DOI: http://dx.doi.org/10.5216/phi.v29i1.78641

On the neutrality and values of artifacts[1]

Daniel de Vasconcelos Costa[2,3]
costa.daniel_4@uerj.br

Pedro Fior Mota de Andrade[4,5]
pefimoan@gmail.com

Abstract: This paper criticizes the thesis of the neutrality of moral values of artifacts, and makes the case for a proposal known as Value Sensitive Design, which states that moral values must be considered in the construction and analysis of artifacts. First, (1) we will present the best defense of the thesis of the neutrality of moral values of artifacts, made by Joseph Pitt. In the following, (2) we will criticize each of the arguments presented by Pitt in favor of the neutrality thesis. Finally, (3) we will consider the Value Sensitive Design proposal presented by Ibo van de Poel and Peter Kroes and explain how it would be suitable for a critique of the values and moral issues that artifacts can represent.

Keywords: Philosophy of technology, Moral analysis of artifacts, Value neutrality of artifacts, Value Sensitive Design.

[1] Recebido: 29-02-2024/ Aceito: 10-05-2024/ Publicado on-line: 30-06-2024.

[2] É professor na Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, Rio de Janeiro, Brasil.

[3] ORCID: https://orcid.org/0000-0001-9724-8229.

[4] É pós-doutorando na Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Rio de Janeiro, Brasil.

[5] ORCID: https://orcid.org/0000-0002-1225-5858.

Introduction

Without a doubt, one of the most enduring questions in the philosophy of technology is whether technological artifacts are value-neutral or not[6]. Are we justified in ascribing moral predicates to artifacts created by human hands such as cell phones, weapons, or, even more abstract ones, such as programs? A common position is the one called by many philosophers of technology as "*value-neutrality thesis*" (VNT) (Whelchel, 1986; Pitt, 2014; Poel e Kroes, 2014; Miller, 2021; Heyndels, 2023). This position states that artifacts cannot be called "good" or "bad" in a moral sense. Their morally good or bad use is due to the people who employ them, and it is these people who should be morally blamed or praised, not the artifacts themselves. The VTN is normally supported by those who create technologies and artifacts, such as engineers and scientists, who also claim that science in general is value-neutral and should not be morally judged (Weinberg, 1977; Whelchel, 1986, p. 3-4). One important component of their discourse is that the discovery of new knowledge or the creation of new artifacts is not to be stopped, since those who should be blamed for the consequences of scientific knowledge or artifacts are the people who use them. This position could be well exemplified by the popular slogan "guns don't kill people; people kill people" (Verbeek, 2008, p. 98-99; Pitt, 2014, p. 89-90).

We argue in this paper that the VNT is not only false, that artifacts do in fact embody some kind of moral aspect and, therefore, can be morally judged by themselves

[6] Whenever we talk about values, we are speaking *only* about *moral* values.

regardless of the intentions of those who use them, but also that engineers, scientists, and anyone engaged in producing knowledge and creating artifacts should guide themselves by moral considerations derived from a specific stance toward any human creation, the *Value Sensitive Design* (VSD) (Friedman, 1996; Cummings, 2006; Friedman & Kahn, 2008; Hoven & Manders-Huits, 2009; Poel & Kroes, 2014; Friedman & Hendry, 2019). Only by considerations of this kind one can truly say that the morally bad consequences of applying scientific knowledge or using an artifact are to be ascribed *only* to those who use them, and not to the knowledge or the artifact themselves too.

We will begin with an analysis of, to our knowledge, the most coherent defense of the position that artifacts are value-neutral and, therefore, avoid any kind of moral evaluation, the one from Joseph C. Pitt (2014; 2023)[7]. After presenting the most relevant points of his argument, we will criticize it in two ways. First, we will show that it is flawed by its own standards, for it employs a definition of moral value that fails to satisfy what Pitt himself thought to be necessary. Second, even if his definition was adequate, the consequences that he tried to draw from it when applying it to artifacts do not work either. Then, we will argue that, in order to understand the relationship between artifacts and morality, we need an appropriate concept of artifact. Following this understanding, in the end, we will claim that the notion of VSD can be defended as a moral principle applied to the moral evaluation

[7] It is worth noticing that Peter Kroes (2020), Boaz Miller (2021) and Sybren Heyndels (2023) undertook a similar task that we will do in this paper. There will be similarities, but also differences between our paper and the others.

of artifacts in general. However, we will not put forward any specific account of the VSD. The remarks provided will only point out the notion that the VSD can have multiple senses and directions, and any one of them can be justified in different contexts, since it seems intuitive to think that, for instance, the kind of VSD considerations required in automobile construction is different of the one required in pharmaceutical manufacturing.

Can artifacts embody values? The Value-Neutrality Thesis defended against their critics

Perhaps one of the most succinct defenses of the VNT was made by Joseph C. Pitt (2014). In his account of the relationship between artifacts and human values, he argued for a dilemma: (1) either we must accept the common thesis that artifacts have no moral value attached to them, that they are not intrinsically bad or good, only humans or their actions can be bad or good and, then, use artifacts in a bad or good way; or (2) the notion of value that could be ascribed to them would be an empty concept given how an artifact is made, and it would have no application, at least, not in the sense that the opponents of the VNT would hope (Pitt, 2014, p. 90).

His argument has two parts. The first part has to do with the conceptual basis of the notions he will use in his argument. The second one concerns the argument itself.

First, he assumes correctly, as we see it, that we should settle on a definition of “value” before speaking of it in the discussion about values in artifacts. If we do not, how would we be able to say that artifacts embody values in the first place? This question could not even be raised, since it assumes the understanding of what values are. This brings us to a second point about the search for the definition of value, which cannot be defined in such a way that it precludes, in principle, artifacts from embodying values. Pitt says:

> If all I did was to assert, as I did above, that values are not the sort of thing artifacts can have, then I win by default. [...] Rather than simply stipulate, the case needs to be made for why values are the sorts of things artifacts cannot have in any meaningful way (2014, p. 90-91).

Again, we think that he is right. If we define value in such a way that artifacts cannot embody them, the sentence "artifacts cannot embody values" becomes an analytic truth, and the whole discussion of whether artifacts can embody values or not turns out to be trivial. However, it is when he tries to define value in this non-empty sense that some of the problems in his argument become clear.

After acknowledging the difficulties of the task of defining "value", Pitt proposes what he claimed to be a "pragmatist account of value":

> a value is an endorsement of a preferred state of affairs by an individual or group of individuals that motivates our actions. Values, on the other hand, as a motivation to achieve a preferred state of affairs, serve as action initiators, directing what we do in one direction rather than another (2014, p. 91).

As we can see, for Pitt, his definition is a pragmatic one because it assumes an intimate connection between values and human actions. A value points to a state of affairs and brings about an action in order to achieve it. There are, then, two parts in his concept of value. First, a value is a kind of *motivational state* capable of guiding or even generating an action. As a motivational state, it is deeply connected with human beings and the different kinds of states that they can have. Most of the philosophers of action assume that, for an action to be carried out, we need, at least, beliefs and desires or intentions, that is, *mental states* capable of serving as a motivation for the action (Davidson, 2001; Bratman, 1987)[8,9]. Second, a value takes a *goal* to be better than others and *directs* the action towards it. This is a structural aspect of the concept of value, without which there would be no direction for the motivation and, therefore, no possible action based on value. As he says: "A preferred state of affairs is a goal to be achieved. [...] Endorsing the goal means *acting* in such a way as to bring it about, this is the pragmatism part" (Pitt, 2014, p. 93).

Based on this definition of value, Pitt says that he does not want to deny that artifacts could embody values, but that this kind of value embodiment could not be used to criticize value neutrality. Attempting to do this would be not only strange but also irresponsible. Thus, his argument consists of three minor arguments. The first argument can be called "the too many values argument" (Pitt, 2014, p. 93-94). The second one, "the empirically unidentifiable argument" (Pitt,

[8] Some could argue that scientists, such as psychologists or neuroscientists, conceive physiological processes as the real motivating force behind all human behavior, be it conscious or unconscious. However, it would be strange to assume that values are a kind of physiological process. In this sense, defending value as a kind of mental state seems the best approach.

