

Technology design as experimental ethics

Peter-Paul Verbeek, University of Twente

Appeared as:

P.P. Verbeek (2013). 'Technology Design as Experimental Ethics'. In: S. van den Burg and Tsj. Swierstra, *Ethics on the Laboratory Floor*. Basingstoke: Palgrave Macmillan, pp. 83-100. ISBN 9781137002921

1. Introduction: Blurring Boundaries

Doing ethics of technology has become a complicated activity after the developments that took place in philosophy of technology over the past decades. Contemporary approaches like Actor-Network Theory and Postphenomenology have convincingly argued that we need to blur the boundaries between human beings and technological artifacts. While Don Ihde has shown that human relations with the world are fundamentally mediated by technologies (Ihde 1990), Bruno Latour claims that we need to give up the separation we make between human and nonhuman beings (Latour 1993). In the meantime, according to various authors, even the field of morality has become a hybrid affair. Moral actions and decisions of human beings are fundamentally mediated by technologies, like turnstiles that intervene in fare-dodging in the subway, and antenatal diagnostic technologies that inform moral decisions about abortion (Verbeek 2011).

This blurring of the boundaries between humans and technologies is a serious challenge for the ethics of technology. It makes it impossible for ethicists to play the role that is typically associated with them: the role of a border guard assessing whether a technology are morally acceptable or not. Ethics cannot defend a boundary between humans and technologies anymore when recent insights from philosophy of technology show that this boundary does not exist.

This does not imply, however, that the ethics of technology has reached its end. As I explained elsewhere (Verbeek 2010) the real challenge is to develop new ways of doing ethics of technology, that shift their focus from 'assessing' technologies toward 'accompanying' their development, implementation and use. Rather than determining whether a technology is morally acceptable or not, the ethics of technology could focus on the question of helping to shape good hybrids. And

rather than taking on the role of an external judge it then plays the role of an engaged participant who brings in perspectives that might otherwise remain underrepresented.

In this contribution I will elaborate what such an 'ethical accompaniment of technology' could entail. First, I will further articulate the notion of 'ethical accompaniment' as opposed to 'technology assessment'. By discussing Michel Foucault's concept of 'limit attitude', I will show the notion of 'limit' can be used not only to assess the desirability of technologies but also to accompany their development. After this, I will propose a framework to read and to design the moral significance of technologies. By linking the approach of technological mediation to design thinking, it becomes possible to expand the realm of ethics from words and ideas to things and technological systems. In three sections, I will indicate three ways in which designers could take the mediating role of technology into account in their work: they can *anticipate* mediations, systematically *assess* them, and deliberately *design* them 'into' a technology.

2. Ethics: from assessment to accompaniment¹

Implicit in many ethical approaches to technology is the model of a struggle between human beings and technologies. While some technological developments can be beneficial, others compose a threat to humanity. The role of ethicists, then, is to assess if technologies are morally acceptable or not. As indicated in the introduction, this model of a struggle between humanity and technology has become highly problematic in view of recent developments in philosophy of technology. According to recent insights, the human being cannot be understood in isolation from technology, just like technology cannot be understood in isolation from humanity. Approaching their relation in terms of struggle and threat is like attempting to give a moral evaluation of gravity, or language. It does not make much sense to be 'against' gravity or language, since they simply form the basis of our existence. In similar ways, technology inevitably helps to shape what it means to be human.

To be sure, this does not imply that all roles of technology in human existence are equally desirable, and that human beings are in fact powerless victims of the power of technology. But it does imply that the 'opposition model' of humanity and technology might not be the most productive model if one wants to change undesirable configurations of humans and technologies.

¹ This section incorporates reworked fragments from my article 'Resistance is Futile', forthcoming in *Technè* 2013-1.

Ethics should not focus on defending the boundaries between humanity and technology but on governing their intertwinement.

