4 OUT OF CONTROL: FLOURISHING WITH CAREBOTS THROUGH EMBODIED DESIGN

Author: Anco Peeters

Author ORCID: 0000-0002-2380-4154

Abstract: The increasing complexity and ubiquity of autonomously operating artificially intelligent (AI) systems call for a robust theoretical reconceptualization of responsibility and control. The Meaningful Human Control (MHC) approach to the design and operation of AI systems provides such a framework. However, in its focus on accountability and minimizing harms, it neglects how we may *flourish* in interaction with such systems. In this chapter, I show how the MHC framework can be expanded to meet this challenge by drawing on the ethics of carebots and embodied design. First, I examine how discussions about flourishing with carebots invite us to consider the extent to which we control our moral character Second, I argue that we can understand the cultivation of moral character in terms of embodied virtues arising from operating in particular ecologies. Third, I demonstrate how this analysis fruitfully informs the design of carebots as supporting reciprocity and empathy.

Keywords: Virtue ethics, roboethics, care robots, embodied cognition, enactivism, embodied design.

4.1 Meaningful human control: an incomplete puzzle

With the proliferation of artificially intelligent (AI) systems in the public sphere, there is a growing need to elucidate the responsibility for the actions of such systems. Consider an autonomous car that causes a traffic accident with a pedestrian: who is responsible for the harms caused? The person in the car – if indeed there is one? The designer, the manufacturer, the pedestrian, the government? Is there a shared responsibility? An emerging approach to clarify these matters frames such questions in terms of 'meaningful human control' (MHC) of AI systems. 'Meaningful' in this context is to be taken as a robust type of control that is morally grounded. To this end, recent accounts of MHC emphasise the capacity for human agents to exert guidance control (Fischer & Ravizza, 1998) in morally laden scenarios: that is, the idea that humans ought to be able to align the actions resulting from the interactions with AI with their own internal values (Santoni de Sio & Van den Hoven, 2018).

While providing novel and useful answers, such attempts to flesh out MHC remain incomplete. In allying the MHC framework with the idea of guidance control, MHC requires the human agent to be in a position to, before or after an action, track and reflect on the reasons that were relevant for performing that action. But many human actions, even morally relevant ones, are automatic and non-reflective. Pressing an agent to explain their actions risks inviting confabulation, which undermines the idealized relation between reasoning and acting. Additionally, there remains an open issue of bridging the gap between knowing what is the right thing and actually doing that thing (Santoni de Sio & Meccaci, 2021). At the very least, this gap implies that MHC accounts relying on guidance control may not cover all scenario's where questions of responsibility involving artificial agents arise.

In this chapter, I aim to augment existing approaches to MHC by drawing on the philosophical tradition of virtue ethics. Briefly put, virtue ethicists consider the question: how do I live a life worth living? In answering this question, they put particular emphasis on the development of an agent's moral character and positive habits, rather than on generalized and abstracted rules and duties (Vallor, 2021). By considering how individuals may morally flourish in diverse situations, virtue ethics provides us with a tool to reverse the ethical polarity of MHC approaches that would otherwise focus solely on harm minimisation. To show how a virtue-based lens enriches MHC-debates, I will initially restrict the scope of my account to a domain that shows the relevance of contemporary virtues in AI technology: healthcare, and more specifically, the role of carebots in the (clinical) care for other humans. My aim here is not to provide a fully fleshed-out set of technical

specifications for the design of carebots, but rather to show how, through examining the design of such robots through the lens of virtuous practices, MHC is usefully expanded.
The development of my virtue ethical augmentation of MHC-accounts unfolds in three steps. In Section 4.2, I present recent discussions on the introduction of carebots into healthcare. By examining how caregivers may flourish or wither in caregiving practices, I highlight aspects such as character and moral awareness that would otherwise remain neglected in current MHC-approaches. I then explain how the virtue ethical perspective identifies reciprocity and empathy as key virtues which are potentially undermined by the introduction of carebots in caregiving. This erosion of caregivers' control over how to flourish in their profession is, however, avoidable when we carefully consider carebot design. In Section 4.3, I investigate the psychological underpinnings of these virtues and their associated vices, thus providing grounds for considering potential positive and supportive interventions. Casting virtue and vice as embodied habitual patterns allows for pinpointing the relevant aspects of the behavioural and technological landscape when making carebot design decisions. Section 4.4 is then informed by the findings from the preceding sections and proposes design recommendations for carebots that take the embodied nature of virtue seriously and support the cultivation of reciprocity and empathy. I conclude by articulating what this case study shows for the further development of the MHC framework and providing directions that future work on virtue and MHC should take.

