Ethics in Progress (ISSN 2084-9257). Vol. 10 (2019). No. 2,
Art. #7, pp. 68-81. DOI:10.14746/eip.2019.2.7

A Softwaremodule for an Ethical Elder Care Robot. Design and Implementation

Catrin Misselhorn

(Georg-August-Universität Göttingen; catrin.misselhorn@uni-goettingen.de)

1. Introduction

Artificial morality is an emerging field in the area of artificial intelligence. Artificial intelligence has the aim to model or simulate human cognitive abilities. Artificial morality, more specifically, explores whether and how artificial systems can be furnished with moral capacities. Machine ethics provides the theoretical and ethical framework for thinking about artificial morality (Misselhorn 2018). Artificial morality becomes an issue because the development of increasingly intelligent and autonomous technologies will eventually lead to these systems having to face morally problematic situations. In these situations, the systems cannot be fully controlled by human operators anymore because that would deprive them of the advantages that they have as autonomous systems

The aim of this article is to develop some strategies on how to proceed when trying to implement moral capacities in an elder care system. Elder care is a highly relevant area of application because due to the shortage of caregivers there is a growing interest in autonomous assistive systems that can be used in geriatric care environments. Yet, most applications do not consider ethical aspects explicitly (e.g. Bajo 2008; Zato Domínguez *et al.* 2013; Parra *et al*. 2014; De Paz Santana 2015).[1] In this area of application the deployment of autonomous artificial systems is, however, highly problematic from a moral point of view. Even if we take it for granted that it is undesirable to substitute machines for human caregivers, there will be many situations in which the behavior of a geriatric care system will, somehow, have to be morally regulated by itself. It will, hence, be important to think about developing machines which have the capacity for a certain amount of autonomous moral decision making. The relevant capacities may be called moral agency in a functional

[1] An exception is the pioneering and still actual work of the Andersons (2008; 2011) which is later discussed in more detail.

sense (Wallach & Allen 2009; Misselhorn 2013; 2015; 2018). Yet, this theoretical debate is not the focus of this paper which concentrates on the more practical issue of developing a software module for a geriatric care robot that can take ethical decisions.

The basic idea is not to provide fixed moral standards for care robots, but it is rather shown how to proceed when trying to determine and implement the relevant moral capacities. Elder care was not just chosen as area of application because of its urgency and great moral relevance, but also because it shows that the implementation of moral capacities depends on the purpose and context for which an artificial system is designed.

The first part of the article is dedicated to different approaches to moral implementation: top-down, bottom up and hybrid. Since bottom-up approaches generate problems of operationalization, transparency and safety in practical contexts, top-down approaches seem to be best in the first instance. The second part, therefore, discusses the most detailed and promising top-down approach to implement moral capacities in a geriatric care system. It goes back to the pioneering work of Anderson and Anderson (2008; 2011). Yet, as the critical discussion will show, a hybrid approach is more apt for the area of geriatric care than a top-down approach. This is due to characteristic requirements of care as an area of application. The Andersons' top-down approach is for methodological reasons not suited to meet these requirements. The third part provides the theoretical foundations for a hybrid approach based on Nussbaum's capabilities view in ethics and describes the steps that have to be taken to implement such a view in a software module for a geriatric care robot. The last part examines in which sense one can ascribe moral capacities to the proposed system and draws a comparison between artificial and human moral agents.

2. Moral Implementation: Top-down, Bottom-up and Hybrid

In the debate about the implementation of moral capacities in artificial systems one distinguishes between top-down, bottom-up and hybrid approaches (Wallach & Allen 2009). All three approaches bring together a certain ethical approach with a certain approach to software design. *Top-down approaches* combine an ethical view that regards moral capacities as an application of moral principles to particular cases with a top-down approach to software design. The basic idea is to formulate moral principles like Kant's categorical imperative, the utilitarian principle of maximizing utility or Asimov's three laws of robotics as rules in a software which is then supposed to derive what has to be morally done in a specific situation.

