

To delegate or not to delegate: Care robots, moral agency and moral responsibility

Aimee van Wynsberghe

Department of Philosophy, University of Twente, the Netherlands

Email: a.l.vanwysberghe@utwente.nl

Abstract. The use of robots in healthcare is on the rise, from robots to assist with lifting, bathing and feeding, to robots used for social companionship. Given that the tradition and professionalization of medicine and nursing has been grounded on the fact that care providers can assume moral responsibility for the outcome of medical interventions, we must ask whether or not a robot can assume moral responsibility for the outcome of its actions. In this paper I discuss the issue of moral agency and moral responsibility in terms of care robots and care contexts. With an understanding that the roles of care robots need to be limited to prevent the delegation of tasks which require moral responsibility, I discuss the design of a robot prototype using a method for the design of future robots tailored to addressing ethical concerns. This approach is called the Care Centered Value Sensitive Design approach and reveals itself to be the most promising method for integrating ethical deliberation into the design of future care robots.

1 INTRODUCTION

The use of robots in healthcare is on the rise, from robots to assist with lifting, bathing and feeding, to robots used for social companionship [1]. The issue of responsibility is of the utmost importance in healthcare contexts and in the therapeutic relationship [2,3]. A human care-giver must be morally responsible for the outcome of care actions. The professionalization of medicine and nursing is grounded on this fact. This begs the question whether or not a robot can be morally responsible for the outcome of its actions.

In this paper, I challenge the claim made by roboticists that robots ought to be endowed with moral reasoning capabilities based on the contexts within which they will be placed and the roles they are assigned [4,5]. In order to argue against the appeal for programming robots to be, or to act as, moral agents I discuss the concepts of moral agency and moral responsibility as they apply to robots. It becomes clear that regardless of whether or not one considers the robot to be a moral agent the robot cannot assume moral responsibility for its actions. This finding has significant repercussions for robots in care contexts. After discussing the significance of moral responsibility in care I show how the robot's role in a care practice can be carefully decided upon throughout the design process using the Care Centered Value Sensitive Design (CCVSD) approach [6,7,8].

2 FROM MORAL AGENCY TO MORAL RESPONSIBILITY

The way the debate is currently framed is whether or not robots can be considered moral agents. Typically speaking, moral agency is required for moral responsibility. Is it possible

then to consider robots as moral agents? That depends on your conception of moral agency and moral status. Here, I take a look at three prominent theories of moral agency: the organic view, the standard conception, and the morally intelligent view. According to the first two views, a robot cannot be a moral agent and therefore should not be delegated actions where moral responsibility is required. According to the third view, moral agency and moral responsibility are separated from one another; you can be a moral agent without assuming moral responsibilities. Therefore, a robot can be considered a moral agent; however, given its lack of intentions, the robot cannot be held morally responsible for the consequences of its actions. In what follows, a closer look is taken at these three possibilities.

From the organic view of moral agency, only a genuine organism (human or non-human animal) may be considered a candidate for intrinsic moral status and thus be considered a moral agent. This has to do with the belief that moral thinking, feeling and action arise organically out of the biological history of the human species [9]. From this, of course robots may have the capabilities for high level reasoning but cannot be considered full moral agents due to their inorganic make-up.

In contrast to the organic view of moral agency, the standard conception of a moral agent refers to: "beings who are capable of acting morally and are expected by others to do so" [10, pg 125]. Thus, "moral agents are beings that are 1. capable of reasoning, judging and acting with reference to right and wrong; 2. expected to adhere to standards of morality for their actions; and 3. morally responsible for their actions and accountable for their consequences" [ibid]. Here, there is no indication as to the physical make-up of the agent but rather solely to the capabilities, expectations and associated responsibilities of a moral agent.

An agent is a moral agent when the intentional states that it cultivates and the subsequent actions it performs are guided by moral considerations. This requires a capacity for moral deliberation, which is reasoning, in order to determine what the right thing to do is in a given situation. A capacity for moral deliberation requires a capacity for reasoning and knowledge of right and wrong. Moral deliberation typically results in moral judgments, which are judgments about right and wrong. It also frequently results in intentions to perform certain actions that are held to be moral, and to refrain from performing actions that are held to be immoral. [10, pg 126].

According to this view, it is once again problematic to include robots within the category of moral agents for a number of reasons. First, robots only have intentions in so far as they have been programmed. It is the intentions of the designers and not the intentions of the robot that are considered for agency. Added to this, a robot cannot be held responsible or liable for its actions insofar as it cannot be punished for bad consequences. As for the

reasoning capabilities of a robot, certain roboticists believe it is possible to program robots to have highly sophisticated reasoning capabilities making the robot intelligent enough to be considered a moral agent. This way of thinking brings us to the morally intelligent view of moral agency.

