

Carebots and Caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century

Shannon Vallor

Received: 9 January 2011 / Accepted: 10 March 2011 / Published online: 31 March 2011
© Springer-Verlag 2011

Abstract In the early twenty-first century, we stand on the threshold of welcoming robots into domains of human activity that will expand their presence in our lives dramatically. One provocative new frontier in robotics, motivated by a convergence of demographic, economic, cultural, and institutional pressures, is the development of “carebots”—robots intended to assist or replace human caregivers in the practice of caring for vulnerable persons such as the elderly, young, sick, or disabled. I argue here that existing philosophical reflections on the ethical implications of carebots neglect a critical dimension of the issue: namely, the potential moral value of caregiving practices for *caregivers*. This value, I argue, gives rise to considerations that must be weighed alongside consideration of the likely impact of carebots on care recipients. Focusing on the goods internal to caring practices, I then examine the potential impact of carebots on caregivers by means of three complementary ethical approaches: *virtue ethics*, *care ethics*, and the *capabilities approach*. Each of these, I argue, sheds new light on the contexts in which carebots might deprive potential caregivers of important moral goods central to caring practices, as well as those contexts in which carebots might help caregivers sustain or even enrich those practices, and their attendant goods.

Keywords Carebots · Capabilities · Virtue ethics · Care ethics

1 Background

In the early twenty-first century, we stand on the threshold of welcoming robots into domains of human activity that will expand their presence in our lives dramatically, extending well beyond the primarily industrial contexts in which robots have for decades already been working alongside or in the place of human beings. Recent

S. Vallor (✉)

Department of Philosophy, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, USA
e-mail: svallor@scu.edu

developments range from the expanding use of robotic “surgeons” to conduct medical procedures with a degree of accuracy and precision well beyond the capabilities of human hands,¹ to the rapid development and implementation of sophisticated robotic drones which now assist or take the place of human soldiers in critical and dangerous military operations,² to the commercial rise of personal robots used in the home to clean our floors, mow our lawns, play with our children, and spy on nannies and contractors.³

This phenomenon naturally gives rise to philosophical questions of various sorts, but among the most common and pressing are ethical questions. Of particular concern to many are the ethical implications of what have been termed “carebots”. Carebots are robots designed for use in home, hospital, or other settings to assist in, support, or provide care for sick, disabled, young, elderly, or otherwise vulnerable persons. The kind of support they may provide varies widely, but generally following Sharkey and Sharkey (2010), we can distinguish between the following actual or potential functions of carebots: performing or providing assistance in caregiving tasks; monitoring the health or behavioral status of those receiving care or the provision of care by caregivers; and providing companionship to those under care.

Motivations for the development of carebots are various, but include demographic pressures in countries like Japan, Germany, and the USA to ensure the ability to care for a rapidly expanding population of aged or otherwise vulnerable persons, economic pressures upon individuals, private or public institutions to reduce the costs of care, social pressures to reduce growing institutional failures to provide quality human care, and recognition of the need to reduce the often-overwhelming physical and psychological burdens placed upon individual caregivers.

Development and implementation of carebots is in a relatively early stage. A few limited applications have found some initial success—AIST’s *Paro Therapeutic Robot*, a bionic fur-covered harp seal, was recently classified in the USA as a Class 2 medical device used to provide companionship in nursing home settings, and has been widely used in Denmark (Tergesen and Inada 2010). Most carebots are still in the development stage, though some are relatively advanced along that path—including the *RIBA* nurse robot developed by Japan’s RIKEN-TRI Collaboration Center for Human-Interactive Robot Research (RTC). *RIBA*, already a second generation upgrade to the earlier *RI-MAN*, can lift a patient from a bed to a wheelchair, and vice versa.⁴ *GeckoSystems International Corporation* is currently conducting limited in-home trials of its *CareBot*. Combining a mobile-service robot base with interlinked AI modules for navigation, tracking, scheduling, and reminders (e.g., medications, food, etc.), and even verbal interaction (“chat”), the *CareBot* is described in *GeckoSystems’* marketing as having a commercial potential that includes both elder care and child care.⁵ Indeed, while it may be quite some time

¹ The most well-known of these is the *daVinci Surgical System* manufactured and marketed by *Intuitive Surgical, Inc.*, but others include *Medsys’ LapMan* and *Titan Medical Inc.’s Amadeus*.

² See Singer (2009) for an excellent discussion of emerging military robotics and their implications.

³ Consider as examples *iRobot’s* popular *Roomba*, *Friendly Robotics’ Robomower*, *Sony’s AIBO*, and *Meccano’s Spykee*.

⁴ <http://www.ribarobotnurse.com/>, accessed 5 Jan 2011.

⁵ http://www.geckosystems.com/markets/consumer_familycare.php, accessed 5 Jan 2011.

before there are practical, safe, and cost-effective carebots available for widespread consumer or institutional use, the development of carebots is likely to continue to advance, if we consider the grave deficiencies of existing social means of providing human care in many developed nations and the ever-expanding commercial appetite for new technologies in many of those same countries. While we would expect carebot developers to offer generous, even unrealistic predictions of the role of such technologies in the future, the roadmap from which they are working is quite explicit, as indicated on the placeholder marketing site for RIBA, which has already secured its commercial domain name: “There might come a time when human caregivers will be replaced by robots. An impossible vision it may seem, but do not be too surprised about human caregivers being assisted by robots in the near future.”⁶

2 Carebots and Ethics

Pronouncements of the latter sort understandably set off philosophical alarms about the ethical dangers of “replacing” human caregivers with machines. We might reasonably regard such statements as commercial hyperbole, yet even the more realistic near-term scenarios for carebots supply ample motivation for ethical reflection and inquiry. What are the various sorts of ethical concerns we might legitimately have? For methodological purposes, let us set aside important questions about the practicality of the grander technological aims of carebot researchers and developers. Let us also set aside well-motivated worries about the safety of such technologies should they be introduced. In the absence of a significant collapse of existing social mechanisms for assuring medical product safety (mechanisms which one must grant, are far from perfect), let us assume for the sake of argument that grave and widespread safety issues can be prevented. There are still a large number of issues that remain.

