

Can Technological Artefacts Be Moral Agents?

Martin Peterson · Andreas Spahn

Received: 30 April 2010 / Accepted: 22 September 2010 / Published online: 8 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In this paper we discuss the hypothesis that, ‘moral agency is distributed over both humans and technological artefacts’, recently proposed by Peter-Paul Verbeek. We present some arguments for thinking that Verbeek is mistaken. We argue that artefacts such as bridges, word processors, or bombs can never be (part of) moral agents. After having discussed some possible responses, as well as a moderate view proposed by Illies and Meijers, we conclude that technological artefacts are neutral tools that are at most bearers of instrumental value.

Keywords Artefact · Moral status · Moral relevance · Moral agent · Verbeek

Introduction

On March 29, 2010, 39 people were killed in a suicide attack in the Moscow metro. According to the Russian police, ‘two female suicide bombers from north Caucasus have been identified’, one of which is 17 year old Dzjennet Abdarchmanova.¹ Despite her youth she was already a widow. Her husband was shot dead by the Russian army on New Year’s Eve 2009.

Although various groups are likely to disagree about who is ultimately responsible for the attack in the Moscow metro, most laypeople would presumably concede that the bomb itself was not a moral agent that could be held responsible. Common sense

¹ The Economist, March 29, 2010.

M. Peterson (✉) · A. Spahn
Section for Philosophy and Ethics, Eindhoven University of Technology, P.O. Box 513,
5600 MB Eindhoven, The Netherlands
e-mail: m.peterson@tue.nl
URL: www.martinpeterson.org

A. Spahn
e-mail: a.spahn@tue.nl

suggests that Mrs Abdarchmanova was at least partly responsible for the attack herself; and some responsibility should perhaps also be assigned to people in Caucasus who helped her to carry out the attack. A small number of people might even be willing to attribute some moral responsibility to Alfred Nobel, the inventor of the chemical substance used in many bombs. However, to assign moral responsibility to technological artefacts, such as bombs and weapons, would be absurd.

Surprisingly, a significant number of scholars working on the ethics of technology disagree with this picture. Some of them believe that, ‘moral agency is distributed over both humans and technological artefacts’.² This is no doubt a radical claim. At first glance, it is not easy to see why we should accept it. However, according to Peter-Paul Verbeek, an important argument is that ‘humans and technologies do not have a separate existence anymore’.³ His point seems to be that moral agency is distributed over (some) technological artefacts because the technological artefacts and humans form—in some sense that remains to be specified—a close unity in which humans and the technological artefact can no longer be separated. This is a challenging idea.

In this article we scrutinise the recent debate over the moral status of technological artefacts. Can they be moral agents? Or can technological artefacts be morally relevant in some weaker sense? Our aim is twofold. First, we aim to clarify the main arguments in the debate by stating them in ways that everyone can understand. This is not a trivial task. Some of the arguments were originally formulated by scholars coming from a tradition in which the divide between analytic and continental philosophy seems to make cross-boundary communication very difficult. Our second aim is to evaluate the arguments and adjudicate which arguments, if any, we ought to accept.

To begin with, it is helpful to distinguish between two different views about the moral status of technological artefacts, viz. the Strong and the Moderate Views. Advocates of the Strong View believe that both humans and technological artefacts can be moral agents, and that one is therefore justified in concluding that ‘technologies embody morality’.⁴ So on this view, it does indeed make sense to give a very prominent role to technological artefacts in ethics. In contrast to the Strong View, the Moderate View does not entail that technological artefacts are, or can be a part of, a moral agent. However, scholars defending this view nevertheless believe that the role of (some) technological artefacts is quite different from that of other artefacts. In this article we will focus in particular on a version of the Moderate View recently defended by Christian Illies and Anthonie Meijers.

The upshot of this paper is a defence of a view that we take to be an attractive alternative to the Strong and Moderate views. We call this new view the Weak Neutrality Thesis. It holds that technological artefacts sometimes affect the moral evaluation of actions, although these artefacts never figure as moral agents or are morally responsible for their effects. The structure of the paper is as follows. In the

² Verbeek (2008, p. 24). Our discussion of Verbeek’s views is based on his 2005, 2006, and 2008. We are aware that other philosophers have also defended similar ideas.

³ Verbeek (2008, p. 14).

