of these games when it comes to social robotics.

3.6 What the Third View Delivers: Robots as Instruments-in-Relation

Paradoxically, by unpacking the human dimension of robotics in its use and development, this position enables us to be critical of anthropomorphizing robots. Exactly because social robots are deeply related to the human and the social, we are able to point to problematic (or good) anthropomorphizing and explain why it is problematic (or good) by pointing to the large whole in which the robot is embedded, for example a particular game about gender. At the same time, and again somewhat paradoxically, the position avoids a naïve instrumentalist position by taking seriously robots’ role as an instrument of meaning making and an instrument that is part of, and co-constitutes, a larger context: robots

are part of what I have called—in analogy to *context*—a ‘con-technology’ [37: 263]: just as texts are part of a large context that shapes their meaning, (other) technologies are also part of a larger human, meaningful, and social-relational environment in which they are embedded and which shapes their meaning. Robots *are* instruments; but they are instruments-in-relation: they are always connected to humans and the social-cultural fields in which they operate. Naïve instrumentalism suggests a superficial, external relation between humans and their tool. The approach I developed rejects this assumption. My deconstruction of the debate concerning anthropomorphism has relied on a critique of the presupposed external relation between humans and robots. Humans and robots are entangled in the many ways outlined; they are internally and deeply related.

4 Implications for Dealing with the Phenomenon of Anthropomorphizing and for Social Robotics as a Project and Field

What does this approach and “third” view on the relation between humans and robots imply for evaluating anthropomorphism in social robotics? What kind of normative position about these phenomena and practices follows from the previous critical discussion?

4.1 Robots are Neither Others nor Mere Machines

First, on the basis of the alternative view developed here, it seems that we have to reject not only the two initial views (naïve instrumentalism and uncritical posthumanism) but also the normative positions related to them. I suggest that anthropomorphism and anthropomorphizing in social robotics neither be condemned as going beyond what a machine should be (a mere machine) nor be (mis)used as a chance to project otherness onto the machine. In the many senses outlined above, social robots—even those that are anthropomorphic—may be experienced as others but are not Others with capital “O”, as if they mysteriously stood outside the human world. Social robots are human-made and are part of larger efforts of human meaning making. Therefore, they cannot be constructed as *totally* other. There is a lot of sameness there, due to the link between the robot and its developmental context and use con-technology. We may perceive them as other, but we should not forget that they are human-made.¹

¹ I could add that they are also not “other” in the sense of friends or companions, since all we know about human friendship and companionship is that their value lies, among other things, in the friend or companion not being created by us and, indeed, in their otherness. This cannot be created, although some degree of otherness can emerge in the interaction. However, I will not develop this point here.

But neither are social robots mere machines. Anthropomorphizing is one of the things we do as meaning-making humans: even if robots are not designed to invite anthropomorphizing, there will always be “anthropomorphizing” going on, in the sense that the meaning of robots, including social robots, will always be linked to wider *human* practices of meaning making and broader horizons of meaning, a ‘form of life’ (to use the Wittgensteinian concept again). The approach presented does not enable us to decide for or against a specific anthropomorphizing by developers/designers or users. But it *does* help us to understand that any discussion about a particular case will have to evoke meanings from beyond the specific interaction: meanings drawn from our society and our culture, with all its dimensions—including a power dimension. This implies that the pro-anthropomorphism position within instrumentalism is also untenable if and in so far as researchers holding this position are unwilling to consider this wider relationality, meaning, and normativity of the anthropomorphization work they do.

4.2 Re-Defining the Goals of Social Robotics

Moreover, the relational and hermeneutic approach outlined should also be applied to our understanding of social robotics as a field. Social robotics should be understood as an interdisciplinary and multidisciplinary field that is not only concerned with the making of assemblages of material artefacts and software called “robots” (instrumentalist position) or with the creation of “artificial others” (posthumanist position) but also always at the same time with the making of human, social, cultural, and political meanings (and the very claim that the robot is an “other” is one of these meanings that has political significance and that must be problematized).

