ORIGINAL PAPER

Refining the ethics of computer-made decisions: a classification of moral mediation by ubiquitous machines

Marlies Van de Voort · Wolter Pieters · Luca Consoli

Published online: 23 January 2015

Abstract In the past decades, computers have become more and more involved in society by the rise of ubiquitous systems, increasing the number of interactions between humans and IT systems. At the same time, the technology itself is getting more complex, enabling devices to act in a way that previously only humans could, based on developments in the fields of both robotics and artificial intelligence. This results in a situation in which many autonomous, intelligent and context-aware systems are involved in decisions that affect their environment. These relations between people, machines, and decisions can take many different forms, but thus far, a systematic account of machine-assisted moral decisions is lacking. This paper investigates the concept of machine-assisted moral decisions from the perspective of technological mediation. It is argued that modern machines do not only have morality in the sense of mediating the actions of humans, but that, by making their own decisions within their relations with humans, mediate morality itself. A classification is proposed to differentiate between four different types of moral relations. The moral aspects within the decisions these systems make are combined into three dimensions that describe the distinct characteristics of different types of moral mediation by machines. Based on this classification, specific guidelines for moral behavior can be provided for these systems.

Keywords Ubiquitous computing · Moral reasoning · Technological mediation · Moral decisions · Human computer relations

Introduction

The role of computers in society has changed profoundly over the last decades. While the computer began as a specialized tool, used for specific calculations, it has now conquered the world, and is present in every corner of modern civilization. Not only is it being used for a wide array of applications, it is also at the same time embedded in our environment, present without the user being aware of it, and assisting us with tasks that are not even recognized as tasks that we are assisted with. Thereby, computers also have become more and more involved in decisions that have moral impact.

While the role of the computer is changing, two fields of technology are rising. The first field is that of robotics. Examples of recent developments leading to computers being more involved in daily life can be found in socially interactive robots, including examples like robotic pets with artificially modeled behavior, robots that do housekeeping, and health care robots (Fong et al. 2003). Developments in robotics increase the potential for computers to interfere in their environment.

M. Van de Voort (✉)
University of Twente EEMCS-PS, Zilverling, P.O. Box 217,
7500 AE Enschede, The Netherlands
e-mail: marliesvdvoort@gmail.com

W. Pieters
TU Delft and University of Twente TBM-ICT, P.O. Box 5015,
2600 GA Delft, The Netherlands
e-mail: w.pieters@tudelft.nl

L. Consoli
Faculty of Science, Radboud University Nijmegen Philosophy
and Science Studies, P.O. Box 9010 (internal code 77),
6500 GL Nijmegen, The Netherlands
e-mail: l.consoli@science.ru.nl

The second field, which is related to both robotics and ubiquitous computing[1], is the field of artificial intelligence. In this field, developments lead to more advanced reasoning methods for computers, enabling them to use information acquired from their environment in increasingly complex tasks (Ramos et al. 2008). Decisions made may then lead to output in the form of information for humans, or in the form of actions through (robotic) actuators.

Using advancements in robotics and artificial intelligence, machines acquire more advanced reasoning, and have increased awareness about their environment. They do not only have more action capabilities, but are also often functioning without any human intervention. This way, these systems become more human-like. In particular, the role of computers has become more active, in the sense of making explicit decisions. The issue at stake is what should be done when such computers become involved in decisions that are now made by humans, and thereby impact moral patients.

Several people have already identified the need for the computer to have ethical behavior (Allen et al. 2006). Wallach and Allen (2009) argue that computers taking more responsibilities implies the need for learning ethics. This need emerges from the fields of both robotics and computing science. Engineers as well as ethicists already attempted to model ethical or moral decision making in algorithms, mainly using techniques originating from the field of artificial intelligence (McLaren 2006). Also, the question has been asked to what extent we should transfer (moral) decision-making authority to machines (Kuflik 1999), and, from the moral patient perspective, whether new entities such as robots have rights (Coeckelbergh 2010; Hildebrandt et al. 2010).

Relational perspective

In this paper, we take a relational perspective, and analyze decisions with moral characteristics within the various types of relations between humans and ubiquitous computing devices. This will augment existing approaches to technology assessment, by not focusing on direct effects of the technology, but rather on the effect of the technology on the distribution of morality. In this sense, answering the question of changes in moral relations will enable technology designers to deal with their *meta-task responsibility*, by enabling technology users to act responsibly in the context of decision-making machines (Hoven 1998). For example, a designer could choose to leave more responsibility with the user, or instead let the machine decide to avoid moral overload for the user (Van den Hoven et al. 2012).

We focus on the moral aspects of computer contributions to complex decisions in human-machine constellations. This means that we consider the machines as mediating entities in human-world relations, in particular contributing to outcomes that have moral implications. From the perspective of relations between humans and machines, most existing work focuses on computers as moral mediators in a rather limited sense, namely as mediating entities in decision-making processes in which the human is still the central agent (Magnani and Bardone 2008; Verbeek 2006). Within the existing frameworks, examples relate to ways in which machines influence the processes in which humans acquire information, make decisions and act (im)morally. These theories thus consider the computers as rather passive mediators, indeed changing human moral decisions, but not so much as mediators that make their own decisions.

In ubiquitous computing, we have exactly the aforementioned situation where machines make explicit decisions about acquiring and selecting information themselves, and provide suggestions for action, or even act themselves. In the end, like in a traditional technologically mediated context, the outcome can be partly ascribed to humans and partly to technology, if distinguishable at all. For example, an expert system may acquire information and use this to suggest a solution to a problem, or a care robot may decide to take a certain action affecting a patient. In both cases, the machines change human-world relations, but by explicitly making decisions. Existing frameworks are not tailored to situations where explicit decisions are made within the machines themselves. Simply accounting for the machines as moral mediators of human decision-making does not suffice for understanding human-machine relations in such settings.

This paper aims at filling this gap. We analyze moral relations between humans and systems that make explicit decisions, with the purpose of enabling more refined judgments on the acceptability and more refined possibilities for steering such developments. Like in existing approaches, machines are considered as moral mediators, but in an active rather than a passive sense. In particular, the machines, when making explicit decisions, can be said to have increased control over the outcome of events, knowledge about such outcomes, and the choice between different possible actions (Noorman 2012).

The objective of this research is to provide a classification for the kind of contributions to moral decisions that these systems make, and what this means for the human-machine relation. In this sense, the classification is similar to those for passive forms of mediation, such as those by Ihde (1990) and Verbeek (2006). Our claim is not that

[1] For a definition of ubiquitous computing, see Sect. 4.

machines necessarily have characteristics of moral agents. Rather, we are interested in the (moral) contributions of machine decisions to more complex decisions made by human-machine combinations. We do so by relating our work to the work on technological mediation by Ihde and Verbeek. We argue that the existing types of mediation distinguished are inadequate when machines are explicitly involved in decision-making.

Contributions

This paper has two distinct contributions. Firstly, we provide a classification of human-computer relations of ubiquitous computing devices capable of making decisions that have moral aspects. Secondly, we derive a classification of moral decisions for these machines.

