Theor Med Bioeth (2015) 36:265–277
DOI 10.1007/s11017-015-9331-y

Artificial agents, good care, and modernity

Mark Coeckelbergh[1]

Published online: 23 May 2015

Abstract When is it ethically acceptable to use artificial agents in health care? This article articulates some criteria for good care and then discusses whether machines as artificial agents that take over care tasks meet these criteria. Particular attention is paid to intuitions about the meaning of 'care', 'agency', and 'taking over', but also to the care process as a labour process in a modern organizational and financial-economic context. It is argued that while there is in principle no objection to using machines in medicine and health care, the idea of them functioning and appearing as 'artificial agents' is problematic and attends us to problems in human care which were already present before visions of machine care entered the stage. It is recommended that the discussion about care machines be connected to a broader discussion about the impact of technology on human relations in the context of modernity.

Keywords Health care · Care · Machines · Robots · Artificial agents · Labour · Modernity · Ethics · Ethics of health care · Care robots

Introduction

Tools and artefacts have always been used in medicine and health care, and most contemporary medical technologies are widely accepted and relatively uncontroversial. Robots and other automation technologies, however, seem to be an

A previous version of this article was presented at *AISB 2014*, Goldsmiths college, London, April 2014.

✉ Mark Coeckelbergh
mark.coeckelbergh@dmu.ac.uk

[1] Centre for Computing and Social Responsibility, De Montfort University, 5.79 Gateway House, The Gateway, Leicester LE1 9BH, UK

exception to this rule and raise new ethical questions concerning the future of medicine and health care. For instance, is it acceptable to replace human nurses with artificial nurses? Should surgery be automated? Can an expert system be trusted to make decisions about life and death? Can elderly people be left at their homes in the hands of a robot?

Today, machines are already used in medicine and health care. Think about surgical robots remote-operated by a surgeon who might be physically located thousands of miles away, or mobile devices that function as interfaces through which doctors and nurses can communicate with their patients. While these machines are certainly not free from potential ethical objections (e.g., with regard to the quality of tele-care as opposed to proximity between caregiver and care receiver), the more extreme examples above seem to derive their controversial nature from what we may call their 'artificial agency'. Such machines are not merely tools used and operated by humans; they also act as agents which carry out tasks usually assigned to humans. They do not merely help the nurse but replace the nurse. They do not just assist the surgeon but replace her. They do not only provide information to the doctor but make doctor's decisions. In other words, they are artificial medical agents.

Although today it seems highly unlikely that in the foreseeable future *all* care givers will be replaced by machines for *all* tasks, given current developments in the field of medical and care technologies towards more autonomy and automation, and given the increased financial-economic pressure on our current health care systems, it is important to reflect on the ethical issues raised by this 'artificial agency' dimension of machines. The main question I address in this article is therefore as follows: Is it ethically acceptable to use artificial agents in health care, and if so, when and under what conditions?

While most philosophers recognize that there is a potential tension between the quality of care and the use of machines in health care, some are more optimistic than others about the possibility of bridging that gap. For instance, some authors, such as Wendell Wallach and Colin Allen, think that the solution lies in trying to create ethical or moral machines [1, 2] which have a built-in capacity to make ethical decisions. Others have voiced concerns about the reduction of human contact in robotic care [3, 4], a kind of care which seems especially problematic in the case of vulnerable people, such as those who are ill, young children, and elderly people [5].

In this article, I side with the latter concerns, but (1) offer a more comprehensive articulation of what good health care is assumed to be in objections to care robots and care machines (e.g., intuitions about the importance of human contact and "warm" care), (2) further discuss what specific role machines can or cannot be given within medicine and health care (when exactly it is not acceptable to use a machine), and (3) draw attention to what has usually been neglected in discussions within the field: the relation between the quality of care and the labour process, and more broadly, the relation between the quality of care and modernity.