One of the challenges that such a software is facing is how to get from abstract moral principles to particular cases. Another problem is that that there is no consensus about moral principles, neither in the general public nor among philosophers. Our moral intuitions do not seem to speak clearly for or against Kantianism, Utilitarianism or any other approach. One might call this – adopting a term coined by John Rawls in the context of political philosophy (Rawls 1993) – the fact of reasonable pluralism.

The most fundamental objection against top-down approaches regarding artificial morality is the so-called frame problem. Originally, the frame problem referred to a technical problem in logic-based artificial intelligence. Intuitively speaking the issue is sorting out relevant from irrelevant information. In its technical form the problem is that specifying the conditions which are affected by a system's actions does in classical logic not license an inference to the conclusion that all other conditions remain fixed. Although the technical problem is largely considered as solved (even within strictly logic-based accounts), there remains a wider, philosophical version of the problem first stated by McCarthy and Hayes (1969) which is not yet close to a solution (Shanahan 2009).

The challenge is that potentially every new piece of information may have an impact on the whole cognitive system of an agent. This observation has been used as evidence against a computational approach to the mind since it seems to imply that central cognitive processes cannot be modelled by strictly general rules. A corresponding line of argument can also be turned against top-down approaches regarding artificial morality. As Horgan and Timmons (2009) point out, moral normativity is not fully systematizable by exceptionless general principles because of the frame problem. Fortunately, full systematizability is not required. Horgan and Timmons admit that a partial systematization of moral normativity via moral principles remains possible. The frame problem is, hence, not a knock-down argument against the possibility of top-down approaches to moral implementation.

The alternative to top-down are *bottom-up approaches* which do not understand morality as rule-based. This view is closely related to moral particularism, a meta-ethical position which rejects the claim that there are strict moral principles and that moral capacities consist in the application of moral principles to particular cases (Dancy 2013). Moral particularists use to think of moral capacities in terms of practical wisdom or in analogy to perception as attending to the morally relevant features (or values) that a situation instantiates. Moral perception views emphasize the individual sensibility to the moral aspects of a situation (Nussbaum 1985). The concept of practical wisdom goes back to Aristotle who underlined the influence of contextual aspects which are induced by way of socialization or training. In order to bring these capacities about in artificial systems classical bottom-up approaches in software design (i.e. artificial neuronal networks) which start from finding relationships or patterns in various kinds of data have to be adapted to the constraints of moral learning.

Bottom-up approaches might teach us something about the phylo- and ontogenetical evolution of morality (e.g. Axelrod 1984). But they are of limited suitability for implementing moral capacities in autonomous artificial systems because they pose problems of operationalization, safety and acceptance. It is difficult to evaluate when precisely a system possesses the capacity for moral learning and how it will, in effect, evolve. There is no component or mechanism to point to which embodies the moral capacities of the system. Since the behavior of such a system is hard to predict and explain, bottom-up approaches are hardly suitable for practical purposes because they might put potential users at risk. Moreover, it is difficult to reconstruct how a system arrived at a moral decision. This will probably diminish the acceptance of such a system's moral decisions by the users. It is important that autonomous artificial systems do not just behave morally, as a matter of fact, but that the moral basis of their decisions is also transparent. Bottom-up approaches should, as a consequence, be restricted to narrowly confined and strictly controlled laboratory conditions.

Top-down and bottom-up are the most common ways to think about the implementation of moral capacities in artificial systems. It is, however, also possible to combine the virtues of both types of approaches. The resulting strategy is called *hybrid approach*. Hybrid approaches operate on the basis of a predefined framework of moral values which is then adapted to specific moral contexts by learning processes. Which values are given depends on the area of deployment of the system and its moral characteristics. Although hybrid approaches are promising, they are still in the early stages of development.

Yet, which approach to moral implementation should one chose? Unlike Wallach and Allen (2009) think, it does not make sense to answer this question in the abstract. It depends on the purpose and context of use for which a system is designed. An autonomous vehicle will demand a different approach to moral implementation than a domestic care robot. As will be shown in the next section by contrast with a top-down approach, a hybrid approach is particularly apt for a geriatric care system that is used in a domestic environment.

