From Moral Agents to Moral Factors: The Structural Ethics Approach

Philip Brey

This is a preprint version of the following article:
Brey, P. A. E. (2014). From moral agents to moral factors: the structural ethics approach. In P. Kroes, & P. P. C. C. Verbeek (Eds.), *The moral status of technical artifacts* (pp. 124-142). (Philosophy of engineering and technology; Vol. 17, No. 17). Dordrecht: Springer Verlag. DOI: 10.1007/978-94-007-7914-3_8

Abstract It has become a popular position in the philosophy of technology to claim that some or all technological artifacts can qualify as moral agents. This position has been developed to account for the moral role of technological artifacts in society and to help clarify the moral responsibility of engineers in design. In this paper, I will evaluate various positions in favor of the view that technological artifacts are or can be moral agents. I will find that these positions, while expressing important insights about the moral role of technological artifacts, are ultimately lacking because they obscure important differences between human moral agents and technological artifacts. I then develop an alternative view, which does not ascribe moral agency to artifacts, but does attribute to them important moral roles. I call this approach structural ethics. Structural ethics is complementary to individual ethics, which is the ethical study of individual human agents and their behaviors. Structural ethics focuses on ethical aspects of social and material networks and arrangements, and their components, which include humans, animals, artifacts, natural objects, and complex structures composed of such entities like organizations. In structural ethics, components of networks that have moral implications are called moral factors. Artifact ethics is the study of individual artifacts within structural ethics. It studies how technological artifacts may have a role as moral factors in various kinds of social and material arrangements as well as across arrangements. I argue that structural ethics and artifact ethics provide a sound alternative to approaches that attribute moral agency to artifacts. I end by arguing that some advanced future technological systems, such as robots, may have capacities for moral deliberation which may make them resemble human moral agents, but that even such systems will likely lack important features of human agents which they need to qualify as full-blown human agents.

Introduction

Recently, a number of authors in the philosophy of technology have argued that some or all technological artifacts can qualify as moral agents. The notion of a moral agent has traditionally been reserved for human beings, and is used to refer to beings which can be held morally responsible for their actions. Such beings have the capacity to know right from wrong and are able to choose their actions freely based upon their considered moral judgments. Yet, some authors have argued, extending the notion of moral agency to technological artifacts is necessary in order to account for the moral role of (some) artifacts, which is in some cases highly similar to that of human agents. In addition, they have argued, doing so will be useful for the attribution of moral responsibility to designers.

In this paper I will evaluate various positions in favor of the view that technological artifacts can be moral agents. I will find that these positions bear important insights about the moral role of technological artifacts, but are ultimately lacking. I then develop an alternative view, which does not ascribe moral agency to artifacts, but does attribute to them important moral roles. I call this approach structural ethics. I will argue that this approach has all the benefits of approaches that ascribe moral agency to artifacts, while maintaining a distinction between the moral agency of humans and the moral roles of nonhuman entities like technological artifacts.

1. The philosophical concept of moral agency

To begin my inquiry, I will give an account of the classical notion of a moral agent as it has been developed in philosophy. In the next section, this account will then be contrasted with extended notions of moral agent that have been developed in the philosophy of technology. The standard notion of a moral agent is a philosophical notion that refers to beings which are capable of acting morally and are expected by others to do so. Although there is no generally agreed definition of "moral agent," existing definitions tend to emphasize three features.[1] Moral agents are beings that are (1) capable of reasoning, judging and acting with reference to right and wrong; (2) expected to adhere to standards of morality for their actions; and (3) morally responsible for their actions and accountable for their consequences. These three features together define what I will call the standard philosophical conception of a moral agent, or, in brief, the *standard conception*.

Which beings qualify as moral agents on the standard conception? Given the three mentioned features, it appears clear that only adult, rational human beings do. Only rational human beings are capable of moral reasoning, and only they are expected to behave morally and are held morally accountable. Adults that are incapable of distinguishing right from wrong are not normally seen as moral agents, and are not held to be morally responsible for their actions. Similarly, young children are not held to be moral agents, nor are animals or inanimate objects. In contrast, we expect "normal" adults to have a developed capacity for moral reasoning, judgment and action, we expect them to exercise that capacity, and we hold them accountable if they nevertheless engage in immoral acts. In short, we hold them to be moral agents.

A moral agent is a special kind of *agent*. An agent, in the philosophical sense, is a being capable of performing *actions*. Actions constitute a special class of behaviors, since not any kind of behavior constitutes an action (Davidson, 1980). Breathing, for instance, is not an action, even if it is something we do. Actions are intentional, they depend on capacities for rational thought and self-interested judgments, and the performance of goal-directed behaviors based on such thoughts and judgments. Typically, actions are explained by reference to *intentional states* of the agent. Intentional states are mental states like beliefs, desires, fears, perceptions and intentions, that have a directedness to something else. For example, an explanation of why John drank the milk (an action) would refer to intentional states of John that explain his behavior, e.g., John's fear that he was dehydrating and his belief that the milk would quench his thirst. In contrast, a mere behavior (e.g., John's blinking, or his falling when pushed over) would refer to mere physical causes like sand getting into John's eye or someone shoving John. Agents, in conclusion, are beings capable of performing actions, which are behaviors caused by intentional states of goal-directed beings.