⁴ *Ibid.*, p. 11.

first section, “[The Strong View](#)”, we present and evaluate Verbeek’s defence of the Strong View. In the second section, “[The Moderate View](#)”, we present and evaluate the Moderate View as it has been defended by Illies and Meijers. Finally, in the last section, “[The Weak Neutrality Thesis](#)”, we present and defend our own view, the Weak Neutrality Thesis. We argue that our view avoids the weaknesses of the Strong and Moderate views, without losing any explanatory power or other theoretical virtues.

The Strong View

Verbeek’s account of the Strong View consists of four separate claims. He formulates the four claims as follows:

- (1) ‘technological artefacts ... actively co-shape people’s being in the world’⁵
- (2) ‘humans and technologies do not have a separate existence any more’⁶
- (3) ‘technologies ... have an intentionality’⁷

-
- (4) ‘moral agency is distributed over both humans and technological artefacts’⁸

Although Verbeek is never explicit about the argumentative role of these claims, it is reasonable to assume that (1)–(3) are premises intended to support the conclusion (4). This means that if Verbeek is right, we should believe (4) *because of* (1)–(3). Moreover, it also seems that Verbeek takes some of the premises to support each other; in particular, it seems that (3) is meant to support (2).

Whether Verbeek’s conclusion (4) really follows from (1)–(3) seems to be a matter of definitions. If all entities that have intentions are moral agents, then the argument would be valid. However, if intentionality is all that is required for being a moral agent then (1) and (2) would be superfluous in the argument. This indicates that this is probably not the notion of moral agency Verbeek has in mind. It also seems that (2) could support (4) on its own. If we assume that humans are moral agents, and that humans and technologies ‘do not have a separate existence any more’, then it follows that moral agency is distributed over both humans and technological artefacts. The problem with this interpretation is, again, that it makes the role of the other premises difficult to understand. Arguably, the best way to assess the argument is to analyse each of its premises in turn, and then discuss what follows from them.

Premise (1) is called ‘mediation’. The key idea is that technological artefacts are not neutral intermediaries; since artefacts affect humans in so many different ways

⁵ Verbeek (2006, p. 364). See also his 2008, p. 14 and 2005, pp. 154–161.

⁶ *Ibid.* See also his 2005, p. 164.

⁷ Verbeek (2005, p. 115). Verbeek (2008, p. 14), claims that, ‘in many cases “intentionality” needs to be located in human-technology associations—and therefore partly in artifacts’.

⁸ Verbeek (2008, p. 24). In his earlier writings, Verbeek position is less clear: ‘Things do not have intentions and cannot be held responsible for what they do. But... things have a moral valence... Things carry morality because they shape the way in which people experience their world’ (Verbeek 2005, p. 216).

they, ‘actively co-shape people’s being in the world: their perceptions and actions, experience and existence.’⁹ The technological mediation of perception is said to always work through the mechanisms of ‘amplification’ and ‘reduction’. Consider for instance a thermal camera. This camera ‘amplifies’ certain aspects of reality while other aspects are ‘reduced’. With the help of such a camera one may thus see e.g. losses of energy at the siding of a house, but there are also other aspects one could only have seen with one’s naked eye, such as the true colour of the front door. Note, however, that it would be incorrect to say that the thermal camera *changes* reality; it just changes our perception of it.

Verbeek maintains that many actions are actively ‘co-shaped’ by the environment. Revolving doors were originally designed with the intention of keeping the wind out but had the mediating side-effect that people in wheel-chairs could no longer enter buildings through such doors. Verbeek argues that because technologies often have these kinds of mediating effects, the ethics of technology should not be technophobic, contrary to what was suggested by e.g. Heidegger or Habermas. On Verbeek’s view, we should accept the fact that mediation is always going on and adjust our actions accordingly.

We disagree with Verbeek about premise (1). In our view, there is no reason to think that his examples of mediation support the claim he takes them to support. It is simply not true that thermal cameras and revolving doors show that ‘technological artefacts ... actively co-shape people’s being’.¹⁰ It is of course correct that the use of technologies often have important effects on us, and it may very well be true that it is impossible or very difficult to foresee those effects. However, this does not mean that technologies ‘actively co-shape people’s being’.¹¹ Technologies are not active in any reasonable sense. They are passive. The entity that is active is the designer or inventor who decides to produce or sell the new artefacts. This is not to deny that technologies have an impact on people’s being. Our point is just that the claim that they ‘actively co-shape’ our being is misleading. Technological objects certainly have an impact on us and our actions, like many other natural and non-natural objects, but this impact is not active in the sense that it is independent of the designer or inventor who decides to produce or sell the new artefacts.