[9] We use the expression "serving as a motivation for an action" so as not to commit ourselves to one of the positions in the discussion about reasons or causes as motivators of the action, for the "motive" serves as a synonym for both the "reason" or "cause".

2014, p. 94-96). The last one is "the turning the question around argument" (Pitt, 2014, p. 96-97).

The too many values argument is the conclusion of the following premises:

1. Each action proceeds from the decision-making process of these people.
2. Each decision-making process considers different values, in Pitt's sense.
3. Each action is value-laden.
4. Artifacts are normally constructed through the actions of different people.
5. Therefore, each artifact embodies the values of multiple people.

From the acknowledgment that one could attach too many values in the creation of an artifact, Pitt continues his argument by asking us whether we could truly pinpoint the value that an artifact represents, since there could be too many values in one artifact – some of them could be even contradictory to each other. As he argues:

> all human decisions are value-laden and that since any artifact will be the result of many decisions, many values will be involved, so many in fact that it becomes impossible to identify the one value that an artifact embodies, were artifacts to embody values (Pitt, 2014, p. 93).

He claims then that any attempt to say that one value is central to an artifact, whereas others are not, is arbitrary: "we find multitudes of value-laden decisions at every step. There are too many to single out any one without a non-arbitrary selection process and we have seen how difficult that is to do" (Pitt, 2014, p. 98).

There is another form of the too many values argument. According to Pitt, not only the process of developing artifacts is value-laden but also the way we interact with them, so that different people assign distinct values to them through their interactions. He gives us some examples to show how artifacts could embody different values at the same time, assuming, of course, that they do. We will focus on two of them. First, he asks us to consider the football stadium of his university, Virginia Tech (Pitt, 2014, p. 94). Which value is central to this football stadium? Those of the university president, since the good record of the university team brings prestige to the university? Or those of the students who play on the university team and have the aspiration to be football players in the major leagues? Perhaps some might feel that a university football stadium is actually antithetical to what a university should stand for as an educational institution. Which of these values does the football stadium embody, he asks us. It would be strange to say that only one of them is the correct one, especially because all these people feel motivated by their different understandings of the value of the football stadium.

Second, the other example used by Pitt was the F-16 fighter jet (Pitt, 2014, p. 98). One can say that we can use it to kill people in wars. However, it is not wrong to say that it is a technological marvel, that the way it maneuvers is incredible, that its design is interesting, that the cruising speed is enviable for those who must fly with commercial aircraft, and much more. In the end, even if we can use it as a weapon to wage wars, we can value it for many different things, and, perhaps, these are the values that matter the most for many, and not that we can use it to fight wars. The reality is, he claims, that different users will value the same artifact for different reasons, and it would be arbitrary to say that one of these values matters more than the others.

The bottom line for Pitt is that each artifact involves different decisions in the process of its making (Pitt, 2014, p. 99-101). These can embody different values, as the way we relate to the artifacts that are already made is by inserting our values into them. Not only each person can see different values in them, but also these values could be conflicting. A person could, for instance, take the university football stadium as valuable for the university because it brings prestige to it, but, at the same time, she could also see it as contradictory to the whole idea of the general purposes of an educational institution. She could find the F-16 fighter jet a truly beautiful piece of technology, but also hate wars and how they were designed and created for waging wars.

The second argument, "the empirically unidentifiable argument", challenges us to identify the values embodied by the artifacts. He appears to take a positivist stance, according to which something exists only if we can empirically identify it, and since any empirical property associated with values can be identified, one can truly say that values exist[10]. Again, he presents his argument through an example (Pitt, 2014, p.

[10] We say "positivist" because the notion that we can only claim that something exists if we can identify its empirical properties was formulated by those associated with the philosophical tradition known, at first, as "Logical Positivism" and later as "Logical Empiricism".

94-95). He tells us about the well-known historical case of the chairman of the Long Island State Park Commission, Robert Moses, and the construction of the Long Island Expressway. Supposedly, this highway was constructed in such a way that it would hinder bus traffic within Long Island, preventing the lower classes and black people from poorer neighborhoods of New York City from frequenting these beaches. Hence, the Long Island Expressway would embody some racist and classicist values.

However, when Pitt analyses this case, he asks us where exactly in the highway or its blueprint we can find those racist and classicist values, even if we could accept that Moses planned this highway to prevent lower classes and black people from frequenting these beaches:

> So Moses' values are embedded in ….what? Are they to be found in the design, i.e., the working drawings, of the LIE [Long Island Expressway]? Where would we see them? Let us say we have a schematic of an overpass in front of us. Please point to the place where we see the value (Pitt, 2014, p. 94-95).

He continues and asks us what the kind of properties values have. After all, if something exists, it must have properties of some sort:

> Likewise for the LIE – if we look at the actual physical thing – the roads and bridges, etc. where are the values? I see bricks and stones and pavement, etc. But where are the values – do they have colors? How much do they weigh? How tall are they or how skinny? What are they? (Pitt, 2014, p. 95)

Pitt does not reject the idea that Moses built the Long Island Expressway with racist and classicist intentions.

However, even if he had those intentions, they are his mental states, not of the highway or any of its parts, and, therefore, the values are also his, not of the highway or any of its parts. When we observe those artifacts, we can find only empirical properties capable of being identified by sense perception or capable of being studied by the empirical sciences.

Moreover, properties in themselves have their own intrinsic and empirical characteristics. We can think of having weight, extension, shape, having a hole in the center, being impenetrable, or being penetrable. Properties can be defined in such a way that we can identify them by some kind of empirical investigation. However, we cannot even define which kind of empirical characteristics the property of value has so that we can identify them. As Pitt puts it:

> locating whose value is embedded in the artifact is very difficult and locating where the value is in the artifact is equally difficult. And these difficulties stem from our lack of identifying characteristics of values such that we could locate them *in* things (2014, p. 96).

The difficulty of pinpointing what kind of property value is or which empirical characteristics it has makes Pitt believe that one could walk the path that G. E. Moore did to defend value as a property (Moore, 1993, p. 1-27; Pitt, 2014, p. 95-96). In his *Principia Ethica*, Moore argued that one of the big mistakes of traditional ethical theories is their commitment to the idea that moral predicates could be reduced to or explained by natural properties of empirical objects. He focused on the hedonistic utilitarian theory that stated that "good" in moral terms just means something conducive to general happiness. Moore argued that these ethical theories committed what he called the "naturalistic fallacy" (1993, p. 62-71). Does this mean that Moore did not believe that there were moral properties? The answer is no. He believed that moral predicates were simple and unanalyzable, that is, they could not be reduced to or explained by any other natural properties, but they were real. But, if they were simple and unanalyzable, what kind of properties were they? According to Moore, moral predicates show, by themselves alone, that there is an intrinsic value in some acts (1993, p. 58-72). Some interpreters have claimed that Moore defended that moral properties were, then, *sui generis* (Hurka, 2021), that is, they simply existed by themselves, as Pitt puts it: "they are of their own kind" (2014, p. 96).

However, even if we abandon the naturalistic approach to moral properties, against which Pitt argued so far, and argue for the existence of some kind of *sui generis* non-naturalistic properties in order to save the hypothesis that artifacts could embody values, he also claims that this will not work. The first problem is where they could be located in the artifact. Even if we accept that artifacts embody *sui generis* values and leave the naturalistic approach to moral proprieties aside, we cannot pinpoint where the values are. At most, we can say: "they are there, even if we do not know where they are". According to Pitt, this would amount to recognizing that these values are purely metaphorical. They do not truly exist. It is only a way of speaking.

Another point raised by Pitt is that when we claim that *sui generis* values lie in artifacts, we are not merely claiming that they embody the values of human beings. We are actually claiming that they have values *by themselves* (Pitt, 2014, p.

96). They carry values that *they* have and not those of the people who created them. He argues that this is begging the question, for now we have a new kind of value that goes beyond the human realm and even the realm of living beings. These would be some kind of "technological values". But this kind of statement needs explanation, so it could not be accepted as an answer to the question as to whether artifacts can embody values or not.