4.2 Widening caregivers' control over reciprocity and empathy

Given the growing costs and use of healthcare in countries worldwide – in particular due to rapidly ageing populations – it comes as no surprise that healthcare professionals and policy makers are collectively considering or even already implementing robots in care (e.g., Bouwhuis, 2016; Tan & Taeihagh, 2021). A paradigmatic example of such a carebot is Paro, a fluffy artificial companion that resembles a seal and has seen widespread use in Japan with the intention to combat loneliness and anxiety among the elderly (Wada et al., 2010). Besides acting as companions, carebots may include robots performing a variety of other functions, like assisting with surgery in operating theatres, delivering or dispensing medicines, and assisting with physical exercises or tasks.
The adoption of carebots in healthcare is by no means uncontested and a diverse constellation of arguments drive debates on the topic back and forth. In favour, we find arguments about the delegation of heavy or repetitive physical tasks to machines well suited for such labours, the improvement of operational efficiency with its concomitant potential for cost reduction, and the consistent display of (faked) supportive emotions (Na et al., 2023). Against, ethicists point out the risks of casting care as a technological problem requiring a technological solution, and of deceiving or infantilizing those that are cared for (Sparrow & Sparrow 2006; Sharkey & Sharkey, 2010).
At least one application of MHC within debates about robots in healthcare has recently been made. Looking at surgical robots for precision medical operations, Ficuciello et al. (2019) emphasize the need for clear attributions of responsibility and accountability in relation to the increasing autonomy of surgical robots. At stake here is the control of the surgeon over their (partially) autonomous operating tool and, to a lesser extent, the control of the patient over their bodily integrity and autonomy. 'Meaningful' in this context is taken to be about patient health and safety, prompting discussion about at which points in the operation surgeons could override the robot – or, potentially, the other way around.
The preceding analysis in the context of surgical robots is, while an important contribution to the literature, also somewhat paradigmatic of the sometimes narrow focus of existing MHC approaches. Instead of focusing on overriding control of another agent (human or robot), *meaningful* control can also be understood in terms of enabling or restraining an agent's behavioural patterns, cultivating character and affecting moral awareness in a more dynamic, ecological manner over longer time-scales. The philosophical tradition of virtue ethics helps highlight precisely these aspects and opens the door to investigating if, and if so, how, flourish with carebots is possible.
Virtue ethics is a moral theory that asks: how can we live our life in the best possible way? The answer includes that we should strive towards developing ourselves to our full potential, by

cultivating a virtuous character. A virtue is a disposition to act in the right way depending on the particulars of a situation. In this light, social robots have recently been argued to be considered as especially ethically significant since their anthropomorphic or animalistic features – such as resembling pets or humanoids, playing on people's emotions, and conversing in natural language – are thought to make more of an impact on the cultivation of good or bad character traits than other types of artefacts (Cappucio, Peeters & McDonald, 2020).[1]

Through its focus on acting morally as a matter of stable dispositions to act, virtue ethics is in the right position to widen the scope of existing MHC accounts. First, virtue ethics allows us to ask not only what can go wrong in our interactions with artificial agents, but also how we may excel in such relations (Peeters & Haselager, 2021). Second, it changes perspective by asking moral actors what kind of person they want to be, for example, whether one wants to be an empathic caregiver (Vallor, 2011). Third, it enables the consideration of the moral dimension of human-robot relations in terms of reflexive – and not merely reflective – behaviour (e.g., Senft et al., 2019). Unfortunately, virtue ethics is largely under-represented in debates on the ethics of carebots (Vandemeulebroucke, 2018).

In one exception to this observation, philosopher of technology Shannon Vallor (2011) points out that, while discussions about the introduction of carebots understandably often focus on benefits for patient health, "by ignoring the positive value of caring practices for caregivers, current scholarly reflections on the ethical implications of carebots remain dangerously one-sided" (2011, p. 254). She advocates for the consideration of the importance of certain caregiving practices for the care*giver*.

By changing the perspective from care receiver to caregiver, a different kind of question arises. What does good care imply for the caregiver? 'Good' is here not defined in quantitative terms such as curing a certain number of patients, or by positing an abstract value such as 'human dignity'. Instead, this involves being able to be there for those in need of care and show them they are not alone, while supporting them in their healing process (Tronto, 1993). These practices are intrinsically good: being there for those in need is good in and of itself, and such interactions give meaning to the kind of work a caregiver does.

In taking this perspective, as Vallor (2011) points out, we may consider how carebots could transform caregiving practices and create space for improving healthcare practices by taking the aforementioned aspects into account. The focus then shifts from asking: "How do we prevent harm to patients from carebots?" to "How can carebots best support caregivers as someone who is there for others in need?" In cases of exemplary care, caregivers know how to judge who is in need and how to be there for them. In other words, they have cultivated a certain moral attitude towards those in need of care. How can we best think about what a moral attitude is and how to cultivate it? In the virtue ethical tradition[2] this is cast in terms of dispositions to act in a specific way. Exemplary dispositions or virtues are 'courage', 'empathy' and 'prudence'. Less praiseworthy dispositions or vices are 'cowardness', 'narcisism' and 'short-sightedness'.

Virtues and vices have a complex relation (e.g., Sparrow 2021), but typically a virtue is understood as the appropriate middle between two extremes, or vices. In such cases, one vice is a lack of the virtue in question while the other vice is an overflow or overabundance of that virtue. Take empathy, for example. A caregiver who constantly forgets the names of patients, shows no emotional bond with and otherwise objectifies them, has a lack of empathy. On the other hand, a caregiver who is too often overwhelmed by the pain of patients and cannot keep a professional distance can be said to have too much empathy. A truly empathic caregiver has emotional bonds with those they care for, while maintaining their own emotional integrity. This delicate balancing act is deserving of constant reinforcement, guided by moral exemplars or mentors who already have attained a certain level of mastery in that particular skill.

Crucially, virtues are taken to be spontaneous, effortless and situation-dependent. They are habituated through repeated practice and eventually become a part of a person's character, become second nature over time and guide a person's actions without conscious effort. Being virtuous is sometimes likened to being good at sports: continuous practice is important to ensure appropriate responses to the dynamically changing situations the world presents us. Virtues are said to be

situation-dependent because the specific expression of a virtue may vary in different contexts. For example, honesty may require speaking the naked truth in one situation, but tact and diplomacy in another. The expression of a virtue is context-specific, and a virtuous person will adapt their behaviour to each situation in a way that is fitting and appropriate.