3. A Top-Down Approach to an Elder Care System

One of the few attempts to create an artificial system with moral capacities was provided by Anderson & Anderson (2008; 2011). They developed a top-down conception of an elder care software on the basis of moral principles. These moral principles are derived in the framework of a *prima facie* duty approach to ethical theory which goes back to W. D. Ross (1930). Ross held that there are a number of ethical principles which state – in contrast to more standard deontological theories like Kantianism or Utilitarianism – *prima facie* rather than absolute obligations. 'Prima facie' means in this context that these duties count in favor of the moral rightness of an action but may be outweighed by other factors. According to Ross, the prima facie duties are generated from the moral intuitions of "thoughtful and well-educated people" (Ross 1930, 41).

This is an important point at which Anderson and Anderson depart from Ross's original account. They do not start from the moral intuitions of considerate ordinary people but take the intuitions of ethical experts as their measure. The experts' moral intuitions are, from their point of view, best captured by the Principles of Biomedical Ethics by Beauchamp and Childress (1979) who provide a slightly modified list of Ross's prima facie duties for bioethics. Anderson and Anderson embrace thus a top-down approach to moral implementation. The principles adopted for their geriatric care system are the *Principle of Autonomy*, the *Principle of Beneficence* and the *Principle of Nonmaleficence*. The Principle of Autonomy requires that the system does – as far as possible – not interfere with the caretaker's autonomy; the Principle of Beneficence states that the patient's welfare should be promoted and the Principle of Nonmaleficence is supposed to preserve the patient from harm. Anderson and Anderson nourish the hope that their approach might provide insight into how moral learning is functioning in human beings.

This approach has three fundamental deficits: The *first* is its reliance on ethical experts. In contrast to experts in other fields, it is not clear how ethical experts distinguish themselves. Because of the reasonable pluralism of ethical doctrines discussed above there is much less unanimity than in other disciplines. Already the assumption that there are experts in ethical issues is not unproblematic. Should we not all be ethical experts with regard to the morally relevant situations that we encounter in everyday life? Beyond that there is the powerful idea of normative individualism which states that moral decisions must be justified in the light of the concerns and interests of the persons who are affected by it (von der Pfordten 2012). The persons concerned in geriatric care are in the first instance those in need of care, their relatives and the care givers. The most immediate and non-paternalistic way to take the concerns and interests of the parties involved into account is by asking them which moral values are most important to them in care environments.

One might argue against this approach by pointing out that those concerned are not the right persons to consult with respect to these issues. One problem is, for instance, that those concerned might have so-called adaptive preference which are deformed by poverty, adverse social conditions or political oppression (Nussbaum 2001). In elder care one might think of old people who do not believe that they deserve certain goods or treatments, or, at the extreme, even have a right to live because they feel useless for society.

However, the right way to avoid this kind of problem is not to let experts decide instead of those concerned. Such an approach does not consider to the fundamental nature of reasonable pluralism in ethical issues. It is constitutive of liberal democracies that people have the right to choose their own conception of the good and that there are many different and incompatible views about what the good is. This restricts the role of "ethical experts" severely. The best way to avoid distortions as adaptive preferences is to not just simply register the views of those concerned but to discuss these problems in focus groups or qualitative interviews with them.

Secondly, the three principles implemented by Anderson and Anderson certainly play a role in geriatric care. But they are not differentiated enough to capture all the morally relevant aspects of situations in geriatric care. One must, for instance, distinguish between physical health and mental well-being. Moreover, there are other relevant moral values like the dignity and self-respect of care-dependent persons, their bodily integrity or the privacy of the data that such a system generates.