Before presenting his third argument, "the turning the question around argument", we must make a clarification. This is not his argument *per se*, but rather an attack on the motivation of those who are contrary to VNT. At first, Pitt argues that ascribing values to artifacts is a way of avoiding blame for whatever may happen with their wrong use. Because we can argue that artifacts have values and could be morally bad, we could claim that whatever bad happens is to be attributed to the artifact's account and not ours. Nonetheless, according to him, if something bad happens, we will always find that the true responsible for this occurrence are those who used the artifacts.

The second possible motivation for arguments against VNT is ideological. His claim seems to appeal to the idea that ideologies, perhaps political ones, would hold some artifacts, such as money or nuclear power plants, or institutions, such as political organizations or capitalistic enterprises, responsible for the moral disorders of the world. However, like values, such allegations are unfalsifiable, because each counter-argument is usually opposed not by evidence, but by a story of how that evidence is hidden by some "higher powers". Ultimately, for Pitt, ideological arguments are "little more than conspiracy theory run amok" (2014, p. 97). Following both reasons presented by Pitt, we see that such motivations are an attempt to escape responsibility for something we have done or for something we have not done but should have.

These are the three arguments Pitt presents against those who argue for the thesis that artifacts embody values. His argument for VNT is mainly a negative one, although he admits that he is less in favor of a stronger version than a weaker version of it, because, after all, all three arguments do not properly deny that artifacts cannot embody values. Consequently, they *must* be neutral. Pitt only claims that artifacts could embody too many values, understood as a feature of how human beings produce or view these objects, and any attempt to claim that only one of them is the most relevant, or central, is arbitrary. Besides that, to attribute any other notion of value to them is simply empirically unidentifiable. Pitt is very clear about his objectives and how these arguments are supposed to work:

> VNT claims that values are not embedded in technological artifacts. That is a very strong claim. A weaker version (2) claims that even if we could make sense of the idea that technological artifacts embody human values, there are so many that would be involved the claim says nothing significant. There is also a third version (3) that says we don't know whether or not values are embedded in technological artifacts because we don't know what values look like. While I have explored some aspects of (3), I have more seriously been elaborating a defense of version (2) (2014, p. 101).

The limits of the Value-Neutrality Thesis, or how not to argue for it

Pitt's efforts are laudable and perhaps could even advance the discussion about the value neutrality of artifacts, however, his argument is quite strange. There are many problems in different parts of it. We can begin with his proposed definition of the concept of value, in which we can find, at least, two problems. Both problems are distinct kinds of begging the question. The first concerns the structure of his definition of "value".

He defined value as a kind of motivational state that points to a preferred state of affairs as the one we really want to see happening. The problem is that this definition of value presupposes already some notion of value, for when we say that in the value we find a *preferred* state of affairs that we endorse, we are saying that something has value because we value it, since it assumed a prior evaluation of this state of affairs as valuable in some sense; otherwise, it would have not been preferred instead of others. There is something in this state of affairs that makes us prefer it over other options, and this means that we had already placed some kind of value in it that makes it better for us than other possibilities, since the notion of preferred state of affairs assumes the existence of a ranking that is ordered by some criteria, and this is an act of evaluation. Hence, this amounts to saying that "value is what we already value".

Furthermore, his definition is not suitable for the notion of *moral* value. When Pitt provides us with his definition of value, it is about value in general. He acknowledges that there are different kinds of value. He speaks of aesthetic, scientific, and moral values as different kinds of values. However, he does not explain how the notion of an endorsed preferred state of affairs could be different relative to specific kinds of value. Perhaps any endorsed preferred state of affairs becomes moral if we consider it a moral one. But this creates a problem, because if any preferred state of affairs we endorse becomes moral in the case the agent sees it as a moral one, now we have to deal with the problem of relativism (Pitt, 2014, p. 92).

Pitt saw this and tried to go around it by adopting a specific moral view to guide the notion of value, namely, the notion that moral theory should lead us to a good life (2014, p. 92-93). At first, it seems like a good idea, since now we can say that the value we have is a *moral* one not because of what we want or believe, but because of a background ethical theory that provides the reasons why we prefer it. However, this does not seem a good way of escaping the threat of relativism, because, if we want to uphold his understanding of value, we must accept the idea that value is *only* a subjective mental state, and the adoption of an ethical theory becomes nothing more than a subjective endorsement for some preferred state of affairs, which brings us again to relativism.

The only possibility to escape this path to relativism is to support the notion that ethical theories could provide us with values that we should follow, and not take them as mere preferences that we could abandon if we wanted to. But if ethical theories could provide us with values that we should follow, this means that values could not only be a *preferred* state of affairs. They must be objective in some sense, since we should comply with its directives. Hence, they cannot be treated as mere preferences. However, Pitt has already criticized the possibility of objective values, so he cannot escape the pit he created for himself.

The other problem is even more insidious for Pitt. As we saw above, he claimed that we need a definition of value that, in principle, does not exclude the possibility of it being related to artifacts. If we do so, we would beg the question and the statement that artifacts cannot embody values would be an analytical truth. This would be an uninteresting solution, and we believe he is right. Nonetheless, the solution he provides does exactly what he said it should not. The moment he defines values as "motivators" of actions, he is already committed to the thesis that artifacts cannot embody values, since only human beings can truly act[11]. If values cause actions, they must be a kind of motivational state themselves. As a motivational state, they are a kind of mental state, and, prima facie, only human beings are endowed with the kind of mental states necessary for intentional action. It is clear that his definition of value begs the question by excluding up front the possibility of artifacts to embody values (Miller, 2021, p. 58-59). Perhaps the biggest mistake in Pitt's definition is that the notion of value to be analyzed should be its *noun* form, and not its verbal form, as Heyndels puts it (2023, p. 5).

Nonetheless, even if we set aside these problems with Pitt's definition of value, his arguments have problems of their own. As we have seen above, Pitt considers all the

[11] We will leave aside discussions about whether other non-human living beings are capable of acting or not, since our focus is on humans and artifacts.

intentional steps equally relevant to the creation process of the artifact, and favoring any of them as the most relevant is arbitrary. However, this makes no sense when we analyze the concept of artifact and its design process.

Just because some people take part in the design and creation process of an artifact, it does not mean that their motivations or considerations about it are relevant to its characterization. Even if people are motivated to create artifacts for different reasons, none of these reasons adequately characterize the artifacts. Even if they go through many steps, different decisions, and distinct motivations, they have specific *functions* in order to achieve a specific *goal* (Franssen, 2008). These are attached to artifacts in a way that the motivations people have to create them are not. People's values and motivations can change, and so will their perception of the artifact as well, but the functions and goals that characterize the artifacts do not, and the general characterization of the artifact will also not change. This is the main problem for the too many values argument. These are the two aspects that are the most important in the characterization of any artifact, and not the reasons why people take part in the creation process of it. These are irrelevant to the definition of the artifact. To say that one artifact has one specific function or a specific goal means that the ascription of any other function or goal can be said to be just an addition, but it is not what defines it.

There is a distinction that must be made between natural kinds and artifacts. It is a mistake to ascribe a function to a natural kind, even if they are used for something, such as water or sugar, for instance. An artifact is not a natural kind.

They were created by humans for some specific goal. In the same way, they can be distinguished from natural physical objects too, that is, physical objects that are to be found in the natural world (Kroes & Meijers, 2006). Stones, trees, iron, etc. are, for instance, natural physical objects. Artifacts are physical objects, but they are not *natural* physical objects as they cannot be found in the natural world. What makes a physical object an *artificial* one is that they are created by beings with *intentionality* (Kroes & Meijers, 2006, p. 1-2; Heyndels, 2023, p. 16-18)[12]. They are created by human beings with a specific purpose in mind.

The too many values argument ignores that an artifact is a physical object created *for* some end. As Kroes and Meijers claim, they have a "for-ness" in them: "artefacts have a purpose or function: they are objects to be used for doing things and are characterized by a certain 'for-ness'" (2006, p. 1). They are a being-for-something that can only be explained by teleological thinking. A mechanical explanation is necessary to understand it as a physical object, but more is needed if we want to understand it as an artifact. If we find some kind of physical object that appears to be an artifact, the best way to understand it is through how it functions and what goals these functions are supposed to achieve.