Mark Coeckelbergh (2009) has helpfully separated what he calls questions of “roboethics” into several categories. The first category pertains to the minds and realities of robots themselves (in this category we would place concerns about whether artificially intelligent robots could ever properly be considered moral agents, whether they might have “rights”, whether they could relate to human beings in ethically significant ways, and so on). The second category pertains to the application of traditional ethical theories to human–robot interactions, allowing us to judge their ethical significance (in this category we might place worries about whether our filial obligations to the elderly can be satisfied through quality care provided by robots, or whether the use of carebots is sufficiently warranted by their collective social utility). Coeckelbergh wishes to redirect our ethical inquiry to a new, third manner of reflection—by moving away from the application of what he calls “external” ethical criteria, and toward consideration of the “good internal to practice” that may be realized through robot–human interaction (217). That is, he wishes to consider how we *experience* such interactions, and what they “do to us as social and emotional beings” (219).

⁶ <http://www.ribarobotnurse.com/>, accessed 5 Jan 2011.

This, too, is my concern. However, I wish to take it in a somewhat different direction from Coeckelbergh's, in order to illuminate something that I think has thus far been missing from philosophical discussions of "roboethics". Before we can do this, however, we need to take a closer look at how scholars have approached the ethical significance of carebots in particular. Though only a handful of scholars have yet turned their attention to this subject, many of the most important ethical issues have been carefully discussed in the existing philosophical literature on carebots. Of course, there remain substantive disagreements about how to evaluate the moral risks and benefits of introducing carebots into actual care settings. In general, these disagreements involve conflicting judgments about whether carebots will improve or degrade the quality of life of the cared-fors, whether they will improve or degrade relations between cared-fors and caregivers, and whether they will result in an improvement or degradation of personal and institutional quality of care provision.

Specifically, concerns have been raised about:

1. The *objectification* of the elderly as "problems" to be solved by technological means (Sparrow & Sparrow 2006, 143; Sharkey & Sharkey 2010).
2. The potential for carebots to either enhance or restrict the *capabilities, freedom, autonomy, and/or dignity* of cared-fors (Borenstein & Pearson 2010; Sharkey & Sharkey 2010; Decker 2008).
3. The potential of carebots to enhance or reduce *engagement* of cared-fors with their surroundings (Borenstein & Pearson 2010; Sharkey & Sharkey 2010).
4. The potential of carebots to enhance or intrude upon the *privacy* of cared-fors (Sharkey & Sharkey 2010).
5. The *quality* of physical and psychological care robots can realistically be expected to supply (Coeckelbergh 2010; Sparrow & Sparrow 2006).
6. The potential of carebots to either reduce or enhance cared-fors' levels of *human contact* with families and other human caregivers (Sparrow & Sparrow 2006).
7. The potential of carebot relations to be inherently *deceptive* or *infantilizing* (Sparrow & Sparrow 2006; Sharkey & Sharkey 2010; Turkle 2006).

All of these concerns are well motivated and demanding of further study. Something has been left out of all of these treatments, however. Though much attention has deservedly been paid to the impact of carebots on those cared for, that is, how well cared-for humans will be in a scenario where carebots are employed, less attention has been paid to their impact on *caregivers* and society more broadly. Where this does get addressed, it is typically limited to a focus on the way in which carebots may help to "reduce the care burden" on caregivers and on society (Sharkey & Sharkey 2010), empowering people to make caregiving a choice rather than an obligation (Borenstein & Pearson 2010, 283–4), allowing caregivers to trade routine tasks for more emotionally meaningful ones (Coeckelbergh 2010, 183), and eliminating or reducing the numbers of undercompensated, overworked, and ill-trained human caregivers, especially in institutional settings (Coeckelbergh 2010, 183; Borenstein & Pearson 2010, 285). However, I will show that by ignoring the *positive* value of caring practices for caregivers, current scholarly reflections on the ethical implications of carebots remain dangerously one-sided. I then examine the value of caring practices for caregivers through the conceptual lenses of three complementary ethical frameworks, each of which are oriented toward the goods

internal to human practices: *virtue ethics*, *care ethics*, and the *capabilities approach*. Rather than defend the superiority of one particular framework, I conclude that each sheds new light on the potential ethical risks and promise of carebots.

3 The Burdens and Benefits of Caregiving Practices

It will be evident to anyone who has ever had to care for an elderly parent, chronically ill spouse, or disabled child for any significant length of time that “reducing the care burden” on caregivers is a desideratum, even a collective moral obligation, assuming it can be accomplished by ethical means. It is also quite clear to anyone who has spent much time in a skilled nursing facility or other comparable institutional setting that as Borenstein and Pearson note, “...it is difficult to support the notion that current caregiving conditions ought to be preserved” (2010, 286). Surrendering caring practices to robots might be risky or ethically worrisome, but so is maintaining the status quo. So before proceeding, let me stipulate that carebots, properly designed and implemented, might be able to improve the lives of both cared-fors and caregivers in ways that would be ethically desirable and, in the absence of acceptable alternatives, ethically mandated.

Still, it deeply concerns me that there has been virtually no scholarly discussion of what value caregiving, as a human practice, might have for *caregivers* rather than *cared-fors*, what the ethical significance of that value might be, and how the value of such caring practices might be impacted by the widespread introduction of carebots. To the contrary, the standing assumption seems to be that caregiving is generally not only a burden upon caregivers (which we can grant) but that it is *nothing except* a burden. Again and again it is described only as a “requirement”, a social “expectation”, a burden people (especially women) are forced to bear (Borenstein & Pearson 2010, 284; Sharkey & Sharkey 2010). Even the provision of companionship and emotional support is characterized as a “task” (Sparrow & Sparrow 2006). Nussbaum (2006) rightly notes that contemporary societies still tend to assume women will give care “for free, ‘out of love,’” (102), her use of scare quotes redirecting us to the image of caregiving as an unfair burden. Yet if there are no *further* dimensions to caregiving practices, it would seem to follow that *in the absence of consequent injury or moral insult to cared-fors*, we could have no reason to hesitate to surrender our caregiving practices to carebots, and to encourage others to do the same.⁷