Dominant proponents of this view include Luciano Floridi and Jeff Sanders who claim that artificial intelligence opens new avenues when speaking of moral agents [11]. Specifically, technologies with highly sophisticated mechanisms for reasoning, capable of interacting with their environment, acting in an autonomous fashion, learning and adapting to their environment, ought to alter the discussion of moral agents. Their goal is to expand the category of moral agents such that it includes sophisticated technical artefacts, rather than to alter the concept of morality such that artefacts and humans engage in a practice of hybrid morality. Within this conception, Floridi and Sanders aim to disentangle the relationship between moral agency, accountability and moral responsibility. They argue that moral accountability is a necessary but insufficient condition for moral responsibility. According to their view, a moral agent, and ultimately a robot, may be considered a moral agent insofar as it may be considered accountable for its actions (and thus subject to censure); however, it may not be held responsible for its actions given that it lacks the intentions guiding it to make said decisions [ibid].

With this view, once again we are left with the issue of moral responsibility open; even if the robot is considered a moral agent given its sophisticated abilities for reasoning, learning from and adapting to its environment, it is still only accountable for its actions but not responsible.

To be clear, there is no agreement among scholars of robot ethics regarding the status of robots as moral agents but all signs point to: no, robots are not moral agents. As such, robots cannot bear moral responsibility and need not be programmed to have ethical reasoning capabilities. However, some robot scholars and roboticists argue that robots can and should be endowed with sophisticated ethical reasoning capabilities given the roles and contexts in which the robots will be placed. If we agree to this point and program the robot with ethical reasoning capabilities then, according to Floridi and Sanders, robots may be allowed into the category of moral agency. This is only possible by separating the concept of moral responsibility from moral agency. Consequently, this changes the focus of the debate. Now we need to understand the relationship that moral responsibility shares with the provision of good care. Is it possible to provide good care without the element of moral responsibility?

3 MORAL RESPONSIBILITY AND CARE CONTEXTS

For care ethicist Joan Tronto, there are certain necessary and sufficient criteria which render care to be *good* care. When discussing the stages of a care practice Tronto refers to the moral element of responsibility in which a human care giver takes the praise and or blame (being liable) for the outcome of events or behaviour elicited by his/her actions or decisions [2]. This conception of responsibility provides a normative element to a care practice – care practices can be evaluated according to whether or not a clear chain of responsibility has been identified.

But as Tronto also points out, care is more complicated than the completion of one care task after another for which the lines

of responsibility can be clearly drawn. Care practices are intertwined with one another blurring the lines of responsibility from one care worker to another. As an example, consider the good of a patient undergoing surgical intervention; a surgeon is responsible for the physiological good of the patient but the surgical nurses as well as the post operative nurses are also responsible for the good of the patient. The surgical nurses are responsible for maintaining a sterile environment and the handling of surgical instruments, among other things. Post operative nurses are responsible for the changing of bandages, bathing of patients, monitoring patient vital signs and in most cases reporting to the surgeon on the progress of the patient. There is a clear chain of responsibility within the care institute delineating who is responsible for what. It is this chain that is used for solving issues of liability and that facilitates the trust society has in the professions of medicine and nursing.

Tronto also states that the concept of responsibility in care is about much more than merely taking the praise and or blame for actions. It is about having caring intentions, *caring about* patients and their wellbeing [2, 3]. Technologies do not have these intentions. While technologies play a crucial role in the provision of good care they are used as tools within the care process, ultimately it is the care providers and institutes that are morally responsible for meeting the needs of care receivers. In short, care must be fulfilled by an agent capable of intentional states and of assuming moral responsibility; a human moral agent.

4 ROLE DELEGATION AND CARE ROBOTS

Consequently, in care contexts, regardless of whether or not the robot is considered to be a moral agent it cannot be delegated tasks which require moral responsibility. This statement has an important outcome for designers of care robots: the care robot must be designed intentionally to avoid being delegated a role for which a full moral agent (i.e. a human care giver) would traditionally be delegated. This means that the robot can still be delegated portions of a care practice but the robot cannot be delegated tasks which require moral reasoning capabilities as the robot cannot assume moral responsibility. How is this to be accomplished?

In 2012 I introduced an approach dedicated to the ethical evaluation and prospective design of care robots named the Care Centered Value Sensitive Design (CCVSD) approach [7,8]. The approach consists of a framework of components (context, practice, actors, type of robot/robot capabilities and manifestation of care values) used for either retrospective evaluations of current care robot prototypes or in the prospective design of future robots¹. The components are intended to orient one from the care ethics stance and to focus the design team towards the ethically relevant dimensions as identified from the care ethics perspective.