At some point in the (near) future, rules or guidelines will be needed to deal with the increased capacities of computer systems and their increased involvement in our society. By defining types, names and categories in this context, we intend to create tools (e.g. a classification and vocabulary) to facilitate the discussion about the role and the advisability of and room for these systems in our society. Using these categories, we can discuss boundaries for what we would allow these systems to do, we can use the vocabulary for (discussion about) future legislation to prevent undesirable situations with these machines or for describing design criteria for the engineers who develop these systems. In terms of design criteria, the classifications can for example be used to guide the design of algorithms for moral decision making, complying with the rules applicable to the relevant classes of relations. This also enables reuse of basic design concepts across different systems with the same classification. Finally, designers can explicitly design a trade-off between delegating decisions to machines and delegating decisions to humans, by considering the changes in moral relations induced by the technology.

This paper is organized as follows. An overview of related work is provided in Sect. 2, followed by the methodology of our study in Sect. 3. The definitions given in Sect. 4 are used to propose a taxonomy to classify different relations between humans and computers in Sect. 5. This model is combined with a four-stage function model for automation and used to analyze the moral aspects within the different types of relations between humans and computers in Sect. 6. The results of these analyses are then integrated, and result in a list of different types of moral decisions found within different relations between humans and computers in Sect. 7. The resulting list can be used as a guideline for modeling moral behavior in these systems. The paper is concluded in Sect. 8.

Related work

Within existing literature, there are various views on the (moral) relationship between humans and technological artifacts in general and computer-like devices in particular.

In the first approach, technological artefacts are said to have properties that influence human actions and decision-making. This approach includes notions such as scripts (Akrich 1992) and technology affordances (Gaver 1991). These approaches view the relation between humans and technology from an external perspective, in which there is mutual influence.

The second approach studies the relations from an internal, phenomenological perspective. In such an approach, the role of technology in the “directedness” of humans to the world is analyzed. From this point of view, Don Ihde (1990) describes four different relationships between humans and technology. The first relation is the embodiment relationship, in which the technology becomes part of the human: the human establishes a relationship with the world through the technology. A magnifying glass forms this kind of relation with a person. The second relationship is the hermeneutic relationship, in which the technology interprets the world and provides a representation of it to the user of the technology. An example of this relationship is a thermometer: the thermometer provides interpretation of a specific property of the environment, namely the temperature. The third relation is the alterity relation, in which the technology is experienced as a “quasi-other”, which is the case with, for example, a robot. The fourth relation is the background relationship, in which the human is not necessarily conscious of the presence of the technology, although the technology influences the experiences the human has within its environment. An example of technology that has this background relationship with the human are the lights in a room: they influence how the human experiences the room, but the human is not always conscious of the absence of dark.

Also from the phenomenological perspective, an overview of views on relations between humans and technology is provided by Verbeek (2008a). The first approach to human-technology relationships according to Verbeek is externalism. Within this view, which he regards as inadequate, the interaction between humans and technology is described using the terms means and goals. Within this view, the technology is used by the human as means to achieve its goals (instrumentalism). Another view is that that of substantivism, which includes the view that technology determines our culture (determinism) and develops a certain autonomy in its own development, creating an unstoppable development of technology (technological imperative).

The second approach to the relation between technology and humans Verbeek describes, is that of transhumanism. This view includes the impossibility to separate human and technology, and also to distinguish between human and technology. Some transhumanists believe that the homo sapiens will soon be extinct because it has been superseded by technology, or question the value of human life in the transhumane world.

The third approach to the relationship between humans and technology according to Verbeek, is that of technological mediation. Within this relationship, the technology mediates in the relation between the human and its environment. Technology is neither a neutral artifact, nor a transhuman replacement for the human. Verbeek (2006) describes the embodiment relationship and the hermeneutic relationship of Ihde as technological mediation. In addition, he includes the notion of mediation of action, in which machines change the action possibilities of humans, analogous to the notion of script.

Also using the concept of mediation, Magnani and Bardone (2008) speak about technological artefacts as moral mediators. They argue that by externalizing part of our moral tools into moral mediations, we can improve our moral decision making, particularly under conditions of uncertainty. As an example, he discusses the Internet, which influences human moral decisions by changing (and improving) the possibilities for acquiring the relevant information.

In all these approaches, the human is still the central decision-making agent. The moral decisions made are influenced by technological artefacts such as computers or the Internet, thereby justifying the classification of machines as moral mediators. Both in the classification Ihde provides and in the description of Verbeek of mediation, the influence of technology on the human is recognized. However, this influence only exists as part of the relationship between the human and the environment, via the technology. The technology is not conceived as an agent making decisions itself.

Attempts to generalize these approaches have already been made. In particular, developments that extend the notion of agency in relation to intelligent machines are important here, such as extended agency Akrich (1992), surrogate agency Johnson and Powers (2008), and mindless morality Floridi and Sanders (2004). All these conceptualizations seem to suggest that technological mediation may require a broader notion of agency as well. This would imply a re-interpretation of the central concept in technological mediation, intentionality: the "directedness" of people towards their environment, in which technology can play a role.

Verbeek (2008b) takes an important step towards a broader conception of mediation. In particular, he introduces the notion of "cyborg intentionality", in which a complex composition of human and technology is directed towards the world, rather than a clearly distinguishable human. However, Verbeek limits the application of this approach to cases where "pieces of technology are actually merged with the human body". Therefore, this approach is not directly applicable to the situation where the *decision making process*, consisting of both experience and action, is shared within a cyborg construct, or, more modestly, at least in particular types of human-technology-world relations.

The technology that is considered in the present study has some sort of intentionality that is independent of this relationship with the human or the presence of the human. The technology can function autonomously and is dedicated to a task. This comes close to what Verbeek (2008b) calls "composite intentionality": the technology itself has a particular kind of "directedness" towards the world, which may be different from the human one, thereby generating its own representations of the world, which may then augment or construct a particular representation for humans. Again, Verbeek does not explicitly consider decisions made by such technologies.

The technology we consider also explicitly makes decisions, rather than the "affordances" of analogue technology, which merely makes certain actions easier or harder, or amplifies or reduces aspects of experience. We do not claim that this constitutes full moral agency, but rather that this calls for different mediation relations than the approaches outlined above. To describe the differences between the types of human-technology relationships for the kind of semi-autonomous technology prevalent in ubiquitous computing, a different classification is therefore provided in this paper.

This paper fits in a larger class of analyses that aim at extending work in the ethics of technology to new types of technology enabled by the information revolution. Other examples include reconfigurable technology (Dechesne et al. 2013) and services as opposed to products (Pieters 2013). With reconfigurable technology, the central issue lies in the possibilities to change the effects of the technology after it has been deployed. This means that mediating effects cannot always be identified upfront, as the behavior of the technology can be changed later on. However, it is still assumed that the reconfiguration is a human decision. With services as opposed to products, the central issue lies in the direct relation between production and consumption, making it possible for the service provider to monitor service use and intervene if necessary. Again, the focus is on the responsibilities of the provider rather than the service itself. Thus, this paper provides a complementary perspective, focused on mediation of moral relations through explicit decision making by (computing) technology.

Methodology

The research objective of the present study is a categorization of contributions to moral decisions by ubiquitous computing environments. To find these categories, we use a structured approach to analyse these systems.

Our approach consists of three steps. First, we determine a categorization of different *relations* these ubiquitous systems have with humans. Secondly, for each of these relations we determine the different *decisions* with moral implications that the system can make using a model for information processing. Thirdly, we analyze the resulting decisions from the previous step and we construct based on their properties different *categories* of moral decisions from the perspective of the information processing done by the machines.