Intuitively, it seems this cannot be right, and despite the fact that their concerns about carebots tend to focus exclusively on potential harms to *cared-fors*, I doubt (though I could be mistaken) that many of the philosophers who have reflected on the ethics of robot care would endorse the above conclusion. Nevertheless, it begs the question why the potential losses to caregivers who surrender such practices have not been explicitly considered. Of course, it is acknowledged that caregiving may be

⁷ The only way to avoid this conclusion, beyond the one I suggest, is to presuppose a deontological framework, perhaps of a Kantian sort, that entails individual moral obligations to provide care *and* dictates that we may not employ indirect technological means of meeting these obligations. I am not optimistic, however, about the viability of such a claim.

freely chosen, indeed, this is presented as the ideal state of affairs. Borenstein and Pearson, following Nussbaum (2006), tell us that “those who provide care in non-emergency contexts should have the freedom to choose the extent, type, and manner of caregiving” (2010, 284). But the reasons *why* one might choose to give care are not explored, and these reasons matter a great deal to our inquiry.

Some such choices might be motivated by a kind of ethical heroism, a willingness to take on burdens of care so that others do not have to. Others might be motivated by a sense of personal and non-transferrable duty, such that one would rather suffer the burdens of giving care than the guilt of shirking them. Still others might be motivated by external rewards that outweigh the burdens of care, such as monetary compensation, or expressions of gratitude and admiration by cared-for and observers. But all of these reasons presuppose that the caregiving practice *itself* is simply a burden, the free acceptance of which must be justified by something external to the practice. If this is right, then it would appear that nearly all of the external conditions that warrant freely giving care could potentially be eliminated with the introduction of carebots—by transferring caring tasks to entities that cannot experience them *as* a burden. Yet I wish to propose another possibility: that there are goods internal to the practice of caregiving that we might not wish to surrender, or that it might be unwise to surrender even if we might often wish to do so.

In our methodology, then, we return to the orientation of Coeckelbergh (2009), who we should recall, wishes to redirect our inquiry away from “external” ethical criteria, and toward the “good internal to practice” (217) that may be realized through robot–human interactions. He thinks we should think more deeply about how we *experience* such interactions with robots, and what they “do to us as social and emotional beings” (219). I agree. But Coeckelbergh’s interest here is narrower than mine. It is what the *robots* do to us that captures his interest, and here he refers to the way in which their “appearances” to us, regardless of the inner realities of robot “minds”, can powerfully shape our own perceptions, emotions, and capabilities (219–220). I think this is a question of tremendous ethical importance. Yet if we apply it to the carebot question, it targets, once again, the impact of carebots on those to whom they primarily appear, namely those cared *for*. Of course, they may interact with human caregivers in the environment too—nursing staff, family members, and so on. But if we focus narrowly on what the robots do to us through their appearances to us, we leave behind the question of what we may do to *ourselves* when we choose to surrender caring practices to robots. If there are goods internal to the practice of caring, then I am surrendering these goods if I give up the practice, and this holds true whether or not I have interactions with the robot to whom I have surrendered it.⁸

What kind of goods could these be? There are at least three philosophical perspectives that fit well with our methodology and might help us conceptualize the nature of these goods. They are (1) a *virtue ethics* approach, (2) a *care ethics* approach, and (3) a *capabilities* approach. Each of these philosophical perspectives

⁸ It is of course possible that caregivers might choose to *share* caring practices with robots rather than *surrender* them, as I will note. However, given the widespread acknowledgment of the burdens of giving care, we cannot discount the possibility that many potential caregivers will, once they trust the safety and skill of carebots, transfer significant caring duties to them.

offers a conceptual framework oriented toward goods internal to practices rather than external ethical criteria, and each can shed light on the specific nature of the goods internal to caring practices. I will not attempt here to determine which of these perspectives, if any, might be superior to the others. This is for two reasons: first, I do not wish to foreclose the possibility that they are complementary or overlapping perspectives. Second, I believe at this early stage of inquiry it is more important to illuminate the subject matter as richly as possible than to identify with precision the conceptual framework most appropriate to it.

4 Carebots and the Virtues

The virtues are conventionally understood as dispositional states of character that can be regarded as excellent, insofar as they promote human flourishing. I have spoken elsewhere (Vallor 2010) of the way in which a contemporary account of the virtues can help us evaluate the moral significance of emerging information technologies. Rather than rehearse those arguments here, I would like simply to outline two important moral virtues that arguably rely upon caregiving practices for their development: *reciprocity* and *empathy*. These are hardly the only important moral virtues cultivated in caregiving practices—we might mention also patience, understanding, and charity, to name just a few. But the virtues I discuss below are sufficient to illustrate the larger points I wish to make—(1) that there are important goods for *caregivers* that are internal to caregiving practices, and that (2) the potential impact on these from the widespread use of carebots merits our careful attention. Following Coeckelbergh's (2009) invocation of a neo-Aristotelian way of thinking about roboethics, one that focuses on the "kind of moral habits and moral character we should develop," (220) let us investigate the way in which caregiving practices help to shape the moral character of caregivers, and the attendant ethical value that may give us reason to pause at the thought of transferring such practices to carebots.

4.1 Reciprocity

I have defined reciprocity elsewhere (Vallor 2011) as a primitive biological impulse which functions as the seed of human sociality, is the unifying feature of all forms of friendship and, with proper moral and cognitive/perceptual habituation, matures into a social virtue. The basic sociality of human beings is non-controversial. Despite our vulnerability to organic pathologies of reciprocity, such as autism, under normal conditions the human person is fundamentally oriented to others. Reciprocity (*antipeponthos*) is also at the core of Aristotle's understanding of ethical relations, both in its intimate and civic forms (1984, NE 1155b34); human persons, he believes, are naturally predisposed to social give and take (1984, NE 1155a).

