How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance

Mark Coeckelbergh[1]

Accepted: 30 September 2020 / Published online: 23 October 2020

Abstract
Social robots are designed to facilitate interaction with humans through "social" behavior. As literature in the field of human–robot interaction shows, this sometimes leads to "bad" behavior towards the robot or "abuse" of the robot. Virtue ethics offers a helpful way to capture the intuition that although nobody is harmed when a robot is "mistreated", there is still something wrong with this kind of behavior: it damages the moral character of the person engaging in that behavior, especially when it is habitual. However, one of the limitations of current applications of virtue ethics to robots and technology is its focus on the individual and individual behavior and insufficient attention to temporal and bodily aspects of virtue. After positioning its project in relation to the work of Shannon Vallor and Robert Sparrow, the present paper explores what it would mean to interpret and apply virtue ethics in a more social and relational way and a way that takes into account the link between virtue and the body. In particular, it proposes (1) to use the notion of *practice* as a way to conceptualize how the individual behavior, the virtue of the person, and the technology in question are related to their wider social-practical context and history, and (2) to use the notions of *habit* and *performance* conceptualize the incorporation and performance of virtue. This involves use of the work of MacIntyre, but revised by drawing on Bourdieu's notion of habit in order to highlight the temporal, embodiment, and performative aspect of virtue. The paper then shows what this means for thinking about the moral standing of social robots, for example for the ethics of sex robots and for evaluating abusive behaviors such as kicking robots. The paper concludes that this approach does not only give us a better account of what happens when people behave "badly" towards social robots, but also suggests a more comprehensive virtue ethics of technology that is fully relational, performance-oriented, and able to not only acknowledges but also theorize the temporal and bodily dimension of virtue.

Keywords Moral standing · Social robots · Virtue ethics · Practice · Habit · Embodiment · Incorporation · MacIntyre · Bourdieu · Performance

1 Introduction

Social robots are designed to facilitate "social" interaction between humans and robots. Sometimes this leads to behavior towards robots that is seen by some people as ethically problematic or even bad or evil. For example, when employees of the robotics company Boston Dynamics kicked a robot to test how stable it is, some people felt uncomfortable or said that it seems wrong to do so.[1] Similarly, when a sex robot named "Samantha" got vandalized at an electronics festival, the engineer behind the robot called the molesters "barbarians."[2] And when "serious abusive behaviours" were documented for children interacting with a robot in a Japanese shopping mall,[3] this also seems problematic.

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

[1] Universitat Wien, Vienna, Austria

[1] https://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/index.html.

[2] https://www.huffpost.com/entry/samantha-sex-robot-molested_n_59cec9f9e4b06791bb10a268.

[3] https://www.vice.com/en_us/article/ezv3ae/when-humans-bully-robots-there-will-be-consequences.

Many observers may not share the strong emotional reactions that some people have when confronted with these phenomena. But how can we justify the intuition that there is at least *something* wrong with these "abuses"? If we look at the properties of the robot, there seems nothing wrong at all, morally speaking, as far as the robot goes. Property may have been damaged, but there is nothing wrong with regard to the robot as "moral patient", with regard to what is due to the robot. The robot does not feel pain, is not conscious, and does not display any other properties that we usually think of as being sufficient for moral standing.

The only way to makes sense of these moral responses and to potentially justify them, then, is to argue by drawing on a form of indirect moral standing: the robot does not have direct moral standing based on intrinsic properties, but indirectly what is done to it matters because of something about the human person. The present paper discusses what it means to use a virtue ethics approach for conceptualizing this, and offers a specific interpretation of this approach in order to account for the relational and bodily-performative aspects of the human–robot interaction and the virtue or vice that may be connected with it.

After positioning the paper in relation to work in philosophy of robotics and HRI, the original contribution of this paper consists of two parts. First, the paper draws on MacIntyre to introduce the notion of a practice and its connection to the training of virtue. It argues that when evaluating what is happening in these cases of "robot abuse", we should not only look at the individual behavior of humans with regard to a particular technological artefact, but consider the entire practice, in which virtue and vice can grow. This is a more social way to understand virtue and vice that can be connected to the use of technologies. Second, it introduces Bourdieu's conception of habit in order to emphasize the temporal and bodily dimension of virtue and vice: they develop in time and as habits are incorporated, that is, embedded in the person as (among other things) a moving body. This is further theorized by introducing the notion of performance and, in the end, connecting to a Confucian notion of becoming. The paper shows what this approach means for the cases and phenomena of robot abuse in order to show the value of this approach to moral standing of robots, and concludes that, more generally, a further development of this approach can give us a better, more comprehensive virtue ethics of technology than we have so far and can be of interest in fields of applied philosophy beyond robot ethics.

In the course of the paper I will refer to, and critically respond to, the work of Vallor and Sparrow, since they offer (a) a view that already goes some way in a relational direction [34] and (b) elaborate arguments about what virtue ethics means in relation to robots [32, 33]. However, the original contribution of this paper does not depend on my assessment of their work and is predominantly geared towards enriching virtue ethics of robotics with notions drawn from MacIntyre, Bourdieu, and my recent work on performance, thus supporting the project of a more relational and less dualistic technology ethics – a project I expect many philosophers of robotics will sympathize with. In addition, I hope that people in the HRI and social robotics communities may find these concepts useful for their own work. The paper may appeal especially to those researchers who already have an interest in the social-cultural context of social robotics and/or link their work to embodied cognition.