Thus, while I have indicated how specific views and their related normative positions are problematic, the present paper does not lead to a specific normative conclusion as to whether or not roboticists should build anthropomorphic robots. At most, one could derive a cautionary advice from it: *if* you design such robots, then beware that you are more likely to join this wider activity of social meaning making with all its power aspects, or at least join this wider con-technology to a larger extent. For this reason, some may conclude that it is safer not to develop such robots. Others will persist in creating these robots, but then they have the responsibility to do so in a way that is aware of, and takes into account, the social-cultural meanings and consequences. This leads me to my next point about the responsibility of roboticists.

4.3 The Responsibility of Roboticists, Educators, and Users-Citizens

Normatively, recognizing this relational, hermeneutic, and political dimension of social robots and robotics in general implies that roboticists, computer scientists, engineers, psychologists, and all others involved in the field also carry responsibility for the meanings they generate and the social-relational consequences they produce. This should be seen not as something marginal but as part of the core business of social robotics. For example, both a designer of robot hardware and a developer of speech recognition software that will be used in a robot must understand themselves not only as creators of a technological product that has some intended functions (e.g., being able to have a conversation with a human in a home context) but also as social actors that—via the robot—intervene in a social playing field that includes challenges that are not *merely* technical (e.g., avoiding or diminishing gender bias). Responsibility of social roboticists then means responsibility for making sure that the robot and human–robot interaction achieves its goals, but also for doing this in a way that exercises responsibility *to* other people: it means to respond *to* the users and other stakeholders that may be affected by the unintended consequences of the design, code, etc., including the potential hermeneutic/semiotic impact. It means to respond to social problems next to technical ones, and to figure out how changes on the technical side can contribute to mitigating what happens at the social level.

Defining the precise scope of responsibility for the potential hermeneutic/semiotic and societal impact of robots is not easy. For example, it is impossible to fully predict the future, and it remains unclear to what extent one is responsible for surprising, unexpected intended meanings. As said, there is no such thing as full, absolute hermeneutic control—neither in robotics nor elsewhere in human life and culture. But one should increase the level of hermeneutic control from intended meanings to unintended and potential (future) meanings as much as possible. One could expect from robotics researchers and designers (and others in the companies and organizations they work for) that they at least try to imagine and assess such intended and potential impacts. Researchers from other disciplines, also in the humanities and the social sciences, could help by developing and transferring methodologies for this, for example methods for creating scenarios and methods for ethical assessment.

Given the link with the social level, responsibility in and for social robotics includes taking responsibility for meanings and consequences of social robotics in terms of power. Technical researchers in robotics may not see themselves as having much to do with power, because power is usually understood as being concerned with relations between

political authorities and citizens. A Foucaultian understanding, however, directs us to power issues in all kinds of places and levels of the social fabric. The robotics lab is not excluded from this pervasiveness of power. Together with others who make decisions about robotics and employ or use robots in various contexts, those who develop and design robots have *some* power to shape the meaning of practices, the stories of people, and the goals we set ourselves as humans. For the development of anthropomorphic robots, this power is perhaps more significant than in the case of other types of robots, since the anthropomorphism supports the social aspects of the robots, including power. With this power comes responsibility.

Power is not only negative: it can be restricting and oppressing, but it can also be enabling. Power can take the form of *empowerment*: power can be given to people to do things, it can increase their potential. In the context of social robotics, the question is then how to empower technical professions to exercise responsibility for their creations. Unfortunately, traditionally engineers and scientists lack the full toolbox to do so, since the focus is on technical knowledge and skills. This needs to change. Methodologically, a broader kind of multidisciplinary but also interdisciplinary and transdisciplinary approach is needed. This could mean, for example, that next to computer science languages and engineering tools, instruments from the humanities and the social sciences are also used. The very project of a “social robotics” should be conceived of in a relational way and be linked to the project of critical evaluation of human culture and society. This is needed since, as I have suggested, the lab of the roboticist is connected to wider social and cultural environments, to language, to power, and so on, and the making and use of robots at the same time shapes this entire relational field as much as it is itself shaped by it. If roboticists take up this challenge and responsibility, and re-define their place within this larger whole, then they can play an important role in shaping the future of our societies—something which engineers always did, from bridges and railroads to electricity and mobile phones—in a way that is ethically and socially responsible.