If there is a struggle between humans and technologies at all, it should be conceptualized in the way Heidegger conceptualized artworks as a struggle between 'earth' and 'world'. A work of art, Heidegger argues in *The Origin of the Work of Art*, brings a world into being on the basis of 'earthly' elements, like cloth and paint, bronze, vibrations of the air, et cetera (Heidegger 1971). Experiencing a work of art is experiencing the coming into being of a meaningful world, out of these material elements. Watching Van Gogh's painting of a pair of peasant shoes, in Heidegger's well-known example, sets into motion a 'struggle' between the cloth and the pigment on the one hand, and the reality of the pair of shoes that arises out of these material element on the other. In this struggle, there is no oppression and liberation, but interaction and mutual shaping. It would be strange to say that the paint oppresses the shoes, or the other way round: the two are necessarily connected. Similarly, forms of human existence and arrangements of society are 'revealed' in the relations between technological materialities and the human beings who design, organize, and use them.

But what can ethics still be, when the boundaries between humans and technologies disappear? In the symmetrical approach of Bruno Latour, where human and nonhuman entities play equal roles, and in the mediation approach of postphenomenology, where human practices and experiences are always technologically mediated, there does not seem to be an 'outside' position anymore with respect to technology. And if there is no outside anymore, from where could we criticize technology?

In order to articulate an alternative model for ethics, it is helpful to connect to Foucault's approach to the phenomenon of 'critique' in his lecture 'What is Enlightenment?' (Foucault 1997a). In his analysis, Foucault is looking for an answer to what he calls 'the blackmail of the Enlightenment'. This blackmail consists in the fact that it is extremely hard to criticize the Enlightenment, since all attempts to do so are typically explained as being *against* it. Anyone who dares to do open a discussion about the Enlightenment raises the suspicion of being against rationality, democracy, and scientific inquiry. Foucault, however, wants to explore if an *alternative* Enlightenment would be possible. This ambition is recognizable in the context of the ethics of technology. Blurring the boundaries between humans and technologies can easily be explained as giving up on ethics: because there is no clear boundary to be defended anymore, it might seem that 'anything goes'. Therefore an alternative model for ethics needs to be developed.

As the title of his lecture suggests, Foucault was primarily occupied with the work of Immanuel Kant. In fact, he proposes an empirical and practical reinterpretation of Kant's ideas on the Enlightenment. Rather than aiming to transcend the empirical world, as Kant did, Foucault reinterprets Enlightenment as an *attitude* within the world. For Kant, as Foucault explains, Enlightenment was primarily a way out of "immaturity", using "reason" rather than accepting "someone else's authority to lead us in areas where the use of reason is called for" (Foucault 1997a, 305). This way out of immaturity requires critique: critique can tell us under which conditions "the use of reason is legitimate in order to determine what can be known, what must be done, and what may be hoped" (Foucault 1997a, 308). Critique, then, according to Foucault, must be understood as an attitude, an "ethos". It is the attitude of always looking for the limits of what seems to be given and self-evident.

Foucault calls this Enlightenment attitude a "limit attitude". This attitude is looking for "the singular, the contingent, and the product of arbitrary constraints" in "what is given to us as universal, necessary, obligatory" (Foucault 1997a, 315). Unlike Kant, though, Foucault does not want to transcend the empirical world into a transcendental realm. While Kant investigated the transcendental conditions of human reason, Foucault reinterprets critique – the 'enlightened' activity *par excellence* – as a form of practical self-inquiry. For Foucault, critique means investigating what has made us the beings that we are, what conditions and has shaped our current form of existence. In order to be critical, he does not *transcend* the limits of the empirical, in order to find an 'outside' position, but rather positions himself *at* the limit. For Foucault, after all, there is no outside position from where to think. The human subject is always situated *within* the world to which it has a relation.

In the context of technology this means that the frameworks from which one can criticize technology are technologically mediated themselves. We can never step out of these mediations. The most far we can get is: to the limits of the situation we are in. Standing at the borders, recognizing the technologically mediated character of our existence and our interpretations, we can investigate the nature and the quality of these mediations: where do they come from, what do they do, could they be different?