The *third* deficit concerns the lack of flexibility of Anderson's and Anderson's approach. Although the system is capable to generate abstract rules from a number of samples, these rules themselves remain rigid. The system displays a certain amount of moral learning, but it is blind to the needs and moral values of the individual care-dependent persons. This is all the more important since people might diverge in their moral evaluation of a care situation, as will become apparent. The step towards a context-sensitive system that individually adapts to the moral values of care-dependent persons is for Anderson and Anderson blocked for methodological reasons. Since they do not involve the care-recipients' moral perspective, but delegate moral judgments to ethical experts, the system cannot be sensitive to the moral values of the individual care-dependent persons.

Because of the shortcomings of Anderson's and Anderson's approach which is the most detailed and promising attempt to implement moral capacities in a geriatric care system, we have to look for an alternative. The deficits of their approach have to do with its top-down character. An obvious alternative is to opt for hybrid approach in contrast. Such an approach has a top-down element as long as it operates within a pre-defined framework of moral values that are relevant in care. At the same time, it might be bottom-up if it is capable of moral learning by adapting to the way in which individual users weigh these values. A hybrid approach is ideally suited to incorporate the perspective of those concerned and to provide the level of differentiation and flexibility that is desirable in such a system.

4. A Hybrid Approach to a Software Module for Geriatric Care

A good starting point in ethics for such a hybrid system is Martha Nussbaum's capability approach (Nussbaum & Sen 1993; Nussbaum 2006). The approach determines capabilities that human beings need for well-being and to which they are morally entitled, e.g., life, bodily health and integrity, but also the exercise of one's cognitive and sensual capacities and the possibility to form a conception of the good. The definition of capabilities which are necessary for well-being also helps to avoid the problem of adaptive preferences discussed in the last section.

Coeckelbergh (2012) was the first to apply the capability approach to care robots. An aspect that has, so far, not been sufficiently considered is the context sensitivity of the capabilities, specifically in dependence of the individual life span (Misselhorn *et al.* 2013). People evaluate different kinds of capabilities in different phases of their life differently. Since it is difficult to anticipate the change of perspective for people who have not yet reached old age, it is particularly important to consider the moral perspective of the individuals concerned in geriatric care. Moreover, not all old people will weigh moral values in the same way. Autonomy, physical health, and privacy are among the moral values that are relevant in geriatric care. But old people may differ regarding the weight that they ascribe to these values absolutely and relatively in comparison to each other (Misselhorn *et al*. 2013).

If the question as to which decision is morally adequate in a geriatric care situation is supposed to be answered from the perspective of those concerned, it has to be determined, first, which moral values they regard as important and, secondly, how they weigh them. A good way to find out about these matters are surveys. Much can be learned in this respect from the paradigm of experimental philosophy (Nichols & Knobe 2008). In contrast to Anderson and Anderson which take the judgment of ethical experts as their input, the aim of experimental philosophy is to find out what ordinary people think with regard to philosophical issues: "Although philosophers quite frequently make claims about 'what people would ordinarily say', they rarely back up those claims by actually asking people and looking for patterns in their responses" (Knobe 2004, 37). The instruments and measures of experimental philosophy were originally not designed to find out which moral values are important for the elderly with respect to care and how they weigh them; nevertheless, these devices are of help. So far, such an approach has not yet been realized but in the next section a roadmap will be outlined which describes the steps that one has to take when attempting to design an ethically appropriate elder care system.

As a *first* step one has to *identify the relevant moral values* in geriatric care, for instance, with the help of qualitative interviews. The results of these interviews should be a list of the moral values that the individuals concerned consider as relevant in geriatric care. The capabilities approach introduced above is an ethical framework that can be used to individuate and categorize the relevant values. A critical discussion with ethicists in focus groups could be used to avoid distortions such as adaptive preferences.