These normative and methodological implications of the approach outlined above, call for transdisciplinarity and for a reform of the education of robotics engineers and computer scientists. When they enter their professional life, they need to be not only willing but also *ready* and *able* to connect their practice and their development and design decisions to a wider societal discussion and cultural discourse, in which they should be able to critically participate by using words and things (robots). But responsibility should not only be ascribed to roboticists. Users-citizens, on their part, also need to take up their responsibility for the future of social robotics and, more generally, the future of technology and the future of society. They need to understand (and be

educated to understand) that the social robots they might encounter in some of the contexts and con-technologies of their daily lives are neither “others” nor mere machines or instruments, but tools that are part of *their* society and culture, for which they also carry responsibility. If we see robots as not opposed to the human sphere but as a crucial part of it, then robots become interesting in ways we may not have imagined before. Connecting people and languages from different worlds through transdisciplinary education may lead to a transformed understanding of what technology can be and do for society and, ultimately, to a profound change of both. This is what students need to be prepared for: students from the sciences and engineering, but also students from the humanities and social sciences.

4.4 General Conclusion

In that meta-sense, then, I suggest that we need *more* anthropomorphizing rather than less. Not necessarily the anthropomorphizing due to design features of the robots and the interactions they enable, which enhances and augments the already unavoidable and existing relations between the robot and its social and cultural environment, and which can therefore be both exciting *and* very problematic (for example because it proliferates particular power relations that we believe are undesirable or unacceptable). This needs scrutiny and discussion by designers and users-citizens, who should carry co-responsibility for the design of such robots. But we need especially more anthropomorphizing of social robots in the sense of more recognition of the very insight that there *is* that link between robots and their hermeneutic environment, that there *is* that deep social and semantic relationality: the understanding—in the lab as well as in wider society—that social robots, like any other technologies, are very human indeed. Next to everything else it does, anthropomorphism in social robotics puts that relationality and that humanness on display: it reminds us that technology has always been already human.

Funding Open access funding provided by University of Vienna.

Declarations

Conflict of interest The author declares that they have no conflict of interest.