Perhaps it is the confusion regarding the dual nature of an artifact, one physical and another intentional, that makes

[12] We said "beings with intentionality" because it is clear that it is not just human beings that create artifacts in the way we defined. Primates, birds, and beavers, just to name a few, are non-human animals that change their environment or create tools in order to survive in nature. Perhaps they do not have the same kind of intentionality that human beings have, but they have some kind of it. However, to argue about it is beyond the point of the present paper.

Pitt ignore the distinction between the *use* and the *function* of an artifact. Its use is not the same as its function. An artifact has a specific function for what it was made, given by its intentional part, but that does not mean that we cannot use it in a different way, because its physical part allows us to do it (Franssen, 2008, p. 25-29). Take a chair, for instance. It is a common artifact used for sitting. However, someone could use it as a weapon, when they throw it at someone else, or they could use it as a means of seduction, when a person pulls the chair for the other to sit at the table in a restaurant. How we can use a chair depends only on our imagination, but none of this can be said to be the function of the chair, for if no one else uses it as a weapon or a means of seduction anymore, the chair continues to be a chair if this physical object was created for others to sit on it. It can also be a bad chair if no one can sit on it, but it would still be a chair if it is an artifact created to function as something to sit on or for the purpose of people to sit on. It may even happen that some artifacts have their goal rethought because they have worked so well for another goal, even if the way they function does not change. Nonetheless, despite this multiplicity of uses that artifacts can have, they are created with their own specific goal that guides their creation and gives them their function. There is a normativity inherent to artifacts. That is why we can say that artifacts are badly used, or that they ought to be used in another way, or even that they are good or bad for the goal they were created for (Franssen, 2009).

On the contrary, natural physical objects have no function, since they were not created for some goal. However, they can be *used* for many different things. One can use a stone, for instance, as a weapon, when one throws it at another person. But one can also use it as a paperweight or as a print plate, as they are used in lithography. There is no specific goal that a stone should achieve and, therefore, it has no function. At most, it can be used *as if* it had one.

The too many values argument would only be correct if we could not distinguish between the use and the function of an artifact, and we clearly can. However, Pitt seems to be unable to do this because not only does he not distinguish between artificial physical objects and natural physical objects, but he also seems to believe that postulating any goal or function to the artifact would be "arbitrary". This argument would make sense only if it referred to physical objects *simpliciter*. Heyndels sees this mistake, even though he emphasizes it against Pitt's second argument: "What underlies Pitt's argument against the empirical identifiability of moral values from technological artifacts is that technological artifacts are to be understood as *mere* physical objects" (2023, p. 10).

When we understand that artifacts are natural physical objects transformed and manipulated to the point that they create something new, which we cannot find in the natural world, and that they fulfill a function in order to achieve a goal, we are able to understand that most moral judgments and concerns raised against artifacts are not about the values that they embody or, at least, not only about values. We can also think about whether the goal it is designed to achieve is morally good or bad or whether the way it functions to achieve it is right or wrong. The moral criticism of artifacts is not just about value, if it is about value at all. It is mostly about the goals they were designed for.

When we morally evaluate an artifact, we consider what it is supposed to accomplish and by what means it does so. An artifact that is supposed to achieve ends that we understand to be immoral can reasonably be seen as morally bad. Pitt would argue that any artifact could be used for a good purpose, even if it was created to do bad things. He is not wrong. However, he fails to see that it can be *used* for something good, but has a bad *function*.

Perhaps the most forceful example of this point is the concentration camps projected by the Nazis. They were created with one specific goal in mind: to exterminate Jews and other ethnicities. They had this function. Eric Katz argues that the concentration camps were clearly designed to kill those who were imprisoned there. The gas chambers were very instructive in this matter. They had only *one* function, namely, to kill the prisoners in the most efficient manner: "the design of the gas chambers and crematoria were meant to maximize the efficiency and secrecy of the killing operations. The victims were brought to one building alive and were gassed and incinerated out of sight from the rest of the camp personnel and prisoners" (Katz, 2005, p. 416).

He claims that the design and function of the concentration camps show that there were values embedded in them. They were instruments of a specific political and social context and embodied the values of this context. But one does not need to appeal to values to see how the concentration camps were wrong. We just need to understand the goal which it was created to achieve. The concentration camp, as an artifact, was created to be an instrument of death. Pitt could not just say: “Well, one can use it for the good. It is the Nazis who should be blamed for using them in a bad manner” or “Through them we learned how evil humanity can be, and they became a negative moral example, in the sense of what we should avoid.” These would be not good answers. There is no other way to use concentration camps, even as a learning experience. Their very creation is morally offensive.

Another problem with the too many value arguments is that even if we assume that Pitt’s definition is a good one, it is not clear that artifacts are not related to values. Even if only human beings could *have* values, to *embody* them is not the same as having them. This can avoid the problem of begging the question above. Perhaps Pitt realized this because the too many values argument seems to accept that values could be *ascribed* to artifacts, even if these could not embody them. Perhaps Pitt is even correct when he claims that artifacts do not embody values. But he is clearly wrong when he assumes that it is through values, at least in the way he understood them, the only way our moral convictions are expressed.

When we observe the history of moral philosophy, we realize that moral predicates have been attributed to different parts of our actions over time. Some can judge our actions morally based on the motives we act. On other occasions, our goals are said to be “right” or “good”, “wrong” or “bad”, regardless of the reasons for which we act, so that we can judge our actions morally based on their goals. Furthermore, instead of our discrete actions, we can judge someone morally for her character and corresponding behavioral pattern.

Not only there are different parts of an action that can be morally evaluated, but it could even not be the action but something else that ties all these actions together as a whole. The same occurs in the different aspects of our creative capabilities. When creating an artifact, we act not only through values, but we also think about the goals these artifacts should achieve, how we can create these artifacts, the consciously acquired habits and unconscious biases that go along with this creation, and much more. Any one of them can be morally relevant and morally evaluated in the artifact creation processes, from its conception until its final stages.

Pitt reduces the whole possibility of moral evaluation in the creation process of an artifact to the notion of value. But this reduction paves the way for his argument at the cost of reducing morality itself. For instance, why should we ignore our biases when morally evaluating our actions? Even if our values, our means, and our goals were good, our actions could still be morally criticized if the actions themselves incorporate some kind of prejudice. Sometimes, it is not the case that we are not aware of the existence, consciously or unconsciously, of prejudices of our actions. It could be a kind of prejudice that society has because this is a shared and accepted view within it. In this sense, even if people are truly thinking about others and carefully choosing the correct means to help them, the biases and prejudices in their actions are not, strictly speaking, due to those people individually, but due to the way society itself is structured. The same could be said about artifacts. Many of the biases and prejudices they embody are not directly related to the people who produce them, but to the context in which the artifacts are made.

One of the widely known cases of biased artifacts in the aforementioned sense was that of facial recognition programs. As some newspaper articles and academic papers make clear, facial recognition technologies were programmed by algorithms that racially discriminated against people (Buolamwini & Gebru, 2018; Najibi, 2020; Johnson & Johnson, 2023). It was shown that their performance was better with White people than with people of different ethnicities, such as Blacks and Hispanics. This has been characterized as "racial discrimination" by some. Others preferred to call it "algorithmic fairness", since the real problem is not racial discrimination, which is a particular instance of a much deeper issue. But the real problem is *how* the algorithms were made. There was something morally wrong with it, not necessarily with those who created it.

Pitt addressed this problem and argued that the real issue is that the biases we find in face recognition algorithms do not come from them, but from the programmers who trained these algorithms with biased instructions and data. He said: "One of the first pieces of evidence of this was the discovery that facial recognition programs created by White programmers had a hard time identifying Black individuals" (Pitt, 2023, p. 16).

The problem with this answer is that Pitt ignores the fact that programmers could have had no bias at all at the time of creating the algorithm. Being biased is having an unconscious propensity to do something. Even though this may be the case sometimes, it is not *necessarily* true in all cases that they are biased in creating the algorithm or providing the necessary data. They may show bias or unfairness even if the entire decision-making process is completely unbiased and fair. One could even say that the person had not put any thought into what she created, that it was not her concern to think about how it would be used, in the same way that Albert Speer, one of Hitler's Architects, claimed when talking about the concentration camps he designed. It is easy to see that artifacts can have consequences not intended by those who created them, and these can be morally bad.