The first section of the article delivers a number of criteria for *good health care*. In constructing these criteria, I briefly refer also to implications for the use of machines in health care. The second section offers guidance for the introduction of *machines acting as artificial agents* in health care. I discuss whether machines as

artificial agents can meet the criteria articulated in the first section. I raise the questions of what technological 'agency' means and what it means to say that a task is 'delegated to' or 'taken over by' a machine. The third section discusses the two previous questions in the context of the labour process and modernity. I will argue that even if machines acting as artificial agents are not (yet?) widely used in health care, there is a sense in which 'the machine' has already arrived, is already there—a sense which is currently being neglected in discussions about care machines within the field of machine ethics and robot ethics. Next to the visible machines in health care, there is also an invisible 'machine' within which care givers work and which raises ethical and philosophical issues regarding the quality of health care.

In the course of my arguments, I will also make remarks about the epistemology I use in my approach, which, in line with my previous work in the field of ethics and ethics of robotics, is one centred on 'appearance'. Influenced by a phenomenological approach (mainly Martin Heidegger, but also contemporary authors such as Hubert Dreyfus and Don Ihde), I have argued that ethics of robotics should not assume an 'objective' view of the robot and technology; our knowledge of technology is always influenced by how it appears to us. The ethics of technology I propose starts from the lived experience of technology—and indeed of care. This is important enough to mention, since anyone writing about the quality of (health) care needs to be aware of the question: from which point of view should the quality of care be evaluated? How can its quality be known? What kind of knowledge can I, as a philosopher, gain about health care? My discussion will not settle these questions or discuss them comprehensively, but indicate the direction in which I try to answer them.

Good care

In order to construct a conceptual instrument for normative analysis and evaluation of artificial agents in care, I will first articulate a normative ideal of good care, which is always a particular way of looking at care. What is good health care?

Note, however, that although this is the logical order of my argument (i.e., first criteria, then evaluation of machines according to the criteria), my articulation of the following criteria is already informed by thinking about the use of technology in care and its ethical implications. Thus, this section already aids thinking about machines in health care and is also an explicit recognition of the influence machines have on our *thinking* about good care (that is, not only on the practice of care).

I propose the following 'working criteria' of good care, many of which are often assumed when philosophers object to using robots in health care. Although it is not my main purpose here to defend and discuss the criteria and to extensively engage with the literature, I will draw on at least two kinds of sources for *articulating* the criteria: (1) discussions about machines (i.e., robots) in health care and (2) theoretical frameworks for thinking about good care, in particular but not exclusively those influenced by the phenomenological and, to some extent, Aristotelian tradition.

Good care attempts to restore, maintain, and improve the health of persons.

This criterion may seem obvious, but it is not. Consider for instance discussions about euthanasia: is euthanasia part of good health care or not? Moreover, palliative care does not aim to restore health but, rather, tries to achieve the best quality of life for patients with an incurable illness. In general, however, the criterion expresses a key point about what good health care, elderly care, etc. is supposed to do, and technology is meant to contribute to this goal. Other goals and other uses of technology are therefore by definition ethically controversial (think for instance about a technology which assists with euthanasia).

Good care is practiced within the boundaries set by bioethical principles and professional codes of ethics.

Good care remains within the ethical boundaries widely recognized in standard bioethical ethics, such as autonomy, respect for (patient) autonomy, non-maleficence, beneficence, and justice [6], and within the boundaries of relevant professional codes of conduct. However, such principles and codes mainly set negative moral limits to a practice. They are not very helpful in articulating a positive ideal of good care, and rarely play a role in discussions about machines in health care. A much more important criterion, which is central to most objections against 'care robots', is the following.

Good care involves a significant amount of human contact.