But reciprocity is more than a natural impulse—it is also something we must cultivate as a virtue—for understanding how to reciprocate *well*, in the right ways, at the right times, and as appropriate to particular circumstances and people, is part of what it means to become a good person. Even Aristotle's notoriously unequalitarian accounts of unequal friendships and marriage, whatever their substantive errors, are meditations on reciprocity and how to ensure that human bonds are not destroyed by

one member of a pair's inability to give benefits as well as receive them. It is the ethical maturation of reciprocity to which Aristotle refers when he claims that between friends, there is no need of justice (1984, NE 1155a26); for justice is a preventative and remedy for breakdowns and gaps in reciprocity; between complete friends of virtue, then, justice is a cure without a disease.

Lawrence Becker (1986) has asserted that reciprocity is in fact among the most fundamental virtues, one that is essential to creating and sustaining the primary goods that allow human lives to flourish (146–150). This further entails, according to Becker, an ethical obligation on our part to maintain social structures that perpetuate reciprocity as a virtue, and to modify social structures that damage it (163). This claim, if true, reinforces the need to consider the potential impact of the introduction of carebots on the human cultivation of reciprocity. For it is evident that reciprocity is cultivated, among other human contexts, in caregiving practices. This may not be immediately evident if we restrict our attention to the asymmetry between the vulnerable cared-for who receives care, and the one who gives it. Yet let us remember that the elderly parents we hand-feed once fed us, that the sister whose radiation sickness we comfort perhaps nursed us through a childhood flu, and even in institutional settings that nurses, doctors, and therapists are themselves, at times, the patients. We learn in our times of need that others are there for us, and as I wish to stress, we may learn in being there for others to trust that someday, others will be there for us. We also learn the importance of giving for the development of our own moral character, the way it facilitates other virtues, such as patience, empathy, and understanding. We learn in caregiving practices to see how reciprocity holds human relations together and allows other kinds of goods to flow across them. To surrender caregiving practices then, *even if* we stipulate that we will surrender them only to robots who can give what the other truly needs—potentially deprives us of an important opportunity to cultivate reciprocity in ourselves, and to understand its centrality to human bonds.

4.2 Empathy

I have defined this elsewhere (Vallor 2011) as an emotive/perceptual capacity that, like reciprocity, develops in most humans from a basic biological impulse, expresses itself fully in the highest forms of friendship, and, when properly cultivated and expressed, constitutes a virtue. Though Aristotle did not himself consider empathy a virtue (1984, NE 1106a4), I have claimed (2011) that it occupies a critical role in the life of virtue Aristotle describes *and* that Aristotle would likely himself have recognized it as a proper virtue had he enjoyed a better understanding of the integration of emotion and reason in human persons, as well as the essential nature of human vulnerability in grounding ethical human relations.

Empathy is an emotive/perceptual capacity to feel *with* another sentient being, to co-experience, in a significant way, the joys and sufferings of another. In empathy, one “grieves and rejoices with (*sunalgounta kai sunchaironta*)” another human being (1984, NE 1166a8). When cultivated as a virtue, it manifests itself as ability to receive such feelings in the appropriate sorts of circumstances and relationships, and to respond to the other in a manner that is also highly attuned and appropriate to these particulars.

Aristotle makes a distinction between those for whom we merely have goodwill, and those with whom we genuinely empathize in the sense described above. The latter relation, unlike the former, involves a genuine intensity of feeling (1984, NE 1166b34). Aristotle says that with those for whom we have goodwill, we wish them well but “we would not *do* anything with them” (1984, NE 1167a2) “nor take trouble for them” (1984, NE 1167a10). I wish to focus on the latter claim as significant. Though Aristotle certainly did not describe empathetic relations taking place in the contexts we currently associate with caregiving practices—he would view these as almost exclusively feminine practices alien to the masculine ethical ideal he attempts to cultivate—a virtue ethical perspective that applied appropriate correctives to his narrowly gendered ideal can take this last remark as suggestive. For those with whom I empathize, I will indeed “take trouble,” and this occurs most notably in caring relations. In caregiving practices I take upon myself the trouble of recognizing, acknowledging, feeling, and responding appropriately to the suffering of another.

Sparrow and Sparrow (2006) stress the importance of empathy in caregiving as a means of showing the irreplaceability of human caregivers by carebots, claiming that “entities which do not understand the facts about human experience and mortality that make tears appropriate will be unable to fulfill this caring role” (154). I do not dispute this claim, but I wish to stress a different implication. It is primarily in fulfilling these caring roles that humans learn to practice and cultivate empathy as a virtue, and without such opportunities, the development of this virtue may be significantly impeded. For most of us, empathy is a quivering flame constantly vulnerable to being extinguished by apathy or cynicism, or our natural desire to protect ourselves from suffering. Many choose, often regrettably, to turn away from circumstances that might provoke an empathic response, and caregiving contexts, especially those that involve our physical presence to suffering, rank very highly among such circumstances. The concern, then, is that the availability of carebots as a substitute for human care may be very appealing to those who have not yet become comfortable with the cultivation of empathy in caregiving practices, leading them to abandon or greatly limit their exposure to such practices. As Sharkey and Sharkey (2010) note, the possibility of having a robot provide care, or offer a way to supply more remote forms of care, may allow us to feel that we are satisfying our caring obligations even when not present to suffering. I suggest that this is not only a potential loss to the cared-for, but to the caregiver.

When caregiving becomes a practice, and when this practice is done *well*, we become more accustomed to, even welcoming of, the emotional weight of empathy. Gradually, we may learn how to be restored rather than drained by sharing feeling. We can find solidarity in it—as Sparrow and Sparrow (2006) note, “Sometimes the only appropriate response to another’s suffering is the acknowledgment that we too share these frailties, as for instance, when our friend’s suffering moves us to tears” (154). It must be noted, however, that this is only possible when caregivers have sufficient resources and support for such practices—as we will emphasize below, caregiving in inadequate circumstances is likely to drain us of emotional power and starve empathic responses rather than cultivate them. As Borenstein and Pearson (2010) note, “merely because a human is assigned the task of providing care, an emotional bond does not necessarily form” (285). Empathy, then, like reciprocity,

must be exquisitely and quite deliberately cultivated if it is to endure and thrive rather than wither on the vine.

