BOOK SYMPOSIUM

Confronting the Moral Dimensions of Technology Through Mediation Theory

Evan Selinger

1 Playing Philosophical Pictionary with Verbeek

Martin Heidegger famously claimed that great thinkers spend their lives exploring a single thought: its history nuances, misappropriations, and implications. While not as narrowly—or, in my opinion, myopically—focused, most contemporary principals in the philosophy of technology pursue recognizable research programs. Since these programs are distinctive, peers and graduate students can associate complex arguments with leading concepts. Such concepts circulate widely enough to become common terms in database searches, and informatics scholars in principle can use them as tools for determining and visually depicting trends that exemplify a field's central preoccupations. Ultimately, the terms become so resonant that they can be the main pieces in a philosophically adapted game of Pictionary. After writing down a qualifying term, teams could compete to give the most robust account of the ideas it designates.

For example, to evoke the "fourth revolution" is to conjure Luciano Floridi's views on how the social integration of new technologies can alter experiences and conceptions of reality. Likewise, to talk about "focal practices" is to bring to mind Albert Borgmann's understanding of how a special relation to technology can support rather than thwart the pursuit of the good life. Similarly, to mention "critical theory of technology" is to invite comments on Andrew Feenberg's ideas about increasing participation in decision-making processes concerning the design and development of technology. On rare occasions, distinctive concepts belong to two metaphorical copyright holders: Asking if there is a "cyborg" in the house can invite Andy Clark and Donna Harraway impersonations.

Dutch philosopher Peter-Paul Verbeek belongs to this group of Pictionary scholars. Although one of the youngest principals, he already is Professor of Philosophy of Technology at the University of Twente, part-time Socrates Chair at Delft University of Technology, member of the Young Academy (which is part of

E. Selinger (✉)
Department Philosophy, Rochester Institute of Technology, Rochester, NY, USA
e-mail: evan.selinger@rit.edu

the Royal Netherlands Academy of Arts and Sciences), recipient of the prestigious VENI and VIDI grants (from the Netherlands Organisation for Scientific Research), Editor of *Technè: Research in Philosophy and Technology*, and member of the Scientific Editorial Board of *Philosophy & Technology*. To bring up the topic of "technological mediation" is to stimulate conversation about Verbeek's "postphenomenological" and "posthumanist" analyses of technology, i.e., his sense, to be described in more detail below, of the profound ways that technology influences the forms human practices exhibit and the correlative interpretations of the world that arise from within technologically oriented activities.

2 Mediation: Linking What Things Can and Should Do

If one accepts Immanuel Kant's reasons for organizing his systematic approach to philosophy by following up the *Critique of Pure Reason*, which aspires to clarify the foundation of knowledge, with the *Critique of Practical Reason*, which aspires to clarify the foundation of ethics, then Verbeek's new monograph, *Moralizing Technology: Understanding and Designing the Morality of Things* (2011), can be construed as the logical sequel to his first book, *What Things Do: Philosophical Reflections on Technology, Agency, and Design* (2005). As a phenomenological investigation, *What Things Do* is not an orthodox epistemological treatise. Nevertheless, it clarifies fundamental ways in which the knowing, perceiving, and action-oriented subject's relation to the world and others is influenced by technological mediation (which was alluded to above and is described more below). *Moralizing Technology* builds directly upon the antecedent work, treating it as foundational—as capable of being supplemented and expanded, but not significantly revised. In the new text, Verbeek's basic task is to analyze how technological mediation affects moral reasoning and gives rise to new possibilities for moral action, only some of which are intended by the designers of technologies and technical systems.

The analogy posited here between Verbeek and Kant is intended only to be superficially illuminating. Whereas Kant's uses transcendental analysis to instill a Copernican Revolution, in *What Things Do*, Verbeek offers a prospective research vision in which philosophers contribute to shaping technological development through a division of labor in which their primary responsibility is to identify the behavior-modifying trajectories embedded in artifacts and technical systems. Slightly altering a term popularized by his former teacher, the eminent philosopher and public intellectual Hans Achterhuis—who modified Richard Rorty's phrase "linguistic turn" and spoke of an "empirical turn" in philosophy of technology— Verbeek initially characterizes his approach as a "thingly turn" (2005, p. 3). In *Moralizing Technology*, this idiosyncratic parlance is dropped for the more common empirical turn idiom, and he attempts to defend the value of two new meta-philosophical orientations: (1) an "ethical turn" that clarifies the causal and normative implications that follow from thinking about concepts like "moral agency," "moral subjectivity," and the "moral self" from the perspective of technological mediation, and (2) a "third turn" of philosophical orientation that putatively does not yet exist, but can be gestured to, in outline form. Ostensibly, such inquiry can be constituted as an "ethics of accompanying technological development" in which philosophers collaborate with

designers, engineers, scientists, and various members of the public in real time by helping invent the "vocabularies and practices" that can allow innovation to be maximally transparent and just. Both paradigm shifts are provocative and complex, and the symposium discussants can address them in detail.

3 Scripting Behavior Through Gracefully Aging Design

In both *What Things Do* and *Moralizing Technology*, mediation theory is exposited primarily through appropriation of phenomenological and actor network concepts, but divorced from naturalist enterprises, such as cognitive science, behavioral economics, and evolutionary biology. Since Verbeek does not explicitly lament "scientism" as many of his phenomenological predecessors did, it is reasonable to infer that occlusion of insights from the sciences of the mind and human nature is the result of contingent analytic preferences, not principled skepticism concerning the limits of reductionism. Although, endorsing Michel Foucault's historically conditioned conception of subjectivity in *Moralizing Technology* invites the question as to whether Verbeek believes that universal and invariant features of the mind exist, such as cognitive bias, cognitive flexibility, and extended cognition, that are causally efficacious in ways that go beyond the social conditioning imposed by disciplinary forces.

In both *What Things Do* and *Moralizing Technology*, Verbeek treats a special couch as a paradigm case of mediation theory applied to industrial design. *Summarizing Verbeek's original analysis thus can go a long way towards illuminating, in phenomenological fashion, the basic idea of technological mediation*.

Dutch design firm Eternally Yours aimed for an innovative solution to the problem of waste as it applies to the context of environmentally unfriendly consumer behavior. Recognizing that "most of our products are thrown away far before actually being worn out," the firm focused its attention on "developing ways to create product longevity" that would minimize the premature disposal typical of conspicuous consumption societies (2005, p. 373). It thus approached longevity in a manner that went beyond the business-as-usual approaches to making product repair and upgrade affordable (a key economic factor in determining a product's lifespan) and favoring durable materials (a key technical factor in determining a product's lifespan). The emphasis upon user behavior led Eternally Yours to create "gracefully aging" couches that can induce two powerful psychological responses, bonding and attachment. Verbeek writes:

[Eternally Yours] searched for forms and materials that could stimulate longevity. Materials were investigated that do not get unattractive when aging but have "quality of wear." Leather, for instance, is mostly found more beautiful when it has been used for some time, whereas a shiny polished chromium surface looks worn out with the first scratch. An interesting example of a design in this context is the upholstery of a couch that was designed by Sigrid Smits. In the velour that was used for it, a pattern was stitched that is initially invisible. When the couch has been used for a while, the pattern gradually becomes visible. Instead of aging in an unattractive way, this couch renews itself when getting old (2005, 373–374).

Verbeek's analysis of the "stimulating" power of Smits's design pivots on the concept of mediation. He sees one significant difference as marking the crucial demarcation separating ordinary couches from gracefully aging ones—namely, Eternally Yours designers "inscribe" a plan of sustainable user behavior into the latter. By "inscribe" Verbeek means that Eternally Yours made deliberate use of what Bruno Latour calls "scripts." Scripts refer those aspects of an artifact or system that "delegate" behavioral prescriptions, which Latour calls "programs of action." In this case, the following script is said to be inscribed into the Eternally Yours couch: "Bond with me. Do not throw me away, I age gracefully."