In discussions about care robots for elderly care, an important concern is that the care robots will reduce human contact. In the field of robot ethics considerable attention has been paid to this issue [3, 4]. Nightmare scenarios are sketched in which elderly people are abandoned and left in the hands of machines, shielded from the rest of the world. For example, Robert Sparrow and Linda Sparrow write:

> We imagine a future aged-care facility where robots reign supreme. In this facility people are washed by robots, fed by robots, monitored by robots, cared for and entertained by robots. Except for their family or community service workers, those within this facility never need to deal or talk with a human being who is not also a resident. It is clear that this scenario represents a dystopia rather than a Utopia as far as the future of aged care is concerned. [4, p. 152]

Even if such scenarios are unlikely to become reality, the normative concern is clear: whether or not society cares for elderly persons, sick persons, or children, good care should always involve a significant amount of human contact.

Good care does not only mean physical care but has also psychological and relational, e.g., emotional, dimensions.

This criterion takes a broad view of 'care' and 'health'. If care giving were only about doing particular physical tasks, and if health were only a matter of the functioning of the 'body-machine', then care could easily be delegated to machines and there would be little controversy about, for instance, care robots. But most

would agree that care and health is more than performance of physical tasks, and includes emotional and relational aspects. This is also partly why in the ethics of robotics, human presence and human touch is emphasized: it is assumed that care includes an emotional aspect, and that only humans can therefore be good care givers. For example, a nurse talking to a patient is not an activity that stands outside the care process but should be part of it. For human beings, having good social relations belongs to what it means to lead a healthy life. Therefore, a robot that is used as a replacement for a human care giver but only administers physical tasks, or a care process that neglects social relations, would not count as contributing to good care.

This view is in line with relational theories of health care and nursing, for example, phenomenological-hermeneutic approaches to care based on Heideggerian thinking [7] and care-oriented and often feminist currents in moral theory, such as 'the ethics of care' [8–10], which focus on social relationships, embodiment, dependency and vulnerability, and emotion.

Good care is not only professional care but should also involve relatives, friends, and loved ones to a significant degree.

Contemporary health care is professionalized to a high degree, and most people would agree that this has—next to obvious advantages—also important disadvantages: it contributes to the impersonal character of care. Most people would think that involvement of relatives, friends, loved ones, etc. is very important for good care and health. This connects again to the relational dimension of good care and to theories of care which highlight and value informal, non-professional forms of care.

Good care is not (only) to be experienced as a burden but can also be (experienced as) meaningful and valuable.

In contrast to machines, people reflect on, evaluate, and give meaning to what they do, and this includes reflection on, and giving meaning to, care for others. Generally, care giving can be assumed to be a meaningful activity, which may be burdensome and difficult at times, but which is a good and meaningful thing to do. This opposes the view that health care is a kind of 'dirty job' which needs to be handed over to robots because no-one wants to do it and no-one should do it (compare to 'dirty' labour in industrial production processes). Again this view of care is in line with thinking about care that values care, rather than seeing it as something that needs to be done and is without moral quality and significance.

Good care involves skilled engagement with the patient (know-how), next to more formal forms of expertise (know-that).

Good care involves not only knowing-that (theoretical knowledge) but also knowing-how. Of course every care giver needs medical knowledge, but an important part of what (s)he needs to be able to do is *practice* good care, and this is a much more implicit kind of knowledge that is learned by doing and embodied. Care givers (professionals and others) have to learn care skills. The knowledge they have to acquire is also and crucially a know-how; they have to acquire an art of caring. Elsewhere, I have called this 'care craftsmanship' [11]. Some machines, or

rather, the persons who use them, may be said to 'know-that'; but it is doubtful if they ever attain sufficient 'know-how', implicit knowledge, or the embodied and social cognition of human care givers.

This view of care is in line with thinking about the importance and value of skilled engagement. In the phenomenological tradition, it is important to study everyday ways of skilful coping, how people as embodied and meaning-giving beings acquire skill in everyday contexts and practices. For instance, Dreyfus has focused on skilled know how, skill acquisition, and ethical expertise (see, for example, [12]), and this has inspired work on skills in nursing (see again [7]).

Good care requires an organizational context in which there are limits to the division of labour so as not to make the previous criteria impossible to meet.