In the same way, we cannot ignore the consequences, intended or not, of artifacts. The means by which artifacts achieve their goal can also be morally judged, that is, the way it is designed. If someone created a tool that, in order to achieve what could be considered a good goal, harms people because of the way it works, this tool can be morally criticized. A clear example of this is hostile or defensive architecture. It describes the kind of construction found in public spaces that pushes certain types of people away from it for the sake of maintaining some public good, such as cleanliness or public safety. A widespread example of hostile architecture is spikes or stones fixed to the ground or benches with sit dividers to prevent the homeless population from resting or sleeping in these places. Some politicians who advocated this kind of construction argued that without it the city would be more unsafe or less clean.

However, this type of architecture is not necessarily created with this thought in mind. One might say that some of the artifacts display a distinctive design that appeals to some. In this case, one could say that those who create these artifacts are, at worst, ignorant. It is not the promotion of safety or cleanliness that is to blame, nor the intention to use the means to make the city safer or cleaner, but the *means themselves* and their *designs* that are problematic, despite the intentions of those who created these hostile architectures.

The too many values argument fails on, at least, two grounds. First, it fails to distinguish between natural and artificial physical objects and between the use and the function of artifacts. This is because, when these distinctions are properly made, we can see that Pitt's so-called "values", which work as the motivational force for taking part in the creation process of an artifact, are nothing more than individual motivations for participating in the process, but they do not characterize the artifact itself. Second, Pitt reduces morality to value and disregards other aspects that are also important. When we take these other aspects into account, we can see how narrow the dichotomy between value-neutrality and value-ladenness really is. What is relevant to the debate is not whether artifacts can embody values or not, even though the notion of values may be important to the question, but whether they can be morally evaluated by themselves, without any reference to a creator or user.

As far as the empirically unidentifiable argument is concerned, we can see that most of the problems it raises can be equally solved if we take into account the same points raised against the too many values arguments. Before we consider them, we must analyze one of Pitt's assumptions in this argument.

Pitt makes a strong metaethical assumption in his empirically unidentifiable argument. According to him, artifacts can embody values *only if* these are their *real properties*. By real properties, we mean that they should be as empirically real as the weight, size, or shape of artifacts. These are mind-independent properties. Since they constitute mind-independent facts about the world, the truth conditions of any statement involving these properties can only be satisfied if they are to be found directly in the world, not in our subjective mental states. In the metaethical literature, this thesis is known as "*moral naturalism*" (Miller, 2003, p. 4-5; Fisher, 2011, p. 55-72). Pitt seems to be against the possibility of moral naturalism about the values of artifacts.

However, he seems to conflate two distinct theses of moral naturalism, which makes his argument go amiss. Even though his critique is different from ours, Boaz Miller sees this too. He says that there is an ontological reading and an epistemological reading about the empirical unidentifiability of values in artifacts (Miller, 2021, p. 63). He is correct, especially when we consider that the background to Pitt's argument is moral naturalism.

First, Pitt does not distinguish between moral naturalism and *moral realism* (Fisher, 2011, p. 5-6). Both moral naturalism and moral realism have points in common. They agree that moral properties do *exist*. They also share the semantic thesis that moral sentences can be true or false. However, the ontological thesis espoused by moral naturalism is quite different from that adopted by moral realism. Moral naturalism claims that moral properties are real in the sense that they are mind-independent facts and, therefore, they depend on natural facts of the world. A moral realist does not need to endorse this notion, though. Although one could rightly say that moral naturalism is a kind of moral realism, the reverse is not true.

Pitt also seems to talk about moral realism when referring to sui generis values, since this kind of values can only be *non-naturalistic*. But there are other types of moral theories that do not need to commit to such a strong thesis that moral properties must be mind-independent facts for them to be objective, and, therefore, capable of being true or false. Miller seems to point out this fact when he argues that there are *social* facts, that is, facts that are possible only because there are beings capable of symbolic interactions. Moral constructivism is one of the moral positions that acknowledge the possibility of a type of cognitivism that is essentially mind-*dependent* (Rawls, 1980; Korsgaard, 2008). Therefore, there are more ontological possibilities than Pitt allows.

Second, Pitt seems to conflate moral naturalism with *cognitivism*, according to which moral sentences are truth-apt, that is, they can be true or false. The cognitivist thesis is not just the semantic consequence of moral realism. It is an independent thesis. One thing is stating that something exists, another thing is stating that we can speak about it in terms of truth or falsehood. It is possible to hold a cognitivist stance without adopting the realist stance.[13] Perhaps it is his positivist stance that compels him to accept this conflation of both theses, since, under such circumstances, one could only talk about truth-aptness in the context of empirical entities.

[13] See, for instance, John L. Mackie's *error theory* (1977). He seems to endorse a moral cognitivist approach to ethics, as he believes that moral statements are structurally truth-apt. However, since there are no moral properties to be found in the world, moral sentences turn out to be systematically false.

The strength of Pitt's empirically unidentifiable argument lies in this assumption. He already rejected the plausibility of moral realism and the thesis that values must be empirically verifiable. That is why we cannot identify them in artifacts. However, when we acknowledge that moral cognitivism differs from moral realism, it becomes irrelevant whether the thesis of moral realism is true or false. We have new possibilities for making truth-apt moral utterances that do not need to resort to values or moral predicates as if they were empirically identifiable moral properties. In our case, the morally relevant point is the goal or the function the artifact must achieve.

We do not need to argue that values must be empirically identifiable in order to know, for instance, that weapons of mass destruction (WMD) are morally wrong in themselves. They were created with a single purpose: to destroy the enemy. It is irrelevant whether or not we can empirically identify values in WMDs to know that they are still terrifying, as the goal they are supposed to serve is morally wrong at the highest level.

It is the goals and the functions of the artifacts that allow us to morally evaluate them. They do not need to be empirically identifiable, at least not in the way Pitt wants them to be, but they are what make artifacts distinctive.[14] They not only show how the artifacts are normally used, but also present the truth conditions for moral statements about them. Moreover, they do not need to be morally judged solely by

[14] We say that they do not need to be empirically identifiable because Heyndels makes a compelling case for it. However, we doubt that Pitt would agree with Heyndel's claims, and our argument was intended to avoid Pitt's possible criticisms.

reference to their goals and functions. When we do not reduce morality to values or moral properties conceived in Pitt's sense, we understand that the design and consequences of artifacts, whether intended or not, are also morally relevant to our judgment, as we have established above. Artifacts do not need to embody values to be morally evaluated or for our moral judgment of them to be true or false. Much less they require empirically identifiable values, since moral cognitivism does not entail moral realism. The demands of the empirically unidentifiable argument made by Pitt are unnecessary.

The last argument against the value-ladenness of artifacts, the turning the question around argument, is not so much an argument as an attack on the supposed motivations that those who deny VNT might have. Pitt speculated about two of them. The first one would be to escape one's own responsibilities about the consequences of their actions by morally blaming artifacts and institutions. The second one would be ideologically motivated in the sense of a desire to criticize artifacts and institutions for whatever bad happens in the world. We argue that these reasons are also wrong.

First, no one would be able to escape responsibility by saying: "Well, it was the machine gun that made me do this". This just does not make any sense. This kind of plea would be useless before a police authority or a court of law. When people criticize that some kinds of artifacts can be easily purchased and this fact leads to criminal occurrences, they are not claiming that those who acted are not guilty or that artifacts have a will of their own. They claim that artifacts can cause certain situations more easily than others. After all, that is what they are made for. No one can deny that school massacres, such as in Columbine, for instance, would be avoided if legal access to high-caliber weapons were more restricted. This is what some people mean when they say that certain artifacts are responsible for something. They allow us to do something that we could not if they did not exist or were created differently (Verbeek, 2006; Morrow, 2014). Our actions are mediated by them. No one can deny that guns without people cannot kill anyone, which is why people who use guns will be held responsible for what happens. However, weapons make what happens worse than it would have been if there were no weapons. That is why people blame guns for some incidents, and the same can happen with any artifact depending on the context.