Since neither Verbeek nor Latour advocates returning to totemic thinking, the couch is not anthropomorphized as presenting a script through propositional injunction. Indeed, neither thinker would characterize the couch, or any object, as literally speaking a human language. Instead, the script's behavioral suggestions are recognizable through a mundane and prosaically describable process: embodied perception. Looking at the couch's unique design can trigger an affective response that makes its distinctive meaning apparent in a manner as direct as the immediate recognition that a smile expresses happiness. Typical Dutch customers who acquire Eternally Yours couches perceive the furniture to be endowed with a special meaning because they see the emergence of a wear-and-tear induced pattern as what ecological psychologist J.J. Gibson calls an "affordance," i.e., qualities in the object that incline users to see additional use as more attractive than disposability and to behave in a manner that accords with this feeling.

In accordance with Don Ihde's conception of "multi-stability," which posits that the meaning associated with a technology is context specific and subject to change with alterations in how and where a technology is used, Verbeek stipulates that the Eternally Yours script is only efficacious because it aligns with underlying cultural values. To concretize this point, Verbeek proposes a thought experiment involving a plastic coffee cup. In a culture that views the material to be worthless, he claims that the cup would have the script, "Throw me away." By contrast, in a culture that places high value on plastic, the cup could have a different script, perhaps "Save me."

4 Moralizing Technology: Main Issues

With mediation theory functioning as the main engine of philosophical analysis, *Moralizing Technology* raises the following questions:

- How should the moral significance of technology be conceptualized? Are the intellectual resources found in mainstream meta-ethics and engineering ethics sufficient for answering this question? What is the most justified way to go beyond the commonplace instrumentalist perspective, which restricts the moral status of technologies to the causal role they play in realizing and impeding human moral intentions?
- What conception of subjectivity is appropriate for understanding who human beings are when they inhabit a lifeworld of ubiquitous technological mediation? Does such a subject possess sufficient autonomy to qualify as a moral agent? Or, is the concept of "moral agency" in a need of rethinking so as to better accord with the

phenomenological facts captured by mediation theory analyses of technological use? Should technologies be recognized as a new category of moral agents?

- Is moral reason giving a sufficient response to the fundamental problems posed by technology? Or, is the conception of the philosopher as the preeminent producer of archive friendly texts outdated and in need of replacement by a materialist ethics of social design?
- How can mediation theory be applied to the emerging fields of ambient intelligence and persuasive technology?
- Do structures of intentionality exist that fall beyond the scope covered by mediation theory? If so, what is their significance?

To address questions like these, three prominent philosophers of technology have been selected to be symposium discussants: Don Ihde, Ibo van de Poel, and Martin Peterson. Ihde is Distinguished Professor of Philosophy at Stony Brook University and is one of the most recognizable thinkers in the field.[1] His early contributions to philosophy of technology emphasized the embodiment of scientific practice through theory constraining instrumentation. More recently, he developed a postphenomenological approach to philosophy that highlights multistable pathways towards technological development. As already acknowledged, to a significant extent, Verbeek's theory of mediation builds upon phenomenological concepts and ideas that Ihde proposed while studying perception and action. Van de Poel, an Associate Professor in Ethics and Technology in Delft University, is widely regarded as a leading international authority on design ethics and the dynamics of technological development. He recently proposed an innovative approach for addressing the social consequences of new and emerging technologies. Distinguishing himself from theorists who offer predictive models calibrated to try to anticipate foreseeable consequences, Van de Poel argues that the introduction of disruptive technology should be analyzed as a form of "social experimentation" that may very well exist outside the bounds of informed consent. To further this research, he received a prestigious VICI grant from the Netherlands Organization of Scientific Research. Peterson is an Associate Professor of Philosophy at Eindhoven University of Technology. He has written extensively on the philosophy of risk, showing, among other things, that the precautionary principle is incompatible with a number of basic rationality postulates. Peterson also has written several articles on the ethics of cost–benefit analysis and discussed the possibility of human friendship with non-human entities from an Aristotelian perspective.

References

Verbeek, P. P. (2005). *What things do. philosophical reflections on technology, agency, and design* (R. P. Crease, trans.). Penn State: Penn State University Press.
Verbeek, P. P. (2011). *Moralizing technology: understanding and designing the morality of things*. Chicago: University of Chicago Press.

---

[1] All participants reviewed a .pdf copy of the manuscript before the text was published. Their references refer to this original copy.

Peter Paul Verbeek and a Postphenomenological Mirror

Don Ihde
Dept. Philosophy
Stony Brook University
dihde@notes.cc.sunysb.edu

Moralizing Technology is Peter-Paul Verbeek's second major original book in the philosophy of technology. Verbeek, already a full professor (b. 1970) at Twente University, widely visible in both Europe and North America, is a major contender in this field.

Verbeek's signature is also one which identifies him with *postphenomenological analysis*. His dissertation, later to be revised and expanded as *What Things Do* (2005), his chapter in Hans Achterhuis's *American Philosophy of Technology: The Empirical Turn* (2001), had already dealt with postphenomenology, often including my formulation thereof. It is for that reason that I have titled this response with a mirror metaphor—our visions of postphenomenology mostly converge, and this poses a problem for my analysis. First, it is hard for me to "get distance" on his position which is so close to my own. Second, a temptation to turn to a "nitty-gritty" analysis which concentrates upon where do we differ must also be resisted as this trivializes what would be important. Third, and worst of all, to end up simply praising Peter Paul for his conclusions where they are close to mine, becomes an indirect form of self-flattery.

Here, the mirror metaphor may help: I have for several decades been interested in *imaging technologies*, in which mirrors play important and diverse roles. Mirrors are "isomorphic" optics in their most ordinary forms, reflecting a "likeness" to the viewer—but not without serious transformations. To my knowledge, there are no perfect, non-neutral isomorphic optics, even mirrors. A very simple phenomenology of mirror-viewing shows this: I stand before the vanity mirror in the bathroom. Already experienced in mirror viewing, I immediately recognize "myself" in the mirror. But, if I am about to shave—as an experienced viewer, I also may no longer explicitly notice this—since the mirror does not present me with what would be a *bodily* face-to-face image, but rather presents me with an image in which right side to right side and left side to left side faces me, a *mirror*, not bodily, image. Thus I have previously had to learn the bodily movements to put the shaving cream on my cheeks by adjusting to this right–left "reversal" as it were. Here we have discovered one of the transformations of mirror optics. Returning now to the mirror, perhaps we do *not note* that a symmetrical up-and-down reversal *also occurs*. [The symmetry remains analogous to the right/left phenomenon insofar as top to top and bottom to bottom remains symmetric. But, I suggest, does the up/down bodily reflection seem less obvious than the right/left reversal?] In short, in both vectors, the mirror does *not* present us with what would be a bodily face to face. I hope to use this phenomenon to at least indirectly get at a non-mirror perspective on Verbeek's book.

5 Verbeek and Postphenomenology

I now return to *Moralizing Technology* which highlights the unique situation Verbeek addresses. This context includes a *historical–geographical* subtheme. As many have

noted—Verbeek included—if one looks at the mostly 20th century origins of a philosophy of technology, most of the forefathers were European philosophers. With a few 19th century predecessors such as Karl Marx and Ernst Kapp, most major thematizations of technology only begin to occur in the 20th century. Friedrich Dessauer, Martin Heidegger, Karl Jaspers, Ortega y Gassett, and others began to realize that the impact—largely since the Industrial Revolution—of technology was having a massive transformational affect upon European culture. Verbeek's *What Things Do* (2005) deals with this set of origins as well. [My own recent, *Heidegger's Technologies: Postphenomenological Perspectives* (2010) continues this discussion.] But these beginnings change in the later 20th century. As most abbreviated histories agree, the often dystopian, usually transcendental originary generalizers of the early to mid-20th century gave way to several transition generations, including philosophers from the Frankfurt School, pragmatism, and phenomenology, and then on to students of those generations who are today dominant. The Achterhuis recognition of an "empirical turn" from technology overall to particular technologies played an important recognition of how philosophy of technology had begun to look quite different by the late 20th century in both Europe and North America. But whereas Achterhuis took this transformation from American philosophers of technology, an equally important set of events was also occurring in Europe—and most explicitly in the Netherlands! I have sometimes claimed that "There are more philosophers of technology per capita in Holland than in any other country of the world!" During a 1998 Dutch lecture tour, I discussed the Dutch scene with a philosopher of science, one G. De Vries at the University of Amsterdam. I asked him what the state of philosophy of science was at that time, and his response was that no one reads it anymore since it does not address lifeworld problems. What then, is read, I queried, and his answer was science-technology studies (STS), social constructionism, and related more empirically tuned studies—"After all, given that the Netherlands was historically 'socially constructed' with its polders and continues today with the politics of Water Boards, we are all and always have been 'social constructionists.'" We are here at the doors of *Moralizing Technology* with hybrid, composite, and other human–technology combinations which are central to Verbeek's thesis. We are also at these doors because there has been a parallel institutional *geographic* shift concerning centers for philosophy of technology in recent years. While there are a number of academic societies which present the research outcomes, including philosophy of technology, for example the Society for the Social Studies of Science [4/S] is the largest STS style society, but the smaller Society for Philosophy and Technology had been the more primary vehicle for philosophy of technology. Its recent history is telling—while it meets only biennially, on alternate years, it meets outside North America—and recently, the country with the most frequent hosting record has become the Netherlands—indeed, the editors of the journal, the governing board, and society officers now include significant numbers of Dutch academicians. With this shift has also come a perceptible growth rate and conference attendance increase. In philosophy of technology, the Dutch are making their mark.