As such, the Foucauldian limit-attitude provides a "way out" of the question if ethics of technology is still possible when we embrace the hybridizing approaches of Ihde and Latour. Rather than letting ourselves be blackmailed by the Enlightenment – fearing that a nonmodern conceptualization of technology and society as interwoven would make it impossible to have a

reasonable and normative discussion about technology – an alternatively-enlightened ‘limit approach’ can offer a different, nonmodern ethical approach to technology. Not the *assessment* of technological developments from outside is the central goal of ethical reflection then, but rather its *accompaniment*, ‘from within’, borrowing a concept from the Belgian philosopher Gilbert Hottois (Hottois 1996).

The crucial question in such a form of ‘ethical technology accompaniment’ is not so much where we have to draw a boundary between human beings on the one hand and technologies on the other. It rather is how we should give shape to the interrelatedness between humans and technology, which has in fact always been a central characteristic of human existence. The limit-attitude leads to an ethical approach that is not preoccupied with the question of whether a given technology is morally acceptable or not, but that is directed at improving the quality of our lives, as lived with technology.

Focusing on the intricate relations between human beings and technologies does not mean, to be sure, that all relations are equally desirable, and that rejection of a technology is no longer possible. Rather it implies that ethics needs to engage more deeply with actual practices of design, use, and implementation. Giving up an external position does not require us to give up all critical distance; it only makes sure that we do not overestimate the distance we can take. The Foucauldian limit-attitude urges us to develop a ‘critique’ from within, engaging with how technological practices actually take shape, and from a situation that is technologically mediated itself.

In line with Michel Foucault’s ethical work (2010), this ‘technology accompaniment’ can be seen as a form of ‘governance’. By deliberately shaping one’s involvement with technology and with the impact technology can have on one’s existence, it becomes possible to give direction to one’s technologically mediated subjectivity. Governance needs to be distinguished sharply from ‘steering’. Governing technological developments implies a recognition of their own dynamics, and of the relatively limited autonomy human beings have in their relations to technology. Human beings are ‘implied’ in technological developments, just like technologies are ‘implied’ in human existence. From this ‘hybrid’ point of view, in which humans and technologies are closely intertwined, the modernist ambition to ‘steer’ technology and to ‘protect’ humanity against technological invasions needs to be replaced with a more modest ambition to ‘govern’ the development of technology by taking its social implications into account, and to ‘govern’ one’s subjectivity in relation to those technologies. By governing the relations between humanity and

technology, we give up the idea that we can control technology; rather, we aim to understand how technology affects us, and explicitly get involved in that process, by critically designing and using technologies from the perspective of their mediating powers in human existence and our technological society.

3. Analyzing the morality of technology

Governing technological developments requires us to be able to understand the impact of technologies on society. Here, the theory of technological mediation, that developed out of Don Ihde's postphenomenological approach to technology, can be a helpful framework (Ihde 1990). Mediation theory approaches technologies as mediators of human-world relations. When used, technologies establish relations between human beings and their environment. These relations have a hermeneutic and an existential dimension: 'through' technologies, human beings are present in the world, and the world is present for human beings. Technologies, in other words, help to shape human experiences and practices (cf. Verbeek 2005). Cell phones help to shape how human beings experience each other, while intelligent speed adaptation technologies help to shape people's driving behavior in cars.

The central idea in mediation theory is that technologies do not simply create *connections* between users and their environment, but that they actively help to *constitute* them. Cell phones are no neutral intermediaries between human beings, but help to shape how humans are 'real' for each other. And likewise, sonograms are not simply 'pictures' of a fetus, but help to shape what the unborn child is for its parents, and what these parents are in relation to their unborn. Mediation does not take place between pre-given entities, but helps to constitute the reality of these entities.

This mediating role of technologies has important implications for the ethics of technology. Mediation theory shows that human actions and decisions are fundamentally technologically mediated. In a 'material way', technologies help to give answers to the moral questions of 'how to act', and 'how to live'. Moral actions and moral decisions are not the product of autonomous human agency, but take shape in close interaction with technological artifacts. Technologies are morally charged, so to speak. They actively contribute to human morality (cf. Verbeek 2011).