An agent is a *moral* agent when the intentional states that it cultivates and the subsequent actions it performs are guided by moral considerations. This requires, first of all, a capacity for *moral deliberation*, which is reasoning in order to determine what the right thing to do is in a given situation. A capacity for moral deliberation requires a capacity for reasoning and knowledge of right and wrong. Moral deliberation typically results in *moral judgments*, which are judgments about right and wrong. It also frequently results in intentions to perform certain actions that are held to be morally good, and to refrain from performing actions that are held to be immoral. For example, a moral agent may deliberate on what to do with a found wallet, in a way that takes into account both moral and nonmoral considerations. He may then arrive at the moral judgment that turning the wallet in to the police is the right thing to do. This may then result in an intention to give the wallet to the police, and a subsequent action of giving the wallet to the police.

Let us now turn to the second feature of moral agents, which is that they are beings that are expected to behave morally. This is a *normative* rather than a *factual* expectation. That is, we believe that people *should* behave morally, that they have a *moral obligation* to do so. We do not expect that they in fact always do. In fact, we know that they often do not. However, our belief in morality, and our knowledge that others are capable of moral actions, results in an expectation that others behave moral-

[1] See Himma (2009) for some definitions of "moral agent".

ly, or at least make every effort to do so. We do not find it acceptable that people either do not engage in moral deliberation in situations that pose moral dilemmas, or do so but nevertheless act immorally.

The third feature of moral agents, their being held morally responsible and accountable, is a corollary of the first and second feature. Because people are capable of acting morally, and because we expect them to do so, we hold them to be *morally responsible* for their actions. That is, we hold that their actions are appropriately the subject of moral evaluation by others, and of particular kinds of reactions based on such moral evaluations. Such reactions particularly include praise and blame, as well as related responses such as reward and punishment. Thus, if someone acts morally, we have a propensity to praise them for doing so, whereas if someone acts immorally, we may blame or condemn them for their actions. Moral responsibility is usually held to presuppose *free will*: persons have moral responsibility to the extent that they can freely choose their acts. *Moral accountability* is a type of moral responsibility that goes beyond it in assuming the existence of shared moral standards in a community that can be alluded to in evaluating someone's actions (Watson, 1996). When there are such shared standards, moral agents can be praised or blamed with explicit reference to such interpersonal standards. They can be judged to either have upheld or have broken these standards, and can be held accountable for doing so.

2. Theories of artifacts as moral agents

Given the prevailing conception of a moral agent in philosophy, it would seem unlikely that anything else but a human being could qualify as a moral agent. It seems particularly unlikely that technological artifacts like lawnmowers and iPods could qualify as moral agents. Technological artifacts are not capable of moral deliberation, they are not expected to behave morally, and they are not held to be morally responsible or accountable. They therefore seem very poor candidates for moral agents. Recently, however, several philosophers have defended the view that the notion of a moral agent should be extended to include technological artifacts.

There are two versions of this view, which I will now lay out. On the first view, which I will call the *moral artifacts view*, all technological artifacts are, or could function as, moral agents. This view was first proposed, although not in very explicit terms, by Bruno Latour (1992). It has subsequently been defended by Powers and Johnson (2004), Keulartz et al. (2004) and Verbeek (2008). On the second view, which I will call the *morally intelligent agents* view, certain highly evolved technological artifacts, namely those capable of autonomous behavior and intelligent information processing, qualify as moral agents. On this view, the class of moral agents includes, next to human beings, things like autonomous robots and software agents. This view was first proposed by Colin Allen, Gary Varner and Jason Zinser (2000), and has subsequently been developed in an influential article by Luciano Floridi and Jeff Sanders (2004). It has also been defended by a number of other authors, including Stahl (2004), Sullins (2006) and Johnson and Powers (2006).

The moral artifacts view asks for a major revision of our concept of moral agency, extending it to mundane artifacts like knives, automobiles and bridges. The second approach asks for a more limited revision of the standard conception. In this paper, my focus will be on the more radical claim, which is the moral artifacts view. Near the end of the paper, I will also briefly go into the moral intelligent agents view.

The moral artifacts view rests on the observation that the roles that technological artifacts play in human affairs are frequently not morally neutral. There seem to be two ways in which technological artifacts can have a moral impact. First, artifacts are capable of *steering moral behavior in humans.* Artifacts may stimulate or enforce morality by making humans behave morally. For example, a car that flashes a warning when driver or passengers do not wear a seat belt stimulates moral behavior by its users. Second, artifacts are capable of *influencing moral outcomes*. Even when artifacts do not influence moral behavior in humans, they may have consequences that can be morally evaluated. For example, a computer network that randomly provides added bandwidth to some of its users is less just than a computer network that gives all its users equal bandwidth. The network that provides equal

bandwidth hence generates a better moral result, even though it does not influence any person to behave morally. Both of these moral roles of artifacts have been used to argue that artifacts are, or can be, a type of moral agent.