2 Are Social Robots Kantian Dogs? Using Virtue Ethics for Thinking About Moral Standing

The past decade has seen a growing philosophical discussion about the moral standing of robots. For example, Bryson [6] has argued that robots are property and that they have no moral standing, whereas Gunkel [22] has argued that we consider the question regarding robot rights (for a recent overview of the discussion, see for example [7]). There is also work in the field of human–robot interaction (HRI) that acknowledges that there might be ethical problems with the way people treat robots, often framed as problems concerning anthropomorphization (e.g., [2, 17, 20, 31]). For example, Darling [18] has conducted a lab experiment that indicates how framing robots by using anthropomorphic language can impact how people treat robots, and Ku et al. [27] have designed a tortoise-like robot which is meant to restrain abusive behavior by children.

But whereas most arguments in philosophy and HRI concern the moral standing of the *robot* based on its properties, sometimes an argument is made for what we may call the "indirect" moral standing of robots. Consider this argument mentioned by Darling [18], which Kant made regarding the abuse of dogs. Formulating the issue from the point of view of the human moral agent, Kant famously argued that we have only 'indirect' duties towards dogs because of the potential implications for the character of the humans involved and, more generally, for cruelty among humans.

> So if a man has his dog shot, because it can no longer earn a living for him, he is by no means in breach of any duty to the dog, since the latter is incapable of judgment, but he thereby damages the kindly and humane qualities in himself, which he ought to exercise in virtue of his duties to mankind … for a person who already displays such cruelty to animals is also no less hardened towards men. [26]

Thus, here one could say that the dog has moral standing, but only indirectly, since its standing only derives from the moral standing of humans, who ought to exercise their duties.

Similarly, Darling suggested, if we treat robots in inhumane ways, we risk to become inhumane ourselves.

This argument can be interpreted as a claim about duties (i.e., using Kant's term indirect duties) or about consequences, in particular about the future behavior of the human agent (present behavior causes cruelty towards humans in the future), but it can also be formulated from a virtue ethics perspective.

Virtue ethics is one of the main normative theories in ethics. It focuses on the character of persons, usually framed in terms of an agent's disposition. Often this is in turn interpreted in terms of habit and the good life, but some moral philosophers may disagree and in general there are different views on what these terms mean and different approaches to virtue ethics (for an overview see, for example, [23]). In any case, applying virtue ethics means that the problem can be formulated as concerning virtue: it is bad to "abuse" robots since it is not virtuous to do so. The "abuse" of robots is not bad because of the robot, but because it is vicious.

Virtue ethics becomes increasingly popular in philosophy of technology. Vallor [34] has argued that living well with technologies requires a theory of the good life and has found such a theory in virtue ethics—in the Western tradition (Aristotle) and beyond. She has helpfully identified some what she calls 'technomoral virtues': 'traits that twenty-first century humans must cultivate more successfully if we want to be able to live will with emerging technologies.' [34] This approach can and has been applied to social robots. Vallor herself has discussed the virtue of care in relation to robots used in relations of human dependence (220–229). Furthermore, in robot ethics, Sparrow [32, 33] has offered sophisticated arguments about what it means to take a virtue approach to robotics. Cappuccio et al. [7] have argued for giving social robots moral consideration that is not based on intrinsic moral dignity and that does not attribute rights, but on the idea that relationships with robots offer to human agents opportunities to cultivate vices and virtues. And in this journal Coghlan et al. [16] have reported that social robots, for example robots in the shape of animals, may affect virtue in terms of effects on the moral development of children and responses to nonhuman animals. Earlier I have discussed virtue and the good life in relation to care robots [12] and environmental technologies [10]. In general, there has been a continuing interest in technology and the good life among some philosophers of technology (e.g., [5]).

In the discussion about moral standing, I [9] have argued that virtue ethics avoids some problems with what he calls the 'properties' approach to moral standing: since it shifts the focus to the subject of moral consideration, we for instance no longer need to know whether or not a robot (or any other entity for that matter) has particular properties, or how we know that these properties warrant a particular moral standing. I have drawn attention to virtue ethics as one way to approach the question whether it is wrong to kick a robot [13]. Picking up the same example, Sparrow [33] has also argued that even if an agent's "cruel" treatment of a robot has no implications for their future behavior towards people or animals, it reveals something about their character and this gives us a reason to criticize their actions. Leaving aside Sparrow's next interesting but for my purposes not so relevant argument that this works for vice but not for virtue (he claims that *good* treatment does *not* reflect on one's character), this argument has the same structure as the Kantian one but now formulated in terms of virtue and vice: bad behavior towards robots is not bad because of the robot, but because it does something with someone's character. It is vicious. This damage to the moral character of the person, not "harm" to the robot, makes the action wrong. Other authors have made similar claims with regard to issues with other technologies such as violence in video games (e.g., [8, 30]).