This brings us to premise (2) of Verbeek’s argument. Its main point is supposed to be that humans and technologies ‘do not have a separate existence any more’.¹² On a straightforward reading of this claim, it seems to be false. Many technologies, such as sun-powered satellites, would continue to exist for decades even if all humans were to suddenly go extinct. This suggest that satellites and humans do in fact ‘have a separate existence’, according to a literal reading of that claim. We take this to show that this is not what Verbeek seeks to claim.

⁹ Verbeek (2006, p. 364). Verbeek (2008, p. 14), writes that, ‘we cannot hold on to the autonomy of the human subject as a prerequisite for moral agency ... we need to replace the “prime mover” status of the human subject with technologically mediated intentions.’

¹⁰ Verbeek (2006, p 364).

¹¹ Op. cit., our italics.

¹² Verbeek (2008, p. 14).

In order to make sense of (2), one needs to interpret Verbeek's claim in some more sophisticated but less straightforward way. Here is a possible interpretation (which is not discussed by Verbeek) that would make the premise true: A technology is not just a material object, it also comprises knowledge about how the objects is constructed and used. Take, for instance, the solar-powered satellite. It seems very plausible to maintain that if all humans were to suddenly go extinct, there would no longer be anyone around who knew how to operate and maintain the satellite. Hence, there is a sense in which humans and satellites are intertwined: Knowledge about the construction and functioning of the technological artefact is stored in human brains. So if all humans were to go extinct, all this knowledge would no longer exist, and in that sense there would no longer be any satellite technology. The problem with this interpretation is that it is too broad. If all that is required for humans and technologies not to have 'separate existence' is that the former have knowledge about the latter, it would follow that e.g. humans and the weather, or humans and the planet Saturn do not have 'separate existence'. If all humans were to suddenly go extinct, all our knowledge about how to predict, explain and respond to good and bad weather (which seems to affect our life as much as technologies) would be lost. The same would be true of our knowledge about the planet Saturn. This indicates that this interpretation of (2), although easy to understand, is of limited interest.

We now come to what we take to be the most promising interpretation of (2). According to this interpretation, the mediating role of technology blurs the ontological distinction between 'subject' and 'object'. In traditional Cartesian ontologies the subject is taken to be active, and can have intentions and agency, whereas the object is passive and can neither have agency nor intentionality. We believe that this is the distinction that Verbeek wishes to question.

On this reading of (2), Verbeek thus seeks to show how modern technologies cannot be adequately interpreted by using Descartes' traditional categories. We have already explained that Verbeek believes that technologies mediate our interpretation of reality, and it is thus not unreasonable to suppose that this mediating role of technology helps to constitute what counts as 'object' and 'subject' in our society. However, the problem with this interpretation of (2) is that it makes the argument for ascribing moral agency to artefacts extremely dependent on a substantial ontological assumption. Ideally, a good argument for ascribing moral agency to artefacts should be compatible with a wide range of basic ontological assumptions. Verbeek's argument does not meet this desideratum. For example, we take it that a large number of philosophers would agree with us that Verbeek's argument blurs the fundamental distinction between our perception of reality and the way things really are. If it is possible for us to hold false beliefs about reality, we need to distinguish between our beliefs about the world and the way things really are. Verbeek might be right that technologies affect the way we perceive reality, but in our view this does not prove anything about how things really are. We are all familiar with cases in which our perception of reality gets blurred, such as when we drink massive amounts of alcohol. Such cases might be interesting from a psychological or epistemic point of view, but their ontological significance is negligible.

By analysing ultrasonic pictures of unborn children, Verbeek concludes that ultrasound scans are not mere tools for making ‘visible an unborn child in the womb’.¹³ Technologies that mediate our experience of the world do not provide a neutral picture of the object in question. According to Verbeek, these technologies rather (i) constitute what counts as reality, or at least (ii) which aspects of reality count as being relevant. Interpretation (ii) is of course much weaker than (i), and can hardly be taken to support (2). We therefore focus on (i). According to Verbeek, an ultrasound represents the foetus in a very specific way: as a potential patient and as a person with moral status. It thus generates ‘a new ontological status of the foetus’.¹⁴ What counts as an object can be *altered* by technologies. If true, we agree that this would support (2).