Bruno Latour (1992) arrives at the view that artifacts are moral agents by asking the question whether morality is only located in people or also in things. He argues that moral laws in a society are not only enforced by humans but also by artifacts. Artifacts make up the "missing masses" that together with humans make up the moral fabric of society. Artifacts enforce moral rules in a way that is similar to that of humans, Latour argues. For instance, a moral (and legal) rule that tells us to drive slowly in a densely populated neighborhood can be enforced either by a police officer who waves cars down, a street sign that tells drivers to slow down, or a speed bump that forces them to slow down. Morality is hence similarly enforced by both humans and artifacts.

Latour argues that both humans and artifacts are bearers of *programs of action* that aim to enforce particular moral or social rules or configurations. A police officer and a speed bump may, for example, both aim to enforce the rule "IF a car drives in this street, THEN its speed is no more than 30 m.p.h." Bearers of such programs of action are called "agents" or "actants" by Latour. Artifacts, on his view, are *moral* agents when they are bearers of a program of action that enforces a moral rule. Typically, such programs of action are inscribed into the design of an artifact. Latour claims that many mundane artifacts enforce moral rules by facilitating, stimulating, or forcing behaviors and situations that comport with everyday morality.

Powers and Johnson (2004) present an alternative account that revolves around the notion of intentionality. They define a view of agency according to which causality and intentionality, but not mentality, are necessary features of it. Artifacts, they observe, are causally efficacious, meaning that their presence and operation has consequences for what happens in the world. In addition, they argue, artifacts are bearers of intentional states. This claim is based on the observation that artifacts have a directedness at phenomena external to them. For instance, a telephone is directed at human fingers and ears and at human verbal communication. These different types of directedness of telephones constitute different intentional states, according to Powers and Johnson. The intentionality of artifacts is bound up with their function, which defines their intended use, and which is a result of the intentions of designers.

According to Powers and Johnson, the intentional states of artifacts allow for reason explanations instead of mere causal explanations or events. For example, if a speed bump slows down a car, its slowing down can be explained by reference to the directedness of speed bumps at cars and their function of slowing cars down, both of which were intended by a designer. A speed bump is hence different from a pothole, which merely happens to cause a car to slow down. On their view, the speed bump is therefore an agent, whereas the pothole is not. A speed bump is moreover a moral agent because it enforces moral rules and has consequences for moral patients. Because artifacts have intentional states and have moral consequences, it is also possible to make attributions of moral responsibility, Powers and Johnson argue. Moral responsibility for the agency of an artifact lies with the human agents who put its intentional states into it, including, most prominently, the designers.

While both Latour and Powers and Johnson conceive artifacts as agents, they also emphasize that artifacts cannot perform any actions independently of users and designers, that is, of human agents. Powers and Johnson emphasize that "the behavior that results [from artifacts] is a combination of the intentionality of the artifact designer, input from the user, and behavior of the artifact" (2004: 22-23). And Latour emphasizes, similarly, that artifacts do not generate moral outcomes by themselves. For instance, it is not just the flashing light in the car that causes one to wear a seat belt: "I, plus the car, plus the dozens of patented engineers, plus the police are making me be moral" (1992: 226).

If artifacts are always dependent on human agents like users and designers for their agency, does this not demonstrate an asymmetry between human and nonhuman moral agency? Isn't it the case that humans are able to perform moral acts autonomously, whereas artifacts are always dependent on human agents? It would seem that Powers and Johnson are willing to accept this asymmetry. Latour and his followers, however, do not. On Latour's view, human and nonhuman agents are both dependent on constellations or networks of agents for their actions. These networks, which Latour calls actor-networks, consist of both human and nonhuman agents (Latour, 1987). Human agency is, on Latour's

view, always the product of multiple agents co-acting with a human agent. Just like the car does not act alone in causing me to wear a seat belt, I do not act alone in wearing the seat belt. My action is caused not just by me but also by the blinking light in my car, the designers behind it, and the police that checks on seat belt use.

Latour hence does not only extend the notions of agency and moral agency to artifacts, he also engages in a major revision of the concept of human agency. Human agency is, on his view, not attributable to agents, but is rather a property of actor networks, in which multiple actors together produce a particular action. Attributing an action to a particular actor (human or nonhuman) is merely a matter of putting the focus on that actor, while we might have also chosen to emphasize the role of other actors in the network. Morality, in this view, is similarly a property of networks consisting of human and nonhuman entities that together generate moral actions and moral outcomes. This position has been further defended by Keulartz et al. (2004) and by Verbeek (2005, 2008), who however rejects the ontological symmetry between people and things proposed by Latour.

We hence have seen several arguments for extending the notion of moral agent to include technological artifacts. The authors who extend the notion in this way have several motives for doing so. They want to give greater visibility to the moral role of artifacts, to better account for the way morality is realized in society, and to allow for better, more ethical design and use of technological artifacts. In the next section, I will evaluate these arguments and discuss whether they provide sufficient reason to broaden the notion of moral agent to include technological artifacts.