If there is a high degree of division of labour in care, then this makes it difficult, for example, to give enough personal attention to patients and talk to them, or it makes it difficult to give meaning to the care giving. Good care therefore seems to need limits to the division of labour in order to maintain the quality of care. This criterion is also strongly related to the professionalization issue: care tends to become an assembly of specialized care tasks, which can only be done by specialists. Care then may become 'machine' like (see also below), a 'care machine' in which non-professionals have no longer a place or are reduced to passive bystanders, entirely dependent on professional care givers. In the section on modern care, I will say more about division of labour and introduce Marxian thinking.

Good care involves an organizational context in which financial-economic considerations are not the only or not even the main criterion in the organization of care.

Of course, financial and economic considerations are important; they need to be taken into account at the household, organizational, and state/society levels. But if such financial-managerial considerations are too dominant, the aims of a practice—in this case, health care—may become secondary or their attainment may come under too much pressure. Arguably this is happening today, to the extent that discussions about the quality of care are reduced to problems of management and finance.

Good care requires the patient to accept some degree of vulnerability and dependency on others.

Articulating good care is not only about making explicit what care givers should do or what attitude they should take. It is also about what care receivers should do. In modern culture, personal autonomy is often valued so much that it becomes difficult for people to accept that at some moments or periods in their lives, they were and will be highly dependent on others for their health and existence, and that however independent and autonomous they may be or feel, their existential condition gives them vulnerabilities which do not disappear by neglecting them. Good care, then, also means acceptance of some degree of dependence and vulnerability by the care receivers.

More generally, relational and feminist approaches to care have emphasized vulnerability and dependence on others, often against a history of philosophy that has not given a central place to these existential facts. This view is prominent in the 'ethic of care' theory mentioned previously, but also for instance in Aristotelian thinking (e.g., [13]). As Alasdair MacIntyre put it,

> We human beings are vulnerable to many kinds of affliction and most of us are at some time afflicted by serious ills. How we cope is only in small part up to us. It is most often to others that we owe our survival, let alone our flourishing, as we encounter bodily illness and injury, inadequate nutrition, mental defect and disturbance, and human aggression and neglect. [13, p. 1]

These criteria of good care are not uncontroversial. There are certainly less broad and less relational definitions of care, and some readers will disagree with one or more criteria. But for the sake of argument I will assume this richer and relational normative ideal of good care (and health). What does this mean for the question regarding machines in health care, in particular, machines that act as *artificial agents* in health care? I already made some suggestions about implications for machines, but now the questions regarding 'artificial agency' need further attention.

Artificial agents and care

By itself, the involvement of technology in care is not problematic, at least if one presumes that it contributes to the first criterion of health care: it clearly *does* usually help care givers restore, maintain, and improve health. As mentioned in my introduction, medicine and health care have always used tools for this purpose. But some uses of technology *are* problematic and need further discussion. The key question in this section is what happens in a situation where machines *appear* as artificial *agents.*

For a start, what does 'agency' mean here? I make a distinction between two forms of technological agency, a 'weak' and a 'strong' form. All artefacts may have a minimal form of 'agency' within the context in which they function, in the sense of having some influence on how things are done and even on what is done. This kind of agency is generally recognized by philosophers of technology and is not very controversial. For example, if I travel by foot, by bike, or by car, the choice of my means of transportation shapes my practice and experience of travelling. I do different things, I experience the world around me in a different way, and so on. Similarly, medical technologies have also changed medical practice. Think about the MRI scanner in the hospital, the computer at the GP's practice, etc.: these technologies change how diagnosis is done, how care givers communicate with care receivers, etc.

The idea that technology has all these forms of 'weak' agency is not obvious. Many people would accept that technology changes the way we do things, including in medicine and health care. But not everyone is prepared to accept that technology also has agency with regard to knowledge and experience. Usually, knowledge and technology, and culture and materiality, are separated. Yet, phenomenology of

technology, for instance, in the work of Ihde [14], has emphasized the latter kind of agency: technology shapes our experience, our 'world' in a Heideggerian sense. For health care, this means that with new medical and care technologies, new experiences—by care givers and care receivers—emerge. For instance, the MRI scanner does not only 'represent' a knowledge of the human body that was already there, but reveals the human body in a new way, lets it appear in a new way. Technology is thus far more 'active', in epistemological and hermeneutical terms, than is supposed by people who think that technologies are mere instruments.