The implication, then, is that these factors must be attended to when thinking about the moral dimensions of robot care. If the availability of robot care seduces us into abandoning caregiving practices 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. On the other hand, if carebots provide forms of limited support that draw us further *into* caregiving practices, able to feel more and give more, freed from the fear that we will be crushed by unbearable burdens, then the moral effect on the character of caregivers could be remarkably positive.

5 Carebots and the Ethics of Care

Care ethics has been developed by a number of different feminist thinkers, including Gilligan (1982), Noddings (1984), Kittay (1999), and Held (2006). Like virtue ethics, care ethics takes its orientation from the goods internal to practices rather than external moral criteria. Yet unlike virtue ethics, it focuses on caring practices, and caring relations, as fundamental. Rather than holding the virtues themselves to be ethical desiderata, care ethics takes virtue to be merely an outgrowth of the caring relation that is the primary ethical good. And given its focus on caring relations and practices, we should expect to find in these perspectives important insights for the ethical implications of carebots.

For reasons of space, I will confine myself here to examining the perspective of Noddings on the ethical significance of caregiving practices. Noddings (1984) asserts that the caring relation is a first and foremost a natural one. Through our memories of being cared-for, we acquire our first and ultimate understanding of the good. The natural caring relation functions as regulative in human social behavior, motivating us to seek to restore the relation when it becomes deficient or disrupted. From the natural caring relation, then, which *in itself* is not a moral one, we acquire an ethical ideal, a notion of the “ethical self” that comes to function as a commitment to meet the other, and in particular those within our intimate kinship circle, from within the caring attitude. Importantly, Noddings notes that “This caring for self, for the *ethical self*, can emerge only from a caring for others” (14). Thus while the memory of being cared-for is required for the development of the ethical self as a regulative ideal, it is only through caring for others that I can care for that ideal—only through a commitment to caring practices can I ensure that my ethical self is sustained and enriched.

Noddings identifies two central criteria for a genuinely caring commitment—namely, *engrossment* and *motivational displacement* (16). Engrossment is an orientation to the reality of the other, as opposed to my own reality. It allows the needs of the other to be foregrounded in my field of awareness, rather than as the background of my own needs and desires, as they stand outside the genuinely caring attitude. Motivational displacement is the resultant feeling that “I must do something” (14), I must act to respond to the other’s reality, “to reduce the pain, to fill the need, to actualize the dream” (14). These criteria, of course, create quite a

high standard for genuine “caring”. Here it is evident that I cannot “care” remotely, or though routine tasks that may “secure credit” (24) for caring from others without requiring my engrossment in the needs and sufferings of the other.⁹ Thus if I transfer my obligations of care to a carebot, I have “done something” to care but not in a manner that will sustain *my* self as a caring being. And yet, Noddings notes that in all caring situations, there is a risk that “the one-caring will be overwhelmed by the responsibilities and duties of the task and that, as a result of being burdened, he or she will cease to care for the other...” (12).

She also tells us that the ethical ideal sustained through caring for others is itself subject to practical constraint (50). Since caring is concrete activity, I can only care within the limits of the possible, and a host of environmental, social, or institutional conditions are among those that can constrain and impoverish my ethical ideal. When we think, then, about the unrealistic emotional and physical burdens often put upon caregivers in contemporary society, it is hard to deny that the ethical ideal is *already* impoverished and constrained for many of us. Engrossment in the reality of the other is not possible when I have had so little sleep that I am at the limits of human capacity just to go through the routine motions of care—changing the bedpan, remembering to give medication, cleaning the wound. What more can I give? What reason can we have, then, to deny that the introduction of carebots in some contexts, and for some purposes, might be *essential* to sustaining the caregivers, so that they need not face the degradation of their own moral being? It is also worth noting that on Noddings’ view, caring relations need to be reciprocal if caregivers are going to be sustained. That is, caregivers need some acknowledgment of the relation on the part of the cared-for, and though one can continue to care as an ethical choice in the absence of such reciprocity, it is far more difficult to remain engrossed in responding to the reality of another who will not, or cannot acknowledge the value of your response. Can carebots help here? Borenstein and Pearson (2010) note that by providing cared-fors with greater independence, their own capabilities to respond to caregivers may be enhanced, promoting a “care recipient’s ability to ‘love those who love and care for us’” (282).

Consider another scenario, however—the one in which carebots appear to us as liberation *from* care rather than liberation *to* care. We have seen the emphasis on the idea that care must be a choice, that the freedom and autonomy of potential caregivers is of paramount importance. This seems to presuppose, however, that some of us will choose *not* to give care, and that perhaps this choice ought not to be regarded as a bad thing, as long as there are carebot resources available to meet the needs of cared-fors. As Noddings notes, we *are* free to reject the impulse to care (51). We have other desires, and other needs with which caring practices compete. And if we are exhorted by carebot marketers or nursing home managers to “let the robot do its job”, to surrender caring tasks to a being which will surely not forget a pill, will not lose patience, will not let a loved one fall, we might be mistakenly led to think that both we and our cared-for are best served by this surrender. Yet Noddings is right, I think, that “I enslave myself to a particularly unhappy task when I make this choice. As I chop away at the chains that bind me to loved others,

⁹ It is worth noting, with reference to our discussions of virtue above, that Noddings identifies engrossment with the capacity for empathy as “feeling with” (30).

asserting my freedom, I move into a wilderness of strangers and loneliness, leaving behind all who cared for me and even, perhaps, my own self" (51).

Thus it is one thing to choose *when* to care, and for whom, and to ensure that I receive the support I need in caring to sustain my self emotionally, physically, and morally. Carebots might help us do better with all of this, and in the absence of alternative solutions, it might be ethically necessary to employ them in certain contexts. But how many of us be seduced by the possibility opened up by carebots of *not* caring, at least not in the sort of intimate, direct relations that presuppose engrossment in the reality of the suffering or vulnerable other? And how can we protect ourselves against such a seduction unless we attend more carefully to the ways in which engagement in caring practices is critical for our own well-being, and not just that of those who need us?