3. Evaluating the moral artifacts view

Proponents of the moral artifacts view present novel conceptions of moral agency that are intended to replace rather than supplement the existing philosophical concept of moral agency. Most authors do not hold that their view is necessarily ontologically more correct, but rather emphasize its pragmatic usefulness in understanding the moral role of artifacts and their relation to humans. Thus, Powers and Johnston say that they have "practical reasons for calling technological artifacts agents" and use this terminology to "highlight that the ways in which artifacts are designed and used have powerful consequences for the moral character of the world we inhabit" (2004, p. 26). Similarly, Keulartz et al. say that they think that "it is useful to speak of artifacts as (possible) moral agents. Not for ontological reasons, but for pragmatist ones" and say that this conceptualization highlights important aspects of the relations between humans, technological artifacts and ethics.

I agree with these authors that concepts should primarily be evaluated on pragmatic rather than ontological grounds. As Wittgenstein, Peirce and others have shown, we do not usually use concepts to describe objective essences, but rather to selectively highlight aspects of things that are important to us in dealing with them. Consequently, a concept is a good (i.e., useful) concept if it highlights important aspects of a thing or state-of-affairs while not obscuring important other ones. So the question for the moral artifacts view is whether it (a) highlights important phenomena that were previously overlooked, and (b) does not obscure other important phenomena.

The main benefit of the moral artifacts view is that it highlights the facts that technological artifacts play an important role in shaping moral actions and outcomes and that they are part of the moral fabric of society. Technological artifacts have been largely overlooked in moral theory, and have only been assigned an instrumental role in human action, as means that make certain actions possible, or make them easier to perform. Because of this instrumental conception, artifacts are normally thought of as morally neutral. All morality is thought to be located in the choosing and acting human subject. Yet, as Hans Jonas has argued, technological artifacts drastically change human action, and this has important consequences for ethics (Jonas, 1984). The moral artifacts view helps us arrive at a better view of technological artifacts that reveals their important role in shaping moral action and moral outcomes.

Another benefit of the moral artifacts view is that it highlights useful similarities between human agents and artifacts regarding their moral role. As Latour has shown, both human agents and artifacts can be used to enforce the same moral norms, both can influence humans to behave morally, and both

can determine moral outcomes. Both, in addition, are dependent on other entities, both human and nonhuman, for being able to play these roles. These important similarities between them have been less obvious in traditional accounts. In addition, as Powers and Johnson have shown, both human agents and artifacts exhibit intentionality. The intentionality in artifacts can be referred to in making reason explanations, or intentional explanations, of moral outcomes, just as it can in humans. It can even be used to make attributions of moral responsibility, just as it can in humans, by linking the consequences of artifacts to the designers who inscribed their intentionality into these artifacts.

These are great benefits of the moral artifacts view. However, I will argue, this view also obscures and obliterates important phenomena. First, it obscures differences between human agency and the agency of artifacts and the unique characteristics of human agency and human action. Actions, unlike mere events, result from the intentional states of goal-directed, interest-bound beings who intend to cause changes in the world, and these intentional states provide reasons for the action in question. Artifacts are not normally goal-directed, they do not have interests, and they do not have intentional states like beliefs and desires that cause their behavior. Replacing the standard conception of agency with an extended one means that these important differences are obscured.

There are at least two reasons why obscuring these differences is a bad idea. First, the classical notion of an agent has an important role in our moral image of a human being. Part of what makes human beings special and valuable is their ability to form intentional states like beliefs and desires, and then choose to act according to them. In this, they differ from things like rocks and coconuts, which can only passively cause things to happen, without reason or intent. As soon as things like screwdrivers are also called agents, these special features of human agency are lost in our understanding of agency, and the moral image of humans is damaged as a result.

Second, the classical notion of agency has an important role in explaining and accounting for events. Actions, and any events following from them, are explained by reference to reasons and intentions, unlike most other events, which are explained by reference to mere causes. Reason explanations provide us with different information than causal explanations. They give us insights into the motives and justifications of human agents. Our responses to actions tend to be different from those to mere causal events: we tend to the beliefs, desires, and other mental states that underlie these actions, and we do not only respond physically, but also morally and socially. However, an extended notion of agency obliterates this distinction between actions and mere events, and hence the special role of actions in our understanding of the world.

Against this point, Powers and Johnson may object that on their account of agency, all actions rest on intentional states and can be explained intentionally. So their account at least seems to preserve the difference between actions and mere events, and hence between intentional and causal explanation. I believe their account is flawed, however, by attributing intentional states to artifacts. Artifacts certainly have intentional *properties*. For example, a speed bump has a directedness to automobiles that it has been designed to slow down. This is an instance of what John Searle (1984) has called *derived intentionality*: a directedness that derives from human intentions that have been inscribed into artifacts. But artifacts do not have intentional *states*. That is, they do not have, as humans do, states like beliefs, desires, and intentions that provide reasons for their actions. It is false to say: “The car slowed down because the speed bump intended it to slow down”.