In addition, there are two different parts to Pitt's claim that denying VNT could be ideologically motivated. The first part is his accusation. He says that these ideological motivations for denying VNT are part of a conspiracy theory. He provides no real argument to support his claims. He just quotes Langdon Winner's book, *The Whale and the Reactor* (1986), which states that organizations form a power structure whose actions and consequences can be detrimental to human beings and the environment as one of the sources that provides an example of this kind of conspiracy theory. He argues that this is an ideological reason because it is grounded on question-begging or unfalsifiable claims, as the lack of evidence for these claims is due to the very power structures of these organizations.

However, there is no lack of evidence of this kind of behavior in organizations. Economics textbooks address externalities as a type of market failure, which is nothing more than imposing some loss on someone who was not a party in an economic transaction (Stiglitz & Walsh, 2006, p. 252-254; Mankiw, 2024, p. 190-193). Pollution, for instance, is a classic example of externality. It is not necessary to assume that there is a conspiracy theory behind the pollution that we see happening around the world. It occurs because it is profitable for industries, which makes them continue to pursue this path. Still, if we analyze some behaviors that seem like conspiracy theories, there are a lot of counterexamples as well. Just remember how many companies were sued for polluting rivers and causing serious illnesses to citizens in nearby cities, and how companies pay for studies against the scientific consensus that goes against their interests (Oreskes & Conway, 2010). These supposedly unfalsifiable claims and conspiracy theories are not, in fact, conspiracy theories and can be falsified.

In the end, Pitt's accusation of the possible conspiratorial motivation of the VNT criticism is nothing but a fallacy. In fact, since he assumes certain behind-the-curtain motivations of VNT critics that cannot be falsified, his allegation is itself a conspiracy.

Value Sensitive Design and the making of a value-embedded artifact

Pitt's arguments fail thoroughly. He cannot show that artifacts are *necessarily* neutral. Sometimes artifacts can be neutral in the sense he argues, but, at other times, artifacts are not neutral, they *do* embody values. The falsehood of the VNT implies its contrary, that artifacts *can* embody values.

However, the falsehood of the VNT cannot neither explain *how* artifacts embody values nor give any normative account of how they *must* be created in order to embody (good) values. In this section, we will defend one of the most famous accounts of how artifacts should be developed so that (good) values can be incorporated into them, namely, the *Value Sensitive Design* (VSD) (Friedman, 1996; Cummings, 2006; Friedman & Kahn, 2008; Hoven & Manders-Huits, 2009; Poel & Kroes, 2014; Friedman & Hendry, 2019).

First, it must be noted that the idea of embedding values into artifacts is not new. However, these are not necessarily moral. Every artifact is created with a specific function that seeks to achieve a specific goal. With this thought in mind, the human-artifact interaction is one of the most important aspects of artifacts. Artifacts are, then, created with the limitations of the human beings in mind. For instance, nowadays, many chairs are created with the thought of being the most comfortable possible, but they are also designed to not damage the human body. The idea of producing a comfortable chair that, at the same time, does not harm the human body already entails thinking about values in design, even if these values are instrumental. This kind of consideration of incorporating instrumental values into artifacts became known as "ergonomics". These values considered by the design of artifacts are how they can be user-friendly or more efficient in achieving their goal. The VSD goes beyond this thought, for it thinks about *moral* values and not any kind of value.

The VSD is an approach supported by Batya Friedman, Peter H. Kahn, and others at the University of Washington, which takes the whole design process and development of artifacts as essential for the embodiment of values in artifacts (Friedman, 1996; Cummings, 2006; Friedman & Kahn, 2008; Friedman & Hendry, 2019). It can also be used to evaluate the values embedded in the artifacts through an analysis of their design and the consequences of this design. This approach became widely known and adopted by many who work in engineering and seek to incorporate values into the artifacts they produce. There are, however, different approaches to the VSD and critiques of it.

Friedman and Kahn consider how certain values can be embodied in artifacts. To do this, they list a series of values that we take as important (Friedman & Kahn, 2008, p. 1.251-1.257). They consider human welfare, privacy, freedom from bias, trust, informed consent, among many others, as the moral values that engineers and designers should have in mind when they are projecting and designing their artifacts. To achieve this incorporation of values into artifacts, the VSD must consider different aspects of their production. Our focus will be just a specific aspect of the incorporation of values into artifacts through their design process. We will propose how the notions of *goal*, *function*, and *use* of artifacts are central to the incorporation of values into artifacts.

Ibo van de Poel and Peter Kroes argue that the central notion of the VSD is that of *extrinsic final value* (Poel & Kroes, 2014, p. 107). Building on the work of others who have dealt with the concept of value, they claim, very convincingly, that the concept of value has at least two axes, so that it is possible to see four kinds of value. First, following Moore, they state that there is something that could be called "intrinsic value". Intrinsic value is that kind of value that, if something has it, exhibits a kind of goodness under all circumstances, that is, it is good regardless of any other consideration of what could happen in the world. It would be valuable for its *own sake* (Poel & Kroes, 2014, p. 105-106).

One way to understand this idea of "good for its own sake" is that they are good regardless of their effects (Poel & Kroes, 2014, p. 106). However, there is also another class of things that we say that is good not for its own sake, but rather for what it can *cause*. These things are valuable *because* they can cause another state of affairs or something else. The goodness ascribed to these things is not to be found *in* them, that is, it is not intrinsic to them. They are only good because we can achieve something else that we want with their assistance. These things are *instrumental* for us to acquire what we want. That is why this kind of goodness was called "instrumental goodness" (Poel & Kroes, 2014, p. 107).

However, as Kant showed us, this idea of good for its own sake is too limited to be the basis of intrinsic value. To exhibit intrinsic value, not only should something be good for its own sake, but this goodness must be *unconditional* (Poel & Kroes, 2014, p. 106). This means that it is good independently of its *relationship* with anything else, including the moral agents. It is not good because of X or Y, it is simply good. There is no further reason. The fact that this good is unconditional made it the perfect basis for ethics. Kant argued that the moral obligation not to lie, for instance, should be respected even in cases where not lying could lead to a tragic outcome (Kant, 1996). This outcome would not remove, or even undermine, its goodness.

On the other hand, some things are conditionally good, even if some of them may be good for their own sake. In this case, what makes something good is its relationship with something or someone else. One could say that instrumental goodness is one kind of conditional goodness that exists, but it is not the only one. Kant claimed, for instance, that happiness falls under this kind of goodness, since, even if it is not to achieve something else, that is, even if it is for its own sake, one's happiness is only good in relation to her, and no one else (Kant, 1997, p. 9-10). In the case of conditional goodness, the foundation of goodness lies *outside* of the thing that is good. That is why Poel and Kroes called this kind of goodness *extrinsic*, rather than intrinsic (2014, p. 107).

There are, therefore, two ways of considering goodness in relation to the object in which it is found. First, whether it is grounded *in* the object or not, whether it is *intrinsic* or *extrinsic*. Second, whether it is good for what it can cause or for its own sake, whether it is *instrumental* or *final*. According to Poel and Kroes, when someone says that an artifact embodies a value, she is talking about extrinsic final goodness (2014, p. 107). Artifacts do not need to exhibit intrinsic value in order to embody some kind of value, as Pitt claimed they did.

These two axes allow us to make some important and clarifying distinctions. First, even if an artifact does not have any intrinsic value, it could still embody some kind of value that is pursued for its own sake, that is, it can have some final value. Second, the notion of final value can be understood through other notions, such as those of goal, function, and use. It is the possibility of explaining the existence of a final value through these notions that allows us to put forward an approach of VSD that does not fall for Pitt's too many values argument.

The distinction between the goal and the use of an artifact also reveals distinctions related to its design and production, and to the achievement of the goal through its use. The first step in the design process is the thought of: "which goal do we want this artifact to achieve?". Poel and Kroes called this "*intended value*" (Poel & Kroes, 2014, p. 119-121). Artifacts designers want their artifacts to achieve a specific goal. To achieve this specific goal, the designers must consider which properties the artifacts they want to create should have. Together, these properties constitute the function of the artifact. Poel and Kroes say:

> *The designed properties of a technical artifact* x *form the resultance base of an extrinsic final value* G *if the following two conditions are met*:
> 1. *The designed properties of* x *have the potential to achieve or contribute to* G *(under appropriate circumstances)*
> 2. *x has been designed for* G
> [...] for F to be the function of a technical artifact x, it is minimally required that (1) F was intended by the designers to be the function of x, i.e. that the designers purposively designed x for F and (2) x has the capacity to realize F in the appropriate circumstances. These conditions entail the above mentioned conditions if G is part of, or identical to, F (2014, p. 118).