Insights in technological mediation and the moral significance of technology can be a basis for an ‘accompanying’ ethics of technology. As I will explain below, mediation theory can help to make a (moral) ‘mediation analysis’ of a technology-in-design in order to inform the activities of designers.

Three levels can be distinguished at which mediation analysis can inform the work of designers. First, performing a mediation analysis can help them to *anticipate* the moral dimensions of the technology-in-design, for instance in order to avoid undesirable mediating effects. Second, mediation analysis can be the basis for *assessing* the quality of expected mediations. Making such assessments, to be sure, does not imply a shift back from ‘accompanying’ to ‘assessing’ technology; it rather should be seen a fully-fledged part of ‘technology accompaniment’. And, third, mediations can be explicitly designed into a technology. In this case, we can speak of an explicit ‘moralization’ of technology, following Dutch philosopher Hans Achterhuis (Achterhuis 1995).

4. Anticipating mediations

In order to anticipate the mediating roles of a technology-in-design, it is important to try and make an analysis of the potential mediating roles the technology could have in the future. Making such an analysis is a complicated thing; it is never possible to make exact predictions about the future. Still, insights from mediation theory can guide the designer’s imagination in order to make an ‘educated guess’. Mediation theory then functions as a ‘heuristic tool’ to look for possible mediation effects.

In order to present such a heuristic tool here, I will integrate various elements of mediation theory that have been developed over the past years into a coherent framework that can be used to anticipate technological mediations. These elements focus on the *locus*, the *type*, and the *domain* of mediations: which ‘point of application’ does the technology have, which form do its impacts have, and which aspect of human existence does it affect?

Points of application

For conceptualizing the locus of mediation we can connect to Steven Dorrestijn’s work on behavior-influencing technology. In his book ‘The design of our own lives’ (Dorrestijn 2012) he distinguishes four points of application from which technologies can have an impact on human

beings. Dorrestijn divides the space around human beings in four quadrants: 'to the hand', 'before the eye', 'behind the back', and 'above the head'.

Mediations 'to the hand' are physical: speed bumps that make it impossible to drive too fast, or turnstiles that force metro users to buy a ticket. In addition to this, mediations 'before the eye' have a more cognitive nature. Technologies give cues or signals here in order to influence our behavior: navigation systems that beep when one drives too fast, or smart energy meters that persuade people to switch off the standby mode of their equipment. Mediations 'behind the back' concern the contextual and infrastructural role technologies can play. Technologies do not influence human behavior and experiences here directly, but rather by shaping an environment in which specific forms of action and behavior can come about. These influences can be located at a high level of abstraction, as Dorrestijn does when discussing how the development of spectacles indirectly contributed to the success of the printing press, because without spectacles a substantial part of the population would not be able to read. But they can also be seen at a more mundane level, that might have more direct relevance for designers, such as the influence that a reliable and easily reachable system of public transport will have on people's decision to take the car or the train to go to work. The fourth quadrant Dorrestijn distinguishes concerns the most abstract type of influence. Here, it is not *technologies* but 'technology' in general that influences the human being, with theories ranging from utopian optimism to dystopian pessimism. Because designers do not design 'technology' as such, but only concrete technologies, we will not take this quadrant into account.

In sum, three main points of application become visible here: technological mediations can be physical, cognitive or contextual. Technologies help to shape human-world relations through the physical-sensorial relation they create between humans and world, through the cognitive relation they create by giving information that can inform actions and decisions, and by creating a material and meaningful infrastructure that indirectly guides human actions and decisions.

Types of mediation

The second element of this heuristic tool for mediation analysis concerns the type of mediation that is involved. Earlier I made a distinction between coercive, persuasive and seductive forms of mediation (Verbeek 2009). Some technologies actually force their users to behave in specific ways, like speed bumps or automatic speed limiters that require car drivers to slow down. But not all mediating have this compelling form. To stick to the example of car driving: an econometer in a

car, which gives feedback about one's fuel consumption, does not force drivers to drive more economically, but rather persuades them to change their driving style. And the optical narrowing of roads – which is common in The Netherlands, where the central lane diving line on roads is replaced with two lines at the sides of the road to create two biking lanes – seduces people to drive more slowly.