6 Moralizing Technology

Moralizing Technology reflects this background. There are two ways in which precisely this historical–geographic background may be detected. First, one of

Verbeek's very long interests has been actor network theory, the style of analysis developed by Bruno Latour and Michel Callon out of France. Although there are varieties of methods used throughout STS, actor-network theory, pioneered by Bruno Latour and Michel Callon, seems to be the dominant style of analysis favored by social scientists in Continental Europe. Verbeek holds that postphenomenology is a complementary type of analysis that can enhance this dominant approach. Second, although less obvious without careful reading, *Moralizing Technology* deals with, discusses, and argues with a high concentration of contemporary Dutch thinkers, more cited here than in Verbeek's previous works. Thus, a distinctive Dutch "tone" runs through this work. This is the case not only with thinkers, but also with many of the technological examples used to illustrate his points therein. Thus *Moralizing Technology* can also serve as an introduction to much contemporary Dutch thinking in the philosophy of technology.

I shall deal with *Moralizing Technology* in three parts; the first places Verbeek at the forefront of a state-of-the-art philosophy of technology. Verbeek accepts and continues the generational shift which Hans Achterhuis claims as the "empirical turn" from earlier dystopian transcendentalism to later examinations of particular and actual technologies. Verbeek, however, calls for a parallel turn—*an ethical turn*—which focuses upon the normative dimensions of technologies, thus "moralization." The first four chapters of the book undertake this task with aplomb and result in an excellent analysis of the current state of affairs in philosophy of technology.

Entailed in the empirical turn, there were two significant moves away from classical philosophy of technology: First, by turning to specific, actual technologies, empirical turn philosophers had to deal with materiality, both the materiality of technological artifacts and the materiality of technological culture. It was not accidental that the groupings of philosophers who took this turn were frequently from the *praxis traditions* of philosophy. Without here dealing in detail with the important moments in the shift to *praxis*, one can note that Heidegger's early inversion of the primacy of science for the primacy of technology played an early role in what later became the empirical turn. This inversion has recently been recognized by Paul Forman (2007) who sees science primacy as modernist; technology primacy, postmodern. Verbeek picks up and analyzes materiality through Langdon Winner's "Do Artifacts have Politics?" with numbers of Latour's socialized artifacts illustrations (speed bumps, door stoppers, etc.) and, of course, with his and my postphenomenological approach which incorporates material artifacts into human–world, human–technology–world relations along with our varieties of "instrumental intentionalities."

What I find in common in the empirical turn generation is not only the incorporation of a philosophical sensitivity to materiality, but also a framework of *inter-relational ontologies*. Verbeek sees these in Latour, postphenomenology, Winner, and I would add with somewhat more emphasis, also pragmatism, Donna Haraway and Andrew Pickering, as well as some post-humanist postmodernists. Amongst all of this grouping, there emerge descriptive analyses which deny "purities" and affirm "hybrids, composites, boundary-bending mixtures." Now while I also have used this style of description, I have to admit that I am not entirely happy with what appears to me to be something of a transitional, compromise result. What is positive in this consensual emergence is the

recognition—very nicely described in *Moralizing Technology*—that in the ways we experience our lifeworld today, myriads of technologies play roles in our experience of our world, others, and ourselves. These must be recognized and understood in their mediational roles which are also inclusive of our moral and ethical concerns and formations. *Moralizing Technology* is rich with examples in which pre-natal ultrasound scans play a repeat role. The technology that produces fetal images, Verbeek shows, also makes a whole range of new human–fetal relations possible. To visually display the fetus—Verbeek recognizes the transformations which any imaging produces—is already to enhance its role in the parental–fetus set of relations. One outcome is the widening of the range of decidability for parents in relation to the fetus, for example to abort a physically flawed fetus, to deciding *not* to undertake repeat images, a variety of changes in parent–fetus and future parent–child relations takes shape. This example is particularly appropriate for Verbeek's call for the third turn, ethics, precisely because this new set of decisional variations is already intrinsically ethical.

What complicates empirical and ethical turn dimensions is the contemporary milieu in which *modernity* is increasingly seen as a terminal background. Here Verbeek enters the conversation which had already begun with late 20th century Haraway, Hayles, Latour, and others, amongst whom Verbeek must locate himself. Verbeek's readers will recognize that he wants to develop an *amodern* ethics, one which is *post-humanistic*, but not *transhuman*, and an ethics which allows for *freedom* but without the earlier overemphasis upon *autonomy* which modernist thought proclaimed. Verbeek tends to emphasize "posthumanism" or non-humanistic ethics. I can only point to a few of the unique tactics Verbeek employs in this self-location: He does an excellent job showing how the two dominant versions of ethical theory—deontologies with Kantian roots and utilitarianisms with British roots—both remain subject/object modernist which continues to keep sociality and technology distinct. Two original interpretations stand out: Peter Sloterdijk pushed non-humanism away from letters and towards a bodily emphasis. Sloterdijk inverts Heidegger, claims Verbeek, by re-affirming the bodily aspect of being human. It is a subtle reversal of the "linguistic turn" while at the same time facing the biotechnological revolution now occurring. A similar innovative interpretation would reclaim the late Michel Foucault concerning ethics and *freedom*. What Verbeek sees in the *History of Sexuality* Foucault is effectively the recognition of ontological inter-relationality between self-construction vis-à-vis systems of power. Drawing also from Gerard DeVries, this is also an approximation to a neo-virtue ethics which is only gestured toward in the concluding chapter of *Moralizing Technology*.

My second look takes account of Chap. 5 and 6 which deal with designing mediations and the smart technologies of ambient intelligence and persuasive technologies. In short, once one recognizes the inter-relationality of technologies and the human, both individual and social, Verbeek's question becomes: can one *design* to some degree moralities into artifacts? There is no doubt here that Verbeek is right up to date on the often still presumptive technologies in question. My own reading suggests that perhaps he comes too close to "believing" the presumptions which on my part I am not yet convinced that such technoutopian fantasies are likely to materialize. But, the real issue is simpler: precisely by recognizing the multistabilities of all technologies, the ambiguities of designer intent gone astray, and the

sheer complexity of the mix could there be any gestalt simple enough to serve as a guide for such designed positive features—I have my doubts. This leaves muddling through….

My third look at Chap. 7 and 8 takes account of the enigma which *Moralizing Technologies* must deal with: if the dominant philosophical ethics, as Verbeek has so convincingly shown, remain caught in modernist, humanist, and thus subjectivist frames, and at least implicitly, if there are no options of returning to any pre-modern ethics, what can a new ethical turn emerge? Verbeek wisely does not offer us a formula. Rather, he offers us hints. As I have suggested, one alternative is a modernized (or postmodernized) set of variants upon virtue ethics. Here the question becomes what kind of person do I want to be? What style of values do I wish to embody? All of this in the context of the complex set of human–technology mediations. On the North American scene, I suppose the two philosophers who have discussed this include Alisdair MacIntyre and Martha Nussbaum, mentioned here but overshadowed by Gerard DeVries and Steven Dorrestijn.

Moralizing Technology remains a tour de force. The third (and fourth?) ethical turn is both appropriate and timely. Verbeek makes a major mark with this book.