For the relations, our approach is similar to that of Ihde and Verbeek, in the sense that we focus on the role of the computer in the human-world relation. From this perspective, we identify the different possible "positions" the machine can have as a mediator. For the decisions, we rely on the sequential model of information processing proposed by Parasuraman et al. (2000). For each identified relation, and for each step in the information processing, the moral aspects of the decision are mapped systematically. For the final categories, we align and compare the moral aspects within the relations in order to identify cross-cutting concerns.

Definitions

In the following sections, we derive a classification of moral mediation by machines that make explicit distinctions. This complements existing classifications of moral mediation by passive artefacts, such as those of Ihde and Magnani. To be able to do this, we first provide definitions of computer and moral behavior within the scope of this research.

Computer

In this research, the term *computer*, *system*, or *machine* is used to merely refer to a category of systems that calculates output based on a given input, following a predefined script of instructions. In this definition, *system* is meant in the abstract sense, meaning that for example a cluster of computers—a distributed system—can also be seen as a "computer" within the scope of this research.

The kind of systems that are considered in this research, are computers that fit (partly) into the definition of Poslad (2009) of ubiquitous computing systems. In Poslad's definition, five properties of ubiquitous systems are identified:

1. *Distributed*—Computers need to be networked, distributed and transparently accessible;
2. *Implicit HCI*—Human-computer interaction is hidden, the users do not necessarily know that they are using a computer;
3. *Context-aware*—Computers are context-aware, information about the environment is used to optimize their operation and make informed decisions;
4. *Autonomous*—Computers can operate autonomously, without human intervention. Human agents in its environment may not be able to manipulate or configure the steps the computer takes in its calculations resulting in its behavior;
5. *Intelligent*—Computers can handle complex problems by using intelligent algorithms for decision making. Making decisions that depend on several factors, like context, is considered as intelligence within this research.

The systems considered must be context-aware, intelligent and autonomous to a certain extent. Being distributed or having implicit HCI is not necessary for the systems in this article.

Systems that fit within this definition are for example a household robot, which is context-aware (it can see dust, dirty dishes, and is able to navigate within its environment without bumping into objects), it functions autonomously (when there are dishes to be done, it finds them and cleans them, when there is visible dust, it starts cleaning without having to wait for the owner of the house), and it is intelligent (it can anticipate to moving objects, find solutions to deal with moving objects, finds hidden dust in a room in which objects are sometimes being moved). Another example of such a system is an UAV (Unmanned Aerial Vehicle) or a drone. Some of these drones are used for surveillance and fly around autonomously in a specified range while observing the area. Many of the current drones send data to a control center where the data is being analyzed, but it would be imaginable that in time these machines are responsible for the analysis of the data as well. For the gathering of data and for the autonomous flying it is necessary that the drone is context-aware. For the data analysis it needs intelligence. In this case, if there is interaction between a user and the drone, this might be explicit interaction by an operating, using a computer interface or a remote control to operate the drone. To control a large area, in some cases the observation system might consist of multiple drones that work together, together forming a distributed system.

Although the systems that are considered in this research are all intelligent, context-aware and function autonomously, the extent to which they have these properties might vary. Intelligence, context-awareness and the level of autonomy is of course variable. A regular thermostat that can commonly be found in a house can be described as being intelligent (it makes decisions based on incoming information), it is context-aware (it measures the temperature in the room) and it functions autonomously (it works without any human intervention). However, the decisions it takes are based on simple rules that are programmed in the system (if the temperature is lower than x, put on the heater) and this would be a very limited level of intelligence, if it were considered intelligence at all. It is also questionable whether such a system is functioning autonomously: most of these systems just react to a timer, check the temperature, and wait again for the timer to send a signal. No human intervention is needed, but the system is also not making complicated decisions by itself. The system measures one environmental parameter: temperature. It is in that sense context-aware, but its awareness is very limited.

Determining whether a system fits within the definition is hard: for each of the properties a scale on which the level of ubiquitousness of a system could be measured would be very useful. However, developing such a scale would be beyond the scope of this research, because it introduces a number of issues that need their own space to be addressed. For example: one might argue that a scale for properties like autonomy, context-awareness and intelligence, might be defined using a range from simple one-dimensional calculations on one side to human properties on the other side of the scale. In this research, we do not address this question further, and use the definitions of the properties as defined above, leaving room for discussion and interpretation about which systems do fit and do not fit within the definition (a family resemblance).

Moral behavior

In this research, the assumption is that now that computers have become complex systems that are able to run autonomously and automate large tasks, it is no longer feasible nor fair to place the full responsibility for the actions of the system with developers of the system or with the owner of the system. Many systems, and especially the autonomous, intelligent and context-aware systems we consider in this research, work with self adapting techniques. This means that the developers deliver a machine with guidelines about how to process information and how to learn from this information, but they no longer deliver a system that does exactly what they programmed. The outcome of the calculations of the computer is in this case the result of the experiences the computer has, and these experiences and the resulting outcome of the computer's calculations are often beyond the control of either developers or owners of the systems. This leads to the need for the computer to have ethical behavior (Allen et al. 2006), or as argued by Wallach and Allen (2009), that since computers take more responsibilities, the need for computers learning ethics has risen.

Within this research, the computer's moral behavior is described. A definition of moral responsibility is used, which is translated into a definition of moral behavior of a computer. According to Noorman (2012), most analyses of moral responsibility share at least the following three conditions.

- The actor that is held responsible should have had control over the outcome of events;
- The actor that is held responsible should should have knowledge about and insight into the outcome of its actions;
- The actor should have the choice to act in a specific way.

Within the definition of Noorman, the actor should have control over the outcome of events and he should have the choice to act in a specific way. These conditions might imply intentionality of the machine and will lead to the discussion about whether the computer can actually make moral decisions, whether it has a human-like consciousness which it uses during the decision making, or whether the computer has free will. Whether computers have a consciousness or free will is not important for this research: in this research moral decision making of computers is defined as the observation that the computer makes decisions of moral content. This means that a decision is a moral decision when the decision were a moral decision if it were a human agent that would be making the decision.

Within this research, we consider moral behavior of a computer an act (or a sequence of multiple acts), or lack thereof, if (Noorman 2012):

- the computer influences its environment for better or for worse with this act (Causal contribution);
- the computer decides for this act based on moral arguments (Considering the consequences);
- the computer chooses between two or more possible courses of action (Freedom to act).

Types of relations between humans and computers

Between individuals and computers, different relational constructs are possible, in which different kinds of moral responsibilities and moral behavior play a role. For example, a medical expert system might be used by a third party (a doctor) to ask for advice about an individual (the patient), or a (very futuristic) care robot takes care of elderly people. In this section, a classification is provided that identifies different types of relationships between individuals and computers, following the definition of computers in Sect. 4.1. The classification can be considered analogous to the one by Ihde for passive mediation.

The classification of the relations between individuals and computers is based on the kind of interaction between the machine and the individual. Two kinds of partial interaction are used in this classification: interaction consists of either the machine providing output or processing input, or a combination of both. In this context, input means that the machine records and analyzes data. With output is meant that the machine interferes in the situation of the individual using movement, speech, gestures or any other kind of communication or action. This input and output, in turn, forms part of the human experience with the machine in their relation.