However, one can correctly say either “The car slowed down because it is the *function* of speed bumps to slow down cars” or “The car slowed down because speed bumps *are intended* to slow down cars”. The former explanation is a functional explanation rather than an intentional explanation, and does not require any attribution of intentional states to artifacts. The latter is an intentional explanation, but it is left implicit who is doing the intending. Surely, however, it is not the speed bump which is doing the intending. Rather, it is the designers and traffic controllers who intend the cars to slow down *by means of* a speed bump. So a full intentional explanation would read: “The car slowed down because speed bumps are intended by designers and traffic controllers to slow down cars”. But this account also does not require any attribution of intentional states to artifacts. Rather, it seems accurate to say that the intentional states belong to the designers and traffic controllers, and it is their actions (the development and installment of speed bumps) that cause cars to slow down.

Next to obliterating the distinction between agents and mere inanimate objects, the moral artifacts view also obscures the difference between moral agents and entities that have mere moral properties or implications. Most importantly, what is lost in the equivocation is the idea of a moral agent as an agent capable of moral deliberation and of actions based on such deliberation, and the idea of a moral agent as a morally responsible and accountable being.

The capacity for moral deliberation in human moral agents is important because it enables a very different mode of interaction than is possible with entities that lack this capacity. Things that lack this capacity but do play a moral role, like speed bumps, can only be interacted with physically. Speed bumps can be physically created, redesigned or removed. We can have a similar physical mode of interaction with human beings. However, because humans engage in moral deliberation, we can also enter into verbal modes with them: we can deliberate with them, bring forward arguments or ideas, try to convince them, threaten them or influence them otherwise. In attempts to influence moral behavior and moral outcomes, a physical mode of interaction is often the last one we choose with human beings. This is because they are sentient beings capable of moral deliberation. This important capacity is obscured, however, when it is no longer held to be a defining property of moral agents.

Removing the notion of moral responsibility from our conception of moral agency is also unappealing. The concept of moral responsibility is important to us because we believe that people should accept that their actions are subjected to moral standards, that they should be able to defend the moral rightness of their actions to others, and that others can appropriately respond to their actions with their own attitudes, judgments and actions that include praise, blame, punishment and reward. Philosophers have put forward two different reasons why such praise and blame (and punishment and reward) should be issued (Eshleman, 2009). The first, expressed by the merit-based view of moral responsibility, is that praise and blame should be issued to moral agents because they deserve such responses from others. Those who act immorally deserve blame and punishment, and those who act morally deserve praise and reward. The second, encoded in the consequentialist view, is that praise and blame should be issued to moral agents in order to encourage future moral behavior and to prevent immoral behavior.

If we change from the standard view of moral agency to a broad view that includes moral artifacts, then notions like intentional action, moral deliberation and moral responsibility are no longer defining features of our notion of moral agency. This, I have tried to argue, is a significant loss. It may be argued that we could still retain these notions and attach them whenever the moral agents we refer to are human. This, however, is an insufficient response. Notions like those of agent and moral agent are fundamental concepts philosophers (and non-philosophers) use to understand and reason about reality. If these notions are restructured so as to lose important features, then these features are obscured in our understanding of reality. They are no longer activated whenever these concepts are activated in our minds, and as a result become less central in our thinking.

For the reasons given above, the gains brought by the moral artifacts view to include them are hence offset by considerable losses that result from important features of the standard conception of moral agency being obscured. As a result, this does not make the moral artifacts view particularly appealing. At the same time, the standard conception of moral agency also has its disadvantages, because has tended to be accompanied by an instrumentalist understanding of technological artifacts that downplays their moral importance and does not reveal the similar roles that human agents and artifacts often play in giving shape to morality. The inadequacy of these two options raises the question whether a third view is possible, one that incorporates the benefits of the standard conception of moral agency as well as those of the moral artifacts view, and does so without having significant drawbacks. It is to such a view that I will now turn.

4. An alternative account

What we have seen is that on the one hand, there are good reasons to retain the traditional notion of moral agent in its basic form but also to upgrade the role of both technological artifacts and networks

in ethics. My proposal is to introduce a new type of ethics, *structural ethics*, next to the familiar *individual ethics* that focuses on human (moral) agents. Structural ethics focuses on the moral aspects of social and material arrangements (structures or networks consisting of humans and nonhumans), including their impact on the actions of human agents. Structural ethics is intended to be complementary to individual ethics. Individual ethics is solely focused on the morality of individual human agents, their actions, and the intentional states and deliberations underlying them. As I will argue, structural ethics requires a new ethical vocabulary that is different from that of the moral artifacts view, which draws too much from the vocabulary of individual ethics.

Structural ethics studies social and material arrangements as well as components of such arrangements, such as artifacts and human agents. It has three aims: (1) to analyze the production of moral outcomes or consequences in existing arrangements and the role of different elements in this process; (2) to evaluate the moral goodness or appropriateness of existing arrangements and elements in them, and (3) to normatively prescribe morally desirable arrangements or restructurings of existing arrangements. In doing so, it also aims to identify, evaluate and prescribe roles of individual elements in these arrangements. Unlike individual ethics, structural ethics hence looks at larger structures and networks with the aim of engaging in social and technological engineering.