A virtue ethics approach thus offers an interesting way to support the intuition that there is something wrong when people "abuse" robots, without having to accept that robots have moral standing on the basis of their intrinsic properties. While it may well be inevitable that some people will ascribe virtue also to the *robot* based on its appearance [12], what counts according to the virtue ethics approach to moral standing is not the robot but the human person and the moral character of that person.

However, there are at least the following methodological problems with this application of virtue ethics. First, in line with the Western English-speaking philosophical tradition in general, existing accounts of virtue and technology are focused on the individual—here: individual virtue or vice, the individual moral character. It is an ethics concerning the individual user of technology (and in addition on the individual robot, as I have argued previously.) This focus on the agent is not necessarily problematic if that agent is understood in a relational way, that is, as a person related to others. But often this wider field of relations remains out of sight in accounts of virtue ethics. In the current virtue ethics of robotics literature, such relational elements are certainly present, but much more work is needed to develop this dimension of virtue and its relation to technology. For example, while Vallor [34] defines human beings as relational beings 'whose identity is formed through a network of relationships' [34, 16] and discusses care practices (226), a systematic account of the role of practice viz-à-viz technomoral virtues and of the link between virtue and practices are missing. The latter can be found for instance in the work of MacIntyre (see below), but this corpus is not engaged with. And while Sparrow [33] does not deny that there is a social dimension of human–robot interaction, his account of virtue is all about individual agents and their virtue. Again this, by itself, is not necessarily problematic. It even has an advantage: by focusing on agents and their virtue (rather than empirical claims about future behavior), Sparrow successfully avoids behavioristic directions and arguments based on the empirical effects. Like other so-called agent-based treatments of virtue, the center of attention is the character of the agent, and this creates room for a virtue ethics-based argument as opposed to a mere consequentialist one. But it remains unclear how the agent and her virtue are related to the agent's social environment. Is virtue really only based in the agent? While already in earlier work Sparrow recognizes the social character of meaning [32], more needs to be said about that social meaning of human–robot interaction and indeed about the social dimension of virtue.

Second, while Vallor includes in her account what she calls 'moral habituation' [34] and Sparrow's distinction between his agent-based account and an interpretations of virtue in terms of future behavioral consequences implicitly acknowledges the time dimension, in general more explicit attention needs to be given to the temporal dimension of virtue and to how moral habituation works. If it is true that virtue is something that is acquired over time, as Vallor acknowledges, then what are the implications for our thinking about how to behave towards robots? It seems that a virtue ethics approach to moral standing should not only be focused on a specific time when the "bad" behavior towards the robot occurs (now or in the future), but should also take into consideration how this behavior started and developed and in which context this happened, and how it could be changed.

Third, in contrast to the postphenomenological tradition and the care ethics tradition to which Vallor responds, which pay attention to embodiment, the bodily dimension of *virtue* is not explicitly thematized. And while Sparrow [32] unsurprisingly mentions the body in his discussion of robot rape, he does not integrate the theme of body and embodiment in his account of virtue. This is a lacuna, also in the debate concerning moral standing, since the individual action of the robot (ab)user is not only engaged in abstract "behavior", but also has a body and hence the (ab)use involves a particular bodily comportment and bodily movement, which is also learned and embedded in a particular context. However, like in the Western philosophical tradition in general, this is not seen as central to virtue.

How can we fill these gaps and conceptualize these aspects of the "robot abuse" problem, and, more generally, of virtue ethics as applied to the use of technologies? After indicating some other potential resources, this paper focuses on MacIntyre and Bourdieu. First I will use MacIntyre's notion of practice to contextualize the use and abuse of robots and the related vice and virtue. Then I will draw on Bourdieu and my recent work on performance and technology to frame the temporal and bodily dimension of what happens here in terms of use, action and virtue/vice.

3 A More Relational Interpretation and Application of Virtue Theory: Using MacIntyre to Frame (Ab)use of Robots as Embedded in Practices

Let me start with giving a more relational twist to the virtue ethics approach to moral standing. In the discussion about the moral standing of robots relational approaches have been proposed [11, 22], which also fit with arguments for contextualizing technology more generally, for example in socio-technological systems [25] or in a game or performative context [14, 15]. However, these proposals do not directly draw on, and often do not engage very much with, the tradition of virtue ethics.

Within the virtue ethics tradition, which in its Western version mainly responds to the work of Aristotle, we could reformulate the challenge for a virtue ethics approach to moral standing (and to technology) briefly as follows: we need a less modern-individualist version of Aristotle, which puts virtue ethics in a social context. Now such a view is provided by MacIntyre, who in *After Virtue* (2007) offers a concept that is not thematized by Vallor but that is very helpful for elaborating the social-relational dimension of virtue ethics applied to technology: practice. According to MacIntyre, the virtue of the person is always related to a practice. One could reformulate this as saying that virtue is embedded in a social-practical environment. Let us look into the details.