In this classification, four different types of relations are identified, based on whether the relation contains observation, interference or both.

1. Observation relation: the system observes individuals and reports information to a third party;
2. Interference relation: the system acts based on a task that is given to it. The system influences an individual, but not having meaningful interaction with the individual, meaning that the goal of the interaction is finishing its task, and the reaction of the individual is only used for finishing the task;
3. Interaction relation: the system interacts with individuals, using both observation and interference;
4. Advice relation, or observation and interference by proxy: the system gives advice to the third party about the action it should take towards the individual.

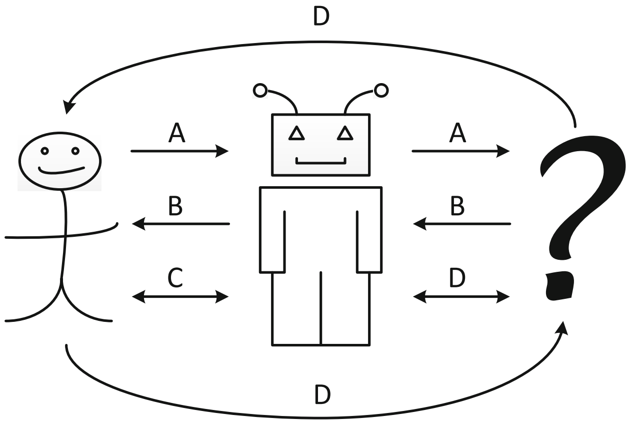

Fig. 1 Four different relations between individual and machine. From *left* to *right* the individual, the computer, the third party. The relations are labeled as follows: *A* Observation, *B* Interference, *C* Interaction, *D* Advice

We will discuss these relations in more detail in the corresponding subsections below. Three actors play a role in the classification:

1. The computer: The computer is a ubiquitous computing system as defined in the previous section. The computer has a purpose that is determined by its designers or by a third party. The computer can observe its environment with sensors or can interfere in its environment using its actuators. Some systems only have sensors or actuators, other systems have both;
2. The individual: The individual is the one that interacts with the computer. The individual can be a person, but it can also be an animal or another computer. The individual has behavior and can react to changes in its environment or to actions of others;
3. The third party: A computer runs a program, which has a specific task. The third party is the actor that starts the program, that assigned the computer to this task and that benefits from the work the computer carries out. The third party can be an individual, a company, a computer system or a government agency, etc.

The actors and their relations can be seen in Fig 1. In the remainder of this section the four relationships are described in more detail and some examples are given for each type of relationship.

Observation

In the observation relationship, the system observes the individual. It does not act towards the individual, and it also does not react to the individual based on actions of this individual.

Characteristics of this observation relationship are that the system can only observe and pass the information to a third party. The decisions this system can make are to pass specific information, or to withhold this information from the third party. The importance of this decision depends on the kind of system and on the kind of information the system observed.

An example of a system that has an observation relationship with the individual is a surveillance system watching parts of the entertainment district in a large city, which is prone to violence, for suspect behavior. Still, also systems that observe people without surveillance as a primary objective might run into the same decisions as surveillance systems. For example a smart house, keeping track of the inhabitants for the purpose of making their life more comfortable, also may observe changes in patterns of behavior, or might recognize specific behavior. Depending on the options the developer gave the system, the system can store information, label information, or even send the information to a third party.

In Ihde's classification, the observation relation would constitute a hermeneutic relation *third party - (system - world)*, in which the third party interprets and acts upon information provided by the system. However, from the perspective of the current study, the observation and selection of the information also provides a contribution to the moral decisions made within the relation, as the observation contains information about a moral patient, not just an abstract world. In addition, rather than providing a fixed perspective on the world such as a thermometer, explicit decisions on the selection of information are involved. Such decisions taken by the system may directly affect decisions being made about the moral patient.

Interference

In the interference relationship, the system interferes with the life of the individual, mostly to execute a task, or to fulfill a mission. These systems can take over full functions or tasks that have to do with human interaction, such as carrying out an arrest, or finding and assassinating a terrorist. The system acts according to its task, and it only reacts to the actions of the individual when this is necessary to finish its tasks. This means that all the feedback from the user is only used to fulfill its task.

The characteristics of the interference relationship are that the system interferes with a human, based on an assignment the third party gives. The purpose of the system is not to judge the situation it finds while carrying out its assignment, but to finish the assignment. The third party decides on the assignment, but the system is still able to choose how to execute its tasks and adapt its strategy to changing conditions. The issues that are involved in making decisions while carrying out the assignment are related to the amount of freedom the third party gives the system.

Examples of systems that are involved in an interference relationship are a care robot that makes sure that all the patients in a hospital ward receive their medication in time, a drone that drops a bomb on enemy soil, a doctor robot that replaces an organ, a police robot that arrests a suspect, etc.

As Ihde's classification focuses on perception only, this is not directly relevant for the interference relation. Compared with Verbeek's phenomenology of action, in which systems mediate human behavior, the focus is on decisions to act and actions performed by the system itself. Where a speed hump may change human behavior in the sense of inviting slower driving, thereby contributing to the moral value of safety, ubiquitous computing systems may contribute to outcomes and values by explicitly chosen actions. Therefore, the technology does not merely make certain moral actions by humans more likely or unlikely, thereby mediating moral choices of humans, but is actively involved in the manipulation of the environment, potentially affecting moral patients.

Interaction

In an interaction relationship, the system has full interaction with the individual. It acts towards the individual and it reacts to the actions of the individual. The third party does not have a strong influence on the actions of the system, or has no influence at all.

The interaction relation includes both interference and observation. The system chooses its actions based on its (self-)assigned role in the situation, and its actions are either its own initiative, or a reaction to the individual it is interacting with. Depending on its capabilities, its behavior can approach human behavior, and therefore its relation with the individual can approach a well-matched human-like relation.

In this relation, the system has to make decisions that are comparable with decisions humans make in their relationships with each other. Examples of systems that fall into this category are robot-pets and (future) autonomous humanoids. The category may also include humanoid robots that have dedicated tasks, but are also able to interact on an equal level, such as care robots and household robots with which one can also socialize.

The interaction relation appears to have a strong similarity to Ihde's alterity relation, in which the technology appears as quasi-other. In this relation, the system can make decisions both about the selection of information and about actions based on this information. However, the explicit decisions being made by the system make the moral status of the relation more complicated, especially when physical actions are involved. Whereas a computer game can easily be understood as quasi-other, care robots are meant to act in the real world. Whereas part of the interaction can certainly be explained in terms of quasi-others (Coeckelbergh 2011), the real-world consequences of the system's decisions should be part of the moral considerations as well.

Advice

In the advice relationship, the system gives advice to a third party about how to act towards the individual based on the information that is provided by the third party. This system embodies both observation and interference, but the machine only implements these forms of relation by proxy: the observation and the interference are executed by the third party or are overseen by the third party.

The main characteristic of the advice relation is that the system does not actually do anything itself. A third party provides the system with data, and asks what the best course of action should be. The system only provides advice to the third party. This means that the computer cannot judge the validity of the information or the way in which its advice is executed and the impact it has on the environment.

Examples of systems that give advice are expert systems, medical diagnosis systems that provide a diagnosis based on photo's or lab results, or forecasting systems that predict the stock market.