Let us consider an example of each of these three types of investigations. The first type can be illustrated with Latour's earlier example of the seat belt. The moral behavior of me wearing a seat belt can be analyzed as the result of not only my actions, but also the (inter)actions of other elements, including enforcement by the police and the behavior of my car, which in turn is the result of actions of engineers as well as safety advocates and policy makers. An analysis of this network of entities that influence my behavior can show how my moral behavior is shaped by them, and it can assign a role to this effect to each of them.

The second type, aimed at evaluation, can be illustrated with cases in which a CCTV surveillance system in a public space is evaluated for its protection of the privacy of citizens. Such an evaluation requires that a whole network of human and nonhuman entities is being considered that play a role in safeguarding privacy. This includes evaluations of, amongst others, CCTV hardware and software, the properties and behaviors of the human operators, the protocols that govern their behavior, the characteristics of the room in which CCTV images are displayed or stored and their accessibility by third parties, and so on. All elements in this network, and their relations to each other, need to be evaluated relative to a set of privacy requirements, for their adequacy in safeguarding personal privacy.

The third type of investigation, aimed at prescription, can also be illustrated with reference to CCTV and privacy. This type of investigation would specify how a network surrounding a CCTV system would ideally be constituted so as to protect privacy and how its different elements would operate. Alternatively, recommendations could be developed for the improvement of an existing network, for example for the improvement of software, training of operators, improvements of facilities or procedures, and so on.

These three types of investigations focus on networks. However, they could also zoom in on particular components of these networks, including technological artifacts, and focus on their moral roles. For instance, it is possible to focus on the role of a particular CCTV software program in ensuring privacy within a particular network. It is also possible to consider this software abstracted from a particular network and consider its privacy-protective properties across a variety of possible networks. More generally, structural ethics can focus on both networks and components of networks, where these components can also be studied independently of a particular network. We may use the term *artifact ethics* for studies in structural ethics that focus on the moral role of technological artifacts in networks or across networks.

Artifact ethics, as a kind of structural ethics, has the advantage over moral agency approaches that it upholds important differences in the moral roles of artifacts and human agents, as discussed above. It moreover has the advantage of being able to attribute moral roles to artifacts, thus avoiding the fallacy that artifacts are morally neutral, while at the same time avoiding the false notion that morality can "reside" in artifacts , independently of their surroundings. In artifact ethics, it can be shown that artifacts sometimes constitute a major cause of morally good or bad consequences, while at the same time highlighting the dependency of these consequences on a larger network of things and humans.

Structural ethics requires a vocabulary to refer to the networks that are being studied as well as the different elements or component that these may contain, their relations to each other, and their behaviors. I will use the term "network" (or sometimes "arrangement" or "structure") to refer to structures of interacting entities that together determine outcomes or actions that are the subject of moral evaluation. The entities in networks include humans, artifacts, animals and natural objects, as well as larger structures composed of such entities. For instance, an organization is a larger structure that is composed of humans who work together towards a common goal, as well as nonhuman entities owned by the organization that are used to further this goal. An organization has itself a network structure, but it can also function as a component of a larger network in which it plays a role.

Relevant for structural ethics is the relative role of these different entities in fixing moral outcomes or behaviors. I propose that we call any entity in the network or arrangement that has a role in fixing moral outcomes or behaviors a *moral factor*. In ordinary English, a factor is an entity or component which contributes to an effect or result. This is the meaning I have in mind. At the same time, the word "factor" derives from the Latin *factor*, "who/which acts", and hence has associations with the notion of an agent. Moral factors shape or influence moral actions and outcomes. They have *moral influence*. The class of moral factors includes both human agents and various kinds of nonhuman entities.

Moral factors can be positive or negative, measured against a moral rule or principle. A *positive moral factor* is one that contributes positively to a moral principle being upheld, whereas a *negative moral factor* contributes negatively. In addition, moral factors can be accidental or intentional. An *accidental moral factor* is one that happens to contribute towards a moral outcome in a particular arrangement. An *intentional moral factor* is one that has been intended to contribute to an outcome in a particular way. For instance, relative to the moral outcome of cars driving safely, a speed bump and a traffic controller would both be intentional moral factors, whereas a pothole that causes cars to drive slowly would be an accidental moral factor.

Whereas intentional moral factors are often positive, intentionally supporting moral principles, they can also be negative and intentionally contribute to violations of moral principles. For instance, relative to the principle of safe driving, a person imitating a police officer who maliciously signals drivers to perform unsafe maneuvers is an example of a human intentional negative moral factor. Oil intentionally spilled on a road is an example of a nonhuman intentional negative moral factor, whereas oil accidentally spilled would be an accidental moral factor. If technological artifacts generate consequences that are positive or negative relative to a moral principle but were not intended by designers or users, then they are accidental moral factors relative to that principle.

Moral factors can be outcome-oriented or behavior-oriented. An *outcome-oriented moral factor* is a factor that contributes positively or negatively to the realization of a moral outcome. A moral outcome is a realized event or state-of-affairs that is the subject of moral evaluation. For instance, an unjust distribution of goods that results from an action or event is a moral outcome, as is harm to a person or a limitation to his or her freedom. Various moral factors can be identified as having caused these outcomes. A *behavior-oriented moral factor* is one that influences the moral behavior or actions of an agent. For example, my wearing a seat belt is a moral action that is influenced by various moral factors, such as blinking lights on my dashboard and police officers who check on seat belt use.