6 Carebots and Human Capabilities

Finally, let us consider the capabilities approach. Capabilities approaches, rooted in the economic and political philosophies of Sen (1999) and Nussbaum (2000, 2006), presuppose that considerations of justice require attention to more than just the availability and social distribution of various external goods, resources, or utilities. Rather, thinking about justice requires attention to basic human capabilities and/or functionings, understood as goods realized in human living through certain kinds of activities and practices. As realization of these capabilities is critical to human flourishing, the requirements of justice demand that we not sacrifice their realization by individuals merely in order to maximize overall utility. While the capabilities approach was conceptualized to respond to deficiencies in other theories of justice (e.g., Rawls, utilitarianism), it can also be used more broadly in the context of ethics, in order to help us conceptualize the good life and the means of its realization.

And in fact it is used by philosophers who weigh the ethical implications of carebots—Borenstein and Pearson (2010, 278) employ this approach in order to show that, for example, carebots might allow caregivers to meet their moral obligations without destroying their own emotional health (a loss that might undermine, for example, their human capacities for love, play or friendship). They also note that the capabilities approach helps us identify how carebots should *not* be used—for example, in ways that would deprive caregivers or cared-for of control over their environment (279). Consider in this regard a carebot that secured the physical well-being of nursing home patients by rigidly confining their movements to prevent any risk of a fall—leaving them socially isolated and deprived of autonomy. Coeckelbergh (2010) also references the capabilities approach as a model for ethical reflection on carebots. Though he does not take the capabilities approach to be exhaustive of the moral issues at stake with carebots (186), he suggests we use it as one way to evaluate whether any given practice of caregiving qualifies as “good care” (185).

I suggest we employ the capabilities approach in one further respect: to help us to conceptualize the goods internal to caring practices for *caregivers*, goods that might be lost should a person choose to surrender some or all of such practices to a carebot. How should we proceed? Following Coeckelbergh (2010) and Borenstein and

Pearson (2010), let us refer to the list of central human capabilities offered by Nussbaum (2006), without suggesting that this list is necessarily complete or definitive. The list identifies and defines the capabilities of life, bodily health, bodily integrity, sensation, imagination and thought, emotions, practical reason, affiliation, [relatedness to] other species, play, and control over one's environment (76–77). While the goods internal to caring practices might be incorporated in all of these capabilities, I wish to focus on just three, *affiliation*, *practical reason*, and *emotion*, which I will argue each represent capacities enhanced for *caregivers* within caring practices. Nussbaum describes these capacities as follows:

6.1 Affiliation

(Part A) “Being able to live with and toward others, to recognize and show concern for other human beings, to engage in various forms of social interaction; to be able to imagine the situation of another.”

Nussbaum (2006, 77)

6.2 Practical Reason

“Being able to form a conception of the good and to engage in critical reflection about the planning of one's life.”

Nussbaum (2006, 77)

6.3 Emotions

“Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude and justified anger. Not having one's emotional development blighted by fear and anxiety. (Supporting this capability means supporting forms of human association that can be shown to be crucial in their development).”

Nussbaum (2006, 76–77)

I assert that caring practices, and more specifically, engagement in such practices *in the role of caregiver as well as one cared-for*, are critical for the development of these capabilities, for both men and women. Let me begin with affiliation. The capability of affiliation described above entails the ability “to recognize and show concern for other human beings.” How do we develop this ability? What human practice is most critical here? I would argue that caregiving practices, while not exhaustive of those through which we develop the capacity of affiliation, are essential to its full development. Certainly I may show concern for the general “other” by means other than giving care—for example, by giving money to charity, through social and political advocacy, or through throwing myself into intellectual and practical activities aimed at addressing important problems that others face. But it is through directly giving care that I learn to recognize in concrete others those expressions of need and desire that first motivate the attitude of “concern” on my

part. Even as a child I come to this attitude through caring practices—by looking after a younger sibling, feeding and watering a pet, fetching Kleenex for a flu-afflicted parent, or patiently listening to the problems of a friend. It is through these interactions that I come to understand what human “affiliation” means, as a way of being with others that is responsive and mutually supportive. It is primarily through repeated concrete exposures to human need, and taking up the practice of responding to need with care, that I develop the abstractive capacity to “imagine the situation of another” outside my own intimate circle of care.

Let us next consider the cultivation of practical reason. Here caregiving also plays a central role. First, it can be argued that no “conception of the good” is complete that does not recognize caring relations as a basic human good.¹⁰ That is to say, we would regard a person as prudentially deficient whose deliberations about action were never informed by an appreciation for the goodness of one human caring for another. Nussbaum herself tells us that we must appreciate “the importance of care as a social primary good” (2006, 2). Now, we might think that such an appreciation could be acquired simply from the experience of being cared *for*, and that the practice of *giving* care, which I have emphasized, is inessential here. I reject this claim, however, for reasons that should be apparent from the above considerations, and which will be reinforced by the discussion of emotional capability that follows. It is through *giving* care that I learn to recognize signs of need in another, that I become habituated to responding to need, and that I come to fully appreciate the goodness of the caring *role* and the importance (and challenge) of caring *well*. As cared *for* I can see the goodness of having my needs met—but I do not fully grasp the goodness of meeting needs until I have taken on that role myself and experienced its challenges and rewards. Without engaging in caregiving practices my grasp of the goodness of caring relations remains impoverished and one-sided.

Secondly, it is at least in part through caregiving practices that I become capable of the prudent “planning of one’s life” that Nussbaum speaks of. How can I possibly plan my life well unless I am aware of the vulnerabilities of sickness, injury, age, or other conditions of life to which I am, sooner or later, going to be exposed? Nor is a merely intellectual grasp of these conditions sufficient for planning. It is not enough to know that I need, for example, health and disability insurance, or adequate savings, or an advanced health care directive on file. I also need to plan for the emotional, physical, social challenges that these conditions may bring, to prepare myself through reflection and conversations with loved ones for how we can together bear those conditions with dignity and grace, if possible. And it is primarily through witnessing others go through those challenges, and more specifically, learning through caregiving practices *how* humans find, or lose, dignity in their vulnerability, that I become able to reflect deeply on how I might secure the dignity of my own future. Thus the capability of prudential reason seems to require for its *full* realization a deep acquaintance with caring relations, not only from the standpoint of the cared-for but also from the standpoint of a caregiver.