MacIntyre's account of virtue, and more generally human good, is an inherently social one: ethics is something that is learned in a practice. But this term "practice" does not just refer to "practical as opposed to theoretical" but also and especially to a social-cultural and communal context. Virtue is not something that isolated moral agents have or do but something that is embedded in a social context, in which we acquire the virtues. Moreover, MacIntyre defines a practice as directed towards human excellence:

> By a "practice" I am going to mean any coherent and complex form of socially established cooperative human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence, which are appropriate to, and partially definitive of, that form of activity, with the result that human powers to achieve excellence, and human conceptions of the ends and goods involved, are systematically extended. [29]

To fully understand this definition, one needs to further discuss the relation between MacIntyre and Aristotle's view. But what interests us here is the 'form of socially established cooperative human activity' that is a condition for realizing the standards of excellence and, ultimately, the virtues that MacIntyre talks about. We can only achieve virtue (or excellence, or *eudaimonia*) by living our lives, and we do that in the context of a practice and a community. We learn virtues from others and by practicing it with others. These others are not only present ones, but also practitioners from the past, to whom we must relate. (194) A practice is related to a tradition. Seeking to achieve the good life, we are socially and historically situated. In this sense, virtue is not a mere individual matter but is always at the same time a social project, or at least one that depends on our relations with others.

Applied to the discussion about moral standing, this means that when we use virtue ethics to respond to phenomena and cases such as kicking a robot or "abusing" a sex robot, our ethical attention should not be limited to establishing the relation between (a) the individual behavior of the user towards the robot and (b) virtue or vice related to that user, but also on how both (a) and (b) are related to the social context and history of that behavior and (potential) vice, that is, how they are related to the *practice* in which they are embedded and in which they have grown. Technological action and interaction are not only a matter of individual use; both the (ab)user and the technology are part of a wider social context. And as a practical context, it has a history. Moreover, virtue and vice become also "socialized" in the sense that they are no longer a question of individual character alone but a feature of an entire practice and the history of that practice. Virtue may well be agent-based, then, but it is always related to, and embedded in, that wider social practical context. In that sense, virtue is also based in a practice as a whole.

For a virtue ethics evaluation of the relevant cases, then, it is important to take into account those practices and histories. For example, the "abuse" of a sex robot may be related to structural risks of abuse within human–human relationship within a particular practice, for example prostitution or sometimes even day-to-day relationships between men and women if these relationships are embedded in specific structures and histories of inequality and oppression, which may tolerate some forms of abuse (e.g., rape within marriage). And when a child kicks a robot, that behavior might have become part of a practice in the sense that there is a history of violent behavior towards robots, animals, other children, and so on, which has been learned in a particular social context (home, school, football club, etc.) and which may be an ethical problem of that practice, context, and history, next to being a problem of that particular individual. Virtue and vice here are then not just a feature of the individual character of a particular person; there is also virtue and vice, moral excellence and the absence of that, in the practice as a whole, which is reflected and realized at individual (inter)active level and in relation to for example robots. Taking a relational turn, we are invited to examine the practices in which these phenomena take place or to which they are related, for instance the relationships between men and women and the relevant educational practices in various contexts.

Thus, if we take seriously technology as a practice, rather than an object to which we "behave", and if we see virtue and vice as being not just about individual character but also as embedded in a practice, then the question regarding the moral standing of the robot gets a more relational dimension. This entails a number of shifts or revisions of the initial virtue ethics approach to moral standing. First we shifted the ethical attention from the potential moral patient or object (the robot) to the moral agent or subject (the human user or abuser). Then we performed a second shift: one from the individual moral agent or subject to the practice as a whole, in which moral excellence may grow and flourish or not. Finally, there also has been a third shift: from the moment of (ab)use to the wider temporal horizon: the history of the practice—and hence the history of abuse.

Let me further elaborate the latter and propose a fourth revision, which concerns the embodiment and performance of the (ab)user.

4 The Incorporation of Virtue Via Habits and the Performance of Virtue: Using Bourdieu to Conceptualize the Temporal and Embodiment Dimension of the (Ab)use of Robots

The initial cases were described in terms that presuppose an agent or person which has a character(Vallor) or a disposition and behavior (Sparrow) and which then (ab)uses technology. What is left out or at least significantly undertheorized in both accounts is (a) how virtue is related to the embodiment and moving body of the one who (mis)behaves towards the robot and, ultimately and more precisely, (b) the incorporation of virtue or vice and its literal movement in and through the person—that is, the performance of virtue or vice.

Since this is relatively new or at least very recent terrain even within the virtue ethics tradition as a whole (for example, MacIntyre of course acknowledges that we have an animal nature—see [28]—but a full account of the relation between embodiment and virtue is still lacking), I propose to look for sources in other places. In philosophy of technology, obvious places to go to are the phenomenological (Husserl, Heidegger and especially Merleau-Ponty) or postphenomenological [24] tradition, which discuss embodiment in relation to technology. More recently, I [15] have proposed to use performance metaphors in order to conceptualize our relation to technology, which, among other things, helps to introduce the moving body into the field. In cognitive science, HCI, and design, there has been attention to performing an act with technology ([36], influenced by Heidegger and others), and of course to embodied interaction (e.g., [19, 35]). The latter approach is influential in contemporary social robotics, whereas usually the performative aspects of human–robot interaction remain unaddressed or undertheorized. However, here I want to start with the work of Bourdieu on habit [3, 4], which stays closer to the virtue ethics tradition and helps to highlight both the temporal and embodiment dimensions of virtue/vice in relation to technological practice. Then I will further develop this by drawing on my use of the performance metaphor.