In a *second step these values have to be operationalized* such that an artificial system is able to recognize them and weigh them according to the moral value profile of the user. This can be done on the basis of scenarios as they are used in the surveys of experimental philosophy. These scenarios will have to deal with the moral aspects of daily routines in geriatric care, particularly situations in which different moral values get into conflict. Typical conflicts with regard to elder care are, for instance: How often and how obtrusively is a geriatric care system supposed to intervene if a care-recipient is not moving? This is a conflict between autonomy and health. Individuals who fear that something could happen to them might welcome such interventions. Persons who set great value on autonomy might, in contrast, get rather annoyed. Should the system monitor and register the medical status of the care-dependent person constantly and report it to a hospital or care facility? The conflict that arises in this case concerns privacy and health. Those who are very worried about their health and less so about issues of privacy might favor this option whereas people who are more concerned about privacy might be sensitive to such attempts.[2]

In a *third step*, the *results have to be implemented* such that an artificial system can recognize the moral value profile of its user and react accordingly. The moral scenarios have to be translated in terms of information processing and must be transformed into algorithms.

Fourthly, the system is set up for use. During a training phase the system asks the user to choose different options with regard to a range of presented scenarios and builds on this basis a model of the moral value profile of the users, i.e., which moral values the user cherishes and how they are ranked in comparison to each other. The most straightforward way to do this is to use a textual interface. It might also be eventually possible to classify the evaluations of the scenarios by the user with the help of an emotion-recognition software. On the basis of this information the system tries to adapt its behavior in new cases to the moral value profile of the user. If it finds out, for instance, that a user ranks autonomy high and is not very worried that something could happen to him or her, it will intervene less obtrusively than in the case of user profiles that are structured the other way around.

Yet, the system's capacity of moral learning should not be restricted to the training phase. The system should constantly adjust its model of the user's moral value profile by interacting with the user and giving him or her the possibility to evaluate the adequacy of the system's decisions. In addition to the capacity for moral learning, the system should also have a self-monitoring function and regularly provide status reports whether it is functioning correctly. If this is not the case the system should inform the user and even turn itself off if there is a serious problem of safety. This is important in order to not put the user at risk if the system is not functioning properly.

Although such a system could be very helpful for elderly people there are three caveats in place. The first one arises regarding the transfer of this model to other areas of application. Although these results might seem quite intriguing at first glance, it would be a mistake to confer them straight-forwardly to other cases. Self-driving vehicles, for instance, operate under quite different conditions. Because of the fact that these devices do not just have effects on their users but affect a much larger group of people, it would be inadequate to make the decisions of such a system dependent on the owner's individual moral value profile as this can be done in the case of the envisaged geriatric care system which is only interacting with one user.

The second caveat is that the proposed system cannot even be used in all geriatric contexts. It is designed for domestic care. The target group would be people who are not cognitively impaired and can still take fundamental decisions regarding their life. Even

[2] An important point to which an anonymous reviewer adverted is that the evaluation of the scenarios might depend on whether they are formulated in way that involves the judging person or as interaction of others. It is an empirical question whether this has an influence on the moral judgments of the subjects or not. If it should turn out that this is the case one has to discuss, though, whether this has an impact on the classification of the subjects' responses as moral evaluations.

so they are physically frail and cannot stay without permanent care in their domestic environment. Since they do not want to move to a residence and do either not have the resources to hire a human care-giver or do not want to have one in their private setting, they would be willing to buy a care system but lack the technical know-how to install such a complex device according to their moral value profile. In contrast to other care systems (e.g. Bajo *et al*. 2008; Zato Domínguez *et al.* 2013) the proposed software module is, hence, not designed for the use in elder care residences.

This also explains why the focus of the approach is on the moral values of the care-dependent individuals and does not consider the relatives or caregivers. It is the care-dependent person who should decide which moral values are realized in his or her care-environment, as long as he or she still has the cognitive capacities to do so. This is required out of respect for the care-receivers' autonomy and takes into account the reasonable pluralism with regard to ethical matters in liberal democracies. The suggestion is by no means to replace human caregivers generally by geriatric care robots. But for people who conform to the requirements specified above, purchasing a care system that is capable of moral decision making and learning in the above explained way might be a viable alternative that allows them to stay in their home and live autonomously as long as possible.

The third caveat concerns the question whether one can really ascribe moral capacities to the outlined system. As will be argued in the next section, such a system may possess moral capacities in a functional sense. Nevertheless, it is important to realize that this does not mean that the system is morally on a par with human moral agents.