A structural ethics approach can account well for the distributed realization of moral norms in society, by showing that these norms are enacted not just by humans behaving according to them, but also by social and material structures being shaped to support these norms. A structural ethics approach can, as we have seen, account for the moral role of artifacts. It can also account for the role of things and humans in the moral behavior of human agents by identifying them as moral factors that are contributory causes of someone's moral behavior. Finally, a structural ethics approach can help solve the problem of distributed responsibility. This is the problem that when a moral outcome is the result of the actions of multiple agents, no single agent can be identified as being solely responsible for the moral outcome. A structural ethics approach can be used to analyze the role of different agents in producing the outcome, directly or indirectly. This analysis can then be used to assign moral responsibilities to these different agents. When technological artifacts are involved, it will only be human beings who are assigned responsibility, since technological artifacts and other items do not bear responsibility themselves, yet can serve as moral factors for which one or more human agents bear responsibility.

Individual ethics has a focus and aim that is different from those of structural ethics. Its focus is on the deliberations and actions of moral agents, instead of on networks or components of them. It has three aims that mirror the descriptive, evaluative and normative aims of structural ethics: (1) to study the moral principles, deliberations, traits and actions of human agents, (2) to evaluate the moral goodness of actions, judgments, and traits of human agents and attribute moral responsibility, and (3) to normatively prescribe what moral actions agents ought to perform, judgments they should hold, or traits they should have. Individual ethics makes use of the standard conception of a moral agent, and therefore concerns itself with human beings.

Structural and individual ethics differ in that they are concerned with the moral dimensions of different phenomena: networks and their components versus human beings and their actions. These phenomena require fundamentally different evaluations and subsequent interventions. In structural ethics, the primary aim of moral evaluation is a better design of social and material arrangements: it is to investigate how components of networks can be rearranged, added or removed, through physical or social redesign, so as to generate better moral outcomes. Many networks have a public status or have public effects, so there is a public interest in their functioning, including their functioning according to public standards of morality.

In individual ethics, the objective of moral evaluation may likewise be to change individuals or their actions, but if so, this objective often does not translate into a plan for redesign but rather into a moral appeal to agents to change their behaviors or convictions. The emphasis on moral appeal in individual ethics stems from the fact that persons are generally believed to have free will and to bear the ultimate responsibility for their actions. In more extreme cases, agents may become the involuntary subject of "redesign", by means of involuntary therapy or treatment, or actions that are deemed immoral or harmful may be prevented through physical restraint or incarceration. In general, however, individual ethics is aimed at affecting moral deliberation in moral agents, whereas structural ethics aims to shape and redesign networks and their components.

Although structural ethics may focus on the moral role of particular components of networks, like artifacts, natural objects, and organizations, structural ethics does not focus on individual persons, since this is already the focus of individual ethics. In structural ethics, persons only appear as network components who have roles as moral factors relative to a nonhuman component which is the object of study or relative to the network as a whole. Individual ethics can be of service to structural ethics by improving its understanding of the moral role of particular humans in networks through an identification and analysis of their moral behaviors, values and beliefs.

Conversely, structural ethics can help individual ethics by identifying moral factors external to an agent that are relevant for the study or evaluation of his or her actions, traits or judgments. This is particularly helpful in moral explanation, moral evaluation and attributions of moral responsibility. Moral explanation may be improved by identifying the position of the agent in a larger network and the moral factors that contributed to an agent's action. For example, an explanation of why an agent committed a murder will be helped by an analysis of the material and social arrangements within which the agent was embedded, and the moral factors that directly contributed to his act, such as the availability of a gun and the presence of other agents who encouraged him. An evaluation of his actions and his responsibility for them may likewise take into account external moral factors that contributed to his action or detracted from it. In a similar way, external moral factors may have a role in the analysis or evaluation of moral beliefs, judgments, and traits.

5. Conclusion

I have argued that the moral artifacts view, according to which technological artifacts qualify as moral agents, brings us a better understanding of the moral role of technological artifacts, but at the same time obscures our understanding of human moral agency by impoverishing the concept of a moral agent. I have proposed an alternative view, which I call structural ethics, which has the benefits of the moral artifacts view but also retains the standard conception of a moral agent, and thus the benefits of

this conception. Structural ethics is supplementary to individual ethics, which focuses on (human) moral agents. It focuses on networks or structures consisting of human and nonhuman entities that have moral implications. I have called such entities moral factors. It can also account for the moral role of any kind of entity in producing moral behaviors and outcomes, including humans, animals, artifacts, natural objects, and complex structures like organizations. Artifact ethics is a division of structural ethics that focuses specifically on the moral role of artifacts, both within particular networks and across a range of possible networks.