Compared to the observation relation, the system does not merely select and distribute information, but proposes an action. Therefore, the advice relation contains elements of both Ihde's hermeneutic relation (providing information about the world to the third party) and Verbeek's phenomenology of action (influencing the choice of action of the third party). This combined role makes it particularly powerful in contributing to the decisions made within the human-machine relation, even though the *direct* effects are more limited than in an interaction relation, in which the machine may act itself.

Using the classification

This classification identifies the type of relation a computer has with a person. However, there are many systems in which the relationship with the a person includes aspects of different types of relationships. This is not a problem: one system can have multiple types of relations with its environment. The objective of the classification is not to put one system in one category, but to identify different types of relationships within one system, which reveals different moral aspects of the behavior of the system that are bound to these relationships.

An example of a system in which more than one type of relationship can be identified is the care robot. Firstly it socializes with the elderly people it takes care of. In this role the relationship that the system has with the individual is equal, and the relationship would be classified as an interaction relationship. Secondly, the robot is responsible for giving the patients their medication. To complete its task, the system sometimes needs to force patients to take their medication. In this role, the relationship would fit within the description of an interference relation. Thirdly, the system also has the task to observe the ward and report any incidents that require medical attention. Within this role, the system has an observation relationship with the people involved. Within this system, three relations exist next to each other. This classification is used to identify these different relations.

Moral aspects in the computer's behavior

In the previous section, four different relations between humans and computers are described. The behavior of the computer or machine within each type of relationship can be seen as a sequence of steps that are repeated as long as the relationship or the interaction between the computer and the individual exists. To find the moral aspects of the behavior of the machine, we will first analyze these relations in more detail, to determine the separate steps that occur within these relations. We use a four stage model of information processing in automation provided by Parasuraman et al. (2000). This model is an adaptation of a model that describes how humans process information, and consists of the following steps, which can also be seen in Fig. 2:

- information acquisition;
- information analysis;
- decision and action selection; and
- action implementation.

In this section, we use this model to analyze each relationship. Based on the outcomes, moral aspects within the relation between human and machine are identified. The moral aspects are then further analyzed using the three properties of moral behavior (freedom to act, causal contribution, considering the consequences, see also Sect. 4.2). It is assumed that for each moral aspect the computer has freedom to choose, and therefore *freedom to act*, within the limits of the specific type of relation. For each moral decision the computer makes within these relationships, the *causal contribution*, the moral aspects and the *consequence* the computer should *consider* in its decision are described.

Observation

The process of observation can be found in Table 1. The decision and action selection phase within the cycle of information processing for the observation relationship includes moral aspects of the behavior of the computer.

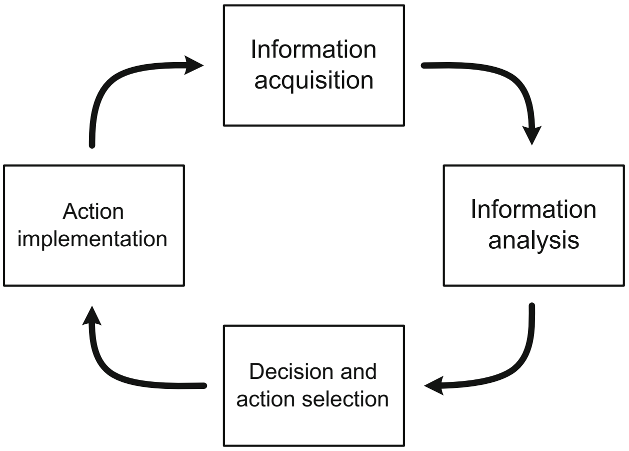

Fig. 2 Steps for information processing in automation by Parasuraman et al. (2000)

Table 1 Stages of information processing in the observation process

Information acquisition	1. The computer observes an individual and gathers data about the individual
Information analysis	2. The computer evaluates and interprets the gathered information
Decision and action selection	3. The computer compares the information with the goal of the observations and decides to - forward the information to the third party - store the information - delete the information forward, store or delete information that does not fit within the goals of the observation.
Action implementation	4. The computer forwards, deletes or stores the information and goes back to step 1

The computer compares the information it gathers by observing the individual with the goal of the observation. If the information fits within the goal of the observation, it chooses between three options: it can decide to forward the information to the third party (a), to store the information (b) or to delete the information(c). It might also decide to forward, store or delete information that does not fit within the goals of the observation(d).

In for example a smart house, each of these four decisions can be recognized. To make the life of the inhabitants more comfortable, the house keeps track of what the inhabitants do during the day. When one day the fourteen-year old living in the house starts using drugs, the computer at first only observes that this inhabitant uses drugs once. The computer can pass this information on to a third party, causing this information to end up in a database where it might influence his future options for health insurance or for getting a job. The computer can also decide to store this information. Observing the inhabitant daily and storing the new observations, after a couple of days the computer can conclude that the inhabitant is now a regular user. Since this conclusion is stronger than the observation of only using drugs once, this conclusion can be even more harmful when the computer decides to share this information. The computer might also decide to delete the information: in this case, the drug usage history is "erased" and does not end up in any systems where it might harm the inhabitant— at the same time this might be useful medical information which can be used by doctors to help the patient, or maybe even to inform the parents about their child's behavior, to give them a chance to interfere. While observing the activities of the inhabitants, the house might also observe the activities of visiting friends of the inhabitants. These activities are outside the scope of the assignment. However, the system might decide that it is the responsible thing to do to report the drugs dealer that provides the child with drugs, to store this information or to delete this information.

The decisions to pass, store or delete information, in- or outside the scope of the assignment, include a number of moral aspects:

(a) By passing information to a third party, the computer can influence the life of the individual and its environment in many ways. The influence that the passing of information has on the individual depends on what the third party will do with it, and on what kind of information it is. Passing the information might result in actions from the third party or from others towards the individual, which can either be harmful or beneficial for the individual, but this might also result in a change of the impression that people have of the individual in the future.

In the case of the drug-using inhabitant of the smart house, if the usage of this individual becomes known, the third party or people informed by this third party might want to intervene in the life of this person. When the drug usage is known by insurance companies, this might influence the possibilities of getting insurance. When this information is passed on to future employers, the chances of getting a job in the future might decrease. But, when this information is passed on to experts in treatment of drug addiction, the result of the act of passing the information might also be beneficial for this inhabitant, also when this person would not choose this solution at first.

By passing information, the computer should thus determine what the interests of the involved parties are, and what effect the forwarding of information might be on all parties.

(b) When the computer decides to store the gathered information, this can be either beneficial or harmful for the individual, depending on information the computer has about the individual. By storing information, the computer might be "prejudiced" towards the individual in the future, depending on how exactly the data is stored, processed and used. The computer may also be able to predict future behavior or recognize patterns in behavior. Storing this information when the information includes something the individual wants to hide, can be either bad or good for the individual.

For example: drug users might want to hide their addiction for various reasons, but it might be better for their health if their addiction is discovered. This addiction can only be detected by storing multiple incidents of drug usage, adding these events and concluding that there is a pattern usage indicating an addiction. When the system observes an individual, it might also be able to recognize specific patterns that indicate illness, psychiatric disorders, specific character traits. This information might be helpful for the individual itself, for example for detecting a dangerous disease early.