However, the concept of machines as artificial agents in care amounts to a far stronger idea of 'agency', one that is similar to human agency. Here, I am considering machines that are (more) autonomous, that *take over* tasks previously done by humans, and that 'work with people' (they are not part of an assembly line in a factory, they work with human patients and human colleagues; consider also, again, the relational aspect of health care). The idea under consideration in this article is that machines could take over care tasks, that tasks previously done by humans are delegated to these machines, which do not merely influence how things are done but also do things themselves; they act as artificial agents. The analogy to transportation technology is an autonomous, robotic car: a car that drives itself, and involves itself in human traffic (and, indeed, mixed human/non-human traffic: it is 'on the road with people'). Such a care is also an 'artificial agent'.

Here, one enters a field of thinking about the agency of machines and the ethics of artificial agents (see, for instance, [2, 15–20]). If a robot has this form of 'strong' agency, then is or can such an agent also be 'moral', and if so, in what sense? Could it be(come) conscious? What kind of ethics would be needed to govern such artificial agents? These are important general questions. However, here, I wish to focus on the problem of artificial agency in relation to good care. The 'artificial agency' dimension raises a number of problems in relation to the criteria just articulated. Let me discuss some of the most important ones.

A first major problem concerns the question about replacement: is it ethically acceptable to replace human care givers by artificial ones? Is it ethically acceptable that machines *take over* human tasks? But what does this mean? Does it mean that all humans should and can be replaced by machines? Even philosophers who do not have principled objections against such 'artificial agent' machines in health care will not want a situation in which there are *only* machines in medicine and health care. But the point seems to be that the 'taking over' itself is problematic. As suggested in the previous section, there is widespread agreement that human contact is necessary, and most people will agree that emotional and relational contact also belongs to good care. Now, if I am right about this ethical consensus, then it means that machines in health care are not acceptable insofar as they *take over* particular human tasks, at least if we assume that machines cannot provide such emotional contact (see below).

But what does it mean to 'take over a task'? For instance, when I use the spell checker of my word processing programme, does the programme 'take over' this task? If I use my navigation device for driving, does the device 'take over' driving? If a robot helps to lift a patient, does that machine 'take over' the task of lifting? In these cases, human agency is still involved in the tasks and the 'taking over' or

'delegation' seems to be a matter of degree. It is not easy to define what exactly happens and how this form of agency must be understood. Moreover, from a phenomenological perspective, whether or not a task is 'taken over' or 'delegated', or the extent to which a task is 'taken over' or 'delegated', should not be defined outside human subjectivity; humans (including the patients of course) experience what happens. Now, in order to avoid having to assess what 'really happens' (as it were from the outside, 'objectively') and to take into account and emphasize the 'internal' point of view, the perspective and lived experience of the actual people involved in care (the care givers and care receivers), I propose to qualify the criterion as follows:

> Machines in care are unacceptable if and *insofar as* the machines *appear* as agents that take over care tasks.

This approach to machines in care is in line with the phenomenological approach I proposed in robot ethics [17, 18]: what counts for understanding and evaluating robots and these forms of agency is not what the robot 'really is' or what the machine 'objectively' is but how it appears to us. From the perspective of Heideggerian phenomenology, we lack direct access to reality; our knowledge and experience of technologies, humans, etc. are always mediated. There are different ways in which something may appear to us; it may reveal itself in different ways. Therefore, whether or not a robot has 'agency' depends on how that robot is perceived in a particular situation, context, and practice.