Design researcher Nynke Tromp, who has extensively studied the various types of influences that technologies can have on human beings, suggests to categorize these influences along two dimensions (Tromp et al, 2011). One dimension indicates the force of the influence (weak versus strong), and the other its visibility (hidden versus explicit). This results, again, in four quadrants. While coercive influences are both explicit and strong, persuasive influences are explicit and weak: one can easily ignore the beep that suggests to wear a seat belt in a car, but one cannot avoid the effects of a speed bump. And while seductive influences are both weak and implicit, 'decisive' influences are implicit but strong. Placing a coffee machine in the hall of a company will seduce people to have more informal interactions, while deliberately designing a multiple-story building without an elevator implicitly but strongly decides for people that they will have to use the stairs. In sum: technologies can force, persuade, seduce, or decide for people.

Domains of mediation

The third and final element in a tool for mediation analysis is the domain of mediation. A first division of domains follows directly from the elementary framework of mediation theory: the existential versus the hermeneutical domain. Technologies help to shape how human beings are in their world, and how the world can be there for human beings; they mediate actions and perceptions, practices and experiences. But beside the existential / hermeneutic distinction, we can also distinguish between individual and social mediations. Technologies do not only help to shape the practices and experiences of individuals, but also inform social practices and frameworks of interpretation. As Jantine Bouma has shown (Bouma et al. 2009), communication systems in cohousing communities have an important influence on social practices in these houses, and digital whiteboards have important implications for practices of teaching and learning in classrooms. In all of these practices, obviously, there are individual human-technology relations at work as well, but the eventual mediating effect reaches beyond this individual level. When designing a whiteboard, not only the effects on the individual experiences and actions of teachers

and pupils is important, but also the effects on the learning process, the roles of teachers and pupils, et cetera.

The table below draws together all of these dimensions of mediation. Going through this table can help designers to stimulate their imagination and anticipate the possible mediating effects of the technology they are designing. Please note that the elements of mediation can be combined in all permutations. Physical, cognitive, and contextual impacts of technologies can have coercive, persuasive, seductive, and decisive forms, at the individual as well as the social level, and both in the hermeneutic and in the existential realm.

Locus	Form	Domain
Physical	Coercive	Individual: experience
Cognitive	Persuasive	Individual: actions
Contextual	Seductive	Social: frameworks of interpretation
	Decisive	Social: social practices

5. Assessing mediations

The second level at which designers can incorporate mediation in their work, is the level of 'mediation assessment'. Here, mediations are not only anticipated but also explicitly evaluated. To be sure: this activity of assessment should be seen as an element of the accompaniment of technology, not as an alternative to it. Such 'accompanying evaluations' of technologies can take

place by using an adapted version of the model that Berdichevsky and Neuenschwander developed for the evaluation of persuasive technologies (i.e. technologies that are designed to have explicit persuasive effects). Berdichevsky and Neuenschwander propose to evaluate persuasive technologies in terms of (a) the intentions of the designer; (b) the methods of persuasion used; and (c) the outcomes of the persuasion (Berdichevsky and Neuenschwander 1999). When we translate this to mediation theory, we could see persuasion as only one of many forms that mediation can take. Also, we can add a fourth dimension: besides intended mediation, there can also be implicit mediations, that occur without having been explicitly intended, but for which designers can feel partly responsible because they could have anticipated it.

To assess the quality of the mediations that are anticipated with the help of the mediation analysis tool described above, then, four steps become visible, which are inspired by Berdichevsky and Neuenschwander, but expand their framework to mediation theory (cf. Verbeek 2011, pp. 106-107):