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

- Fong T, Nourbakhsh I, Dautenhahn K (2003) A Survey of socially interactive robots. *Robot Auton Syst* 42:143–166
- Dautenhahn K (2007) Socially intelligent robots: dimensions of human-robot interaction. *Philos Trans R Soc Lond B Biol Sci* 362:679–704
- Duffy BR (2003) Anthropomorphism and the social robot. *Robot Auton Syst* 42:177–190
- Ishiguro H (2006) Interactive humanoids and androids as ideal interfaces for humans. In *Proceedings of the 11th international conference on Intelligent user interfaces (IUI '06)*. Association for Computing Machinery, New York, NY, USA, 2–9
- Bartneck C, Kulic D, Croft E (2008) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int J Soc Robot* 1:71–81
- Zlotowski J, Proudfoot D, Yogeewaran K et al (2015) Anthropomorphism: opportunities and challenges in human-robot interaction. *Int J of Soc Robotics* 7:347–360. <https://doi.org/10.1007/s12369-014-0267-6>
- Kiesler S, Hinds P (2004) Introduction to this special issue on human-robot interaction. *Hum-Comput Interact* 19(1):1–8
- Reeves B, Nass C (1996) *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, Cambridge
- Kiesler S, Powers A, Fussell SR, Torrey C (2008) Anthropomorphic interactions with a robot and robot-like agent. *Soc Cogn* 26(2):169–181
- Hegel F, Gieselmann S, Peters A, Holthaus P, Wrede B (2011) Towards a typology of meaningful signals and cues in social robotics. In: *Proceedings of the IEEE international workshop on robot and human interactive communication*, pp 72–78
- Kuchenbrandt D, Eyssel F, Bobinger S, Neufeld M (2013) When a robot's group membership matters. *Int J Soc Robot* 5(3):409–417
- Krägeloh CU, Bharatharaj J, Kutty SKS, Irmala PR, Huang L (2019) Questionnaires to measure acceptability of social robots: a critical review. *Robotics* 8(4):88. <https://doi.org/10.3390/robotics8040088>
- Peca A, Coeckelbergh M, Simut R, Costescu C, Pintea S, David D, Vanderborght B (2016) Robot enhanced therapy for children with autism disorders: measuring ethical acceptability. *IEEE Technol Soc Mag* 35(2):54–66
- Eyssel F (2017) An experimental psychological perspective on social robotics. *Robot Autonom Syst* 87:363–371
- Darling K (2017) Who's Johnny? Anthropological Framing in Human-Robot Interaction, Integration, and Policy. In: Lin P, Jenkins R, Abney K (eds) *Robot Ethics 2.0*. Oxford University Press, New York
- Turkle S (2011) *Alone together: why we expect more from technology and less from each other*. Basic Books, New York
- Scheutz M (2012) The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey G (eds) *Robot ethics*. MIT Press, Cambridge, MA, pp 205–222
- Mori M (1970) The uncanny valley. *Energy* 7:33–35
- MacDorman KF, Ishiguro H (2006) The uncanny advantage of using androids in cognitive and social science research. *Interact Stud* 7:297–337
- Freud S (2003) *The uncanny*. Penguin Classics, London
- Richardson K (2018) Technological animism. In: Swancutt K, Mazard M (eds) *Animism beyond the soul: ontology, reflexivity, and the making of anthropological knowledge*. Berghahn, New York, pp 110–128
- Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: *Proceedings of the 12th IEEE international workshop on robot and human interactive communication (ROMAN 2003)*, 2003, pp 55–60
- Coeckelbergh M (2018) How to describe and evaluate “deception” phenomena: recasting the metaphysics, ethics, and politics of ICTs in terms of magic and performance and taking a relational and narrative turn. *Ethics Inf Technol* 20(2):71–85
- Flusser V (1999). *Shape of things: a philosophy of design*. Reaktion Books, London
- Bryson J (2010) Robots should be slaves. In: Wilks Y (ed) *Close engagement with artificial companions*. John Benjamins, Amsterdam, pp 63–74
- Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. *Minds Mach* 16:141–161
- Haraway D (2000) A cyborg manifesto. In: David B, Kennedy BM (eds) *The cybercultures reader*. Routledge, London, pp 291–324
- Latour B (1993) *We have never been modern*. Harvard University Press, Cambridge, MA, Catherine Porter (trans)
- Gunkel DJ (2007) *Thinking otherwise: philosophy, communication*. Purdue University Press, West Lafayette, Technology
- Gunkel DJ (2017) The other question: can and should robots have rights? *Ethics Inf Technol* 20:87–99
- Ihde D (1990) *Technology and the lifeworld*. Indiana University Press, Bloomington
- Coeckelbergh M (2012) *Growing moral relations: critique of moral status ascription*. Palgrave Macmillan, New York
- Coeckelbergh M (2017) *Using words and things: language and philosophy of technology*. Routledge, New York
- Coeckelbergh M (2019) *Moved by machines: performance metaphors and philosophy of technology*. Routledge, New York
- Coeckelbergh M (2011) ‘You, robot: on the linguistic construction of artificial others’ in: *AI & Society* 26(1): 61–69
- Johnson DG (2006) Computer systems: moral entities but not moral agents’. *Ethics Inf Technol* 8:195–204
- Coeckelbergh M (2018) Technology games: using wittgenstein for understanding and evaluating technology. *Sci Eng Ethics* 245:1503–1519
- Coeckelbergh M, Reijers W (2016) Narrative technologies: a philosophical investigation of narrative capacities of technologies by using Ricoeur's narrative theory. *Hum Stud* 39:325–346
- Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186
- Lemaignan S, Fink J, Mondada F, P Dillenbourg (2015) You're doing it wrong! studying unexpected behaviors in child-robot interaction. In Tapus A., André E., Martin JC., Ferland F., Ammi M. (eds) *Social Robotics. ICSR 2015. Lecture Notes in Computer Science*, vol 9388. Springer, Cham
- Coeckelbergh M (2015) The tragedy of the master: automation, vulnerability, and distance. *Ethics Inf Technol* 173:219–229
- Marx K (1867/1976) *Capital (Vol. I)*. Penguin Books, London (reprint in 1990).
- Foucault M (1980) *Power/knowledge: selected interviews and other writings 1972—1977*. Colin Gordon ed. Pantheon Books, New York

44. Foucault M (1980) Power/knowledge. Gordon C (ed) Gordon C, Marshall L (trans). Pantheon Books, New York
45. Foucault M (1988) Technologies of the self: a seminar with michel foucault. Martin LH, Gutman H, Hutton PH (eds). The University of Massachusetts Press, Amherst

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