As mentioned, Vallor already touches upon what she calls 'moral habituation'. Starting with Aristotle's term *hexis*, which is often translated as habitus/habit, she explains that habits arise from the repetition of a pattern of action, are done for motivating reasons, create expectations, and shape cognitive and emotional states [34]. This is helpful, as it suggests that virtue has a temporal dimension. But what is missing is an account of the bodily dimension of *hexis* and a more precise explanation of how this concept of *hexis* helps us to connect the agent or subject of virtue to the social environment.

This is offered by Bourdieu's conception of habit. Habit is of course a temporal matter, as explained by Vallor. But with Bourdieu it gets a social *and* bodily dimension. First, as a sociologist, Bourdieu was specifically interested in conceptualizing the relation between individuals and their social environment. The concept of habitus offers this: it captures how the social order becomes habit. For example, social class becomes a habit of taste [3]. But it is not only the mind that is shaped by society and culture; the body is also shaped by the social-cultural environment. Habitus or *hexis* means that the social structure is *incorporated*: feelings, skills, and ways of bodily comportment become embodied in and through habitus. In *The Logic of Practice* (1990) Bourdieu explains that this happens without conscious aiming. Social organization is thus enabled, but in a way that is not so explicit. There is a collective orchestration without a conductor (53). Whereas Vallor—in line with Aristotle and much of Western modern thinking—stresses reasons, motivation, and 'states' of the mind, Bourdieu points to the dimension of implicit know-how and behavior without aiming. We are regulated without obedience to rules [4].

Bourdieu uses a music performance metaphor here. But to put more emphasis on the moving body, we can also—in line with Coeckelbergh [15]—use a dance metaphor: we are socially and culturally choreographed via habit. Habituation implants "dispositions" in our mind, if you wish, but it also makes us disposed to move and comport in certain ways rather than other ways. Habituation is not only about acquiring a way of thinking; it is also a way of moving. And this way of moving is socially orchestrated (Bourdieu) or choreographed (Coeckelbergh). Habit is performed, and by using this performance metaphor [15] we can highlight (a) its bodily and social dimension and (b) ask the question who or what shapes that performance. Who or what makes us habitually move in certain ways rather than others (e.g., abusive comportment and movements)?

This leads us to the question concerning power. With Foucault we can highlight the power dimension of habituation. Foucault did not use the concept of habit but that of disciplining and (in his last lectures) 'technologies of the self', by which he meant that the social order affects 'bodies and souls, thoughts, conduct, and way of being' [21] The technologies of the self are thus also technologies of the body, which is also shaped by the social order. To put it in terms of the dance metaphor: the social conductor or choreographer exercises power, not only over our minds but also over our (moving) bodies. In the case of robot (ab)use, what we do to the robot is socially choreographed. The (ab)user might exercise power over the robot, but the way she does that is regulated and shaped by the social environment, where she finds herself in a web of power relations that discipline and shape her subjectivity, self, and body. Foucault teaches us that usually there is not one choreographer who exercises power in an obvious and visible way (say, an authoritarian figure who forces particular habits upon us); instead, there are many subtle and invisible power relations, which we are usually unaware of. If we consider only the individual mental disposition or behavior (in the present or in the future) or even only the human–robot relation as a power relation, we miss this wider field of social power and how it shapes us and our incorporated habits—and thus the human–robot relation. More work is needed to reveal these social power environment(s) of human–robot interaction. For this purpose, we do not only need psychology or philosophy that psychologizes the technology user and her habits; we also need the social sciences and a philosophical framework that theorizes the social and performative [15] dimension of what we do with and to technology.

With regard to virtue, then, we can conclude that the acquisition of virtue understood as habituation must be seen as a temporal process that involves both explicit aiming and implicit organization, and that is not the sole project of the individual but also happens to us as social and cultural beings, whose habitual nature is firmly entangled with our social nature. Moreover, the training and acquisition of virtue through habituation is also a bodily and performative affair: it involves bodily comportment, ways of moving, and so on. For example, exercising the virtue of "care" is not only a matter of having certain habits in terms of thinking and action or behavior, understood in an abstract way; it is only a virtue if it is acquired in a practical-social context and if it is translated or (better) takes the form of concrete bodily performances. For example, taking care of a sick person can only be virtuous if it is embedded in standards of excellence that exist within a care practice and tradition and if it is *performed* in bodily comportment and movement towards the patient—performances which may also involve the exercise of power of some people over others, intended or not.

For a virtue ethics approach to moral standing, this understanding of virtue means that the "abuse" of robots must be seen as not just a matter of individual ethics but a social problem—having to do with how our society organizes us via habituation and exercises power over us—and as a matter of lacking virtuous bodily performances or involving vicious bodily performances. Where and when a robot is "abused", virtuous habituation has been lacking and/or a vicious habituation has taken place in a particular *praxis*. Vice is an individual moral failure and it may well be based in agents and perhaps individuals, but it is also at the same time a failure of the social environment to organize the exercise of virtue. And it is not just a vicious "mental" disposition but is also incorporated in the person's body and its movements and comportment. The kicking is incorporated. And it is choreographed. Habituation, understood as incorporation and performance, is socialized. This can render it difficult for someone to become less vicious and more virtuous, but at the same time it also means that social environments can support changes of habit and the learning of virtue (and the unlearning of vice). It is important to recognize that the robot "abuser" need not stay vicious and that there is the possibility of change. The social order renders this change both difficult and possible.