So can no artifact ever qualify as a moral agent? Let us return to the moral intelligent agents view. Intelligent agents have a greater resemblance to human moral agents than any kind of artifact. They behave autonomously, they interact with an environment, they have a certain degree of intelligence and capacity for reasoning, they have intentional states of some sort, and they can be equipped with goals and moral categories and principles. So can they be moral agents? They could be if they meet the three criteria for moral agency that I outlined earlier. The first of these was a capacity for moral deliberation. Most intelligent agents do not have this capacity, so most would not qualify as moral agents. Progress is being made, however, to equip intelligent agents with capabilities for moral decision-making (Wallach and Allen, 2008). So I will assume that some intelligent agents will be able to meet the first criterion (though see Johnson, 2006 and Himma, 2009).

The second criterion, being expected to adhere to standards of morality, is relatively easy to meet. It only requires that people expect intelligent agents not to act immorally. This expectation may already be in place, as we generally do not expect technological artifacts to be designed so as to produce unethical results. So I assume that intelligent agents can also meet the second criterion. The third criterion, that of moral responsibility and accountability, is the one that is hardest to meet. Most proponents of the moral intelligent agents view agree that it makes no sense to attribute moral responsibility to an artificial agent, and to praise or blame it for its actions. This is true both because artificial agents do not have free will and because they do not have the capacity to experience feelings like pleasure and pain (Johnson, 2006; Himma, 2009).

If this is true, there may be two ways to salvage the moral intelligent agents view. The first would be to redefine "moral agent" by dropping the moral responsibility requirement. This is what Floridi and Sanders (2004) propose. I find this solution unsatisfactory because moral agency has traditionally been identified strongly with moral responsibility. For intelligent agents who meet the first two criteria for moral agency but not the third, it would seem better to introduce a new term, such as "quasi-moral agent", which underwrites that these artificial agents are similar to, but in important ways different from, moral agents. A second way to salvage the moral intelligent agents view is by arguing that some intelligent agents can in fact be morally responsible. Intelligent agents can for example be programmed to explain the moral deliberations behind their decisions, and to accept user input on the morality of their actions. This can be seen as a kind of responsibility or accountability. Still, as Stahl (2006) has argued, even such advanced systems lack some of the properties of full-blown moral agents, such as free will and a capacity to experience or feel blame and praise. He proposes that intelligent agents can at best have a sort of quasi-moral responsibility.

So it seems that some intelligent agents can significantly resemble moral agents, without fully qualifying as such. Such artificial agents, which may be called quasi-moral agents, have capacities for moral decision-making and possibly for responsibility or accountability through an ability to provide and receive feedback on their moral deliberations and actions. The vast majority of technological artifacts, however, including the vast majority of intelligent agents, do not qualify as moral agents, but do qualify as moral factors in the framework of structural ethics.

References

Allen, C., Varner, G., and Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3):251–261.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Eshleman, A. Moral Responsibility. *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/moral-responsibility/>.

Floridi, L. and Sanders, J. (2004). On the Morality of Artificial Agents. *Minds and Machines*, 14(3), 349-379.

Himma, K. (2009). Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent? *Ethics and Information Technology* 11(1):19–29

Jonas, H. (1984). *The Imperative of Responsibility: In Search of Ethics for the Technological Age* (trans. H. Jonas and D. Herr), Chicago: University of Chicago Press.

Johnson, D. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology* 8, 195-204.

Keulartz, J., Korthals, M., Schermer, M. and Swierstra, T. (2004). Pragmatism in Progress: A reply to Radder, Colapietro and Pitt. *Techné: Research in Philosophy and Technology* 7(3), 38-48.

Latour, B. (1987). *Science in Action*. Cambridge, MA: Harvard University Press.

Latour, B. (1992). Where are the Missing Masses? The Sociology of a Few Mundane Artifacts. In Bijker, W., and J. Law (eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge: MIT Press.

Latour, B. (1995). A door must be either open or shut: A little philosophy of techniques. In Feenberg, A. and A. Hannay (eds.), *Technology and the Politics of Knowledge*, Bloomington: Indiana University Press, 272-81.

Powers, T. and Johnson, D. (2004). The moral agency of technology. Paper presented at the *2004 Workshop on understanding new directions in Ethics and Technology*, University of Virginia, 2004. Unpublished, 28 pp. Available online at http://www.sts.virginia.edu/E&T2004/pdf/MAT.pdf.

Searle, J. (1984). Intentionality and its Place in Nature. *Synthese* 61 (1), 3-16.

Stahl, B. (2004). Information, ethics, and computers: The problem of autonomous moral agents. *Minds and Machines*, 14(1):67–83.

Stahl, B. (2006). Responsible computers? A case for ascribing quasi-responsibility to computers independent of personhood or agency. *Ethics and Information Technology* 8, 205-213.

Verbeek, P.P. (2005). *What things do—philosophical reflections on technology, agency, and design*. Penn State: Penn State University Press.

Verbeek, P.P. (2008). Obstetric Ultrasound and the Technological Mediation of Morality - A Postphenomenological Analysis, *Human Studies* 31(1), 11-26.

Wallach, W. and Allen, C. (2008). *Moral Machines: Teaching Robots Right from Wrong*. USA: Oxford University Press.

Watson, G. (1996). Two Faces of Responsibility, *Philosophical Topics* 24: 227–248.