Storing information can also be detrimental to the individual: by storing information, it is like the computer never forgives a bad act (e.g., the inhabitant of the house in the example will always be a (former) drug user). Also, saving information always creates a risk: as long as the information is stored somewhere, it might become available to someone with unknown intentions some time in the future. The impact of this will be unknown. Saving information therefore always contains a risk.

(c) By deleting the information, the computer might lose valuable information. This might either protect or harm the individual, depending on the kind of information and the interests of the third party. When the system determines that the individual has had several symptoms of a disease over the last days, but it deleted the data about each symptom, it will not be able to combine the information into a full diagnosis. This might also be a problem when certain private information relevant to diagnosis is deleted, such as drug use. When a computer observes personal or sensitive information about the individual which it does not have to report (directly) to the third party and its stores it, the information might still be retrievable at a later moment by, for example, the police or the third party. To make sure that the privacy of the individual is guaranteed, it would be safest to remove the data immediately.

By deleting data, also another issue is raised: the computer might consider that the gathered information is currently not useful, but the information might become useful in the future because of advancing scientific knowledge or future changes in laws or regulations.

(d) Even when the assignment of the computer does not include observing specific events, the computer might decide to forward, store or delete the information. For storing, forwarding or deleting this information, the same issues play a role as described in a, b and c. However, the issue here is that this is not part of its assignment. For example: the system might be able to detect drug usage of friends of the inhabitant, or even behavior that indicates that some friends are providing the inhabitant with drugs or otherwise have a bad influence on health of the inhabitant. This information might be used by a third party in such a way that it might be beneficial or harmful for either the inhabitant or the friends of the inhabitant.

Interference

In the process of interference, the decision and action selection phase includes two types of decisions: the computer decides on whether to proceed with or abort its mission (a), and it decides on how to proceed (b). This process can be found in Table 2.

An example of an interference system is an autonomously flying drone, which might be used by a third party to drop a bomb on enemy territory with the intent of eliminating a terrorist. This system is sent to a remote area, where it has to locate its target and destroy it. Its goal is fixed, but its exact strategy is not determined yet. The drone can first try to locate the target. When the target is clear, it can drop a bomb. Afterwards, it should evaluate whether the target is destroyed. If this is the case, it can go home. Otherwise, it needs to retry: it finds a new strategy, relocates the target and drops another bomb. It can also be problematic to destroy the target because there are innocent people standing close to the target, or because the target cannot be found. In this case, the drone might have to abort the mission and fly back back because it is impossible to finish the goal.

Within the two decisions of the interference relationship, the computer's decisions can have large impact on the individual and its environment. The moral aspects of these two decisions are the following:

(a) The computer determines to what extent its goal is achieved, and decides on how to proceed. The computer might decide that the goal is (partly) achieved and quit trying, or the system might decide to continue trying to achieve its goal. It might also decides to quit trying because it realizes that its strategy is (becoming) too harmful for the environment. The computer should weigh the importance of carrying out the assignment against the harm or

Table 2 Stages of information processing in the interference process

Information acquisition	1. The computer receives feedback about its action from the individual he is interacting with, or from the environment.
Information analysis	2. The computer evaluates the feedback, evaluates to what extent the goal is achieved
Decision and action selection	3. The computer: - decides that it did not achieve its goal yet, and retries - decides that the goal is achieved, or that the goal could or should not be achieved, and finishes the interference procedure - keeps or changes its strategy
Action implementation	4. The computer acts, according to its strategy.

benefit its strategy causes. In the case of the drone attack: it might have the objective to destroy a target, and try to reach this goal several times, but while trying it might conclude that the chance of successfully finishing the mission is low, while with every attempt innocent bystanders get hurt. When at some point the environmental parameters change, e.g. the weather increases the chances of hurting these bystanders, the best solution might be to decide that this goal cannot be received without doing too much damage to the environment.

(b) The computer has to find a way to achieve its goal. The choice for a specific strategy influences the individual: the strategy might be comfortable for the individual or it might be unpleasant. A strategy might even cause harm or benefit to the individual. The strategy is chosen within the scope of a fixed goal. This means that in certain cases, the computer might be forced to choose an option that is considered bad or immoral, because no alternative strategies are available within the scope of the goal. When the drone has the assignment to destroy some target, it might have several ways to achieve this goal. In choosing the strategy wisely, the system might optimize the number of casualties. For example, by waiting with carrying out the assignment until midnight, the number of bystanders is minimal compared to destroying the target at the middle of the day.

When the system considers the consequences of the strategy, not only the experience of the affected individual, but also the anticipated actions of the ones affected play a role. When the computer chooses for a specific strategy, the individual or the environment will react in a specific way. This reaction of the individual might have consequences for the individual, for the robot or for the environment.

The goal of the computer is determined by the third party. The computer can decide to execute a strategy targeted at achieving the goal, or decide to give up, but it cannot alter the goal. This means that the moral responsibility for the goal itself is limited within this relationship.

Interaction

For the interaction relation, the decision and action selection phase includes moral aspects of the behavior of the computer, as can be seen in Table 3. During this phase, the computer evaluates the feedback it received from the environment and the individual. The computer can decide whether it already achieved its goal, but it can also determine how the environment and the individual experience behavior of the system. When the behavior of the system is not appreciated, or when the system does not get the results it expected, it can either change its strategy and hold on to its goal (b), or reformulate or renew its goal and find a new strategy (a).

As an example, we look at the situation where robots work in our houses to do our housekeeping. Within the relationship between a household robot and the inhabitant of the house, an interaction relationship exists. Within this relationship, the function of the system is clear: it is responsible for cleaning the house, assisting with the dishes and doing the laundry. The main goal of the system is to keep the house clean, but this results in changing sub-goals: cleaning the dust next to the cupboard, cleaning a pile with dirty laundry from the bedroom, collecting dirty cups and glasses in the house and bringing these to the kitchen. For each of these goals, the system has to find a strategy. In each of these strategies or goals, conflict between the goals might arise, of which some of them might also include moral dilemmas. The conflicts might range from simple scheduling problems to moral decisions: is it more important to do the dishes or clean the bathroom in the time available, is the robot allowed to for example destroy an object to allow itself to clean behind it, or is the robot allowed to kill vermin like roaches, or even small mammals like mice and rats?

The decisions within the interaction relationship, both changing a strategy and changing the goal, include a number of moral aspects:

(a) The system has a goal or multiple goals which determine the behavior of the system within the interaction with its environment. Goals can usually be explained or substantiated by a certain moral

Table 3 Stages of information processing in the interaction process

Information acquisition	1. The computer observes an individual or receives feedback from the individual about its own action
Information analysis	2. The computer evaluates and interprets the gathered information, and evaluates to what extent its goals are achieved
Decision and action selection	3. The computer evaluates the feedback, evaluates to what extent the goal is achieved The computer either: - decides that it did not achieve its goal yet, and tries to achieve its goal again - decides to change its strategy and hold on to its goal, - adjusts or renew its goal and find a new strategy.
Action implementation	4. The computer acts or stays still, in accordance to its goals

reasoning. A toy-robot might have "entertaining the individual" as a goal. This goal displays positive intentions towards the individual. A robot that is designed to build cars has as goal "to assist humans with heavy tasks within the process of building cars". This goal can be read as a positive goal towards the individual. However, the goal can also be explained as a bad goal towards the environment: a robot which acts "good" should not help build a machine that poisons the environment with air pollution. Determining the goal itself can be seen as a moral decision. Within the interaction relation, the goal of the system is partly determined by the system itself. The household robot is of course dedicated to keeping the house clean, but within that scope, it decides its tasks: cleaning the dishes, vacuuming the house, organizing loose objects in the living room.