References

Achterhuis, H. (2001). *American philosophy of technology: the empirical turn*. Bloomington: Indiana University Press.

Forman, P. (2007). The primacy of science in modernity, of technology in postmodernity, and of ideology in the history of technology. *History and Technology* 23(1/2), 1–152.

Ihde, D. (2010). *Heidegger's technologies: postphenomenological perspectives*. New York: Fordham University Press.

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

Verbeek, P. P. (2011). *Moralizing technology: understanding and designing the morality of things*. Chicago: University of Chicago Press.

Technology as Moral Agent

Ibo van de Poel
Delft University of Technology
i.r.vandepoel@tudelft.nl

In his new book *Moralizing Technology: Understanding and Designing the Morality of Things* (2011), Peter-Paul Verbeek discusses the implications for the ethics of technology of the mediation theory that he developed in his previous book *What Things Do: Philosophical Reflections on Technology, Agency, and Design* (2005). His approach is not only original but also rich and multi-faceted, seemingly easily moving from abstract philosophical discussions about Heidegger and Foucault to the practice of engineering design and the development of ambient intelligence and persuasive technology. Part of the value of the book lays no doubt in this

combination of philosophical argument and analysis of concrete technological developments. The book also has important practical implications for the design and development of new technologies, augmenting current approaches in engineering ethics and science and technology studies. At the core of the book, however, is a philosophical argument about how current ethical theory should be adapted to do justice to the moral significance of technologies. It is this argument I will be focusing on. There is a slight danger in this focus because it is likely to stress the points in which I disagree with Verbeek rather than the points at which I agree with him. Let me therefore at the outset stress that I agree that technological mediation is an interesting and important phenomenon that requires more attention in the ethics of technology. I also largely agree with the practical implications that Verbeek sketches for the design and use of technologies. In addition, I liked his Foucauldian treatment of freedom. What I doubt, however, is whether all these things require the kind of drastic changes in ethical theory that Verbeek proposes. Surely, ethicists have tended to ignore the role of technologies in morality and ethics, and surely doing justice to their moral significance may require certain changes in ethical theory, but I am not sure whether a redefinition of moral agency is the right target for such changes. But rather than getting ahead of my argument, I should start with explaining why Verbeek thinks that doing justice to the moral significance of technology requires a new approach to ethics.

7 Why We Need a Posthumanist Ethics

The phenomenon of technological mediation draws attention to the fact that technologies actively shape the actions and perceptions of human beings. According to Verbeek, however, "this moral role of technology is hard to conceptualize" because of "the humanist orientation of the established frameworks of ethical theory" (p. 33). Therefore, "the humanist foundations of ethics need to be broadened" (ibid.). As Verbeek explains, his criticism of humanism is not directed at the values articulated by humanists, but rather at the metaphysics of humanism, "which has its roots in modernism, and its radical separation of the human subject and nonhuman objects" (ibid.). His criticism of modernism is similar to that of Latour in *We have never been modern* (1993), although Verbeek also draws on other sources like Heidegger, Slotendijk, and Foucault.

The consequence of the modernist divide between object and subject, according to Verbeek, is that "ethics now suddenly has to be located in one of the two domains. And almost automatically, that domain is the one of the subject, which asks itself from a distance how to act in the world of objects" (p. 42). However,

> Technologies appear to be able to "act" in the human world, albeit in a different way than humans do. By doing so, technologies painlessly cross the modernist border between subject and object. A humanist ethics, as Harbers put it, is founded on a "human monopoly on agency" (Harbers 2005, p. 259). Because of this, it is not able to see the moral dimension of artifacts, which causes it to overlook an essential part of moral reality. (p. 45)

What is needed then is a posthumanist ethics in which some form of moral agency is attributed to technology.

8 Moral Agency

The notion of moral agency brings us to the core of the posthumanist ethics that Verbeek tries to articulate to do justice to the moral roles of certain technologies. He sets out to redefine moral agency "in a direction that makes it possible to investigate the moral relevance of technological artifacts in ethical theory" (p. 57). He basically employs two argumentative strategies to do so. The first strategy is to argue that neither technological artifacts nor humans in isolation possess moral agency but that agency is rather a property or characteristic of *ensembles* of humans and technology. The second strategy is to redefine the notion of moral agency so that objects possess more and humans less agency than in customary philosophical thought, while recognizing that humans and objects are not entirely the same in this respect.

The first strategy argues that neither humans nor objects exist in splendid isolation, but that they always exist in relation to each other. Humans and technological objects constitute each other, and they are the resulting ensembles, associations, or hybrids that possess moral agency. This line of reasoning sees moral agency as a *result* of the interaction between humans and technology rather than as an intrinsic property of humans as in traditional philosophy.

The second strategy is a bit more complicated. Verbeek distinguishes "two criteria that are usually seen as *condition sine qua non* for moral agency" (p. 70); these are "*intentionality*—the ability to form intentions—and the *freedom* to realize its intentions" (ibid., emphasis in original). He sets out to "show that these two criteria can be reinterpreted along postphenomenological lines in such a way that they also pertain to nonhuman entities" (ibid.).

With respect to intentionality, he admits that most technological artifacts cannot deliberately form intentions, but they nevertheless direct our actions and experiences and thus "can "have" intentionality in the literal sense of the Latin word "intendere," which means "to direct," "to direct one's course," "to direct one's mind" (p. 73). He concludes that it "seems plausible, then, to attribute a form of intentionality to artifacts—albeit a radically different form than human intentionality" (p. 74).

With respect to freedom, Verbeek starts with noting that if we take the theory of technological mediation seriously, absolute freedom cannot be a precondition for moral agency, unless we are willing to give up the idea of (human) moral agency altogether. After all, in a technologically mediated world, human beings are never completely free from external influences. He therefore proposes to understand freedom not as the absence of external constraints or powers but rather as the ability to relate oneself to those external constraints or powers. He admits that this "redefinition of freedom…does not imply that we need to actually attribute freedom to technological artifacts" (p. 77). Nevertheless, he maintains that "like intentionality appeared to be distributed over the human and nonhuman elements in human-technology associations, so is freedom" (p. 77).

9 What Are the Bearers of Agency?

The combination of the two argumentative strategies employed by Verbeek raises the question whether Verbeek wants to argue that only ensembles can possess agency (the first strategy) or that also technological devices can possess some form of moral agency, albeit maybe different from human agency (the second strategy). At several places, he states that only ensembles can possess agency, but sometimes he seems to suggest that also humans or objects can possess agency. For example, in Chap. 5, in relation to design, he talks about three forms of agency at work: that of the designer, the user, and the technology. Human agency is, however, different from technological agency: "it does not make sense to consider technologies fully fledged moral agents in the way human beings are moral agents" (p. 132).

What then is the relation between human agency, technological agency, and the agency of human–technology ensembles? The following quote about intentionality may give a hint:

> Intentionality is "distributed" over human and nonhuman entities, and technologies "have" the nonhuman part. In such "hybrid intentionalities", the technologies involved and the human beings who use the technologies share equally in intentionality (p. 72).

Here, the argument seems to be that the constituting parts—the human and the technological object—have intentionality *because* their association has intentionality, and a similar argument may apply to agency according to Verbeek. There are, however, two problems with this suggestion. First, the argument is invalid because it is based on a fallacy of division: parts do not necessary have the same properties as their association. From the fact that I can think, it does not follow that my leg can think.

The other problem is that the argument seems to suggest that intentionality and agency are *equally* distributed over humans and objects, while Verbeek at other places explicitly says that technological agency, unlike human agency, is not full-fledged agency. When it comes to moral responsibility, his position is even more asymmetrical: "The fact that technologies cannot bear moral responsibility themselves does not take away the fact that both users and designers can have a specific moral responsibility for technologically mediated action" (p. 133). Although he states that "it is the amalgam of humans and technologies…that bears moral responsibility" (ibid.), apparently this moral responsibility does, according to him, not distribute equally over humans and technologies.