Verbeek moreover claims that ‘subjectivity’ is also mediated through technology. In his defence of this claim he refers to Foucault’s analysis of the ‘subject’ in his *History of Sexuality*.¹⁵ Foucault’s main idea is that the (moral) subject is not time-independent, in the sense that each temporal period and each ethical system seems to have its own idea about what counts as subject and what constitutes a subject. ‘Ethics is done by ‘subjecting’ oneself to a specific ethical code, and by doing so people constitute themselves as specific moral subjects’.¹⁶ In our time, technology plays a very important role for the way in which the subject constitutes herself. Technology thus becomes relevant for ethics because of the way technology constitutes the subject. We take Verbeek to claim that this line of thought offers additional support for (2).

In our view, neither of the two lines of reasoning outlined above support (2). The claim that technology constitutes the object (because what counts as object is altered by technology) as well as the claim that technology constitutes the subject (because what counts as subject is also altered by technology) are problematic. These claims confuse the distinction between the genesis of our ideas, i.e. claims about how we do actually reach moral and epistemic judgements, and the justification of these claims. Verbeek’s argument can at best show that our beliefs, opinions, and moral judgments may depend to some extent on what technologies are available in society. However, this does not lend support to the ontological claim that, ‘Humans and technologies do not have a separate existence any more’. Our beliefs, opinions, and moral judgments also seem to depend partly on what beliefs, opinions, and moral judgments are expressed by the Pope in Rome (For some people there might be a positive correlation, for others a negative one). Fortunately, the fact that our beliefs, opinions, and moral judgments depend partly on what the Pope says does not prove that the Pope and we, ‘do not have a separate existence any more’.¹⁷

The least plausible of Verbeek’s premises is that technologies have intentionality, i.e. premise (3). We of course agree that if Verbeek could really show that technologies have intentionality, this would indeed lend support to the claim that

¹³ Verbeek (2008, p. 14).

¹⁴ *Ibid.* p. 16.

¹⁵ See Foucault (1976–84).

¹⁶ *Ibid.* p. 18.

¹⁷ *Op.cit.*

moral agency is distributed over both humans and technological artefacts, i.e. to (4). We also acknowledge that the argument Verbeek gives for thinking that technologies have intentionality is true: His argument is that technologies sometime influence people's behaviour, e.g. by forcing people to change their behaviour or by persuading them to choose certain actions. Verbeek mentions an example suggested by Ihde, who points out that one writes slowly with a fountain pen. This makes it easier to think over the sentence carefully. If one instead were to use a word processor it would become easier to rewrite the text and alter individual words. According to Verbeek, this shows that 'these writing technologies ... play an active role in the relation between author and text. They have an intentionality, a trajectory that promotes a specific kind of use'.¹⁸ We disagree with this. Although it is true that technologies tend to influence people's behaviour, we do not think that one should use the term 'intentionality' to characterize this influence. Consider, for instance, a man-made path through a valley in the Alps. One could of course say that the path has a 'trajectory' and 'promotes a specific kind of use', but it would be a mistake to take this to show that the path can have intentionality. A path is by no means an agent that performs intentional actions. To claim otherwise would be a deeply misleading use of philosophical terminology.

Verbeek rightly emphasises that the function of a technological artefact cannot always be traced back to the intentions of the designer. The function of a technological artefact can change over time and it can be used for different purposes in different contexts. This does not entail that there is any intentionality within the artefact, however. The telephone was originally invented and designed for the hard-of-hearing.¹⁹ That it is nowadays being used for a different purpose does not prove much. In order to explain the change of function, no technological intentionality or joint human-artefact-intentionality is needed. All that needs to be said is to simply point out that technological artefacts may be used for different purposes in different situations. This does not mean that they have intentions of their own. We should therefore reject premise (3).

We conclude that Verbeek's defence of the Strong View seems to be based on false premises. Premises (1)–(3) are very implausible assumptions, which we have no reason to accept. Verbeek's attempt to justify the conclusion (4) is therefore not convincing.

The Moderate View

Illies and Meijers defend what we call the Moderate View. They describe the Moderate View as, 'an intermediate position that attributes moral relevance to artefacts without making them morally responsible or morally accountable for their effects'.²⁰ Illies and Meijers never define their notion of moral relevance (and they sometimes use the term 'moral significance' instead of 'moral relevance' without

¹⁸ Verbeek (2005, p. 115). See also Ihde (1990, pp. 140–143).

¹⁹ Black (1997, p. 18).

²⁰ Illies and Meijers (2009, p. 437).

clearly stating whether these terms are synonymous). However, it seems that the notion of moral relevance they have in mind needs to be a quite strong one, that goes beyond the observation that technological artefacts sometimes affect the outcome of our actions. Otherwise the Moderate View would simply collapse into the uncontroversial claim that technological artefacts sometimes play a causal role in a chain of events in roughly the same way as e.g. storms, volcanoes, and other natural phenomena.