- (1) The first step is rather obvious: if designers are explicitly working on a behavior-influencing technology, they could assess the *intended mediations* of the technology-in-design, i.e. the mediations that are deliberately designed into the technology. The central question here is: what arguments can be found in favor and against these intended mediations, and the intentions behind them?
- (2) More interesting, though, is the assessment of the mediations that are *implicit* in the design. The heuristic tool for mediation analysis that was elaborated above can serve as a basis for this. It enables designers to anticipate unintended mediations that the introduction of the technology might bring about. And therefore, it also makes these mediations open for moral discussion: what arguments can be given to support or avoid these mediations?
- (3) A third element in assessing mediations concerns the *forms of mediation* involved. As indicated above, mediations can be strong or weak, and explicit or hidden. In specific circumstances, specific forms of mediation might be more desirable than others. For many, seducing car drivers to slow down in specific zones without them explicitly being aware of it will be less problematic than secretly seducing customers to buy much more than they actually intended by means of subliminal stimuli, like emotion-evoking smells and colors.
- (4) Fourth, the eventual *outcomes* of the technological mediations – the actions and decisions that eventually get shape, as well as the social practices and frameworks of interpretation – can be assessed. All explicit and implicit mediations have effects, both at the individual and at the social

level. These effects might be radically different from the original intentions of the designer. Speed bumps, for instance, will not only mediate the driving behavior of car drivers, but can also attract skateboarders, whose activities do not necessarily enhance road safety.

6. Designing mediations²

The third and last level of the ethical accompaniment of design is the actual *design* of mediations 'into' technologies. Here, the role of designers becomes more socially invasive. Rather than checking for unwanted mediations, or explicitly assessing the implicit and explicit mediations involved in the design, they can also deliberately design-for-mediation. To which degree can this form of 'accompanying technology' be seen as morally desirable?

Moralizing technology

An approach like this was proposed in the 1990s already by Dutch philosopher Hans Achterhuis, in his article 'De moralisering van de apparaten' ('The moralization of devices'; Achterhuis 1995). Achterhuis argued that we need to end the constant moralizing in the environmental discourse. If all ideals of some environmental activists would be realized, even the smallest details of our existence would be subject to moral reflection, Achterhuis stated. The power of the lights we use in the house, the length of time we spend taking a shower, the fuel consumption related to our driving style – if everything becomes morally charged and subject to constant reflection, ordinary life will become impossible. Instead of moralizing each other, Achterhuis states, we need to start moralizing our technologies. We should delegate specific moral tasks and responsibilities to technologies, knowing that they are widely supported and that we are too weak and too limited to put all moral responsibilities on our own shoulders.

Achterhuis' approach was heavily criticized, especially because people were afraid that it would threaten human freedom. If we delegate morality to things, we seem to gamble with the crown jewel of humanity: our capacity to make autonomous decisions, and to take moral responsibility for our actions. I do not share this criticism, as might be clear on the basis of the first section of this article. When taking the phenomenon of mediation seriously, all human actions are

² This section incorporates reworked fragments from my article 'Politics at Issue: On Art and the Democratization of Things' (forthcoming in *Open: Cahier on Art and the Public Domain*, Fall 2012).

technologically mediated. Not the explicit design of mediating technologies is immoral from this point of view, but refusing to take responsibility for these mediations.

Libertarian paternalism

Precisely at this tension between being steered and maintaining autonomy, Richard Thaler and Cass Sunstein seem to have found an answer. In their book *Nudge* (Thaler and Sunstein 2008), they make a case for designing our material surroundings in such a way that it influences us in a positive sense without taking control away from us. A nudge is a tiny push, a small stimulus that guides people's behavior in a certain direction. Our material world is full of such nudges, Thaler and Sunstein claim, varying from photocopying machines with a default setting of single-sided copies to urinals with a built-in image of a fly to seduce men to aim for it. Thaler and Sunstein propose that we design these nudges in an optimal manner, so that we can guide our own behavior in directions that are widely considered beneficial.

The central idea in their approach is that human decisions are to a considerable extent organized and pre-structured by our material surroundings. When we make choices, two systems are at work in our brains, which Thaler and Sunstein call an 'automatic system' and a 'reflexive system'. Most of our decisions are made automatically, without explicit reflection. But for some decisions, we really have to stop and think: they require reflection and critical distance.