5. Artificial and Human Moral Agents

The system outlined has basic moral capacities: it learns by training and interacting with the user to recognize and weigh moral values as specified by the capabilities approach in alignment with the user's moral value profile and adapts its behavior accordingly. It is able to learn what is morally good or bad and even treats persons in line with moral standards that these people endorse. The system, therefore, meets the three basic criteria for agency that were developed by Floridi and Sanders (2004) for artificial systems: It is interacting with its user, it is initiating actions without an external stimulus (it reminds the user, for instance, to take his or her medicine when it finds that the user has not taken it in due time), and it displays adaptivity, since it is able to change its moral evaluations in order to conform to the user's moral value profile.

One has to be aware, though, that this primitive form of moral agency which the artificial system displays does not amount to full moral agency as it pertains to human beings. Wallach and Allen discriminate between different levels of moral agency along two dimensions (Wallach & Allen 2009, 26): *autonomy* and *ethical sensitivity*. A simple tool like a hammer does neither possess autonomy nor ethical sensitivity, but a child safety lock does involve a certain ethical sensitivity despite lacking autonomy. Its ethical sensitivity is entirely owed to the designer and user of the object. For this reason, Wallach and Allen speak of *operational morality*. Generally, both dimensions are independent of each other. There are, on the one hand, systems which possess a high degree of autonomy, but no (or only few) ethical sensitivity, e.g. an autopilot. On the other hand, there are systems with a high degree of ethical sensitivity but no (or a very low degree) of autonomy, e.g., the platform „MedEthEx" (Anderson *et al.* 2006) which is a computer-based learning program in medical ethics and communication skills.

The envisaged geriatric care system may claim *functional morality* which requires moral information processing. Yet, the proposed system does not attain full moral agency as mature human agents, not even in a functional sense. This is partly due to the fact that its moral competence is restricted to geriatric care situations with a particular user. Human morality has, in contrast, a much wider scope and potentially applies to any context. Full moral agency as assigned to human beings involves not just having the capacity of moral information processing, it encompasses a whole lot of intentional attitudes, phenomenal consciousness, free will and the capacity to make moral attitudes themselves the object of reflection and justification.

Phenomenal consciousness is important to get acquainted to the affective aspect of morality and to feel moral emotions like empathy, sympathy, guilt or shame. It concerns the way how these emotions feel like. Although there have been attempts to furnish artificial systems with states that are functionally equivalent to emotions like guilt, remorse or grief (Arkin 2007), what is left out is the feeling of the nagging quality of these moral emotions. Phenomenal consciousness is an important aspect of the richness of human moral life. The artificial system at issue can at most provide a functional equivalent of the cognitive aspects of human moral life in a highly restricted area.

Some argue that the availability of morally relevant information in the global workspace is a necessary condition for moral responsibility (Levy 2014). But it is not sufficient. An important aspect of moral responsibility is the capacity to deliberate about one's moral reasons and the envisaged system cannot do this. Even if one is willing to speak of moral reasons in this case (Misselhorn 2013; Misselhorn 2018) such a system is not capable of reflecting about its moral reasons, not to mention of changing them by itself. It cannot justify the normative basis of its moral decisions and is incapable of taking into account other possible courses of action.

The capacity of such higher-order reasoning is often regarded as a central tenet of freedom of the will (Frankfurt 1971) which the system is hence lacking. This is not a weakness but a strength in practical contexts. A system that were able to question its moral decisions or even to choose immoral courses of action would put the user severely at risk. This is not to say that it is impossible to build systems with these advanced moral capacities. Yet, they are not necessary nor desirable in a system with moral capacities that is suitable for the practical demands in contexts of domestic geriatric care.