Determining what exactly is fitting and appropriate requires one to be an situational expert, or to be *practically wise* in contexts that, for example, require balancing honesty and tact. Practical wisdom is "a form of moral intelligence that enables the skilful, creative, and adaptive modulation of moral judgment and habit to novel or rapidly changing contexts and circumstances" (Vallor, 2021, p. 87). A practically wise person knows how to act in the right way depending on the particulars of a situation, such as the giving of care, and is therefore assumed to understand the relevant moral context.[3]

It is challenging to determine how to be practically wise about interacting with social robots because of the complexity and rapidly evolving nature of these agents. The behaviour of social robots, generally being complex agents, is often difficult to predict and comprehending their abilities and limitations can be hard for their users, even if they are robot experts (as many users will not be).[4] In the next section, I will show why a more embodied approach to virtues will further elucidate how to be practically wise in our interaction with social robots.

Vallor (2011) emphasizes the virtues of reciprocity and empathy as paradigmatic for the domain of healthcare. She takes the virtue of reciprocity as striking the right balance between taking from and giving to others. This virtue involves a sensitivity to the needs and perspectives of others, and it requires individuals to be aware of the give-and-take dynamic involved in social relationships. By striving for reciprocity, individuals build strong, mutually supportive relationships with others. At the same time, this virtue requires individuals to be mindful of their own needs and to ensure that they are not being taken advantage of. Pivotally, Vallor emphasizes that practising reciprocity teaches us that we may need to be there for others and that, when the need arises, others may be there for us. Thus, the virtue of reciprocity strikes a balance between the demands of self-interest and the needs of others, allowing individuals to live and interact in a way that strengthens human relations and grounds other virtues such as empathy, thus not only supporting the flourishing of the caregiver but also the people she interacts with.

Empathy can be understood as the capacity to feel *with* another human being. It is central to the domain of healthcare as it is through the caring for others that we learn to relate to others in a way that is neither too close nor too distant. This requires an integration of emotion and reason, as a caregiver has to practise how to remain receptive to the needs of others, without being overwhelmed by their feelings, for instance, when a painful treatment is required. Vallor (2011) identifies empathy somewhat poetically as "a quivering flame constantly vulnerable to being extinguished by apathy or cynicism, or our natural desire to protect ourselves from suffering" (p. 259). She thus casts empathy between the associated extremes of not feeling with others because we do not care, or because we care *too* much and shield ourselves from these feelings to prevent over-identification.[5]

The risk that carebots then present, is that they might remove the opportunity for caregivers to learn and develop the practice of reciprocity and, especially, empathy. Granted, carebots might support caregivers in some straightforward ways, such as when a carebot helps lift a patient that a caregiver might otherwise have been unable to do. However, were carebots to take a central role in caregiving practices and increase the distance between the care receivers and the caregiver "before we have had sufficient opportunities to cultivate the virtues of empathy and reciprocity, among others, the impact upon our moral character, and society, could be quite devastating" (Vallor 2011, p. 260). Connecting this to MHC approaches, I propose that 'meaningful' control in this context ought to signal the (lack of) capacity for caregivers to practice these virtues when carebots are introduced in their professional spaces. The challenge then is to consider how the design of carebots might enable rather than undermine reciprocity and empathy, and, consequently, caregivers' moral control over their character, in caregiving practices.

4.3 Understanding embodied dimension of reciprocity and empathy

Developing design recommendations for carebots that support reciprocity and empathy requires a better understanding of the behavioural and psychological aspects of said virtues. In this section, I develop an account of moral character in human-technology interaction that sees the cultivation of virtue as being restrained or enabled by the sociotechnical environment an agent operates in. These enabling and restraining conditions map onto the typical understanding of virtue as residing between a 'vice of overabundance' (e.g., recklessness in the case of courage) and a 'vice of lack' (e.g., cowardice).

Articulating such an empirical account is a theoretical challenge as, though there is a recent interest in virtue within the social sciences, thoroughly integrated and interdisciplinary approaches to the psychology of virtue remain lacking (Kristjánsson, 2018). Fortunately, some useful building blocks to construct such a bridge have recently been provided by Mark Coeckelbergh (2021). He proposes to frame the use and abuse of robots as the practising of habits within a social and cultural context, and draws attention in particular to the embodied and situated nature of virtue.

Embodiment is the bridge by which Coeckelbergh connects the ethical concept of virtue to the social and cultural environment in which habits unfold. In doing so, he distances himself from Vallor (2016), saying she "stresses reasons, motivation, and 'states' of the mind" (Coeckelbergh, 2021, p. 37). Instead, Coeckelbergh aims for an explanation of virtue as arising from habits of implicit embodied knowing-how. A useful example would be the knowledge I have of riding a bicycle: this knowledge is hard to transfer through verbal articulation, but fairly easy to *show* (if I have a bicycle handy). Somewhat unsatisfying though, is his conclusion that more work is needed to better understand how habits, within a social environment, are influenced by human–robot interaction. He suggests that we require "a philosophical framework that theorizes the social and performative dimension of what we do with and to technology" (p. 37). I agree and argue that such a framework needs to account for the regulatory psychological aspects that sets virtue apart from mere behaviour.