10 Is Moral Agency the Right Target?

I think the ambiguities alluded to above are not coincidental, but point at a broader issue. This is the question whether moral agency is the right place to start when one wants to conceptualize the morality of technology. Verbeek at several places suggests that redefining moral agency is the only way in which one can do justice to the moral significance of technologies in ethical theory. This suggestion, however,

seems me to be based on a false dilemma. This false dilemma is not only apparent in Verbeek's work, but seems to characterize the debate about the moral agency of technology more generally. As Deborah Johnson notes in relation to the moral agency of computers:

> The debate seems to be framed in a way that locks the interlocutors into claiming either that computers are moral agents or that computers are not moral. Yet, to deny that computer systems are moral agents is not the same as denying that computers have moral importance or moral character; and to claim that computer systems are moral is not necessarily the same as claiming that they are moral agents (Johnson 2006, p. 195).

There are, moreover, several other kinds of entities, besides technology, that are morally significant without being a moral agent. Typical examples are social structures and laws. Both co-shape human behavior like technologies do and do so in other ways than just through mute physical causation. Both social structures and laws can be morally evaluated and are shaped by humans. It is not incidental, I think, that some philosophers of technology have conceived of technologies as being somehow similar to social structures (Sclove 1995) or to laws (Winner 1983).

I am also inclined to believe that moral agency is not the most appropriate place to start to conceptualize the moral significance of technology. The main reason for my doubts is that, as Verbeek recognizes, the notion of moral agency is closely connected to moral responsibility. It seems that if technological artifacts can have moral agency, they can, at least in principle, also bear moral responsibility, a conclusion that Verbeek does not seem willing to draw although it remains a bit unclear how he sets out to avoid it. Of course, one could give up the conceptual connection between moral agency and moral responsibility, but I doubt whether that is a price worth paying because, after all, one would still need an argument why technological artifacts, unlike humans, cannot bear responsibility. Apart from that, the argument that technologies can bear some form of agency is my view not very convincing. Although Verbeek offers a very interesting notion of freedom, i.e., as the ability to relate oneself to external constraints, which in my view convincingly shows that human moral agency is very well possible in a technologically mediated world, he admits that on this notion, technologies cannot be called free. If technologies have agency, it then seems to be a kind of derived agency that is based on the agency of human–technology ensembles. But even if we grant that such ensembles can have agency, it does not follow that technological objects have agency as we have seen. For all these reasons, I think that moral agency may not be the best candidate for understanding the moral significance of technology.

References

Harbers, H. (2005). Epilogue: political materials-material politics. In H. Harbers (ed.), *Inside the Politics of Technology*. Amsterdam: Amsterdam University Press.

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology, 8*,195–205.

Latour, B. (1993). *We have never been modern*. New York: Harvester Wheatsheaf.
Sclove, R. E. (1995). *Democracy and technology*. New York: The Guilford Press.
Verbeek, P. P. (2005). *What things do. philosophical reflections on technology, agency, and design* (R. P. Crease, trans.). Penn State: Penn State University Press.
Verbeek, P. P. (2011). *Moralizing technology: understanding and designing the morality of things*. Chicago: University of Chicago Press.
Winner, L. (1983). Techne and politeia: the technical constitution of society. In P. T. Durbin and F. Rapp (Eds.), *Philosophy and technology* (pp. 97–111). Dordrecht: D. Reidel.

Three Objections to Verbeek

Martin Peterson
Eindhoven University of Technology
M.Peterson@tue.nl

Peter-Paul Verbeek defends a radical view about the moral status of technical artefacts. To put things briefly, he argues that (1) artefacts, such as 'binoculars, thermometers, and air conditioners' (p. 56), have a special and morally relevant form of intentionality and that (2) we should, partly as a consequence of this, ascribe moral agency to hybrid entities consisting of technical artefacts and ordinary humans (pp. 58–61).[1] In addition, Verbeek claims that (3) this view about technical artefacts has important implications for traditional ethical theories, including consequentialism, Kantian duty ethics, and virtue ethics (pp. 61–63).

In what follows, I shall discuss claims 1–3 in some detail. This is because these ideas seem to be central to Verbeek's view as a whole. My hypothesis is that 1–3 are either false or misleading.

Having said that, I would like to emphasize that I think Verbeek's work is interesting and deserves respect. In my view, he has managed to raise a number of important questions that are worth taking seriously. This holds true no matter how good or bad my objections turn out to be. Many philosophers have attempted to formulate coherent views about the ethics of technology, but to do so is no easy task. It seems that no one (including the author of this article) has managed to provide a convincing account of exactly how and when artefacts enter the moral domain. Verbeek's view should be judged in light of this lack of well-developed alternatives. The ethics of technology is still a somewhat immature sub-discipline, and more promising theories will hopefully be developed in the future.

11 First Objection: Artefacts Have no Intentionality

As explained above, Verbeek believes that binoculars, thermometers, and air conditioners have some morally relevant form of intentionality (pp. 55–58). How should we understand this claim? Verbeek is careful to point out that he does not

---

[1] All quotes are from Verbeek (2011).

take artefacts to be conscious beings. Obviously, binoculars, thermometers, and air conditioners have no mental states, so if only entities having such states could have some form of intentionality, then artefacts would have no intentionality. However, Verbeek believes that he has found a way around this problem. According to Verbeek, 'The concept of intentionality actually has a double meaning in philosophy' (p. 55). In ethics, intentionality refers to 'the ability to form intentions' (ibid.), whereas in phenomenology, 'intentionality indicates the directedness of human beings toward reality' (ibid.).

This seems to be a neat distinction. On the one hand, we have the ordinary notion of intentionality, according to which intentionality is closely linked to mental states. Then there is also another notion, which is less demanding than the ordinary one. According to this alternative notion, intentionality just requires 'directedness… toward reality' (ibid.), and an entity can have such directedness without having any mental states.

As far as I can see, it might very well be plausible to claim that technical artefacts have the second, weak form of intentionality (given that the phrase 'directedness toward reality' can be spelled out in sufficiently great detail). It is less attractive to claim that binoculars, thermometers, and air conditioners have the first, strong form of intentionality. Unfortunately, Verbeek does not reject the strong notion of intentionality altogether. He explicitly claims that 'these two meanings of the concept of intentionality augment each other' (ibid.), but he never explains how or in what sense. Moreover, he never offers us a clear explanation of what is left of the first, strong notion of intentionality in his own hybrid notion of technological intentionality. All he says is that 'The ethical implications of the second meaning of the concept of intentionality are closely related to those of the first'. (ibid.) The following passage is probably the best summary of Verbeek's view:

> Binoculars, thermometers, and air conditioners help to shape new experiences, either by procuring new ways of accessing reality or by creating new contexts for experience…This implies that a form of intentionality is at work here—one in which both humans and technologies have a share. And this, in turn, implies that the context of such "hybrid" forms of intentionality, technologies do indeed "have" intentionality—intentionality is "distributed" among human and nonhuman entities, and technologies "have" the nonhuman part. In such "hybrid intentionalities", the technologies involved and the human beings who use the technologies share equally in intentionality. (p. 56)

As far as I can see, the problem with this story is that it does not give Verbeek what he needs. It might very well be true that binoculars, thermometers, and air conditioners help to shape new experiences, but the moral relevance of this is ambiguous. If the ability to 'help to shape new experiences, either by procuring new ways of accessing reality or by creating new contexts for experience' (p. 56) is all that is needed for being endowed with the weak form of intentionality, then many natural artefacts would have such intentionality. This indicates that there is nothing special with technical artefacts from a moral point of view.

Let us take a closer look at this objection. What I am claiming is that if all that is required for having a morally relevant form of intentionality is the ability to 'shape new experiences, either by procuring new ways of accessing reality or by creating

new contexts for experience' (ibid.), then many natural artefacts seem to have such intentionality. Imagine, for instance, that you are about to climb the Matterhorn, which is one of the most famous peaks in the Alps. It seems hard to deny that a mountain such as the Matterhorn can sometimes 'help to shape new experiences, either by procuring new ways of accessing reality or by creating new contexts for experience' (ibid.). Moreover, the Matterhorn has a form of 'directedness…toward reality' (p. 55)—the north face is the most difficult one to climb. But does this really show that the Matterhorn has any *morally relevant* form of intentionality? I believe even Verbeek would agree that the answer is no.

Anthonie Meijers has pointed out to me—and I believe this is a point that has been made by several Dutch philosophers working on the philosophy of artefacts—that there are many philosophically important differences between natural objects and technical artefacts. For instance, unlike natural objects, technical artefacts have an intentional history. Moreover, artefacts typically have certain effects only when used by some agent in certain ways. We can only see or measure certain things with a microscope or thermometer when we interact with them under the right kind of circumstances. Earthquakes and tsunamis are not used by agents for any specific purpose, and it seems implausible to argue that humans somehow interact with them.