5 Further Discussion

Keeping in mind Sparrow's [33] distinction between agent-based versions and behaviouristic version of the virtue argument, it is important, however, to understand this argument about the link between virtue and performance/incorporation (which is socially choreographed) in a way that distinguishes it from empirical claims about future behavior. To understand moral virtue and character as being incorporated and performed does not imply that, in order to condemn a particular behavior towards the robot, one needs to make a claim about future behavior towards humans (the behavioristic, effects interpretation of the Kantian argument Sparrow wants to avoid). Instead the focus is on revising our understanding of what happens in the present by putting that present act in the context of a history and future of habituation. The focus remains on the character and disposition of the agent, at least if this is contrasted with a consequentialist argument about the causation of cruel habits. The "only" modification is to understand this character and disposition in a relational way and as not a matter of mental cognition alone but also as embodied, performed, socially shaped, and requiring temporal processes of habituation. This performativity enables persons to potentially change their character *and*, since the virtue or vice is performatively incorporated, expressed, and constituted, this enables us, as third-person evaluators, to say something about the person's virtue or vice (e.g., call her "cruel") and to meaningful interpret and respond to the "abuse" in the first place. Meaning is indeed social, but like virtue, vice, and habituation it is also embodied and performed.

This does not mean that virtue or vice are *all* about bodily performance; there is more going on in human–robot interaction and human (inter)action in general. Performance and virtue-as-performance is never "just" bodily but involves body/mind. I do not use the concept of performance as a way to draw attention to the body as *opposed* to the mind, but rather to overcome such dualistic understandings of virtue/vice, habituation, and human–robot interaction in general. To put it using a more familiar term: to say that cognition is embodied, does not imply that everything that is going on is merely bodily. The same is true for virtue and vice understood as performance. Moreover, it could be that a particular virtue or vice is expressed in, constituted by, and incorporated via, more than one performance. But we always perform. Moreover, there is no such thing as "private virtue" or "private vice", just as there is no such thing as a "private performance" (in analogy to Wittgenstein's term "private language"). As Vallor and Sparrow would be ready to acknowledge (but do not theorize) and as I argued in the previous pages, both the person's moral character and our evaluation of it are linked to, and shaped by, the social environment. The "cruelty" and the vice as well as our reaction to it are both performances that are choreographed by our social environment, which employs methods of organization and technologies in order to make us more virtuous or more virtuous. Again, this is not done by one, authoritarian conductor or choreographer; the social habituation of virtue or vice usually happens in implicit and invisible ways.

Furthermore, this social and bodily embeddedness of virtue and vice, understood via a more comprehensive understanding of the concept of habit and habituation, does not relieve people from the exercise of virtue and responsibility. Nor does it remove the individual agent from the picture. Instead, it puts the individual moral agent in a relational context and understands virtue as incorporated and performed, instead of a merely "mental" disposition. It therefore explains how difficult it is to change vice into virtue, and at the same time offers ways to do this: individual persons need to be supported by their social environment and pay attention to their bodily movements, next to other things. The road towards virtue is not just about changing one's "mentality" but also about changing one's "performativity". By making us aware of this social-performative context, the proposed does not seek to abolish personal responsibility but rather makes possible its exercise in a way that is more aware.

Finally, this way of thinking about virtue does not only respond to the Western tradition (it suggests a revision of

Aristotle) but also resonates with Confucian thinking about persons, ritual, and (what in the Western tradition might be called) virtue—a tradition which is also helpfully drawn upon by Vallor in her book and which is increasingly included in robot ethics discussions. One way to make a link between Confucianism and what I have been saying here is to consider the concept *ren*. Often rendered as 'benevolence' or 'goodness', Ames and Rosemont [1] translate the term in a radically relational way. If human being is relational and a matter of becoming, then *ren* is not a thing but a process of growing relationships. Moreover, they argue against psychologizing the term: 'Ren is not only mental, but physical as well: one's posture and comportment, gestures and bodily communication.' (49) Similarly, according to the approach I articulated, virtue or vice should not be reified or psychologized but be understood in a relational, embodied, and "becoming" way. Virtue and habitus are not only a matter of mental dispositions but also of posture, comportment, gesture, and performative communication. And as socially shaped processes of habituation, they are not only ways of being and ways of *becoming*.

More work is needed to further develop this point and, more generally, to connect (relational approaches in) virtue ethics to other, non-Western philosophical traditions. It would also be interesting to compare notions of becoming in Asian philosophy to notions of becoming in Western process philosophy, and then explore what this means for robot ethics. But this is beyond the scope of this paper.