Following Poel and Kroes, we can say, for instance, that if the goal of a knife is to cut, then its function is to cut.

We can see that Poel and Kroes regard goals as if they were always morally laden. They say that every artifact is designed for an extrinsic final value. However, this is not always true. The goals of artifacts can be evaluated as morally neutral or as morally relevant. They agree with this point (Poel & Kroes, 2014, p. 115). Some artifacts have some goals deemed to be morally good for our social environment. For instance, speed bumps are designed to slow cars down by making them drive slowly over them. This is their goal, and they are made to achieve it. This goal of slowing down cars does not come without thought, though. Something made for slowing down cars *simpliciter* would be very strange. However, when we think that speed bumps are normally placed in residential areas or areas with a high probability of car crashes, we realize that there is perhaps a deeper goal connected with the goal of speed bumps, the goal of *traffic safety*. Speed bumps were created for the people's safety.

Likewise, we can morally criticize artifacts when they are designed to harm people's lives. In general, different kinds of weapons can be criticized in this way, especially those that are highly destructive, such as assault rifles or WMDs. No one can truly say that these weapons were created for any good goal. Even if it is argued that WMDs are necessary as a method of deterrence, they are only needed as such because other threatening sides already have WMDs to coerce or enforce their demands. Some artifacts cannot even be meaningfully defended without provoking a deep sense of disgust, such as the gas chambers in Nazi Germany.

We can see, then, that goals can be *directly* good or bad when they are designed to achieve goals that are considered good or bad. However, there is another way of considering the morality of artifacts that goes beyond the goals they are intended to achieve. As we noted above, there is a distinction between the goal and the use of artifacts. Artifacts designed for one goal can be used for another. Thus, even if it was designed for some good goal, it can be used with bad intentions.

Furthermore, artifacts that are supposed to have morally good or at least morally neutral goals can be designed in such a way that they can have morally bad side effects. The idea of facial recognition software, for instance, can be made with the intention that it would be good for society. However, as stated above, many of these software are designed in such a way that they exhibit biases in their face recognition.

That is why the goal is not the only aspect that should be morally evaluated in artifacts. The design process should too. When designers are not careful or transparent about how they create their artifacts, they can create morally bad artifacts, even with the best intentions. Creating artifacts without carefully supervising all stages of development, without considering their possible effects on society, can be morally objectionable, because the artifacts they designed could be truly bad. Some say, for instance, that AI is being developed too quickly, without any concern as to whether it will have positive or negative effects on society.

Given these possibilities, Poel and Kroes thought of the VSD as containing three distinct steps and feedback relationships for creating morally good artifacts and for diagnosing the bad ones. The first step is the *intended value*. In the design process, designers must consider the goals of the artifact they want to create, and how they can achieve these goals. However, it is not enough to design and create an artifact with the intention that it should achieve a specific goal. They also have to consider how to create these artifacts so that their design can actually achieve their intended goals. This consideration is what they called "*embodied value*" (Poel & Kroes, 2014, p. 119-121). Artifacts should not just be made to achieve a goal deemed to be morally good, they *must* achieve this goal. Without this, there are only good intentions in the design of the artifact. Also, they must consider the possible negative effects of artifacts for value to be adequately embodied in them.

However, after the design process and creation of an artifact, designers can also find out that the artifact they designed can be misused, even though it can still be used as normally intended to achieve its goal. This is another possibility that designers must take into account in their design process. Poel and Kroes called it "*realized value*" (2014, p. 119-121). The realized value is the consideration of whether the artifact can be used *only* for the goal it was thought to achieve. This is, in a sense, the final step in the case the artifact achieves the goal for which it was created.

Artifacts can, then, not only have unforeseen negative effects but they can also be used for purposes other than those for which they were created. In these scenarios, there must be a return to the design process to correct the production of the artifact. These are feedback processes inherent to the entire design and creation processes (Poel & Kroes, 2014, p. 120-121). Designers have the moral responsibility to release their products only after this process, when their artifacts are designed in such a way that they can achieve only the goals for which they were created, that is, they achieve the goals for which they were created *and* they *cannot* be used in another way (Poel & Kroes, 2014, p. 120-121; Poel, 2001;

Koepsell, 2010; Morrow, 2014, p. 341-342).

Of course, the approach of the VSD advocated by Poel and Kroes, which we believe to be the most defensible, does not postulate that all artifacts ought to be created with the intention that they must achieve some goodness or represent some value. Rather, it only requires that artifacts be designed in such a way that they do not cause morally bad outcomes. We should not think, for instance, that cell phones must *always* serve some moral goal. After all, they were created for communication purposes, and one can communicate with other people for a variety of reasons, from moral to neutral ones. Cell phones can be used as a means of saving people when they become part of a communication center with paramedics. But they normally do not serve any moral goal when people use them just to talk to each other. On the other hand, cell phones can also be used for immoral goals when someone uses them, for instance, as a means of communication to commit bank robberies. It is impossible to prevent this kind of use, but designers can create ways to make it more difficult it when they allow cell phones to be tracked by the police. Even though cell phones must be protected against hacking by others, there should be legal ways for authorities to have access to them in order to prevent crimes. A fully access-proof cell phone can protect the privacy of people who use it, but it will also allow free communication between those who want to commit crimes. To avoid this, a cell phone should not be completely access-proof. Designers must think about the privacy of people who use cell phones legally, but also think that official authorities must be able to prevent and search for those who commit crimes.

Conclusion: "People kill people, and guns kill people too"

We have shown that Pitt's arguments for the value neutrality of artifacts are not convincing. They are perhaps the best defense of the value neutrality of artifacts, but he ignores numerous possibilities for understanding how morality and artifacts could be connected. Furthermore, he bases his argument on a very strict comprehension of morality, one that assumes that all values should be objective in a very strong sense. According to him, moral values must be real, and we must be able to identify them through sense perception. However, this is not the only way to understand morality.

Pitt also ignores how the artifact design process is carried out. The postulation that there are "too many values" involved in it is the result of a conceptual confusion. The people involved in creating artifacts may have different reasons for their participation in their conception and creation. However, it does not mean that the artifact will embody different values. His inability to distinguish between goal, use, and function is at the heart of the problem here.

It should be noted that the failure of his arguments is not in itself an argument for accepting the possibility of value-ladenness of *all* artifacts. What Pitt does not seem to understand is that when the value neutrality of artifacts is denied, and that artifacts can embody values or be judged morally, it does not entail that artifacts *have* values, or they should be morally blamed *simpliciter*, but that they can contribute to bringing about some state of affairs in society.

Those who support value neutrality in its strongest version make the mistake of assuming that those who deny it argue that *all* artifacts are value-laden. In this sense, the VNT states that if something is an artifact, it is *necessarily* value-neutral. This is not correct, though. The only assumption they make is that some artifacts, in some contexts, can be morally criticized. Against the VNT, they simply assert that it is *possible* for an artifact to be value-laden.

A point that could also be made is that sometimes, when one argues for the neutrality of artifacts, it seems that they do not want to argue that they are neutral, but that they are actually good, even if they have moral reservations to say so. Typically, those who say that "guns don't kill people, people kill people" seem to be gun enthusiasts. They really see something good in weapons, and do not consider them merely neutral.

We also argued that the best way to analyze the values of artifacts and how to create artifacts that are morally good, or at least that avoid morally negative outcomes, is through the VSD thesis. The VSD does not imply that all artifacts have moral consequences, it only states that artifacts can embody values and have moral consequences. Its focus on the design process makes it possible not only to evaluate already created artifacts morally, but also provides designers with a way to think morally about their artifacts and what they should consider to be relevant when they create them.

The VSD approach advocated by Poel and Kroes allows us to structurally consider the notions of goal, function, and use in the artifact design process. It also fits very easily with the criticism of Pitt's arguments. It also allows for moral considerations beyond the strong moral realism suggested by Pitt as the basis for the values embedded in artifacts. We only made a few points about the VSD. A stronger foundation for it is still needed. However, it was enough to show how the notion of artifact neutrality is, perhaps, a more ideological position than Pitt would like it to be.