However, although I accept all these differences between technical artefacts and natural objects, the problem is that Verbeek's account of hybrid intentionality fails to pick up on them. Humans sometimes interact with natural objects, such as the Matterhorn, in ways that make them fulfil Verbeek's criteria of hybrid intentionality. The Matterhorn shapes new experiences and creates new contexts for experience (which cannot be achieved without climbing the mountain) in roughly the same ways as microscopes and thermometers enable us to see and measure things we could not have measured otherwise. Therefore, if the shaping of new experiences is what matters for Verbeek's notion of hybrid intentionality, then the fact that mountains are different from technical artefacts in many other respects seems to be irrelevant. We all agree that the Matterhorn has no intentional history, but Verbeek's criterion of hybrid intentionality makes no reference to the intentional history of technical artefacts (and if it did, this would of course be somewhat question begging).

The Matterhorn fulfils Verbeek's hybrid condition of intentionality, but this hardly shows that mountains are of any interest to ethics. This, in turn, shows that Verbeek's theory of technological intentionality is either false or misleading. I think the best way to illustrate this is to simply substitute the talk about technology with the word 'mountain' in the passage quoted above.

> [Mountains] help to shape new experiences, either by procuring new ways of accessing reality or by creating new contexts for experience…This implies that a form of intentionality is at work here—one in which both humans and [mountains] have a share. And this, in turn, implies that the context of such 'hybrid' forms of intentionality, [mountains] do indeed 'have' intentionality—intentionality is 'distributed' among human and nonhuman entities, and [mountains] 'have' the nonhuman part. In such 'hybrid intentionalities', the [mountain] involved and the human beings who use the [mountain] share equally in intentionality. (p. 56)

Not convinced by this argument? Then you should also reject Verbeek's claim that technical artefacts and human beings have a hybrid form of intentionality. The mere fact that technical artefacts, unlike mountains, are created (intentionally) by humans does not change this. An intention is no infectious disease. And even if it were, it seems that humans interact at least as much with natural artefacts that shape new experiences. Although claims about the intentional history of an entity could help us to distinguish technical artefacts from natural objects, Verbeek has by no means managed to explain why such claims about the past would make any moral difference.

12 Second Objection: Artefacts Are No Moral Agents

Very few people think of binoculars, thermometers, and air conditioners as (parts of) moral agents. Verbeek is the exception that proves the rule. His moral analysis of cases in which someone fires a gun at a fellow citizen runs as follows:

> Without denying the importance of human responsibility in any way, we can conclude that when a person is shot, agency should not be located exclusively in either the gun or the person shooting, but in the assembly of both. (p. 64)

This view about moral agency is deeply puzzling. In order to see this, note that Verbeek clearly believes that when someone fires a gun at another person, we should locate the moral agency of this event in, 'the assembly of both [the gun and the person shooting]' (ibid.). However, Verbeek *also* claims—very explicitly—that this ascription of moral agency to the assembly of the gun and the person shooting can be done '[w]ithout denying the importance of human responsibility' (ibid.). How could these two claims ever be compatible with each other?

It seems that the only way of making sense of Verbeek's position is to maintain that the there are *two* moral agents responsible for firing the gun: first, we have the ordinary human being (so we do not have to deny 'the importance of human responsibility in any way'); then we also have a novel kind of moral agent, viz. the assembly of the gun and the person shooting.

This innovative reading would perhaps make Verbeek's position come out as logically consistent, but this is hardly an attractive approach to moral agency. To start with, if the fact that the assembly of the gun and the person shooting entails that the person shooting is a moral agent, why not then also ascribe agency to the gun itself, i. e. argue that there are not two but *three* agents involved in this example? What is the philosophical advantage of ascribing moral agency to exactly *two* entities, viz. the person shooting and the assembly of the gun and the person shooting?

A possible reply, suggested to me by Anthonie Meijers, could be that it is common to ascribe agency to an organisation or a corporation, such as a bank. For instance, most people probably agree that a bank is responsible for ensuring that money deposited by its customers are not stolen, and this of course entails that the chairman of the bank is responsible for taking certain actions to prevent this from happening. To ascribe agency to a bank does not preclude us from also ascribing agency to its chairman.

In reply to this objection, I wish to make it clear that what I believe is a problem for Verbeek is that he tries to locate the agency of *one and the same action* in both the person shooting and the assembly of the gun and the person shooting. It might of course be true that the person shooting was responsible for buying the gun in the first instance, thereby enabling the assembly of the gun and the person shooting to carry out an atrocity, in the same way as the chairman of the bank was responsible for ensuring the bank as a whole implements a policy for preventing various types of fraud. But this is not the issue that is at stake here. Verbeek wishes to avoid the accusation that he cannot account for the importance of human responsibility as the assembly of the gun and the person shooting carries out an atrocity, and his strategy is to also ascribe agency to the person shooting. In the bank case, this would mean that all decisions about, say, bonuses, are taken not just by the bank, but by the bank *and* the individual members of the board of the bank. No matter what your view about collective agency happens to be, I can assure you that this inflationary approach would hardly be welcomed by any participants in this debate.[2]

Verbeek goes on to relate his view on moral agency to various views about freedom. He takes the plausible view that the assembly of the gun and the person shooting can only be a moral agent if it can exhibit some morally relevant sort of freedom, that is, Verbeek accepts the common idea that moral agency requires an ability to choose freely among alternative actions. He then goes on to claim that 'Technologies cannot be free agents as human beings are' because 'freedom requires the possession of a mind, which artefacts do not have' (p. 59). Despite this very reasonable remark, he nevertheless ends up in the wrong corner: he insists that binoculars, thermometers, and air conditioners are parts of complex entities that qualify as free moral agents (pp. 58–61).

Verbeek's argumentative strategy is straightforward. In summary, he argues that although technologies cannot be free if we take the word 'freedom' to mean what it usually means, there is an alternative notion of freedom. According to Verbeek, this alternative notion is relevant to discussions of technical artefacts, and it is also sufficiently strong for warranting the conclusion that artefacts can be (parts of) moral agents. Here is a representative quote:

> Rather than taking freedom from (technological) influences as a *prerequisite* for moral agency, we need to reinterpret freedom as an agent's ability to *relate* to what determines him or her…Technologies 'in themselves' cannot be free, but neither can human beings. Freedom is a characteristic of human-technology associations. (p. 60, his italics)

Again, the words used by Verbeek do not mean what they normally mean. The word 'freedom' apparently means something radically different, just like in George Orwell's novel *Nineteen Eighty-Four* (1949). The problem with Verbeek's Newspeak account of freedom is that it does not seem to be relevant to our original problem: can technical artefacts (in combination with humans) have the kind of freedom that is often taken to be a prerequisite for the ordinary, Oldspeak account of moral agency? Since Verbeek operates with quite unusual notions of freedom and agency,

---

[2] See, e.g. Pettit and Schweikard (2006) and Hindriks (2009).

it is even hard to tell what his answer to this seemingly simple question would be. This is a serious objection. To put things briefly, it is hardly interesting to discuss whether there is a Newspeak account of freedom that enables artefacts to be (parts of) Newspeak-style moral agents.

However, based on his innovative account of freedom, Verbeek goes on to argue that technical artefacts should play a prominent role in discussions of moral agency. His conclusion is as follows:

> By rethinking the concepts of intentionality and freedom in view of the morally mediating roles of technology, I have dispatched the major obstacles to including technological artifacts in the domain of moral agency....The position I have laid out in this chapter is based on the idea that the moral significance of technology is to be found not in some form of independent agency but in the technological *mediation* of moral actions and decisions—which needs to be seen as a form of agency itself. (p. 61, his italics)

Apart from the fact that Verbeek's redefinition of freedom is questionable, there is also another problem with this position. In analogy with what I claimed above, it seems that if we accept the 'Newspeak account', it follows that mountains would have the same kind of freedom and moral agency as technical artefacts. This is quite peculiar, and Verbeek fails to explain what the moral difference could be between mountains on the one hand and technical artefacts on the other. It is certainly true that there are *some* differences, but what is lacking is an analysis of how we should think of those differences. For instance, as pointed out above, technical artefacts tend to be created intentionally (in the ordinary Oldspeak sense), whereas mountains are not. This could perhaps be claimed to be an important moral difference. But recall that Verbeek operates with his own, weak notion intentionality, which is applicable to both technical artefacts and mountains. Therefore, that simple move is not open to him. As far as I can see, all the key claims Verbeek makes about the moral properties of technical artefacts seem to be equally applicable to mountains. It would perhaps be tempting to use some Oldspeak distinctions for explaining the difference, but once you have started to speak Newspeak, you have to stick to that vocabulary. If we use Verbeek's Newspeak vocabulary, it is simply impossible to make interesting and substantial distinctions between thermometers and mountains.