In order to correctly understand the Moderate View it is paramount to note that it is not a claim about the moral relevance of individual actions, but a claim about what Illies and Meijers call Action Schemes. An Action Scheme is defined as, ‘the set of possible actions with different attractiveness that is available to an agent or group of agents in a given situation’.²¹ So rather than focussing on how artefacts affect individual actions, Illies and Meijers focus on how artefacts affect the *set* of actions available to the agent. Their key point is that technological artefacts are morally relevant in the sense that they sometimes affect the attractiveness of the possible actions that make up an Action Scheme. Note that this notion of moral relevance is a rather weak one. In light of our remark above, it may actually be *too* weak: Virtually all kinds of entities, including cars, flowers and volcanoes will sometimes affect your options for actions; hence, everyone would certainly have to admit that technological artefacts are morally relevant. We suspect that this is not the notion of moral relevance Illies and Meijers have in mind.

As explained by Illies and Meijers, ‘the introduction of the mobile phone has extended our range of possible communicative actions’.²² A mobile phone is thus morally relevant in the sense that it affects what alternative actions are available to us. This point can also be spelled out by talking about reasons. By introducing new actions people get new reasons for action that they did not have before. For example, since it is nowadays easy to communicate with one’s near and dear even from remote places in the world it could be claimed that the introduction of mobile phones and other technologies have given us a reason to keep in touch with our relatives that we did not have before. If things become possible, that were not possible before, this might change our moral obligations. If we do not have a moral obligation to do *x* because it is impossible for us to do *x*, but to do *x* would be morally required if *x* was possible, then a change in the realm of ‘can’ implies a change in the realm of ‘ought’. This notion of moral relevance seems to be the one that Illies and Meijers have in mind.

We are now in a position to give a more precise account of the sense in which Illies and Meijers take technological artefacts to be morally relevant. Consider the following two slightly different principles for evaluating sets of actions, which we will call *S* and *S'*, respectively:

S: ‘We might even see it as better if, ceteris paribus, people have more rather than fewer options for actions’.²³

²¹ *Ibid*, p. 427.

²² *Ibid*, p. 427.

²³ *Ibid*, p. 431.

S': 'a situation S_1 is morally better than situation S_2 if people have the option to do something morally good in S_1 that they do not have in S_2 '.²⁴

Note that according to principles S and S' the mere possibility to perform a good action is morally relevant. By creating new possible actions one may thus turn a bad situation into a good one. Technological artefacts are thus morally relevant in the sense that they often create new options for actions.²⁵

Illies and Meijers explicitly claim that the binary relation 'better than' is a relation that holds between situations, not between actions. This means that their position cannot be conceived of as an alternative to traditional ethical theories such as consequentialism or Kantianism. Traditional moral theories evaluate individual actions, not situations or sets of alternatives. (Illies and Meijers however correctly point out, but do not explore, the link between the Moderate View and the capability approach advocated by Sen and Nussbaum.)

Should we accept the Moderate View? We think one could raise at least three objections against it. The first objection is a moral one. As pointed out above, the Moderate View entails that we should assess situations (or Action Schemes) rather than individual actions from a moral point of view. This entails that actions that are possible to perform, but never actually performed, may sometimes affect our moral verdicts. To take an extreme example, suppose the Nazis had invented Cyclone C in 1940. Cyclone C is, we assume, a gas that would have had caused slightly more pain and death than Cyclone B. (The latter is the chemical substance used by the Nazis for killing six million Jews during World War II).

Would it really have made a moral difference if the Nazis could have used Cyclone C in addition to Cyclone B for killing six million Jews? Principle S states that, 'we even see it as better if, *ceteris paribus*, people have more rather than fewer options for actions'. In our view, this is morally absurd. It simply makes little sense to claim that one would have made things better by inventing Cyclone C and thereby had given the Nazis 'more rather than fewer options for action'.

Principle S' is more reasonable than S, since it at least takes the moral features of the options added to the Action Scheme into account. That said, it is nevertheless deeply problematic to claim that one situation is morally better than another just because people have the option to do something morally good in the first situation that they do not have in second. In order to see this, suppose that Cyclone C was actually a much less harmful substance than Cyclone B (so much less harmful that the choice between the two was of significant moral importance). Now Illies and Meijers are committed to the implausible claim that (i) the situation in which both Cyclone B and Cyclone C are available but only Cyclone B is used is morally better than (ii) the situation in which only Cyclone B is available and used. The mere possibility to kill people in a less harmful way would thus make a moral difference to the atrocities carried out by the Nazis. This is implausible. If both Cyclone B and

²⁴ *Ibid.*

²⁵ Note that this view presupposes a rather thick notion of action. Some people may of course doubt that making a phone call is an action. On a minimalist account of action, one rather uses one's fingers for pressing some buttons. On this minimalist account technological artefacts thus never make new actions available.