The embodied, enactive approach to cognition provides helpful tools along the lines suggested by Coeckelbergh. Although providing a full account falls beyond the scope of the present paper, one key element of this approach is that it recognizes the role of the body and environment in shaping cognitive processes. In doing so, it reserves a co-constituting role for technological artefacts, emotions and social practices in constraining or enabling the ongoing formation of virtuous or vicious habits and emphasizes their context-specific and dynamic nature (e.g., Di Paolo et al., 2017; Ramírez-Vizcaya & Froese, 2019). By considering how individuals' bodies and brains are shaped by their interactions with their environment, the embodied, enactive approach acknowledges that virtues are grounded in active engagement of brain, body *and* environment, rather than solely in abstract neural information-processing activity (Thompson, 2007; Hutto, Myin, Peeters & Zahnoun, 2018).[6]

In cognitive science, the study of conditions than enable or constrain behavioural patterns has been best articulated in the respected, yet somewhat niche tradition originating in the work of J.J. Gibson (1979). Gibson famously coined the phrase 'affordance', with which he captured the notion that agents directly perceive opportunities for action through the invariant structures that their environment provides them. Importantly, these opportunities for action cannot be captured by studying either the environment or the agent, but only emerge from a coupled agent-environment system.

A famous example of such an affordance is that of diving gannets: while hunting for fish, these seabirds plummet with literal break-neck speed into the water. It is crucial for them to retract their wings at precisely the right moment when doing so. Being too fast risks losing one's prey, while being too slow risks breaking one's bones. A, now classic, study has shown that the higher-order variable τ is a better predictor for the moment at which the gannets adjust their wings than competing theories that explain the process in terms of computations over mental representations (Lee & Raddish, 1981). It turns out that τ is the ratio of the size of the oncoming surface on the gannets' retina and the speed with which this image grows (or shrinks). This enables the gannet to

have an immediate perception of the moment at which it needs to act, that being the moment that the surface of the water fills a certain amount of its eye (see also Chemero, 2009, pp. 123-125).
This elegant yet powerful concept of affordance has since found its way in areas outside of cognitive sciences, not in the least in technology design as a way of thinking about which aspects of a given technology enable (or constrain) actions given the embodiment of its potential user (see, e.g., Norman, 1988; Peeters, 2021). By extension, it inspires the consideration of MHC as being socioculturally embedded: if technologies afford certain actions in the Gibsonian sense, then meaningful control over both the technologies and our (moral) selves crucially depends on the environmental elements that co-constitute, for example, human-robot relations.

Figure 4.1: Traditional paper flight strips as described by Mackay et al. (1998).

To illustrate the crucial importance of the embodied and embedded dimensions of our interactions with artefacts, consider an exemplary case from the late 1990s. In a striking study, Mackay, Fayard, Probert and Médini (1998) investigated the introduction of new electronic air-strips at a Parisian air traffic control station. The traditional paper-based system (Figure 4.1) was being replaced with a sophisticated computer-based one, but traffic controllers resisted the change. As Michael Wheeler (2010) notes, “one is inclined to focus, naturally enough, on the information carried by these strips. But this is not the only contribution of the strips” (p. 33). Crucially, the paper strips contributed more than mere alphanumeric information: their physicality and the space in which they were used enabled them to be held as a reminder, placed at an angle to signal potential collisions, and used for signalling colleagues through body language. The particular physical aspects of the strips enabled kinds of interaction that could not be captured by the electronic strips that merely translated the written, symbolic information. Thus, this study shows the importance of considering the affordances and other non-information-related uses of tools in technology design and implementation.
What does this imply for our present investigation of reciprocity and empathy of caregivers in the context of carebots? First, it forces us to take into account how reciprocity and empathy are either constrained or enabled by our embodied and embedded interactions with the potential physical design of such carebots. As mentioned, virtues are typically seen as an ideal mean between two extremes or vices (Vallor, 2021). Those vices are defined by either signalling a lack of virtue or an overabundance of virtue. For example, in the case of courage, having a lack of courage entails being cowardly, while having an overabundance of courage implies being reckless.
I propose that the way vices are traditionally understood in terms of lack or overabundance maps onto how certain behaviours can be restrained or enabled by environmental conditions. Proceeding with the illustration of courage, Joshua Skorburg (2019) describes how augmented-reality smart

glasses could enable soldiers and police officers to make more accurate threat assessments in the field (for example, by showing potential hidden weapons), arguing that their "cognitive and affective processing" (p. 2343) would be transformed. Enforcers who will use these devices for an extended period of time will integrate them into their innate cognitive system and are thus potentially enabled to become more practically wise in dealing with danger and act more courageously. At the same time, overly relying on such devices risks becoming reckless by cultivating an overabundance of courage.

Second, it acknowledges that virtue, and by extension moral character, is co-constituted by the environment that an actor operates in, which includes the technological devices therein (Alfano & Skorburg, 2017). Provocatively put: this implies that meaningful control over the cultivation of one's moral character and awareness therefore inherently lies partly beyond our immediate control. Realising these limitations could provide a healthy 'check' on how much in control we should consider ourselves to be and how much responsibility we can shoulder as individuals.

In Tables 4.1 and 4.2, I list a number of potential constraining and enabling conditions that illustrate how empathy and reciprocity in care might be shaped (see Marck, 1990; Neufeld & Harrison, 1995; Mercer & Reynolds, 2002). Naturally, these deserve further investigation, but for present purposes they provide enough of a handle to inform the next section for a proof-in-concept. Thus, in the following section, I consider how these constraining and enabling conditions on empathy and reciprocity inform design recommendations for carebots.