Thus, if the care robot is perceived as a tool, then the (perceived) agency remains on the side of the human. The agency of the human care giver and the corresponding responsibility to provide human contact and 'care' in the sense of 'care about', or concern, remains intact. Then the care giver can be said to (give) care *with* the tool, *with* technology. The machine mediates the care action but does not itself have agency. However, insofar as the machine is perceived as 'taking over' the task of care and as taking on the role of the human care agent, then, if the ideal of care articulated above is assumed, it seems that something would be expected from the machine that the machine cannot give: the machine cannot take up this role and responsibility, cannot care in the ways defined above. It may appear to have emotions [18] but unable to fulfil what is needed for care as articulated above. Then it is not acceptable to use the machine in this way—that is, in such a way that the appearance of 'artificial agency' is produced—and ways have to be found to develop and use machines as 'mediators' and 'assisting tools' rather than machines that appear to 'take over'.

Of course, this argument assumes that machines do not have emotions, cannot be really concerned, cannot exist as embodied and relational beings, etc. Some philosophers might disagree, but they have the burden of proof. Note also that this argument against using machines in care that appear as artificial agents does not necessarily exclude the use of robotic pets or similar artefacts, which may function and appear as agents indeed, but do not *necessarily* appear as taking over the *role and task* of the care giver. For instance, they may appear as *recipients* of care, care receivers (for example, the robot PARO is often perceived and used in this way). This may of course raise different ethical issues, which I shall not discuss here (for

instance, the issue of deception has been discussed elsewhere—including by the author).

Questioning modern care

A second major problem that needs to be discussed, and which so far has been neglected in the machine ethics literature, concerns not so much the use of machines in itself but more the context of care—with or without machines. As was already suggested in the section on good care, these reflections on the role of technology attend us to the ways health care is organized and practiced today and invite us to critically reflect on them. Even without explicitly considering the role of technology or machines, current health care is highly professionalized, is often experienced as a burden, seems to have an emphasis on formal forms of expertise rather than know-how and craftsmanship, is usually done in an organizational context in which there is a high degree of division of labour, and is often discussed in financial-economic terms alone.

I propose to understand this development (or tendency) as part of the cultural-material process and a way of framing experience that is usually named "modernity", which has its brighter and its darker sides. Contemporary health care as 'modern' health care means, for instance, that it is more efficient care based on science and regulated by well-meant regulations, but it also means that care giving work is divided into small units, is calculated, is professionalized, formalized, and regulated. As Marx, Weber, Heidegger, and other classic theorists of modernity have pointed out, this modern way of organizing labour and of looking at things inevitably leads to objectification (or reification) and alienation. Let me explain.

Marx has argued that because of division of labour, humans lose the power to determine the purpose of the work. It *automatizes* the worker, who does 'very one-sided and machine-like type of labour' and is thus being 'reduced to a machine' [21]. Workers thus become alienated from their labour, but also from the products of their labour. If they have to work in this way, the workers can no longer feel satisfied about the results of their labour; the workers are estranged from the products of their work. Moreover, Heidegger has argued that in our modern technological age, we frame the world as a 'standing reserve' [22], things ready for us to use and control. One could conclude from this that workers have also become part of this 'standing reserve': they are *human resources* to be used in industrial production processes.

These criticisms are not only relevant for thinking about industrial production (e.g., Marx about capitalism) or about energy production (Heidegger's example). In modernity, all kinds of practices become shaped by this kind of thinking and this way of organizing work, even if they do not *literally* resemble industrial production processes or assembly lines. For health care work, it means that under modern conditions, care work has become 'labour', which (1) is wage labour (care is something one does for money), (2) involves modern employment relations (with professionalization, disciplining, formalization of the work, management, etc.), and

(3) involves relations between care giver and care receiver in which the receiver is in danger of appearing to the care giver as an *object* (a thing rather than a human person, a subject) and in which care is made into a commodity, a *product* or a *service*. Patients and other vulnerable people are managed and processed. This not only degrades the care receiver but also alienates the care giver from her work and the care receiver. Because of the division of labour, the care giver becomes an automaton: to the extent that (s)he does work that can be described in Marx's words as 'very one-sided and machine-like type of labour', (s)he is thus being 'reduced to a machine'. The care receiver encounters only a 'robot' care giver, a 'robot nurse' or 'robot doctor' who does only her small part of labour in the larger health care machine, in care organized as a production process. Insofar as this happens in contemporary health care, that is, insofar as contemporary health care is *modern* care, the quality of care is already seriously jeopardised even without considering the role of the (visible) machine. Even without robots, 'the machine' has already arrived, but we do not tend to notice its presence because, as care givers and care receivers, we work and live in it.