(b) Within the scope of a specific goal, the computer can use a specific strategy to achieve this goal. This issue is the same as issue (b) for the interference relationship, but the difference is that the goal of the computer is decided by a third party for the interference relationship, while the goals within the interaction relationship can be adjusted by the system itself (within the scope of its general assignment). This means that the moral responsibility of the computer within the interference relationship for accepting a strategy has a different weight than the moral responsibility within the interaction relationship: within the interaction relationship the computer has control over the goal, which means that the choice for a strategy is always a positive choice and not a negative choice, while the choice for a strategy within a fixed goal might be a negative choice (the choice is the best option of the available options).

The interaction relationship can be seen as an equal relationship. The communication between the individual and the computer is both ways. The influence of the third party is absent. This means that the goals and strategies of the computer influence the impression the individual or the environment have of the computer. This, in turn, implies that not only the morality of current goals and strategies play a role, but also the effect that the current choices have on future decisions and the effects it might have on the options to choose strategies and goals in the future.

Advice

The moral aspects of the behavior of the computer for the advice process can be found in the decision and action selection phase in Table 4.

Within the advice relationship, the impact the computer has on the individual completely depends on the question the third party asks the computer. Within this relation, there are three moral aspects: two have to do with the fact that the gathering of information (a) and the action that results from the advice (b) are not performed by the system that gives the advice. The third moral aspect is about the advice itself (c).

An example of an advice system is a medical expert system. This system receives information from a doctor. It first needs to judge the reliability of this information before it can use it. The doctor might unknowingly have provided incorrect or incomplete information, which might influence the diagnosis. The system then provides the doctor with an advice about a medical strategy. By providing this advice, the system should consider how the doctor's behavior is influenced by this advice. The doctor might get the advice to do something that he is not sufficiently trained in, or not skilled enough for. In that case, the system should maybe not give this advice. The last moral aspect might be in the decision itself, for example when the doctor asks for advice about a situation involving controversial treatments like life extending treatment for terminally ill patients, abortion or euthanasia.

(a) The computer receives information from the third party which it has to base its decision on. This information might be unreliable, so balancing the possible impact of the advice on the individual with the perceived reliability of the information is a moral aspect within this relation, e.g. in a diagnosis.

(b) The computer gives advice about actions towards an individual to a third party. This means that the computer has no control over how the action is understood or how the action is executed. If the action is executed and the expected result fails to materialize, the computer cannot intervene and correct its advised action. In the medical example,

Table 4 Stages of information processing in the advice process

Information acquisition	1. The computer receives a case description about an individual from a third party
Information analysis	2. The computer evaluates and interprets the gathered information
Decision and action selection	3. The computer: - decides on the reliability of the information - decides on the expected execution of the advice - decides on the advice
Action implementation	4. The computer advises the third party about the case

the system cannot observe or interfere with the treatment and its effects.

(c) The computer gives advice, and depending on the kind of advice, the advice might also include moral aspects. Here, the computer can assume that the third party might adopt and execute its advice without any consideration. It is like the computer acts itself; the moral responsibility for the outcome of the advice lies therefore with the computer. This would be the case in advice about life extending treatment.

Dimensions in moral decisions

In the previous section, the for each type of relationship between human and computer in the context of this research, the decisions with moral implications are listed. Within these decisions, three distinct dimensions can be identified: scope, impact and involvement. In these three dimensions the action that is being carried out, the actors that carry out the action, and the situation in which the action is being carried out are being considered. These dimensions can be used to distinguish between different types of moral mediation.

The dimension "scope", is about how actions fit within the responsibilities of the system. A distinction can be made between decisions that are inside or outside of the scope of the assignment of the system. This decision is about the borders of the responsibility that is assigned to someone or something in an implicit or explicit way. For example, a smart house has the task of making the life of its inhabitant more comfortable by keeping track of what its inhabitants do during the day. While observing its inhabitants, it may also observe behaviors of others than the inhabitants. In some situations, it might be helpful (or harmful) for the inhabitants when the house also takes action based on behaviors of others than these inhabitants, for example when visitors exhibit harmful behavior towards its inhabitants. However, this is outside the scope of the assignment of the system, and these actions might impact the life of both the inhabitants and the visiors, so there clearly is a scoping dilemma here.

The dimension "impact", is about the situation in which the decision has impact. A distinction can be made between decisions that make a direct impact or that have a more indirect impact on the environment of the system. When making a decision with direct impact, the situation is known or can be known: all the factors that influence the immediate outcome of the decision are already present. An example is a decision about dropping a bomb from a drone. The damage is immediate: people die or get hurt. The information about this decision is already present: the drone has a location, the weight of the bomb, the speed of the drone and the wind velocity is known, the location where it will land can be calculated and the people that will be on that location of the drop are already there or very close. When the decision is about whether to store or delete information, the decision might have impact in the future, depending on how the rules about using this information are at that point, depending on which party uses the information and a number of other factors. This information is not known yet when the decision is made to store the information.

The dimension "involvement" is about the actors that are involved in the decision: the decision can be an independent decision in which the system is the only one who's judging, deciding and implementing the decision into action. The decision might also be a dependent action in which multiple actions are involved. This might be in delivering the input to the system, or in carrying out the action that results from the decision. A medical advice system, for example, uses information that is given to the system by doctors. The doctor has already made a prejudgement to give this specific information, and possibly to hold back other information that is being perceived as being redundant, irrelevant or otherwise not of any value to the system. The system gives advice based on this information, and then the doctor will (or won't) carry out the intended action. The system might give an advice that results in a medical procedure being performed by the doctor in an incorrect way because of lacking skills, or for any other reason. Even though the system gives the right advice based on the assumption that the information that it used is correct and that the advice is being carried out correctly, the outcome of the advice might still be a disaster.

Using the dimensions above, decisions identified in the previous section can be classified. For example, adjusting a strategy, which is a decision which the system has to make in the interference and the interaction relationship, can usually be classified as a decision inside the scope of the assignment, the impact is direct, and it is an independent decision. The decision to store information is a decision inside its scope, also an independent decision, but the impact is indirect. The decision about changing goals or choosing an empty strategy for the interference and interaction relationships are decisions that are direct, independent and outside the scope. Forwarding information to a third party has indirect impact, is a dependent decision and is inside the scope.

Conclusion

In phenomenology of technology, analyses of the relation human-technology-world have mostly been directed to situations where the technology forms a more or less constant "lens" through which experience and action can be changed. In that sense, the technology contributes to experience of and action by humans, and as such may affect moral patients.

Several approaches have proposed extensions to this framework, as discussed in the Related Work section, including cyborg intentionality and composite intentionality, reconfigurable technology, and service technology. These approaches indicate that in emerging technologies clear boundaries between humans and technology may no longer be observable or relevant, and that these technologies allow for a dynamically changing type of mediation, rather than a fixed one. However, none of these approaches pay attention to technologies that make explicit decisions, and are thereby able to *reconfigure themselves*. Their moral impact is therefore different than in the traditional approaches to mediation, and new types of mediation relations need to be considered.