Cyclone C are available, although only Cyclone B is used, that situation can hardly be any better than the situation in which only Cyclone B is available and used. The fact that the Nazis had the opportunity to use a less harmful chemical, but did not take that opportunity, either made things worse or made no moral difference. It certainly did not make the situation *better*.

Similar examples can be constructed in which only positive values are at stake. Imagine, for instance, that a group of clever engineers have invented a new technology that would, if used, provide enormous quantities of safe and environmentally friendly energy. Also suppose that we know for sure—never mind how—that this technology will never be used. According to Illies and Meijers, the introduction of this new technology was nevertheless a change for the better. They are logically committed to the claim that actions that are never performed may make a situation morally better. This conclusion however strikes us as implausible. For why would the introduction of a merely possible action make a situation better?

At this point Illies and Meijers could reply that it is important to distinguish between first and second-order responsibilities. In their vocabulary, our first-order responsibility is to carry out morally right actions, but our second-order responsibility is to make sure that there are some morally good actions to choose, i.e. to bring about a good Action Scheme. The problem with this reply is, however, that bringing about an Action Scheme is in itself an action. We thus have a first-order responsibility to bring about good Actions Schemes rather than bad ones. This means that what appears to be a second-order responsibility is actually a first-order responsibility. The distinction between first- and second-order responsibilities, to which Illies and Meijers frequently appeal, does not seem to take care of the objection it is designed to take care of.²⁶

We now turn to our second objection to the Moderate View. Illies and Meijers explicitly claim that one should make moral appraisals of entire situations, including Action Schemes. This seems to be a category mistake. As far as we can see, it makes little sense to claim that one situation (or Action Scheme) could be morally better than another, at least if ‘morally better’ is taken to mean something like, ‘has the property of being morally preferable to’. In our view, the relevant concept to analyse in moral inquiries is the moral status of individual actions. Consider, for instance, consequentialism. This theory provides a clear and unambiguous answer to the question: ‘Under what conditions is an action morally right?’. Consequentialists claim that an action is morally right if and only if its consequences are at least as good as those of all its alternatives, but it would make no sense to say that an alternative action in one set of alternatives is better than some other alternative action in some *other* set of alternatives. The consequences of an action are evaluated

²⁶ Let us also mention another moral worry with the Moderate View: If we admit that technological artefacts are somehow morally relevant in a non-causal sense, it might be tempting to blame artefacts rather than humans for atrocities such as the Holocaust. That is, if we were to accept that bombs and chemical substances could be held morally responsible for the death of six million people, or at least affect the moral evaluation of a situation, there is a risk that we start paying too little attention to the moral responsibility of humans. By including artefacts in moral discussions, we may shift the focus from humans to artefacts in an inappropriate way. This is of course not an objection against the internal coherence of any of the positions discussed here; it is purely pragmatic remark.

relative to the consequences of its alternatives—‘inter-situationistic’ comparisons have no meaning, nor do moral claims about entire situations or sets of alternative actions.

A possible objection to the claim that Illies and Meijers make a category mistake in evaluating entire Action Schemes is that similar appraisals are legion in e.g. the literature on freedom of choice.²⁷ The contemporary discussion of freedom of choice is, however, mainly concerned with evaluating how much *freedom* there is in a set of actions, not its overall moral goodness. It seems perfectly coherent to claim that freedom of choice only has moral implications in appraisals of individual actions, but not when appraising sets of actions.

Although it is of course possible to argue that the moral categories discussed and analysed by consequentialists (as well as many non-consequentialists, such as Kantians) are too narrow, it remains to explain what would be gained by such a radical departure from one of the most basic assumptions in moral philosophy. Consequentialists (as well as many non-consequentialists, such as Kantians) can make interesting moral evaluations of new technologies by focusing on their *actual*, *foreseen*, or *intended* effects. For them, there is no need to make moral evaluations of entire situations or Action Schemes. Illies and Meijers fail to explain what they believe could be achieved with their approach, that cannot be achieved with the traditional approach.