Table 4.1: Reciprocity

Factors leading to a lack of reciprocity	**Factors leading to an overabundance of reciprocity**
Physical exhaustion: Caregivers who are overworked and exhausted may find it difficult to provide the same level of care as they did previously. They may also feel physically limited and be unable to perform certain caregiving tasks.	*Lack of self-care:* Caregivers who neglect their own needs and well-being may become overly focused on caregiving and prioritize it over everything else.
Emotional burnout: Caregivers who are constantly giving and not receiving enough support in return may experience emotional burnout. This can lead to feelings of hopelessness, helplessness and apathy.	*Emotional overextension:* Caregivers may feel a strong emotional connection to the people they care for and feel compelled to always go above and beyond in their care.
Conflict with other responsibilities: Caregivers who have multiple responsibilities, such as work and family, may struggle to balance their caregiving duties with other commitments. This can lead to reduced time and energy for caregiving and a decrease in the quality of care provided.	*Role confusion:* Caregivers may not have a clear understanding of their role and responsibilities, leading them to overcompensate in their care.
Financial stress: Caregivers who are struggling financially may find it difficult to meet the demands of caregiving, especially if the person they are caring for has significant health needs. They may also feel constrained by the financial costs of caregiving.	*Fear of rejection or loss:* Caregivers may feel a strong need to be accepted and appreciated, which could drive them to always prioritize the needs of others over their own.
Social isolation: Caregivers who lack support from family, friends, and community may feel isolated and alone in their caregiving duties. This can lead to feelings of hopelessness and decreased motivation to continue caregiving.	*Cultural or social pressure:* Caregivers may feel pressure from society or their community to always be selfless and put others first, even if it means over-extending themselves.

Table 4.2: Empathy

Factors leading to a lack of empathy	**Factors leading to an overabundance of empathy**
Cognitive overload or burnout: Caregivers may feel overwhelmed or exhausted from the constant demands of caregiving, leading them to become detached and unresponsive to the emotions and needs of others.	*Emotional burnout:* Caregivers who are already emotionally exhausted may become more susceptible to over-identification, as their own emotions become more entangled with those of others.
Emotional numbing: Caregivers who have experienced trauma or have dealt with challenging circumstances for an extended period of time may develop a defence mechanism of emotional numbing, making it difficult for them to feel empathy for others.	*Emotional sensitivity:* Individuals who are highly sensitive to emotions may become overwhelmed by the emotions of others, leading them to become overly invested in their experiences.
Personal biases or prejudices: Caregivers may have unconscious biases or prejudices that affect their ability to	*Social and cultural norms:* Caregivers may be expected by their community or society to show a high degree of

empathize with certain individuals or groups.	empathy, which could drive them towards over-identification.
Lack of emotional regulation: Caregivers who struggle with managing their own emotions may find it difficult to connect with and understand the emotions of others.	*Lack of boundaries:* Caregivers who have poor boundaries may struggle to separate their own experiences from those of others, leading to over-identification.
Limited perspective or life experience: Caregivers who have limited life experiences or a narrow worldview may struggle to understand or relate to the experiences and emotions of others.	*Personal experience:* If a caregiver has experienced similar emotions or situations to the person they are caring for, they may be more likely to become overly identified with their experiences.

4.4 Building blocks for virtuous carebot design

This section aims to shows how an embodied approach to reciprocity and empathy in caregiving practices usefully informs carebot design recommendations. Keeping in mind the general structure of virtues, these tentative recommendations aim to either steer caregivers away from a lack of reciprocity (or empathy), or away from an overabundance of reciprocity (or empathy). Thus, they create space for caregivers to meaningfully engage with the cultivation of these virtues. This proposal is not intended to stand on its own, but rather as informing a democratic and ethical design process (Verbeek, 2011; van Wynsberghe & Robbins, 2014). Ideally, that involves a diverse range of stakeholders including, but not limited to ethicists, policy makers, caregivers and, particularly, patients.

In what follows, I dovetail two recent developments in technology design theory. First, I draw on a recent proposal for ethical design approaches that place virtue centre stage (Reijers & Gordijn, 2019).[7] Second, I apply insights from the embodied design approach as recently articulated by Christopher Baber (2021). His departure from "task ecologies" and its grounding in embodied cognition and design practice fits well with the environment-involving direction I previously articulated. Though these lines of thought have different points of departure, originating in either ethics or embodied cognition, they are converging on similar ideas. These approaches have not yet explicitly engaged with each other; an omission I seek to rectify.

Mirroring my preceding analysis, Virtuous Practice Design, or VPD, proceeds from the identification of relevant practices and virtues in the design of new technologies. This is done by including the larger social and cultural niche that a technology might become part of. Pivotally, Wessel Reijers and Bert Gordijn (2019) advance VPD as widening the scope of technology design to go beyond technical specifications and include prescriptions for the technical *practices* that a particular technology will be embedded in. In doing so, Reijers and Gordijn highlight that not only the specific use of the technology deserves consideration, but also how this practice is educated and regulated. Therefore, in articulating design recommendations for carebots the impact on caregiving practices and how they are taught and instituted, requires attention (see also Bedaf et al., 2016).