Retrospective evaluations using the CCVSD approach revealed the relationship between care robot capabilities and the manifestation of care values [7]. In particular, a difference in the autonomy of a lifting robot altered the overall practice of lifting and the manifestation of care values such that a human-operated robot (e.g. a wearable exoskeleton) was shown to be the more

¹ For a theoretical explanation and justification of the components of the CCVSD framework please see references [7], [8], and [9]

ethically sound choice for lifting in a care institute over an autonomous robot or even the traditional mechanical lifts used in many hospitals today. This was revealed after careful consideration of the relationship this practice shared with the establishment of the therapeutic relationship.

The component of “context” was significant here in that the same robot may not be considered the ethical choice in a home context in which a relationship has already been established. Instead, the autonomous robot for lifting may provide a superior alternative for the provision of dignified care in a home setting in which a patient would prefer to have greater autonomy rather than relying on loved ones for such a vulnerable task.

Using the CCVSD approach for the prospective design of future robots revealed an additional strength of the approach, namely, it captured the relationship between care robot capabilities and the responsibility delegated to the care robot [6]. The CCVSD approach engages the design team in a deliberation of care robot capabilities and how these capabilities realize care values when used in context. Through this deliberation it is also possible to envision the robot in its context of use and to make explicit the chain of responsibility as well as the potential for misuse. Consequently, by making the chain of responsibility explicit it is possible to track and limit the amount and type of responsibility delegated to the robot. By ‘kind’ of responsibility I am referring to a difference between moral responsibility and other forms that bear no moral impact. An action bearing no moral responsibility may be the kind of jello given to a patient. An example of an action that bears moral responsibility may be decisions regarding whether or not to intubate a patient, whether or not to maintain a patient on life support, or other triage decisions.

The example used to demonstrate the prospective use of the CCVSD approach was a robot prototype named the “Wee-bot” [6]. To be clear this is not an actual robot but a suggestion I made based on current robotics technologies and my experience in care contexts observing the practices and needs of care workers. The “Wee-bot” robot can be used for the collection of urine samples in pediatric oncology wards of a hospital to ensure the safety and wellbeing of the nurses in this ward². Through deliberation of the potential capabilities of the robot it was possible to make clear how these capabilities changed the amount and type of responsibility delegated to the robot. Three robot scenarios were discussed presenting robotic platforms with varying capabilities:

- i. A mobile robot that can travel through the hospital corridors and elevators avoiding obstacles via its sensors (thus, autonomously) but that is human-operated for urine retrieval and testing
- ii. A mobile robot that not only travels throughout the hospital autonomously but also travels inside the patient’s room and collects the urine sample autonomously
- iii. A mobile robot that acts autonomously for travel and sample collection the nurse must identify themselves to the robot prior to its entry into the hospital room

² For a detailed description of the robot prototype please reference [6]

When discussing the capabilities of the human-operated robot in scenario (i) the robot’s role is to collect the urine sample but the responsibility for the safe and successful completion of the task lies with the human care giver. A problem does exist when one considers that the robot could minimize the efficiency of the care practice and detract the attentiveness of the nurse from the patient to the robot.

The second scenario (ii) was suggested to mitigate concerns of the nurse’s attentiveness but raised significant concerns of responsibility. In this scenario the robot takes over the entire role of urine collection and testing and thus the associated responsibility. As such, the robot is delegated the same amount and kind of responsibility as was originally delegated to the human care giver. The robot is responsible for: deciding when to do urine retrieval and testing; informing and interacting with the patient whose urine was being collected and tested; collecting and testing the urine sample; and, passing on the test results to the appropriate oncologist.

Given this scenario it is then possible to assume that the nurse may not feel needed for this practice and may not be present when the robot is acting. What would happen if the test is not taken at the right time or the results are not sent to the appropriate oncologist? The administration of chemotherapy drugs are a crucial part of cancer treatment and as such need to be monitored closely. If this step were not completed there could be life and death consequences. Would the nurse even know that something went wrong? In the instance that something goes wrong and we can place accountability on the robot who is to blame and thus liable?

With these concerns in mind the final scenario of the paper (iii) was presented to balance the distribution of responsibility within the nurse-robot network; the robot is endowed with autonomous capabilities for travelling throughout the hospital as well as for sample collection and urine testing but it is necessary for the nurse to identify themselves to the robot before the robot can begin their task.