Note that I used the term 'appear' again. I am not so much, and certainly not only, interested here in what could be said from an external point of view, 'objectively' or 'scientifically', about care, if such a project makes sense at all (and in a way, this is exactly what Marxism, for instance, tried to do). According to the approach I take here, what crucially matters for good care is also, and rather, how care givers and care receivers *experience* care. The point is that care is only good to the extent that it can avoid these types of alienating experiences.

What does this mean for the role of machines in health care, and in particular, machines acting as 'artificial agents'? In industrial contexts, machines are usually used to *automate* the 'production'. The division of labour into small units is accompanied by, made possible, and increased by means of machines. Now this also seems to happen in health care, insofar as humans are being replaced by machines, that is, insofar as machines act as 'artificial agents'. As in historical labour processes, the automatization of health care means that human workers are replaced by machines. I already argued in the previous section that this 'replacement' is ethically problematic. But now I can expand this argument with the proposed analysis of health care as a *modern* practice and as involving modern labour. Here, the worry is that the machine is used to automate health care as part of its further modernization, and that this has the alienation effects mentioned. This is not about the machine 'taking over'; it is about humans *becoming* machines.

Thus, my argument against machines as 'artificial agents' in health care now has two parts. A first part focuses on replacement. If automatization by means of 'artificial agents' means that the machine takes up the role of a care *agent*, then insofar as that happens (the appearance of replacement is produced), the criteria of good care are not met. Perhaps machines can be given different roles, but then it would be better for those roles or functions not to involve 'agency', or at least not the 'strong' form of agency defined above. A second part focuses on alienation. The problem here is not so much with the agency of the *machine*, but with automatization understood as part of the continued and increased modernization of health care. The machine is then seen as an element within a larger process of

modern organization of labour and of care, which turns *humans* into 'machines'. On the one hand, machines become more 'human' in the sense that they are given agency; on the other hand, humans become more 'machine-like' in the sense that under modern conditions, they have to work, care, and be cared for as though cogs in a machine—cogs in a *care machine*.

Finally, to the extent that this condition of modernity depends on *humans* who sustain or increase it (and I think it does to a significant degree; 'modernity' is always acted out and embodied and exists only in these performances—it is not a 'thing' out there), not only care givers (and especially their managers, the politicians who organize health care, etc.) but also *care receivers* play a crucial role in the modernization of health care. For instance, accepting vulnerability and dependency on others seems to be a criterion of good health care, and perhaps it is a precondition of any human care to 'work', but under modern conditions, many care receivers are very keen to keep what they see as their autonomy and independence. In fact, we are so attached to our autonomy, that some of us would *prefer* 'machine care' over human care. The assumption, then, is that independence from other human beings has to be preserved at all costs—even at the cost of being totally dependent on a machine. By taking this route, however, we risk losing the humanity and dignity we were seeking to preserve, not because we are dependent on a machine (we always make ourselves dependent on machines; this, by itself, is not ethically problematic unless they gain strong agency) but because by refusing human care and replacing it with machine care, we deny the social dimension of our being. Being dependent on other human beings is part of what it is to exist as a human being. Moreover, as I suggested in the section on criteria of good care, it also belongs to good care. According to a relational approach to care, influenced by Heideggerian and Aristotelian thinking, human existence crucially involves being vulnerable and being dependent on others in coping with this vulnerability. To think that the best type of care is self-care in isolation from others is to undermine the very notion of care.