6 Conclusion: Towards a More Relational and Comprehensive Virtue Ethics of Technology

By drawing on virtue ethics, this paper has offered an approach to deal with "abuse" of robot cases and, more generally, to the moral standing of robots. But we have come a long way from the initial problem formulation; I have proposed significant revisions. Initially, the problem of moral standing framed in virtue ethics terms was about a robot that was being "abused". Using a virtue ethics approach, I then shifted the focus to the human (ab)user of the robot and her virtue or vice. But this was not sufficient to really understand and evaluate these kinds of cases: I added that this particular vicious (inter)action must be understood as embedded in a social practice and history, and as involving performance—including bodily performance. We can conclude that instead of talking about the moral standing of the robot, these phenomena and cases invite us to ask (a) about how virtue can be habituated and incorporated in, and performed by, its human users (and how vice can be dishabituated etc.) and (b) about the moral quality not only of the character and life of the person, but also that of the practice and social-cultural environment in which these performances and persons are embedded. We thus arrived at the basis for a performance-oriented and more social-relational virtue ethics approach to *moral standing*.

This approach enables us to more accurately formulate our moral intuitions and responses to the phenomena mentioned in the beginning of the paper. What is potentially wrong about molesting sex robots, for example, is not that harm is done to the robot and not *only* that this individual behavior might badly reflect on the moral character of the person in the sense of mental dispositions, but also with the social context in which this abuse arises: with the practice of using women for sex (that is, as if they were a machine), and with the way current intimate relationships, understood as a practice, are sometimes organized in particular contexts. Moreover, the notions of habit and performance alert us to the developmental and bodily-kinetic dimension of these kinds of problems: there is something wrong with the gestures and the performance, and behaviors such as kicking someone or treating someone else as an object are learned in a particular context and become incorporated and are performed. The problem is not just situated at the time when the "abuse" occurs, and moral change is not just a matter of changing "mental" disposition but also require learning different ways of performative comportment.

This approach may be of interest to philosophers, but researchers from HRI and related fields in social robotics may also find this work helpful when they discuss some of the phenomena mentioned in the beginning of this paper. The notions (practice, habit, performance) and relational approach introduced here could complement their conceptual toolbox (often filled with notions and theories from psychology) or connect to it in various ways (e.g., via embodied cognition). The approach offered in this paper could thus contribute to, and further stimulate, ongoing interdisciplinary conversations between ethics of robotics and HRI.

Beyond thinking about moral standing of robots, the proposed approach can help us to move towards a more relational *virtue ethics of technology* that theorizes the link between virtue and practice and takes into account the temporal and bodily dimensions of virtue, that is, virtue in its history and its concrete, bodily performances. Like their technologies, the human (ab)users must be understood as embedded in social and cultural environments, and what they do with technologies involves their bodies as much as their minds. Such a relational and embodied-performative approach may well be acceptable to, or appeal to, many philosophers currently working in the field of technology ethics, including Vallor and Sparrow; my main aim was not to criticize their work but rather to open up a conceptual space that enables the further development of work on virtue in the philosophy of technology community in a more relational direction.

Moreover, the proposed approach could also support relational directions in environmental virtue ethics and/or virtue ethics applied to animals. For example, based on the proposed approach one could argue that whatever direct moral standing animals might have in virtue of their intrinsic properties, a virtue ethics approach would give them at least *also* indirect moral standing via the relations they have with human beings and their habits and practices. Treating these beings well is a matter of individual virtue but also the moral quality and excellence of how we organize ourselves, and exercising that virtue and reaching that excellence is not just about a change of mind but also requires changes in our bodily performances and a change of our practices. Whatever other reasons there may be for protecting them and for caring about their flourishing (and I personally think there *are* other good reasons), virtue ethics gives us at least *one* good reason to do so: for our virtue's sake. And that virtue is not only about individual motivation and action: the approach argued for here gives us some extra reasons (if we need any) for why we should not shoot the Kantian dog or, more generally, kill or abuse animals. At the same time, the proposed approach helps us to understand why we often do kill animals or have them killed, even if we do not really intend to or are hardly aware of the consequences of our actions: the problem is not just the individual dispositions but also and perhaps mainly the incorporated habit and the social practice, for example eating meat and its related bodily performances and (often hidden) social-technological practices. Neither the acquisition of virtue nor effective moral change depends on individual motivations and "mental" dispositions alone. We have to change an entire practice and—perhaps the most difficult of all—change our habits.

It thus turns out that engaging with the question concerning the moral standing of robots is not a "marginal" thing to do but invites the traditional, very central question concerning virtue and the good life. Responding to valuable work already done on virtue ethics in philosophy of robotics by Vallor and Sparrow and drawing on a range of conceptual resources that are not yet fully used in this field, I have proposed to interpret that question and hence the concept of virtue in a more social, relational and performance-oriented way. Ultimately, this is not only a philosophical exercise and a reflection on phenomena discussed in HRI but invites us to questioning our personal habits and practices: not just our "mental"-cognitive dispositions but also our comportment/performance and our relation to our social-practical environment.

But what about the "abuse" of robots? If this virtue ethics argument is right and if my interpretation of what virtue means makes sense, then what we (not) do to robots matters and should matter: if not to them, then at least to and for *us*. It matters to and for humans as social and embodied-performative beings who are continuously invited to grow our relationships, evaluate our performances, and *become* more virtuous.

Funding Open access funding provided by University of Vienna.