¹⁰ Noddings argues for a stronger view that caring relations are the first and ultimate source of our conception of the good (1984, 99). Even if this view is rejected as too strong (which I do not assert), the weaker claim that an appreciation of caring relations is *essential* to any adequate conception of the good has intuitive plausibility.

Finally, how are the emotional capabilities Nussbaum speaks of realized through the practice of giving care? To see this, it may help to reflect first upon the moral psychology of the sort of person who never becomes comfortable or skilled in caregiving. This person refuses to take on a caring role out of fear of emotional exposure and anxiety about becoming “too attached” to the being who is cared-for, too vulnerable to the possibility of their loss and the attendant grief, too pained by unrequited longing for the cared-for to be made whole; or too angry, frustrated, or exhausted by the burdens that care entails. Most of us, I think, will find some or all of these fears recognizable—we have seen them expressed by others, or had such fears ourselves. In such a condition, one either takes on the caring role in spite of such fears (the reasons why need not matter here) or one abdicates that role—leaving it empty *or* leaving it to be filled by someone else. What difference does this choice make? A great deal of difference. The person who entirely succumbs to these fears lives as an absent, “checked-out” parent, especially during a child’s illness or disability. They find excuses to avoid sick friends recuperating at home. They cite a hatred of hospitals as a reason to limit visits to dying parents. They become withdrawn and distant with a terminally ill spouse. They may attempt to compensate by “assisting” in ways that do not require much emotional exposure—working overtime to help with medical costs, sending flowers and food baskets by courier, buying an expensive van with a wheelchair lift. But as long as they continue to avoid *caregiving*, that is, offering the full range of their emotional and physical being for the good of one who is suffering, dependent, or vulnerable, their capacity for a rich emotional life remains stunted and impoverished.

It is not her fears that consign such a sad person to her fate—we likely all have such fears. It is the abdication of caring practices that engraves those fears as the dominant feature of such a person’s emotional landscape, crowding out all other possibilities. Meanwhile, the person who takes up caring practices—even, perhaps, out of guilt or in response to social expectation—is forced to confront those fears with action. Caring actions will indeed bring precisely what the carer fears—grief, longing, anger, emotional exhaustion. But they can also allow those emotions to be gradually incorporated into a whole life, into living and sustaining attachments with other things and persons. Caring practices bring other emotions too—such as the gratitude Nussbaum speaks of, whether it be for that one day the clouds of dementia lifted a little and a parent’s recognition reappeared, or the smile of an arthritic spouse after a gentle massage, or gratitude for the child who begins to show maturity, independence, and her own capacity to care.

Of course, caring practices in the absence of adequate resources and support can have a powerfully destructive impact on our emotional, prudential, and social capabilities—the unmitigated stress, anxiety, physical, and mental exhaustion can result in emotional withdrawal, a gradual numbing of affect, or extreme emotional volatility—any or all of which can result in fractured relations with cared-fors and others, severe depression, motivational apathy, impaired judgment, even complete emotional collapse. Borenstein and Pearson (2010) are correct that caregivers “should be able to fulfill their obligations without simultaneously relinquishing their own prospects for human flourishing” (278). Yet by not attending explicitly to the way in which sustaining caring practices may be an integral part of that flourishing,

the potential of carebots to support human caregivers and relieve their burdens appears unequivocal, without the concern that must complement it—that humans may (deliberately or unwittingly) surrender opportunities for giving care that are critical to their *own* development and well-being. Thus the proper question is not whether carebots are inherently destructive or supportive of the development of central capabilities in caregivers. The proper question is, *in what particular, concrete contexts will the use of carebots enhance caregivers' capabilities by allowing them to sustain caring practices, and in what contexts will their use encourage us to gradually surrender those practices, perhaps blind to the way in which such a surrender may negatively impact our fundamental capabilities?*

Repeatedly, discussion of these issues has skewed toward the impact on cared-for rather than caregivers. Where the impact of carebots on caregivers is made central, it is entirely in terms of the potential benefits of being liberated from caregiving obligations, without consideration of the attendant risks of such liberation (Borenstein & Pearson 2010). For example, consider the dispute about whether carebots will in fact come to gradually replace human caregivers, or whether by freeing us from routine tasks they will “make room for more and deeper human care” (Coeckebergh 2009, 183), helping us to “care better”. This is an important question, and the answer depends upon a number of social, economic, cultural, political, and technical variables that I will not attempt to enumerate. Some authors’ predictions about the future of carebot–human relations are distinctly pessimistic (Sparrow & Sparrow 2006), others more cautiously optimistic (Sharkey & Sharkey 2010; Borenstein & Pearson 2010). But in either scenario, the risks and benefits are weighed in a manner that ignores or minimizes the potential value of caregiving to caregivers. For example, there is a consensus that “the elderly need contact with fellow human beings,” and that if the use of carebots led to depriving them of human contact, this would be ethically problematic (Sharkey & Sharkey 2010; Borenstein & Pearson 2010). But no one asks whether we (potential caregivers) need contact with the elderly! Likewise, if using carebots to help us “care better” is seen as an ethically significant gain, it is seen as a gain in terms of the quality of care given to cared-for—they receive “deep” rather than “shallow” care, which is clearly better for them (Coeckebergh 2009). But it has not been asked whether being freed to “care better” might also be better for caregivers, by allowing them greater access to caregiving practices that sustain or enrich their emotional, prudential, and affiliative capabilities.