Conclusion

To conclude, using machines in health care is in principle acceptable, but insofar as they function and appear as 'agents' that take over some essential care tasks, roles, and responsibilities, they threaten the fulfilment of the criteria of good care I articulated on the basis of my interpretation of the discussion about care robots and the influence of Heideggerian and Aristotelian currents in philosophy. These criteria are already hard to meet in a modern care context, which tends to alienate both care givers and care receivers in various ways. The design and use of 'care machines', but also the organization of medicine and health care, should therefore be re-directed to avoid these problems. If, in care, machines are to be used in ethical ways, artificial agents should not be made more ethical (as proposed, for instance, by Wallach and Allen); instead, the problems of modern health care and modern human relations should be tackled by finding a better role for machines. This is not only a task for care givers, for developers of medical technology, and for those who

organize health care. It is also a matter of re-thinking and re-defining care *receiving*, and care receivers can contribute to this project by reflecting on their attitudes towards care, dependence, vulnerability, and of course, machines.

References

1. Anderson, Michael, and Susan Leigh Anderson. 2007. Machine ethics: creating an ethical intelligent agent. *AI Magazine* 28(4): 15–26.
2. Wallach, Wendell, and Colin Allen. 2009. *Moral machines: teaching robots right from wrong*. Oxford: Oxford University Press.
3. Sharkey, Amanda, and Noel Sharkey. 2010. Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology* 14(1): 27–40.
4. Sparrow, Robert, and Linda Sparrow. 2006. In the hands of machines? The future of aged care. *Minds and Machines* 16: 141–161.
5. Whitby, Blay. 2011. Do you want a robot lover? In *Robot ethics: the ethical and social implications of robotics*, ed. Patrick Lin, Keith Abney, and George A. Bekey, 233–249. Cambridge, MA: MIT Press.
6. Beauchamp, Tom, and James F. Childress. 2001. *Principles of biomedical Ethics*, 5th ed. New York: Oxford University Press.
7. Benner, Patricia (ed.). 1994. *Interpretive phenomenology: embodiment, caring, and ethics in health and illness*. Thousand Oaks: Sage.
8. Gilligan, Carol. 1982. *In a different voice*. Cambridge, MA: Harvard University Press.
9. Noddings, Nel. 1982. *Caring: a feminine approach to ethics and moral education*. Berkeley: University of California Press.
10. Tronto, Joan. 1994. *Moral boundaries: a political argument for an ethic of care*. New York: Routledge.
11. Coeckelbergh, Mark. 2013. E-care as craftsmanship: virtuous work, skilled engagement, and information technology in health care. *Medicine, Healthcare and Philosophy* 16(4): 807–816.
12. Dreyfus, Hubert L., and Stuart E. Dreyfus. 1991. Towards a phenomenology of ethical expertise. *Human Studies* 14(4): 229–250.
13. MacIntyre, Alasdair. 1999. *Dependent rational animals: why human beings need the virtues*. Chicago: Open Court.
14. Ihde, Don. 1990. *Technology and the lifeworld: from garden to earth*. Indianapolis: Indiana University Press.
15. Asaro, Peter M. 2006. What should we want from a robot ethic? *International Review of Information Ethics* 6(12): 9–16.
16. Coeckelbergh, Mark. 2009. Virtual moral agency, virtual moral responsibility. *AI & Society* 24(2): 181–189.
17. Coeckelbergh, Mark. 2010. Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology* 12(3): 209–221.
18. Coeckelbergh, Mark. 2010. Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology* 12(3): 235–241.
19. Torrance, Steve. 2008. Ethics and consciousness in artificial agents. *AI & Society* 22: 495–521.
20. Sullins, John. 2006. When is a robot a moral agent? *International Review of Information Ethics* 6(12): 23–30.
21. Marx, Karl. 1959 [1927]. *Economic and philosophic manuscripts of 1844*. Moscow: Progress Publishers.
22. Heidegger, Martin. 1977. The question concerning technology. In *The question concerning technology and other essays*, trans. W. Lovitt, 3-35. New York: Harper & Row.