Identification could be through voice commands, facial recognition, retinal scans, finger print analysis etc. This gives the robot permission to enter the room but also ensures that the nurse is present and responsible for the practice. To strengthen this interaction, the robot could be programmed with semantic links endowing the robot with the capacity of knowing ‘why’ it must ensure the presence of the nurse. The robot may also be designed such that when it leaves the hospital room it must also interact with the nurse prior to sending the information to the oncologist. What’s more, once the information has been sent to the oncologist the robot requires that the nurse ‘sign-off’, in a manner of speaking, before the robot is able to leave the room. [6, pg 438]

In this scenario the amount of responsibility delegated to the robot is minimized and the kind of responsibility is no longer in terms of the robot making any decisions regarding urine testing. The robot’s role is to collect and test the urine sample in a safe and efficient way but this time under the oversight of the nurse. All decision making responsibilities pertaining to the patient’s preferences remain in the domain of the nurse. What’s more, the responsibility of the nurse is still to ensure that: the collection of urine is taken, the test is made, and the results are passed onto the appropriate oncologist.

The practice of urine collection and testing in the pediatric oncology ward differs now from before the robot was introduced

in that the nurse no longer has to jeopardize their own safety or wellbeing in order to complete the practice. By distributing the roles throughout the socio-technical system of the care team (the nurse and the robot along with the hospital room, the toilet, the urine sample etc...) it is possible to limit the amount, and kind, of responsibility that is delegated to the robot through carefully deciding on, and limiting, the role of the robot.

5 CONCLUSION

This paper was intended to bring to light the problems associated with discussing the moral agency of a robot, in particular of a care robot. The discussion of moral agency and moral responsibility here highlighted the impossibility of claiming that a robot is a moral agent capable of assuming moral responsibility for the outcome of an action. This does not mean that the robot has no ethical impact. Quite the opposite, the robot carries a significant ethical force by its impact on the actions and decisions of the socio-technical network of the care practice, care team and care institute.

For roboticists hoping to program a care robot with sophisticated ethical reasoning capabilities it was revealed that such a robot may be considered a moral agent but even still could not be capable of assuming moral responsibility. Given the necessity for moral responsibility in the delegation of care tasks, a robot cannot be solely responsible for providing good care. The only reason a robot would need to have moral reasoning capabilities is if it were the sole provider of care. Therefore, there is no need to program a care robot with the capabilities for ethical reasoning.

The fact that a care robot cannot assume moral responsibility poses limitations to the kinds of roles the care robot may be delegated. With this in mind, the goal of this paper was to make clear the relationship that a robot's capabilities has on the role assigned to the robot. By understanding how a shift in one capability changes the amount of responsibility assigned to the robot it becomes possible for robot engineers and designers to limit the amount of responsibility delegated to the care robot through a careful selection of its capabilities. A method for such deliberations throughout the design process, the CCVSD approach, was presented here along with the case study of the 'Wee-bot' robot. In this paper I intended to show how the CCVSD approach explicitly addresses the issue of responsibility delegated to the robot and as such remains the strongest and most encompassing ethical approach for the design of future care robots.

ACKNOWLEDGMENTS

The ideas presented in this paper are original but draw on previous material written and published by the author. For access to these articles please see the references.

Thank you to CTIT and EUCogIII for funding to make it possible to participate in the MEMA-14 Symposium.

REFERENCES

- [1] N Sharkey, & A Sharkey. The Rights and Wrongs of Robot Care. In P. Lin, K. Abney, & G. Bekey (Eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (pp. 267-282). Mit Press. (2011)
- [2] J Tronto. *Moral boundaries : a political argument for an ethic of care*. New York: Routledge. (1993)
- [3] J Tronto. Creating Caring Institutions: Politics, Plurality, and Purpose. *ETHICS AND SOCIAL WELFARE* , 4 (2), 158-171. (2010)
- [4] M Anderson and S L Anderson. Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine* 01; 28:15-26. (2007)
- [5] W Wallach and C Allen. *Moral machines : teaching robots right from wrong*. New York; Oxford: Oxford University Press. (2010)
- [6] A van Wynsberghe. A method for integrating ethics into the design of robots. *Industrial Robot* 40(5). 2013
- [7] A van Wynsberghe. Designing Robots for Care: Care Centered Value-Sensitive Design. *Journal of Science and Engineering Ethics*, 4. (2013)
- [8] A van Wynsberghe. *Designing robots with care: Creating an ethical framework for the future design and implementation of care robots*. Enschede: University of Twente. (2012)
- [9] S Torrance. Ethics and consciousness in artificial agents. *AI & Society* , 22 (4), 495-521. (2008)
- [10] Brey P. From moral agents to moral factors: the structural ethics approach. In *The Moral Status of Technical Artefacts*. (pp125-142) Springer, Netherlands. (2014)
- [11] L Floridi and J Sanders. On the Morality of Artificial Agents. *Minds and Machines* , 14 (3), 349-379. (2004)