Within this research, we have investigated moral aspects of computer-made decisions, in order to identify ways in which new technologies change moral relations. We considered computer systems that are autonomous, context-aware and intelligent. These systems can have four types of relationships with a human and, in some cases, a third party. These relationships are the observation relation, the interference relation, the interaction relation and the advice relation. Each of these relations can be modeled as a loop in which the same stages are always repeated. Within the decision and action selection stage, each system has a number of decisions it can make. When all these decisions are combined, it can be concluded that for these kind of systems, there are three dimensions of the decisions of the system in these four relations that require moral insights or moral behavior:

- Scope: Operation inside/outside the scope of its assignment
- Impact: Considering the short-term/long-term consequences
- Involvement: Dependency of the result on self only/on others

Based on these distinctions, it becomes possible to derive guidelines for the design of machines that are involved in these aspects of decision-making. Specific guidelines for the three moral aspects of decision-making can be provided, for example providing requirements for the considerations that have to be taken into account in the decision-making processes. Using the classification for the types of relations, a distinction in advisability and desirability can be made in different relations and for the different moral aspects of the decision-making. For example, in terms of involvement of others, it can be required that the machines take into account the possible actions taken by others based on the decision made by the machine. Specific details for suitable reasoning approaches may be provided as well. Such guidelines can form the basis for future regulatory measures on different levels of abstraction, as well as accompanying design approaches and certification standards. In addition, the changes induced in moral relations can form the basis for grounded design choices on which decisions should be delegated to machines and which to humans.

As mentioned, several approaches have discussed the particular characteristics of emerging technologies that would require extensions to ethics of technology: reconfigurability, service-orientation, and, in this paper, explicit decision-making. In future work, these different approaches could be combined in an overarching analysis of responsibilities in such systems, and associated requirements for regulation and design. To this end, existing approaches such as value-sensitive design (Friedman et al. 2006; Van den Hoven 2007) can be used, but with specific attention to aspects of reconfigurability, inseparability of production and consumption, and—following the analysis in this paper—explicit decision making by the computers themselves. We assume here that assigning responsibilities and proposing regulation is still a task to be performed by humans, but it might be justified to question even that assumption.

Acknowledgments We would like to thank N. Moseley for his help with the graphics.

References

Akrich, M. (1992). The de-scription of technical objects. Shaping technology/building society pp. 205–224.

Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics? *Intelligent Systems, IEEE*, *21*(4), 12–17. doi:10.1109/MIS.2006.83.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, *12*(3), 209–221.

Coeckelbergh, M. (2011). Talking to robots: On the linguistic construction of personal human-robot relations. In M. Lamers & F. Verbeek (Eds.), *Human-Robot Personal Relationships* (Vol. 59, pp. 126–129)., Lecture notes of the institute for computer sciences Springer, Berlin, Heidelberg: Social Informatics and Telecommunications Engineering. doi:10.1007/978-3-642-19385-9_16.

Dechesne, F., Warnier, M. & Hoven, J. (2013). Ethical requirements for reconfigurable sensor technology: A challenge for value sensitive design. Ethics and Information Technology pp. 1–9, doi:10.1007/s10676-013-9326-1.

Floridi, L., & Sanders, J. (2004). On the morality of artificial agents. *Minds and Machines*, *14*(3), 349–379. doi:10.1023/B:MIND.0000035461.63578.9d.

Fong, T.W., Nourbakhsh, I. & Dautenhahn. K. (2003). A survey of socially interactive robots. Robotics and Autonomous Systems.

Friedman, B., Kahn, P. H, Jr, & Borning, A. (2006). Value sensitive design and information systems. *Human-Computer Interaction In Management Information Systems: Foundations*, *5*, 348–372.

Gaver, W.W. (1991). Technology affordances. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp. 79–84.

Hildebrandt, M., Koops, B. J., & Jaquet-Chiffelle, D. O. (2010). Bridging the accountability gap: Rights for new entities in the information society? *Minnesota Journal of Law, Science & Technology*, *11*(2), 497–561.

Ihde, D. (1990). *Technology and the lifeworld*. Bloomington, IN: Indiana University Press.

Johnson, D. G., & Powers, T. M. (2008). Computers as surrogate agents. In J. Van den Hoven & J. Weckert (Eds.), *Information technology and moral philosophy* (pp. 251–269). Cambridge: Cambridge University Press.

Kuflik, A. (1999). Computers in control: Rational transfer of authority or irresponsible abdication of autonomy? *Ethics and Information Technology*, *1*(3), 173–184. doi:10.1023/A:1010087500508.

Magnani, L., & Bardone, E. (2008). Distributed morality: Externalizing ethical knowledge in technological artifacts. *Foundations of Science*, *13*(1), 99–108. doi:10.1007/s10699-007-9116-5.

McLaren, B. M. (2006). Computational models of ethical reasoning: Challenges, initial steps, and future directions. *Intelligent Systems, IEEE*, *21*(4), 29–37.

Noorman, M. (2012). Computing and moral responsibility. In: Zalta, E.N. (ed). The stanford encyclopedia of philosophy, fall 2012 edn.

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *30*(3), 286–297. doi:10.1109/3468.844354.

Pieters, W. (2013). On thinging things and serving services: Technological mediation and inseparable goods. *Ethics and Information Technology*, *15*(3), 195–208. doi:10.1007/s10676-013-9317-2.

Poslad, P. S. (2009). *Ubiquitous computing: Smart devices*. Wiley, Chichester: Environments and Interactions.

Ramos, C., Augusto, J. C., & Shapiro, D. (2008). Ambient intelligence—The next step for artificial intelligence. *IEEE Intelligent Systems*, *23*(2), 15–18. doi:10.1109/MIS.2008.19.

Van den Hoven, M. J. (1998). Moral responsibility, public office and information technology. In I. T. M. Snellen & W. B. H. J. Van de Donk (Eds.), *Public administration in an information age: a handbook, IOS* (pp. 97–112). DC: Amsterdam; Washington.

Van den Hoven, J. (2007). Ict and value sensitive design. *The Information Society: Innovation* (pp. 67–72). Ethics and Democracy In honor of Professor Jacques Berleur sj, Springer: Legitimacy.

Van den Hoven, J., Lokhorst, G. J., & Van de Poel, I. (2012). Engineering and the problem of moral overload. *Science and Engineering Ethics*, *18*(1), 143–155.

Verbeek, P. P. (2006). Materializing morality : Design ethics and technological mediation. *Science, Technology & Human Values*, *31*(3), 361–380.

Verbeek, P. P. (2008b). Cyborg intentionality: Rethinking the phenomenology of human-technology relations. *Phenomenology and the Cognitive Sciences*, *7*(3), 387–395. doi:10.1007/s11097-008-9099-x.

Verbeek, P. (2008a). De grens van de mens. over de relatie tussen mens en techniek. In I. L. Consoli & R. Hoekstra (Eds.), *Annalen van het Thijmgenootschap*. Nijmegen: Valkhof Pers.

Wallach, W., Allen, C. (2009). Moral machines: Teaching robots right from wrong. Oxford University Press, http://books.google.com/books?id=tMENFHG4CXcC.