In *Embodying Design* (2021), Baber capitalizes on Gibson's ecological approach by considering the design of technologies, not on their own, but as moving parts of larger systems. Accordingly, he takes affordances not as properties of artefacts, but as emerging from the ongoing dynamics of a human-artefact-environment system. Baber identifies three constraints that drive these dynamics (p. 48). First, the organism or agent defines the shape and size of the body that is engagement with the task. Second, the task provides (or sometimes enforces) the socio-cultural norms that presents some action outcomes as more desirable than others (e.g., Brancazio, 2019). Third is the physical environment, defined in terms of how bodies act in relation to the laws of physics.

To illustrate, consider the action landscape of a playground as seen through the eyes of an adult and a three-year old. A small slide might afford the child a brief thrilling experience. The same slide might not afford the adult the same thrill, either by the restrictive size of the construction or through the norms that restrict the adult in a more subtle way. Such norms can restrict in more than one sense. Perhaps the adult does not consider it proper to go down a toddler-sized slide (a shame for sure). Or, more insidiously, they do not even conceive of the very possibility (even more of a shame). It is precisely at this point, when examining the role of socio-cultural norms, that VPD complements Baber's theoretical framework. VPD helps capture the relevant norms in terms of the

practices and virtues that are deemed relevant in a given context, such as empathy and reciprocity for caregiving.

Conjuring up a poetic image, Baber (2021) argues that the task of the designer who takes embodiment seriously, is to engage in ‘seamful design’ (p. 86). That is, to identify what seams appear in the human-artefact-environment system when articulating different scenarios that enable different types of affordances. These scenarios can be varied by considering variations in the three main constraints, for example, by considering agents with different body sizes, under different socio-cultural norms in different physical spaces.

With the building blocks provided by an embodied approach to virtuous design, we can now turn to formulating a number of design recommendations. By nature of the iterative process outlined above, these recommendations remain, for now, tentative, awaiting prototyping and empirical testing in supervised scenarios. They serve both as a launchpad for further research and as a proof-of-concept for expanding our thinking about MHC.

Let us start by looking at reciprocity and the context in which it is practised and educated. I earlier identified emotional burnout or overextension as factors for, respectively, leading to a lack or overabundance of reciprocity. As a care receiver's physical and mental health lies at the core of caregiving practice, fastidious or overworked caregivers run the risk of neglecting their own health. One potential recommendation might therefore be to have carebots not only monitor a patient’s health, but also the physical and emotional health of caregivers. A carebot could for instance present caregivers with questions for reflection, to check whether emotional burnout or overextension is rising, and advise on potential avenues for avoiding further development of such issues.

Offering a regularly returning mental mirror to caregivers provides the added benefit of raising awareness of the risks associated with a lack or overflow of reciprocity. This in turn supports check-ins amongst (human) colleagues as well. For example, in contexts where there is a shortage of caregivers, those that are on the job could feel compelled to go above and beyond in their care, neglecting their own needs. Through reminding caregivers of this risk, carebots push back against perceived socio-cultural norms that might otherwise impede the cultivation of reciprocity and create space for affordances that were previously not perceived because of work pressure.

Another interplay of affordances and norms can be considered in the context of empathy. In some cases, a religious denomination or a past traumatic experience inclines a care receiver to refuse physical care from persons of a specific gender. However, it might not always be possible to accommodate said preference. Resulting conflicts between the care receiver and the caregiver on call potential further personal biases or prejudice in the latter. This factor puts pressure on the practice of empathy. When a carebot enables the affordance of, for example, assistance with a bath with maintaining the care receiver’s privacy, the opportunities for the caregiver to cultivate empathy remain as communication stays open.

These are but a few considerations for supporting reciprocity and empathy. Other affordances are available, for example, when carebots provide medical information so as to advice the caregiver in emergency cases. Yet, these design possibilities can only support virtuous care if there is institutional support for them, and necessary conditions that ensure, for instance, patient privacy or caregiver job security are addressed. Once more this goes on to show that meaningful interactions with autonomous agents like carebots, depends heavily on the practices and spaces that they are embedded in.

4.5 Conclusion

This chapter considered how MHC can be understood, not only in terms of oversight and conscious deliberation, but also in terms of creating space for agents to flourish. Agents, such as caregivers, who are provided this space thus receive more control over their own moral development. In charting the behavioural landscape of the virtues of reciprocity and empathy in caregiving practice, I showed how virtue ethics can help MHC widen its scope to address this issue. By considering the factors that can enable or constrain these virtues in caregivers, I made tentative design

recommendations for carebots aimed at supporting the cultivation of reciprocity and empathy in the context of caregiving.

While the virtue approach offers a useful framework for understanding the behavioural dynamics of caregiving, it is important to acknowledge its limitations within the context of carebot design. For example, the automatic and situated nature of ethical virtues might raise some fundamental questions about the extent to which these virtues can be intentionally cultivated through design. Moreover, the context-specificity of ethical behaviour means that carebots will need to be adaptable to the unique needs and requirements of each caregiving scenario.

Despite these limitations, the proposed design recommendations for carebots are an important step forward in the development of the MHC framework. By considering the behavioural dynamics of reciprocity and empathy in the caregiving context, these recommendations provide valuable insights into how carebots can be designed to support the flourishing of caregivers without taking away from them what is crucial to the practice of care. Future research in this area could focus on several key areas. Firstly, by examining the potential long-term effects of carebot usage on the emotional, social and psychological well-being of both caregivers and care recipients. Secondly, by studying the ways in which the design recommendations can be implemented. This could involve the use of such carebots in supervised, pilot settings to study their impact in a controlled setting and integrate user feedback.