That said, we of course admit that there is an uncontroversial sense in which some Action Schemes might be morally better than others: By bringing about Action Schemes comprising morally attractive options (by e.g. developing new technologies) one makes it possible to perform those morally attractive options, and this is of course morally praiseworthy. Note, however, that on this account the bringing-about of an Action Scheme is a mere tool for enabling the realisation of morally attractive actions. Many objects, such as penicillin, blood transfusions and perhaps Action Schemes, can be valuable in this instrumental sense. However, this is hardly sufficient for showing what Illies and Meijers wish to show.

Our ontological objection could also be spelled out by appealing to a weak version of Occam’s razor. Principles S and S’ require that entities such as sets of actions and situations have the same moral properties as actions. A set of elements can be very different from its element. If all the books in your library happen to be green, it does not follow that the set of all books in your library is green. On the contrary, it is widely agreed that sets of books have no colour at all. The property of being green is not a property that can be reasonably ascribed to a set of books. Therefore, by claiming that sets of actions (Action Schemes) can have moral properties just because the elements of the set can have those properties, Illies and Meijers implicitly commit themselves to a rather strong ontological claim: their theory puts certain restrictions on their ontology, which may turn out to be unnecessary. No matter what kind of entity moral goodness is, it seems clear that not all objects can have that property. A wall cannot be morally good, nor can a prime number or a set of green books. However, by assuming that sets of actions can be

²⁷ We would like to thank an anonymous reviewer for raising this concern. See e.g. Gustafsson (2010) for an interesting measure of freedom of choice.

morally good and bad, Illies and Meijers implicitly commit themselves to certain ontological assumptions about what a set is. From an ontological point of view, this seems unnecessary. Given that there is a more simple and less demanding ontology available, according to which only individual actions can have moral properties, we ought to prefer that ontology, every thing else being equal.

Our third objection to the Moderate View is that it is too broad. Even if we were to accept the Moderate View, it would not give us what Illies and Meijers say and believe it gives us. According to Illies and Meijers, ‘This analysis sheds new light on the responsibilities that engineers, researchers, developers and the producers of artefacts have’.²⁸ However, if one were to accept the controversial claim that sets of actions, rather than individuals actions, are morally good or bad, then new technological artefacts would be in no sense unique or different from a large number of other entities, including natural artefacts. Imagine, for instance, that you are planning to sail from Syracuse to Athens. Because of the work of a group of clever engineers it is now possible to do this trip in a carbon fibre 40 ft sailing yacht with North 3DL laminated sails, which will enable you to cover the distance faster than before. One would thus be justified in concluding that the invention of carbon fibre boats has improved the Action Scheme available to the agent. In the bad, old days one had to use slow boats made of wood.

The problem is that natural phenomena, such as new islands created by volcanoes, or storms, or big waves created by a storm, also affect the Action Scheme available to the sailor. These natural phenomena seem to be at least as important for what can and cannot be done, and they are at least as unpredictable and difficult to control as are new technologies. Illies and Meijers believe that there are important moral differences between technological artefacts and natural phenomena, but their analysis fails to articulate what those differences consist in.

The same is true of the distinction between old and new technologies: Illies and Meijers frequently say that new technologies are somehow special from a moral point of view. But according to the analysis they provide, it seems that in the vast majority of cases the fundamental difference is between the situations in which you have access to *some* technology that solves the problem, no matter whether it is a new or old technology. For example, if you wish to sail from Syracuse to Athens, the crucial point would in many cases be whether you have a boat or not, not whether it is an old or new one.

The Weak Neutrality Thesis

We have argued above that both the Extreme and Moderate Views are implausible. At the other end of the spectrum we have the Neutrality Thesis, which is dismissed by Illies and Meijers. The Neutrality Thesis holds that, ‘artefacts are merely neutral means to the ends agents pursue’.²⁹ Illies and Meijers argue that the Neutrality Thesis ‘has little support’ because of the immense effects technological artefacts

²⁸ *Ibid.*, p. 438.

²⁹ *Ibid.* p. 437.

have on society and on our daily life.³⁰ They claim that '[Technological artefacts] are able to change our relationship to the world in quite fundamental ways and to introduce potentially serious moral consequences that go beyond those of their designers' intentions.'³¹

We agree that all these claims about the nature and effects of technological artefacts are true. It is indeed difficult to imagine what the world would have been like without mobile phones, nuclear weapons, and sailing boats. Technological artefacts have a huge impact on our daily lives and on society at large. However, in our view, this does not suffice for showing that the Neutrality Thesis is false. On the contrary, we believe that the Neutrality Thesis is by far the most plausible standpoint in this debate.