6. Conclusion

The goal of this article was to show how an assistive system in geriatric care could sensibly be furnished with moral capacities. The first part of the article discussed three basic approaches to implementing moral capacities in artificial systems: top-down, bottom-up and hybrid approaches. They bring together a certain ethical approach with a certain approach to software design. Top-down approaches combine an ethical view that regards moral capacities as an application of moral principles to particular cases with a top-down methodology in software design. Bottom-up approaches, in contrast, bring together a particularist view in ethics with bottom-up approaches in software design (i.e. artificial neuronal networks). Hybrid approaches try to unite the virtues of the other two approaches while avoiding their shortcomings. They operate on the basis of a predefined framework of moral values which is then adapted to specific moral contexts by learning processes. Which approach to moral implementation one should choose depends on the purpose and the context of use for which an artificial system is designed.

As became apparent hybrid approaches are particularly apt with regard to artificial systems that are designed for the context of geriatric care. The discussion of the approach of Anderson and Anderson (2008; 2011) showed the weaknesses of top-down approaches in the area of geriatric care. It was criticized because of its reliance on experts, its coarseness and lack of flexibility. Alternatively, a hybrid approach was developed which uses the tools of experimental philosophy in order to identify, discuss and operationalize the moral values regarding geriatric care of the persons concerned. A roadmap was suggested that outlines the steps that one has to take in order to design such a hybrid geriatric care system with moral capacities. The system under consideration is supposed to be able to identify the morally relevant aspects of geriatric care situations and act accordingly. It learns to build a model of the user's moral value profile in a training phase and constantly adjusts this model by interacting with the user.

In the last part, a comparison between the system's moral performance and human moral agency was drawn. Although such a system does not possess moral agency as it pertains to human beings, it is an artificial moral agent that is not just able to recognize what is morally good and act accordingly. It is even capable of treating persons in line with moral standards that these persons endorse. These capacities might one day be of great importance for autonomous artificial systems that are taking care of elderly people who are still cognitively alert but physically impaired such that they can stay in their domestic environment and live autonomously as long as possible.

References

Anderson S., Anderson M., & Armen, Ch. 2006. "MedEthEx: A Prototype Medical Ethics Advisor." *Proceedings of the Eighteenth Conference on Innovative Applications of Artificial Intelligence, AAAI,* Boston, Mass*.*

Anderson S. & Anderson M. 2008. "ETHEL: Toward a Principled Ethical Elder Care Robot." *Proceedings of the AAAI Fall 2008 Symposium on AI in Eldercare: New Solutions to Old Problems,* Arlington, Virginia.

Anderson S. & Anderson M. 2011. "A Prima Facie Duty Approach to Machine Ethics and Its Application to Elder Care." *Proceedings of the AAAI Workshop on Human-Robot Interaction in Elder Care,* San Francisco.

Arkin R. 2007. "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture." *Technical Report GIT-GVU-07-11.* College of Computing. Georgia Institute of Technology.

Axelrod R. 1984. *The Evolution of Cooperation*, New York: Basic Books.

Bajo J. *et al.* 2008. "GR-MAS: Multi-Agent System for Geriatric Residences." *Frontiers in Artificial Intelligence and Applications* 178 ECAI 2008:875-876.

Beauchamp T. & Childress J. 1979. *Principles of Biomedical Ethics*. Oxford, NY: Oxford University Press.

Coeckelbergh M. 2012. "How I Learned to Love the Robot." In: I. Oosterlaken & J. van den Hoven (Eds.), *The Capability Approach. Technology and Design* (pp. 77-86). Dordrecht: Springer.

Dancy J. 2013. "Moral Particularism." In: *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), E. N. Zalta (Ed.). URL = <http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>.

De Paz Santana J. F *et al*. 2015. "An Integrated System for Helping Disabled and Dependent People: AGALZ, AZTECA, and MOVI-MAS Projects." *Advances in Intelligent Systems and Computing* 333:3-24.

Floridi L. & Sanders J. W. 2004. "On the Morality of Artificial Agents." *Minds and Machines* 14:349-79.

Frankfurt H. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68:5-20.