References

Alfano, M. (2014). What are the bearers of virtues? In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 73-90). Bloomsbury. https://doi.org/10.5040/9781472594150.ch-004

Alfano, M. & Skorburg, J. A. (2017). The extended and embedded character hypothesis. In J. Kiverstein (Ed.), *The Routledge handbook of philosophy of the social mind* (pp. 465–478). Oxford: Routledge.

Baber, C. (2021). *Embodying design: An applied science of radical embodied cognition*. Cambridge, MA: MIT Press.

Bedaf, S., Draper, H., Gelderblom, G.-J., Sorell, T. & de Witte, L. (2016). Can a service robot which supports independent living of older people disobey a command? The views of older people, informal carers and professional caregivers on the acceptability of robots. *International Journal of Social Robotics, 8*(3), 409–420.

Bouwhuis, D. G. (2016). Current use and possibilities of robots in care. *Gerontechnology, 15*(4), 198–208. https://doi.org/10.4017/gt.2016.15.4.003.00

Brancazio, N. (2019). Gender and the senses of agency. *Phenomenology and the Cognitive Sciences, 18*(2), 425–440. https://doi.org/10.1007/s11097-018-9581-z

Cappuccio, M. L., Peeters, A., & McDonald, W. (2020). Sympathy for Dolores: Moral consideration for robots based on virtue and recognition. *Philosophy & Technology, 33*(1), 9–31. https://doi.org/10.1007/s13347-019-0341-y

Chemero, A. (2009) *Radical embodied cognitive science*. MIT Press.

Coeckelbergh, M. (2021). How to use virtue ethics for thinking about the moral standing of social robots: A relational interpretation in terms of practices, habits, and performance. *International Journal of Social Robotics, 13*(1), 31–40. https://doi.org/10.1007/s12369-020-00707-z

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. E. (2017). *Sensorimotor life: An enactive proposal*. Oxford University Press.

Fischer, J. M., and Ravizza, M., 1998. *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.

Ficuciello, F., Tamburrini, G., Arezzo, A., Villani, L., & Siciliano, B. (2019). Autonomy in surgical robots and its meaningful human control. *Paladyn, Journal of Behavioral Robotics, 10*(1), 30–43. https://doi.org/10.1515/pjbr-2019-0002

Friedman, B., & Kahn, P. H., Jr. (2003). Human values, ethics and design. In J. Jacko & A. Sears (Eds.), *Handbook of human-computer interaction* (pp. 1177–1201). Mahwah: Lawrence Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Psychology Press.

Hursthouse, R. & Pettigrove, G. (2022). Virtue Ethics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy.* Stanford University. https://plato.stanford.edu/archives/win2022/entries/ethics-virtue/

Hutto, D. D., Myin, E., Peeters, A., & Zahnoun, F. (2018). The cognitive basis of computation: Putting computation in its place. In M. Sprevak & M. Columbo (Eds.), *Handbook of the Computational Mind* (pp. 272–282). Routledge.

Kristjánsson, K. (2018). Virtue from the perspective of psychology. In N. E. Snow (Ed.), *The Oxford handbook of virtue* (pp. 546–568). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199385195.013.26

Lee, D. & Reddish, P. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature, 293,* 293-294. https://doi.org/10.1038/293293a0

Mackay, W. E., Fayard, A.-L., Probert, L., & Médini, L. (1998). Reinventing the familiar: Exploring an augmented reality design space for air traffic control. *Conference Proceedings on Human Factors in Computing Systems 1998* (pp. 558–565). ACM Press/Addison-Wesley. https://doi.org/10.1145/274644.274719

Marck, P. (1990). Therapeutic reciprocity: A caring phenomenon. *Advances in Nursing Science 13*(1), 49–59.

Mercer, S. W., & Reynolds, W. J. (2002). Empathy and quality of care. *British Journal of General Practice*, *52*(Suppl), S9.

Na, E., Jung, Y., & Kim, S. (2023). How do care service managers and workers perceive care robot adoption in elderly care facilities? *Technological Forecasting and Social Change, 187,* 122250. https://doi.org/10.1016/j.techfore.2022.122250

Neufeld, A., & Harrison, M. J. (1995). Reciprocity and Social Support in Caregivers' Relationships: Variations and Consequences. *Qualitative Health Research*, *5*(3), 348–365. https://doi.org/10.1177/104973239500500306

Norman, D. A. (1988). *The design of everyday things*. Basic Books.

Peeters, A. (2021). Hoe nemen we culturele affordances waar? [How do we perceive cultural affordances?]. *Algemeen Nederlands Tijdschrift Voor Wijsbegeerte*, *113*(3), 393–397.

Peeters, A., & Haselager, P. (2021). Designing virtuous sex robots. *International Journal of Social Robotics*, *13*(1), 55–66. https://doi.org/10.1007/s12369-019-00592-1

Peeters, A. (2019). *Thinking with things: An embodied enactive account of mind–technology interaction* (Doctoral thesis in Philosophy, 185 pp.). University of Wollongong. https://ro.uow.edu.au/theses1/806/

Reijers, W., & Gordijn, B. (2019). Moving from value sensitive design to virtuous practice design. *Journal of Information, Communication and Ethics in Society, 17*(2), 196–209. https://doi.org/10.1108/JICES-10-2018-0080

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, *34*, 1057–1084 https://doi.org/10.1007/s13347-021-00450-x

Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI, 5*. https://doi.org/10.3389/frobt.2018.00015

Schwartz, B. and Sharpe, K. (2010). *Practical Wisdom: The Right Way to Do the Right Thing*. Penguin.