Compliance with Ethical Standards

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

References

1. Ames RT, Rosemont H (1998) The analects of confucius: a philosophical translation. Random House, New York and Toronto
2. Bartneck C, van der Hoek M, Mubin O, Al Mahmud A (2007) "Daisy, Daisy, Give me your answer do!"—switching off a robot. In: Proceedings of the 2nd ACM/IEEE international conference on human–robot interaction, Washington, DC, pp 217–222
3. Bourdieu P (1984) Distinction: a social critique of the judgement of taste (trans: Nice R). Harvard University Press, Cambridge
4. Bourdieu P (1990) The logic of practice (trans: Nice R). Polity Press, Cambridge
5. Brey P, Briggle A, Spence E (eds) (2012) The good life in a technological age. Routledge, New York
6. Bryson J (2010) Robots should be slaves. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins, Amsterdam
7. Cappuccio ML, Peeters A, McDonald W (2020) Sympathy for dolores: moral consideration for robots based on virtue and recognition. Philos Technol 33:9–31
8. Coeckelbergh M (2007) Violent computer games, empathy, and cosmopolitanism. Ethics Inf Technol 9:219–231
9. Coeckelbergh M (2010) Robot rights? Towards a social-relational justification of moral consideration. Ethics Inf Technol 12:209–222
10. Coeckelbergh M (2011) Environmental virtue: motivation, skill, and information technology. J Environ Philos 8:141–170
11. Coeckelbergh M (2012) Growing moral relations: critique of moral status ascription. Palgrave Macmillan, New York
12. Coeckelbergh M (2012) Care robots, virtual virtue, and the best possible life. In: Brey P, Briggle A, Spence E (eds) The Good life in a technological age. Routledge, New York, pp 281–292
13. Coeckelbergh M (2016) Is it wrong to kick a robot? Towards a relational and critical robot ethics and beyond. In: Seibt J, Nørskov M, Andersen SS (eds) What social robots can and should do. IOS Press, Amsterdam, pp 7–8
14. Coeckelbergh M (2018) Technology Games: using Wittgenstein for Understanding and Evaluating Technology. Sci Eng Ethics 24:1503–1519
15. Coeckelbergh M (2019) Moved by machines. Routledge, New York

16. Coghlan S, Vetere F, Waycot J, Barbosa Neves B (2019) Could social robots make us kinder or crueller to humans and animals? Int J Soc Robot 11(5):741–751
17. Darling K (2016) Extending legal protection to social robots: the effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In: Calo R, Froomkin M, I Kerr (eds) Robot law. Edward Elgar Publishing, Cheltenham and Northampton, pp 213–232
18. Darling K (2017) Who's Johnny? Anthropomorphic framing in human–robot interaction, integration, and policy. In: Lin P, Bekey G, Abney K, Jenkins R (eds) Robot ethics 2.0. Oxford University Press, Oxford
19. Dourish P (2004) Where the action is: the foundations of embodied interaction. MIT Press, Cambridge
20. Duffy BR (2003) Anthropomorphism and the social robot. Robot Auton Syst 42(3–4):177–190
21. Foucault M (1988) Technologies of the self: a seminar with Michel Foucault. The University of Massachusetts Press, Amherst
22. Gunkel D (2018) The other question: can and should robots have rights? Ethics Inf Technol 20:87–99
23. Hursthouse R, Pettigrove G (2016) Virtue ethics in stanford encyclopedia of philosophy, Retrieved 25 September 2020 from https://plato.stanford.edu/entries/ethics-virtue/
24. Ihde D (1998) Expanding hermeneutics: visualism in science. Northwestern University Press, Evanston
25. Johnson DG (2006) Computer systems: moral entities but not moral agent. Ethics Inf Technol 8:195–204
26. Kant I (2012) Lectures on anthropology. In: Wood AW, Louden RB (eds) Cambridge University Press, Cambridge
27. Ku H, Choi JJ, Lee SJ, Do W (2018) Shelly, a tortoise-like robot for one-to-many interaction with children. In: Proceedings of the ACM/IEEE international conference on human–robot interaction, pp 353–354
28. MacIntyre A (1999) Dependent rational animals: why human beings need the virtues. Open Court, Chicago
29. MacIntyre A (2007) After virtue: a study in moral theory, 3rd edn. University of Notre Dame Press, Notre Dame
30. McCormick M (2001) Is it wrong to play violent video games? Ethics Inf Technol 3:277–287
31. 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, pp 205–221
32. Sparrow R (2017) Robots, rape, and representation. Int J Soc Robot 9:465–477
33. Sparrow R (2020) Virtue and vice in our relationships with robots: is there an asymmetry and how might it be explained? Int J Soc Robot https://doi.org/10.1007/s12369-020-00631-2
34. Vallor S (2016) Technology and the virtues. Oxford University Press, New York
35. Varela FJ, Rosch E, Thompson E (1991) The embodied mind: cognitive science and human experience. The MIT Press, Cambridge
36. Winograd T, Flores F (1986) Understanding computers and cognition: a new foundation for design. Ablex, Norwood

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

Mark Coeckelbergh is Full Professor of Philosophy of Media and Technology at the Department of Philosophy of the University of Vienna and Vice Dean of the Faculty of Philosophy and Education. He has been President of the Society for Philosophy of Technology and member of the High-Level Expert Group on AI of the European Commission. He is the author of 12 books and numerous articles in the area of ethics of technology.