Horgan T. & Timmons M. 2009. "What Does the Frame Problem Tell Us About Moral Normativity?." *Ethical Theory and Moral Practice* 12:25-51.

Knobe J. 2004. "What is Experimental Philosophy?." *The Philosophers' Magazine* 28:37-39.

Levy N. 2014. *Consciousness and Moral Responsibility*, Oxford: Oxford University Press.

McCarthy J. & Hayes P. J. 1969. "Some Philosophical Problems from the Standpoint of Artificial Intelligence." In: B. Meltzer & D. Michie (Eds.), *Machine Intelligence* (pp. 463-502). Edinburgh, UK: Edinburgh University Press.

Misselhorn C. 2013. "Robots as Moral Agents?." In: F. Roevekamp (Ed.), *Roboethics, Proceedings of the Annual Conference on Ethics of the German Association for Social Science Research on Japan* (pp. 30-42). München: Iudicum.

Misselhorn C. 2015. "Collective Agency and Cooperation in Natural and Artificial Systems." In: C. Misselhorn (Ed.), *Collective Agency and Cooperation in Natural and Artificial Systems. Explanation, Implementation and Simulation. Philosophical Studies Series* 122 (pp. 3-25). Dordrecht: Springer.

Misselhorn C. 2018 (3rd ed. 2019). *Grundfragen der Maschinenethik*, Stuttgart: Reclam.

Misselhorn C. *et al*. 2013. "Ethical Considerations Regarding the Use of Social Robots in the Fourth Age." *GeroPsych – The Journal of Gerontopsychology and Geriatric Psychiatry Special Issue: Emotional and Social Robots for Aging Well*? 26:121-133.

Nichols S. & Knobe J. (Eds.) 2008. *Experimental Philosophy*. Oxford, NY: Oxford University Press.

Nussbaum M. 1985). "Finely Aware and Richly Responsible. Moral Attention and the Moral Task of Literature." *Journal of Philosophy* 82:516-529.

Nussbaum M. 2006. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, London: The Belknap Press of Harvard University Press.

Nussbaum M. 2001. "Adaptive Preferences and Women's Options." *Economics and Philosophy* 17:67-88.

Nussbaum M. & Sen A. 1993. *The Quality of Life*. Oxford: Clarendon Press.

Parra V. *et al.* 2014. "A Multiagent System to Assist Elder People by TV Communication." *ADCAIJ*: *Advances in Distributed Computing and Artificial Intelligence Journal* 3:10-16.

Rawls J. 1993. *Political Liberalism*. New York: Columbia University Press.

Ross W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.

Shanahan M. 2009. "The Frame Problem." *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), E. N. Zalta (Ed.). URL = <http://plato.stanford.edu/archives/ win2009/entries/frame-problem/>.

Von der Pfordten D. 2012. "Five Elements of Normative Ethics – a General Theory of Normative Individualism." *Ethical Theory and Moral Practice* 15:449-71.

Wallach W. & Allen C. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford, NY: Oxford University Press.

Zato Domínguez C. *et al*. 2013. "Virtual Organizations of Agents for Monitoring Elderly and Disabled People in Geriatric Residences" (pp. 327-333). *Information Fusion (FUSION) 2013*. 16th International Conference, Istanbul, Turkey, July 9-12 2013.

Catrin Misselhorn (Göttingen)

A Softwaremodule for an Ethical Elder Care Robot. Design and Implementation

Abstract: The development of increasingly intelligent and autonomous technologies will eventually lead to these systems having to face morally problematic situations. This is particularly true of artificial systems that are used in geriatric care environments. The goal of this article is to describe how one can approach the design of an elder care robot which is capable of moral decision-making and moral learning. A conceptual design for the development of such a system is provided and the steps that are necessary to implement it are described.

Keywords: artificial morality; machine ethics; moral implementation; assistive systems in geriatric care; care robots; capabilities; experimental philosophy.

Ethics in Progress (ISSN 2084-9257). Vol. 10 (2019). No. 2, Art. #7, pp. 68-81.

DOI:10.14746/eip.2019.2.7