13 Third Objection: Verbeek's Theory Does Not Square Well With Virtue Ethics

Let me finally discuss Verbeek's claims about the implications of his position for traditional ethical theories. Although he seems to think that his radical view about intentionality and moral agency could in principle be rendered compatible with (non-standard) accounts of Kantianism and utilitarianism, Verbeek argues that it fits particularly well with virtue ethics. The following quote is a representative summary of Verbeek's take on virtue ethics:

> From a *virtue-ethical* position, it is much easier to incorporate the moral roles of technologies. As Gerard de Vries has noted (de Vries 1999), this premodern

form of ethics does not focus on the question of "how should I act" but on the question of "how to live." It does not take as its point of departure a subject that asks itself how to behave in the outside world of objects and other subjects. It rather focuses on "life"—human existence, which inevitably plays itself out in a material world. From this point of view, it is only a small step to recognize with de Vries that in our technological culture, not only ethicists and theologians answer this question of the good life but also all kinds of technological devices tell us "how to live". (p. 63, his italics)

There are several problems with this passage. First, and this is a minor issue, it is misleading to give credit to Gerard de Vries for having noted that virtue ethics does not focus on the question of *how should I act* but on the question of *how to live*. This point has been made many times in the literature (by, e.g. Anscombe 1958 and Foot 1985) long before de Vries mentioned it in 1999. My second and more important worry is that virtue ethics seems to be the ethical theory that fits *least* well with Verbeek's claims about intentionality and moral agency. As Verbeek correctly points out earlier in the book, it seems possible to reformulate Kantianism and utilitarianism such that they could become claims not about what a human being ought to do, but claims about what human–technology associations ought to do. For instance, if you read Verbeek and then decide to become a Kantian, it seems reasonable to think that it is not the person holding the gun, but rather the assembly of the gun and the person shooting that should perform actions which are compatible with the categorical imperative. Moreover, if you read Verbeek and then decide to become a utilitarian, it seems reasonable to think that it is not the person holding the gun, but rather the assembly of the gun and the person shooting that should perform actions that maximize overall utility. Given that we accept the quite implausible claim that the assembly of the gun and the person shooting is a moral agent, everything seems to be fine from a Kantian as well as a utilitarian perspective.

However, now imagine that you read Verbeek and then decide to become a virtue ethicist, as Verbeek thinks we should. Then it is no longer sufficient to believe that the assembly of the gun and the person shooting is a moral agent; then we also have to accept one additional and highly controversial claim, viz. the claim that the assembly of the gun and the person shooting is an entity that *has a life*. As mentioned above, virtue ethics is a claim not about what the moral agent should do, but a claim about how the moral agent should live. Therefore, since the moral agent in Verbeek's theory is the assembly of the gun and the person shooting, rather than just the person shooting, Verbeek must concede that the assembly of the gun and the person shooting can be more or less virtuous and lead a life that is either good or bad from a virtue ethical point of view. This is a very uncomfortable concession to make for a philosopher.

References

Anscombe, G.E.M. (1958) Modern moral philosophy. *Philosophy*, *33*(124): 1–19.
De Vries, G. 1999. *Zeppelins: Over filosofie, technologie en cultuur*. Amsterdam: Van Gennep.
Foot, P. (1985). Utilitarianism and the virtues. *Mind*, 94(374):196–209.

Hindriks, F. (2009). Corporate responsibility and judgment aggregation. *Economics and Philosophy*, 25(2):161–177.

Orwell, G. (1949), *Nineteen eighty-four*. London: Secker and Warburg.

Pettit, P. and D. Schweikard (2006). Joint actions and group agents. *Philosophy of the Social Sciences*, 36 (1):18–39.

Verbeek, P (2011) *Moralizing technology*. Chicago: University of Chicago Press.

The Irony of Humanism: On the Complexities of Discussing the Moral Significance of Things

Peter-Paul Verbeek
Dept. Philosophy
Twente University
p.p.c.c.verbeek@utwente.nl

14 The Blackmail of Humanism

Without any doubt, Sigmund Freud would have tremendously enjoyed the current discussion about the moral significance of technology. Even though only few people would still call themselves 'Freudians' nowadays, there are some ideas in his work that will never die. Among these is definitely the thesis that modern science has caused humanity various 'narcissistic offences', humiliating human beings with scientific insights. Copernicus showed that not the earth but the sun is the center around which the heavenly bodies circle. Darwin showed that the human being is not the central entity in God's creation, but a mammal that developed from the ape. And Freud himself showed that in many cases, not conscious decisions but unconscious factors shape our behavior.

The fierce resistance against the idea that human morality is interwoven with nonhuman entities might very well reveal yet another narcissistic wound. Discussing the moral significance of technological artifacts evokes fears about giving up on human autonomy and responsibility, and even raises doubts about the rationality of one's mind. If material objects are allowed to play a substantial role in morality, how can human beings possibly be responsible for their actions? Don't we end up in a kind of premodern animism, if we seriously consider an active role of nonhuman objects in the realm of human subjects? Human dignity itself seems to be at stake when we have to recognize that things are involved in our moral actions and decisions.

This dispute reminds of Michel Foucault's discussion of Immanuel Kant's ideas on the Enlightenment. In his lecture *What is Enlightenment*, Foucault discusses how, for Kant, enlightenment meant a way out of immaturity—immaturity being defined as "a state of our will that makes us accept someone else's authority to lead us in areas where the use of reason is called for." Some of the objections against *Moralizing Technology* seem to imply a similar logic. It seems that we have to make a choice between using reason or letting things decide for us what to do. And this mirrors what Foucault, in the same text, calls 'the blackmail of the Enlightenment': if you are not entirely with it, you are supposed to be against it.

The most interesting aspect of these objections is that they in fact reproduce the error they claim to identify. They typically fight against the idea that things can be moral agents—which is depicted as an absurd idea that threatens human responsibility. Humans, not things, do ethics; who would blame a car for an accident or a gun for a shooting? This is a fight against windmills, though. Hardly anybody who studies the moral significance of things would claim that things in themselves have moral agency. And if this claim is made at all, it is done at a level of abstraction that makes it much less absurd, like Floridi and Sanders did in their inspiring article 'On the morality of artificial agents' (Floridi and Sanders 2004). *Moralizing Technology*, in any case, does not intend to defend this idea at all. In fact, the idea that artifacts can be moral agents originates from the very modernist metaphysics that *Moralizing Technology* intends to help overcome.

Only from a metaphysics in which humans and things are radically separated— humans being active and intentional, nonhumans being mute and inert, as Bruno Latour has elaborated so convincingly—it becomes relevant to ask if things, just like humans, can be moral agents. When we give up this separation, a completely new picture emerges. The central question then is not if artificial agents can have morality too, like Floridi and Sanders did, but how morality is distributed over human and nonhuman entities. Recent insights from philosophy of technology urge us to see moral agency as the outcome of complex interactions between humans and things. Indeed, things do not have moral agency—but the most crucial point is: neither do humans. Morality is a hybrid affair that cannot be located exclusively in things, but not in humans either. Each in their own way—distinct, but never separated—humans and things contribute to moral actions and decisions.

Moralizing Technology is an inquiry into the nature of these interactions of humans and things, and into their implications for ethical theory and the ethics of design. My approach, therefore, should not be read as a defense of *animism*— bringing spirit to things—but as a critique of *humanism*. I do not claim that material objects are 'spirited', but I want to move away from ethical approaches that isolate and immunize human existence from its material conditions and contexts.