To a significant degree, our automatic system is organized by our material surroundings. To use one of Thaler and Sunstein's examples: when fried snacks are within reaching distance in a company's canteen and the salads are hidden behind refrigerator doors, it is very likely that many people will choose the less healthy food. The layout of canteens gives nudges in a certain direction. If we want to take responsibility for such situations, we must learn to think critically about nudges. If we can design them better, we in fact design our automatic system in a more desirable way. Thaler and Sunstein, therefore, call such design activities 'choice architecture': the design of choice situations. We need to rewrite the default settings of our material world.

But these activities of choice architecture should never close down the reflexive system. For Thaler and Sunstein, it is extremely important that nudges always remain open to reflection and discussion, and can move from the automatic to the reflexive system. This is why they indicate their approach as 'libertarian paternalism'. It is paternalistic because it explicitly exposes people to nudges in a direction that is considered desirable. But it is libertarian as well, because these

nudges can always be ignored or undone, in all freedom. Just like everyone is currently free to use both sides of the paper when copying, even though the standard setting is one side, no one should be forced to eat a salad and pass up the croquettes in a 're-nudged' cafeteria.

Thaler and Sunstein's way out of the dilemma between influencing behavior and respecting autonomy, therefore, is the "opt-out": by drawing on our reflexive system, we should always be able to move away from the nudges. Every act of paternalism is compensated by the explicit possibility to take a libertarian stance toward it.

The question, however, is whether this libertarian-paternalistic attempt to tame the morality of things is a real solution. Both the libertarian and the paternalistic elements of Thaler and Sunstein's approach take a separation of humans and technologies as a starting point, with a libertarian focus on saving human autonomy. On the one hand, nudges are the result of paternalistic human design, while on the other hand the people subjected to this paternalism always have the libertarian possibility of ignoring it. The phenomenon of technological mediation has no place whatsoever in Thaler and Sunstein's approach: nudges are discussed as instrumental interventions of paternalistic designers that can be either accepted or rejected by critical users. The boundaries between humans and technologies remain fully intact here; rather than critically *interacting* with technology to shape one's existence, the primary form of criticism is to *opt out*.

Design as Material Ethics

If libertarian paternalism is no real option, what could be the alternative, then? Do the boundary-blurring approaches of actor-network theory and postphenomenology simply urge us to accept paternalism? Should designers be allowed to steer our actions and decisions behind our backs, without giving us the possibility to reject these interventions in our lives? The answer to this question is both yes and no. Yes, because contemporary approaches in philosophy of technology show that human actions and decisions are *always* technologically mediated. There is no way in which designers could avoid having an impact on human existence. But at the same time the answer is no, because this impact is not necessarily exerted behind our backs.

Once we see the phenomenon of technological mediation, we can always develop a critical relation to it – in the Foucauldian sense of critique. Not to step out of the field of mediations, but to stand at the boundaries of that field, in order to find out which forces are exerted upon us, and how we can shape our own lives in interaction with these forces. While we cannot conceive of

ourselves as autonomous beings anymore, because of the fundamentally mediated character of our lives, we can still develop a free relation to these mediations. Without being able to undo or ignore all of them, we can critically and creatively take up with them. Being a citizen in a technological society requires a form of 'technological literacy'. Not in the sense that every citizen needs to understand all technical details of the devices around them, but in the sense that we develop a critical awareness of what technologies do in society.

This technological literacy on the part of technology users, then, is a necessary complement to the responsible ways of dealing with mediations that we can ask of designers. While designers need to take responsibility for the mediating roles of their products by anticipating, assessing, and designing the implicit and explicit mediations that are involved, users need to take responsibility for their own, technologically mediated existence.

Conclusion: an example

Let me conclude by giving an example of such a 'moralizing technology'. It is a telecare technology for patients with Chronic Obstructive Pulmonary Disease (COPD), a lung disease that dramatically reduces one's lung capacity, and that is potentially lethal. Patients with this disease need to have their lung capacity checked on a regular basis, and need to continuously adapt their activity pattern very carefully to the situation of their lungs and their physical condition. The problem for COPD patients is to find the right balance between training one's lungs enough to slow down the progression of the disease on the one hand, and not demanding too much of oneself on the other. The 'COPD.com'³ system aims to help patients to find this balance. COPD.com is primarily a 'disease management system'. Its basis is a so-called 'Body Area Network' that integrates various sensors to monitor one's activity level and one's physical condition. The data generated by this network are translated into a coaching program that is accessible via a web portal. Patients can log in, find information about their condition, and get advice about the optimal exercise they should have.