In order to explain and defend the Neutrality Thesis it is helpful to distinguish between two versions of it, which we shall refer to as the Strong and the Weak Neutrality Thesis, respectively. The Strong Neutrality Thesis holds that technological artefacts (i) never figure as moral agents, and are never (ii) morally responsible for their effects, and (iii) never affect the moral evaluation of an action. Advocates of the Weak Neutrality Thesis hold (i) and (ii) to be true, but reject (iii)—on this view, technological artefacts sometimes affect the moral evaluation of actions. The difference between the Moderate View and the Weak Neutrality Thesis is that advocates of the latter position accept a fourth condition that is inconsistent with the Moderate View: (iv) Actions, unlike sets of actions, are morally right or wrong, or good or bad.

We accept the Weak Neutrality Thesis but reject the Strong Neutrality Thesis. Consider, for instance, a terrorist who intends to kill ten million people in a big city by blowing up a small atomic bomb hidden in a suitcase. Compare the possible world in which the terrorist presses the red button on his suitcase and the bomb goes off, with the possible world in which he presses the red button on the suitcase but in which nothing happens because there was actually no bomb hidden in the suitcase. In the first example ten million people die, but in the second no one is hurt. Consequentialists, as well as advocates of every minimally plausible version of non-consequentialism, admit that the action of pressing the button in the first example was wrong (given certain plausible empirical assumptions), whereas the same need not be true of the action in the second example. Perhaps the consequences of pressing the button if there is no bomb in the suitcase would turn out to be fairly good. In fact, even if both actions are performed with exactly the same intention, and intentions are much more important than consequences, part (iii) of the Strong View still seems to come out as false. Anyone who thinks that consequences are of some *moral* importance should accept this. (Almost no one, not even contemporary deontologists, think that consequences are entirely irrelevant. See e.g. Kamm (2007).)

The bomb example is designed to show that the mere presence of a technological artefact, viz. a bomb in a suitcase, can affect the moral evaluation of an action. In the first example it is wrong to press the button and right not to press it, but in the second example it does not matter from a moral point of view (given certain

³⁰ *Ibid.*

³¹ *Ibid.*

empirical assumptions) whether the terrorist presses the button. This shows that the Strong Neutrality Thesis is false. But the truth of the Weak Neutrality Thesis is not jeopardized.

It might be objected that our example, and perhaps even the formulation of the Weak and Strong Neutrality Theses, presuppose that inter-situationistic moral comparisons are possible. This is because we claim that *one and the same* action can be right or wrong depending on whether there actually is a bomb in the suitcase, and as we imagine that the bomb is no longer in the suitcase, the agent is no longer facing the same situation as before. This seems to render our defence of the Weak Neutrality Thesis incoherent.

In reply to this objection, we wish to point out that the Weak Neutrality Thesis can, at the expense of some terminological inconvenience, be restated in a way that does not require any inter-situationistic moral comparisons. The situation in which there is a bomb in the suitcase is clearly different from the situation in which there is no such bomb present. However, the two particular acts we are comparing, pressing-the-button-in-the-first-situation and pressing-the-button-in-the-second-situation are instances of the same generic action. The Weak Neutrality Thesis can therefore be restated as a claim about the moral status of generic actions: technological artefacts (i) never figure as moral agents, and are never (ii) morally responsible for their effects, but may sometimes (iii) affect the moral evaluation of generic actions.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Black, H. (1997). *Canadian scientists and inventors: Biographies of people who made a difference*. Markham, Ontario: Pembroke Publishers.
- Foucault, M. (1976–1984) *The history of sexuality, Vol 1-3*. London, UK: Penguin Books.
- Gustafsson, J. E. (2010). Freedom of choice and expected compromise. *Social Choice and Welfare*, 35(1), 65–79.
- Ihde, D. (1990). *Technology and the lifeworld*. Bloomington, Indiana: Indiana University Press.
- Illies, C., & Meijers, A. (2009). Artefacts without agency. *The Monist*, 92(3), 420–440.
- Verbeek, P.-P. (2005). *What things do: Philosophical reflections on technology, agency and design*. Pennsylvania, US: Pennsylvania State University Press.
- Verbeek, P.-P. (2006). 'Materializing morality: Design ethics and technological mediation'. *Science Technology & Human Values*, 31, 361–380.
- Verbeek, P.-P. (2008). Obstetric ultrasound and the technological mediation of morality: A post phenomenological analysis. *Human Studies*, 31, 11–26.