This is the irony of humanism: in its radicality, it loses what it aims to defend. A hybrid approach to the relations between humans and things does not reduce human morality, but adds to it; it shows dimensions that normally remain underexposed. It does not excavate human responsibility by blaming cars for accidents, but expands the ways in which we can design, implement, and use technologies in responsible ways. Let me make this clear by replying to a few of the topics raised in the thoughtful and provocative discussions of *Moralizing Technology* by Don Ihde, Martin Peterson, and Ibo van de Poel.

15 Mediated Intentionality

Martin Peterson's thorough review is a fine example of radical humanism. He first explains why things cannot have intentionality by showing how absurd it would be to speak of the intentions of a mountain. Things do not have intentions, Peterson claims, but an intentional *history* that he finds lacking in my argumentation. Not the fact that

things help to shape practices and experiences makes them morally significant—in that case, non-technological entities like mountains would be moral agents too—but the fact that they have their origins in human intentions. Therefore, my theory is "either false or misleading". Moreover, he feels that I do not use the "ordinary" notion of intentionality—which, for him, is "closely linked to mental states"—but a "weak" form of intentionality that "just requires 'directedness toward reality.'"

Apart from his remarkable characterization of the most central concept in phenomenology as "not ordinary" and "weak", Peterson fails to see here how I aim to make visible a close connection between human intentions for action and the technologically mediated character of human perception and experience. His critique shows how hard it is, from a modernist–humanist point of view, to conceptualize the interwoven character of humans and technologies. Peterson suggests that I attribute moral intentionality to technologies, while my central claim is simply that, in mediating how humans are directed at reality, technologies help to shape moral intentions.

What is misleading here, in fact, is Peterson's reductionist assumption that only human 'input' can make technologies morally significant. The central example in the book clearly shows the limitations of this humanist approach. Ultrasound was not explicitly developed for medical diagnostic purposes and certainly not to change abortion practices. Still, it helped to create a new form of responsibility, because expecting parents now explicitly have to make a decision about the lives of their unborn children. This situation is not the result of somebody's explicit decision; no intentional history will help us to understand it exhaustively.

And what is worse: refusing to conceptualize the active moral role of obstetric ultrasound in abortion practices makes it harder to give this technology an intentional *future*, in practices of design and redesign. And that is exactly what *Moralizing Technology* does. Its analysis of the moral significance of technology and the mediated character of human morality culminates in an attempt to expand the ethics of design, enabling designers to anticipate, assess, and design moral mediations in technology.

16 Hybrid Agency

Even more puzzling, from a radical humanist point of view, is the role of technologies in moral agency. Peterson develops a complicated system to understand my claim that the hybrid character of moral agency does not urge us to give up human responsibility. He states that my position implies that there are in fact *two* moral agents responsible for an action—firing a gun, in his example. One would be the human being shooting the gun and the other the assembly of human and gun, as a composite agent. And, of course, this would be absurd: one and the same action cannot be the product of two different entities.

My position is much less complicated, though. When we give up the *idée fixe* that only autonomous beings can be moral agents, it becomes perfectly possible both to acknowledge that one's moral actions and decisions are technologically mediated *and* to take responsibility for these actions and decisions. Technological mediations do not make human beings powerless. To the contrary: they make it

possible to live our lives in specific ways, while we also have the ability to develop an active and critical relation to these mediations. Nobody has to choose to have an abortion when an ultrasound scan reveals a serious disease. Still, the mere possibility to have a scan inalterably conditions our existence: we now *have* to make a decision.

This also relates directly to Peterson's comments on my concept of freedom, which I sharply distinguish from 'autonomy'. While autonomy indicates the absence of external influences, the concept of freedom that I defend takes these influences—or technological mediations—as its starting point. Freedom, then, does not come about by liberating oneself from the influences of technology, but by engaging in an explicit relation with them. Developing a free relation to technology implies: understanding its mediating power and getting actively involved in the impact these mediations have on one's existence. It is remarkable that Peterson dubs this definition 'newspeak', complaining that "the words used by Verbeek do not mean what they normally mean". In fact, my definition builds upon 'oldspeak' conceptions of freedom by Isaiah Berlin and Michel Foucault. Moreover, developing new concepts has always been the core business of philosophers.

Also Ibo van de Poel is worried about my redefinition of the concept of agency. He wonders whether such a drastic change in ethical theory is actually needed to understand the moral significance of technology. Why make such a conceptual mess, when there are much tidier solutions, he seems to ask himself. Also for Van de Poel, an example of the problems I run into is my seemingly double definition of agency. On the one hand, I claim that agency needs to be located in *assemblages* of humans and technologies, while on the other hand I seem to ascribe agency to things as well, as parts of these assemblages. Adding the latter form of agency, Van de Poel claims, is a fine example of the 'fallacy of division': from the fact that I can think and that my leg is a part of myself, it does not follow that my leg can think. Things do not have agency; rather, human agency is co-shaped by things.

This critique is certainly interesting. In fact, van de Poel addresses the question of how to characterize the mediating role of things in moral agency. But again, my central claim that moral agency is a hybrid affair does not imply that things are moral agents just like humans are. This is an example of what, in turn, could be called the fallacy of reduction. While agency cannot be limited to humans, acknowledging that things have a share in moral agency does not make them moral agents.

17 Responsibility

Van de Poel argues that the main criterion here should be that agency requires *responsibility*—which things obviously cannot take. But in fact, the very possibility to take responsibility is one of the main reasons to take the role of things in moral agency very seriously. On the one hand, we do not take responsibility in a vacuum but in a thoroughly mediated situation, as the ultrasound example shows. Sonograms do not act' on themselves, but nevertheless, they fundamentally shape what we can

feel responsible for and how we can take on that responsibility. And what is more, acknowledging this moral role of technologies makes it possible to take responsibility for it and to help shape it in practices of design, implementation, and use. We can only reorganize practices around technologies when we understand the precise role technologies have in them. Seeing the moral significance of technologies makes us *more* responsible, rather than less.

Precisely at this point, however, Don Ihde has his doubts. We should not overestimate our possibilities to organize and design the moral significance of things, he states. We need to recognize the multistabilities of all technologies and therefore the ambiguities of designer intent. Technologies typically end up in different relations with human beings than their designers expected, and therefore, their mediating power is hardly predictable.

Ihde is completely right here. Technological mediations are always part of human–technology relations, which are principally unpredictable. Still, this does not make us entirely powerless. Industrial Designers, as I show in the book, have developed various methods to make an educated guess about possible use practices. And once we see the phenomenon of mediation, including its moral dimensions, it becomes our moral responsibility to make such an educated guess and to design 'for the good'.

18 Mirrors and Symmetry

Don Ihde's use of the mirror metaphor is highly interesting in this context. This metaphor does not only make it possible to create an interesting combination of nearness and distance between our positions. It can also be expanded to shed new light on the issue of 'symmetry' between humans and nonhumans that has proven to be the core theme in discussions about the moral significance of technology.

The concept of symmetry comes from Bruno Latour, who claims that we should analyze humans and nonhumans in symmetrical ways. Rather than making an a priori distinction between them, we need to see the continuity between them. From such a symmetrical approach, not only humans but also things 'act'—and from there, it is only a small step towards defending that things can be *moral* agents as well.

Yet, this symmetry is not what I am after. And this is what distinguishes my 'post-phenomenological' account from Actor-Network Theory, despite the many forms of kinship that are there as well. Symmetry is what you get when you use a mirror: a mirror image is completely symmetrical to its original, and it derives all of its main characteristics from that original. In my approach, however, there is no symmetry but interaction. Things are not symmetrical to humans, but together, humans and things constitute myriad entities. In this approach, it remains very relevant to make a distinction between humans and things—it is not the *distinction* between humans and technologies that we need to depart from, but their radical *separation*.

Moralizing Technology, then, ultimately intends to show that Kant's "way out of immaturity" implies that we acknowledge the moral dimension of artifacts. A mature approach to technology does not exclude technologies from the realm of

ethics. Only by taking seriously their fundamental role in moral actions and decisions can we better understand the character of human morality. And, more importantly, can we take responsibility for the material world in which we live our lives.

References

Floridi, L. and J.W. Sanders. (2004). On the morality of artificial agents. *Minds and Machines*, *14*(3), 349–79.