Resumo: O presente artigo critica a tese da neutralidade de valores morais de artefatos, e defende uma proposta conhecida como Value Sensitive Design, que afirma que valores morais devem ser considerados na construção e análise de artefatos. Primeiro, (1) apresentaremos a melhor defesa da tese da neutralidade de valores morais de artefatos, realizada por Joseph Pitt. Em seguida, (2) criticaremos cada um dos argumentos apresentados por Pitt para a defesa da tese da neutralidade. Por fim, (3) consideraremos a proposta do Value Sensitive Design apresentada por Ibo van de Poel e Peter Kroes e explicar como ela seria adequada para uma crítica dos valores e questões morais que artefatos possam representar.

Palavras-chave: Filosofia da tecnologia, Análise moral de artefatos, Neutralidade de valores dos artefatos, Value Sensitive Design.

References

BRATMAN, Michael E. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press, 1987.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, v. 81, p. 1-15, 2018.

CUMMINGS, Mary L. Integrating ethics in design through the value-sensitive design approach. *Science and Engineering*

Ethics, v. 12, n. 4, p. 701-715, 2006.

DAVIDSON, Donald. Actions, reasons, and causes. In: DAVIDSON, Donald. *Essays on Actions and Events*. 2. ed. Oxford: Oxford University Press, 2001. p. 3-20.

FISHER, Andrew. *Metaethics: An introduction*. Durham: Acumen, 2011.

FRANSSEN, Martin. Design, use, and the physical and intentional aspects of technical artifacts. In: VERMAAS, Pieter. E., et al. *Philosophy and Design: From engineering to architecture*. Dordrecht: Springer Science+Business Media, 2008. p. 21-36.

FRANSSEN, Martin. Artefacts and normativity. In: MEIJERS, Anthonie. *Philosophy of Technology and Engineering Sciences*. Amsterdam: Elsevier, v. 9, 2009. p. 923-952.

FRIEDMAN, Batya. Value-sensitive design. *Interactions*, v. 3, n. 6, p. 16-23, 1996.

FRIEDMAN, Batya; HENDRY, David G. *Value Sensitive Design: Shaping technology with moral imagination*. Cambridge: The MIT Press, 2019.

FRIEDMAN, Batya; KAHN, Peter H. Human values, ethics, and design. In: SEARS, Andrew; JACKO, Julie A. *Human Computer Interaction Handbook: Fundamentals, evolving technologies, and emerging applications*. 2. ed. New York: Lawrence Erlbaum Associates, 2008. p. 1241-1266.

HEYNDELS, Sybren. Technology and neutrality. *Philosophy* *and Technology*, v. 36, n. 4, p. 1-22, 2023.

HOVEN, Jeroen van den; MANDERS-HUITS, Noemi. Value-sensitive design. In: OLSEN, Jan Kyrre Berg; PEDERSEN, Stig Andur; HENDRICKS, Vincent F. A *Companion to the Philosophy of Technology*. New York: Blackwell Publishing, 2009. p. 477-480.

HURKA, Thomas. Moore's moral philosophy. *The Stanford Encyclopedia of Philosophy*, 2021. Disponivel em: <https://plato.stanford.edu/archives/sum2021/entries/moore-moral/>. Acesso em: 26 fev. 2024.

JOHNSON, Natasha N.; JOHNSON, Thaddeus L. Police facial recognition technology can't tell black people apart: AI-powered facial recognition will lead to increased racial profiling. *Scientific American*, 2023. Disponivel em: <https://www.scientificamerican.com/article/police-facial-recognition-technology-cant-tell-black-people-apart/>. Acesso em: 26 fev. 2024.

KANT, Immanuel. On a supposed right to lie from philanthropy. In: KANT, Immanuel; GREGOR, Mary. *Practical Philosophy*. New York: Cambridge University Press, 1996. p. 605-616.

KANT, Immanuel; GREGOR, Mary. *Groundwork of the Metaphysics of Morais*. New York: Cambridge University Press, 1997.

KATZ, Eric. On the neutrality of technology: the Holocaust death camps as a counter-example. *Journal of Genocide*

Research, v. 7, n. 3, p. 409-421, 2005.

KOEPSELL, David. On genies and bottles: scientists' moral responsibility and dangerous technology R&D. *Science Engineering Ethics*, v. 16, n. 1, p. 119-133, 2010.

KORSGAARD, Christine M. Realism and constructivism in twentieth-century moral philosophy. In: KORSGAARD, Christine M. *The Constitution of Agency: Essays on practical reason and moral psychology*. New York: Oxford University Press, 2008. p. 302-326.

KROES, Peter. Moral values in technical artifacts. In: GARNAR, Andrew Wells; SHEW, Ashley. *Feedback Loops: Pragmatism about science and technology*. Lanham: Lexington Books, 2020. p. 127-140.

KROES, Peter; MEIJERS, Anthonie. The dual nature of technical artefacts. *Studies in History and Philosophy of Science*, v. 37, n. 1, p. 1-4, 2006.

MACKIE, John L. *Ethics: Inventing right and wrong*. London: Penguin Books, 1977.

MANKIW, N. Gregory. *Principles of Economics*. 10. ed. Boston: Cengage, 2024.

MILLER, Alexander. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press, 2003.

MILLER, Boaz. Is technology value-neutral. *Science, Technology, & Human Values*, v. 46, n. 1, p. 53-80, 2021.

MOORE, G. E. *Principia Ethica*. 2ª Revised. ed. Cambridge: Cambridge University Press, 1993.

MORROW, David R. When technologies makes good people do bad things: another argument against the value-neutrality of technologies. *Science and Engineering Ethics*, v. 20, n. 2, p. 329-343, 2014.

NAJIBI, Alex. Racial discrimination in face recognition technology. *Science in the News*, 2020. Disponivel em: <https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>. Acesso em: 26 fev. 2024.

ORESKES, Naomi; CONWAY, Erik M. *Merchants of Doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. New York: Bloomsbury Press, 2010.

PITT, Joseph C. "Guns don't kill, people kill"; values in and/or around technologies. In: KROES, Peter; VERBEEK, Peter-Paul. *The Moral Status of Technical Artefacts*. Dordrecht: Springer, 2014. p. 89-101.

PITT, Joseph C. Value-free technology? In: ROBSON, Gregory J.; TSOU, Jonathan Y. *Technology Ethics: a philosophical introduction and readings*. New York: Routledge, 2023. p. 14-17.

POEL, Ibo van de. Investigating ethical issues in engineering design. *Science and Engineering Ethics*, v. 7, n. 3, p. 429-446, 2001.

POEL, Ibo van de; KROES, Peter. Can technology embody values? In: KROES, Peter; VERBEEK, Peter-Paul. *The Moral Status of Technical Artefacts*. Dordrecht: Springer Science+Business Media, 2014. p. 103-124.

RAWLS, John. Kantian constructivism in moral theory. *The Journal of Philosophy*, v. 77, n. 9, p. 515-572, 1980.

STIGLITZ, Joseph E.; WALSH, Carl E. *Economics*. 4. ed. New York: W. W. Norton & Company, 2006.

VERBEEK, Peter-Paul. Materializing morality design ethics and technological mediation. *Science, Technology, & Human Values*, v. 31, n. 3, p. 361-380, 2006.

VERBEEK, Peter-Paul. Morality in design: design ethics and the morality of technological artifacts. In: VERMAAS, Pieter E., et al. *Philosophy and Design: From engineering to architecture*. Dordrecht: Springer Science+Business Media, 2008. p. 91-104.

WEINBERG, Steven. Reflections of a working scientist. In: TEICH, Albert. H. *Technology and Man's Future*. 2. ed. New York: St. Martin Press, 1977. p. 41-58.

WHELCHEL, Robert J. Is technology neutral? *IEEE Technology and Society Magazine*, v. 5, n. 4, p. 3-8, 1986.

WINNER, Langdon. *The Whale and the Reactor: A search for limits in an age of high technology*. Chicago: The University of Chicago Press, 1986.