Senft, E., Lemaignan, S., Baxter, P. E., Bartlett, M., & Belpaeme, T. (2019). Teaching robots social autonomy from in situ human guidance. *Science Robotics, 4*(35), eaat1186. https://doi.org/10.1126/scirobotics.aat1186

Sharkey, A., & Sharkey, N. (2010). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology, 12*(4), 225–234. doi:10.1007/s10676-010-9234-6

Sparrow, L., & Sparrow, R. (2006). In the hands of machines? The future of aged care. *Minds and Machines, 16*(2), 141–161. https://doi.org/10.1007/s11023-006-9030-6.

Sparrow, R. (2021). Virtue and Vice in Our Relationships with Robots: Is There an Asymmetry and How Might it be Explained? *International Journal of Social Robotics, 13*(1), 23–29. https://doi.org/10.1007/s12369-020-00631-2

Senft, E., Lemaignan, S., Baxter, P. E., Bartlett, M., & Belpaeme, T. (2019). Teaching robots social autonomy from in situ human guidance. *Science Robotics, 4*(35), eaat1186. https://doi.org/10.1126/scirobotics.aat1186

Sparrow, R. (2021). Virtue and Vice in Our Relationships with Robots: Is There an Asymmetry and How Might it be Explained? *International Journal of Social Robotics, 13*(1), 23–29. https://doi.org/10.1007/s12369-020-00631-2

Skorburg, J. A. (2019). Where are virtues? *Philosophical Studies, 176*(9), 2331–2349. doi:10.1007/s11098-018-1128-1

Tan, S. Y., & Taeihagh, A. (2021). Governing the adoption of robotics and autonomous systems in long-term care in Singapore. *Policy and Society, 40*(2), 211–231. https://doi.org/10.1080/14494035.2020.1782627

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge.

Vallor, S. (2011). Carebots and Caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century. *Philosophy & Technology, 24*(3), 251–268. https://doi.org/10.1007/s13347-011-0015-x

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Vallor, S. (2021). Twenty-First-Century Virtue: Living Well with Emerging Technologies. In E. Ratti and T. A. Stapleford (Eds.), *Science, Technology, and Virtues* (pp. 77–96). Oxford University Press. https://doi.org/10.1093/oso/9780190081713.003.0005

Vandemeulebroucke, T., Dierckx De Casterlé, B., & Gastmans, C. (2018). The use of care robots in aged care: A systematic review of argument-based ethics literature. *Archives of Gerontology and Geriatrics, 74*, 15–25. https://doi.org/10.1016/j.archger.2017.08.014

van de Poel, I. (2013). Translating Values into Design Requirements. In D. P. Michelfelder, N. McCarthy, & D. E. Goldberg (Eds.), *Philosophy and Engineering: Reflections on Practice, Principles and Process* (Vol. 15, pp. 253–266). Springer Netherlands. https://doi.org/10.1007/978-94-007-7762-0_20

van Wynsberghe, A., & Robbins, S. (2014). Ethicist as Designer: A Pragmatic Approach to Ethics in the Lab. *Science and Engineering Ethics, 20*(4), 947–961. https://doi.org/10.1007/s11948-013-9498-4

Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press.

Wada, K., Ikeda, Y., Inoue, K., & Uehara, R. (2010). Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot Paro. *19th International Symposium in Robot and Human Interactive Communication*, 533–538. https://doi.org/10.1109/ROMAN.2010.5598615

Wheeler, M. (2010). Minds, things and materiality. In L. Malafouris, & C. Renfrew (Eds.). *The cognitive life of things: Recasting the boundaries of the mind* (pp. 29–37). McDonald Institute for Archeological Research.

Notes

[1]For an excellent recent introduction that contrasts virtue ethics with other ethical theories, such as consequentialism or deontology, in relation to emerging technologies, see Vallor (2021).
[2]Specifically on agent-based accounts, see Hursthouse and Pettigrove (2022).
[3]For an examination of practical wisdom with in-depth examples in a variety of contemporary societal domains, I recommend Schwartz and Sharpe (2010).
[4]For this reason, Robert Sparrow (2021) argues that we can only be vicious towards social robots, and not virtuous, as being virtuous implies having the relevant practical knowledge. A virtuous person, being practically wise, in their interaction with social robots would realise that such robots cannot be the proper recipient of virtues like empathy, as robots have no inner emotional or conscious life. A person being vicious towards a robot dog faces no such requirement. I have previously disagreed with this line of reasoning (Peeters, 2019, Ch. 4).
[5]Elsewhere, I proposed a kindred conception of empathy in the context of romantic relations with social robots as lying between being self-obsessed and being 'other'-obsessed (Peeters & Haselager, 2021).
[6]This aligns with emerging work in related approaches to embodied cognition (e.g., Alfano, 2014; Skorburg, 2019).
[7]Existing approaches to ethics in design seek to develop new technologies as supporting a value or group of values (Friedman & Kahn, 2003). Recent work has seen focus on translating values into design requirements through a pyramid-style model of translation layers, from abstract values, to norms and to concrete requirements (van de Poel, 2013). While providing useful directions, this Value-Sensitive Design approach does not quite align with the virtue ethical perspective we took in the preceding, focusing more on the design of a technology rather than on how it impacts a stakeholder's moral character. This makes the competing Virtuous Practice Design approach a better fit for present purposes (Reijers & Gordijn, 2019).