This impacts how we think about assessing the ethics of particular carebot uses—we focus on strategies that emphasize asking the elderly what levels of human/robot interaction, respectively, they prefer (Sparrow & Sparrow 2006; Sharkey & Sharkey 2010). The preferences of cared-for are obviously critical here, and must be consulted as part of any ethical approach. But by ignoring the value of caregiving practices for caregivers, we open ourselves to a disturbing (though perhaps unlikely) scenario. Imagine a century from now that carebots have been developed which are not only cheap, safe, and skilled, but good at keeping the elderly, children, and convalescent persons *happy*—amused, comforted, and distracted from their needs or pains. Imagine that a great many people in this future would *prefer* to be cared for by carebots than human caregivers, who still at times become impatient, dull, or angry in their company. How would widely respecting those preferences, and greatly

decreasing human involvement in caregiving practices, impact the development of the central human capabilities we have discussed here? How would this impact the flourishing of human beings in the long run? Alternatively, imagine a scenario where the elderly or otherwise vulnerable prefer for aesthetic or tradition-bound reasons *not* to have robots involved in their care, but empirically we find that (1) limited uses of carebots for routine, physically demanding or tedious tasks free caregivers to sustain and enrich their own central human capabilities to a significant degree, and (2) that such uses have no negative impact on the central human capabilities of cared-fors. Is it *clear* that the preferences of cared-fors would be ethically definitive in such a case?

There is yet another question we must pursue, related to the way in which new technologies can radically change our contexts of practice in such a way as to alter the meaning of the goods internal to those practices. For example, Coeckebergh (2009) tells us that interactions with robots may not only support the development of a central human capability (such as play) but may also “redefine that capability” (220). This requires careful reflection. What would it mean to “redefine” the various capabilities that come together in, and are cultivated through, caregiving practices? How plastic is care? Can we stretch the definition of our capabilities in such a way that we could accept the marketing of a carebot as “enabling friends and family to care from afar” (Sharkey & Sharkey 2010)? Could the goods realized in traditional caring practices also be realized through “remote caring”? Or does this stretch the meaning of a caring practice beyond reason, detaching it from the sort of goods or capabilities that sustain human flourishing? My aim in this paper is simply to shed light on the importance of asking these questions.

Deciding whether our primary moral concern about carebots and caregivers should be the potential loss of opportunities to cultivate virtues like reciprocity and empathy, or potential constraint and impoverishment of the ethical ideal cultivated in the caring relation, or impairment of the capabilities of affiliation, practical reason, and emotion cultivated in caring practices, is a task for moral philosophers. This may not be their most important task, however—partly because these three perspectives, by focusing on the goods internal to caring practices, are already deeply complementary in many key respects. But beyond this, it is critical that the task of perfecting our conceptual framework not occlude the concrete, empirical developments upon which we hope our framework will shed light, and which give our attempts at theoretical precision their sole meaning. It has yet to become clear what carebots will be able to offer us beyond their current incubatory stage, or when they will become a genuine choice for potential caregivers as opposed to a matter of fascinating speculation for philosophers of technology. Nor do we know precisely which cultural, economic, demographic, and psychological conditions will be operative to influence caregivers’ motivations and choices if and when that day comes. It will be important for philosophers to attend closely to future developments in these respects, and to adjust their range and degree of moral concern accordingly. In the meantime, philosophers can make a distinctive contribution to preparing the moral community for such choices by encouraging reflection on the value of caring practices, not only for those cared for, but for caregivers as well.

References

- Aristotle (1984). *The complete works of Aristotle: Revised Oxford translation*. J. Barnes, Ed. Princeton: Princeton University Press.
- Becker, L. C. (1986). *Reciprocity*. Chicago: University of Chicago Press.
- Borenstein, J., & Pearson, Y. (2010). Robot caregivers: harbingers of expanded freedom for all? *Ethics and Information Technology*, 12(3), 277–288.
- Coeckelbergh, M. (2010). Health care, capabilities and AI assistive technologies. *Ethical Theory and Moral Practice*, 13(2), 181–190.
- Coeckelbergh, M. (2009). Personal robots, appearance and human good: A methodological reflection on roboethics. *International Journal of Social Robotics*, 1(3), 217–221.
- Decker, M. (2008). Caregiving robots and ethical reflection: The perspective of interdisciplinary technology assessment. *AI & Society*, 22(3), 315–330.
- Gilligan, C. (1982). *In a different voice*. Cambridge: Harvard University Press.
- Held, V. (2006). *The ethics of care: Personal, political, global*. Oxford: Oxford University Press.
- Kittay, E. F. (1999). *Love's labor: Essays on women, equality, and dependency*. New York: Routledge.
- Noddings, N. (1984). *Caring: A feminine approach to ethics and moral education*. Berkeley: University of California Press.
- Nussbaum, M. (2006). *Frontiers of justice: Disability, nationality, species membership*. Cambridge: Harvard University Press.
- Nussbaum, M. (2000). *Women and human development: The capabilities approach*. Cambridge: Cambridge University Press.
- Sen, A. (1999). *Development as freedom*. New York: Knopf.
- Sharkey, A., & Sharkey, N. (2010). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology*. doi:10.1007/s10676-010-9234-6.
- Singer, P. W. (2009). *Wired for war: The robotics revolution and conflict in the 21st century*. New York: Penguin Press.
- Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16(2), 141–161.
- Tergesen, A., & Inada, M. (2010). It's not a stuffed animal, it's a \$6000 medical device: Paro the robot seal aims to comfort the elderly, but is it ethical? *The Wall Street Journal*, June 21, 2010, <http://online.wsj.com/article/SB10001424052748704463504575301051844937276.html>. Accessed January 8, 2011.
- Turkle, S. (2006). *A nascent robotics culture: New complicities for companionship*. July 2006: AAAI Technical Report Series, http://web.mit.edu/sturkle/www/pdfsforstwebpage/st_nascentroboticsculture.pdf. Accessed January 8, 2011.
- Vallor, S. (2010). Social networking technology and the virtues. *Ethics and Information Technology*, 12(2), 157–170.
- Vallor, S. (2011). Flourishing on Facebook: Virtue friendship and new social media. *Ethics and Information Technology*. doi:10.1007/s10676-010-9262-2.