This system is quite invasive: it monitors one's activities in a detailed way, and advises patients about what to do. If ethical reflection would limit itself to assessing if this technology is morally desirable or not, the only relevant questions would be if this invasiveness stays within acceptable norms, and if the system is safe and reliable. From the point of view of mediation theory, though,

³ COPD.com is developed by Roessingh Research and Development, the University of Twente, and Medisch Spectrum Twente; for more information see <http://www.copddotcom.nl>.

the most interesting questions relate to the impact of the system on the daily lives of patients. A life with COPD takes on a new shape when COPD.com starts to play a role in it. Using the heuristic tool described above can result in various anticipated mediations. For instance: what does the continuous monitoring of one's activity level and one's physical condition do to the self-understanding of patients? The system could result in a far-reaching medicalization of people's lives. Also, responsibilities will start shifting. Rather than nurses and doctors, now the patient him- or herself becomes a central agent in the treatment of the disease. On the one hand, this enhances the autonomy of the patient, but on the other hand, it also makes patients more responsible when things go wrong. Third, the interaction between care giver and patient will change. Part of the work of the nurse will shift from having conversations with patients about their condition to making the system work optimally so that the system can have these 'conversations', for instance.

These are only a few examples of what a mediation analysis could reveal when designing a technology like this. An obvious next step, after assessing the quality of these mediations, would be to incorporate the results of this analysis in the design process itself. Designers could choose to make it impossible for patients to monitor their physical condition more than twice a day, for instance, in order to prevent an unwanted medicalization to occur. Also, protocols for using COPD.com could require a regular visit to a nurse or a doctor to ensure that all relevant aspects of the life of the patient are known, to keep up an interpersonal relation of care, and to see how the interaction between the patient and the system takes shape.

COPD.com is definitely a 'moralizing technology', in the sense that it actively prescribes its users how to behave. Patients who use the system, though, are not reduced to slaves of the machine. They can be helped by nurses and doctors to develop a critical relation to the system: to evaluate the quality of the advices it gives, to take the liberty to ignore or modify the advices the system gives, and to find a good way of dealing with the new lifestyle and self-image that the system introduces.

In a very modest and minimal way, this example shows how much ethical space there actually is on the lab floor. Rather than assessing this technology from an external perspective, by focusing on the question if it should be considered a desirable technology or not, the proposed method for 'accompanying technology' makes it possible to get involved in the practices of design and use that surround this technology. First, by making a mediation analysis of the COPD.COM it becomes visible in which ways this technology helps to shape the activities and experiences of COPD

patients and the people in their environment. Where does it exert its influences: is it physical, cognitive or contextual? Are the influences coercive, seductive, persuasive, or decisive? And to they take place at the individual level or at the social level? Second, this analysis can be the basis of a careful assessment of the quality of these various mediations: what are their effects, and how can these be valued? On the basis of this, a third step can be made: how to design desirable effects into the technology? How invasive can a technology be, without being experienced as an obstacle rather than an aid? How to design successful mediations?

Taking seriously the idea that technology and society continuously help to shape each other, therefore, does not imply the end of ethics, but rather a new beginning. The design of technologies itself has become an intrinsically moral activity. Responsible design requires the anticipation, assessment and explicit design of the mediations that the technology will introduce in society. Designing mediations is inevitably a jump into an unknown future, and will always have an experimental character. But by systematically anticipating and assessing the mediations involved in the design, we at least organize these experiments as responsibly as we can. Ethics on the lab floor does not only involve processes of scientific innovation, but also of technological design.

In order to give such an ethics of 'technology accompaniment' a firm basis, two strategies can be followed. On the one hand, further research needs to be done into the specificities of this accompanying form of ethics. By using empirical methods from sociology and cultural anthropology it becomes possible to investigate in which ways users and designers make technologies morally significant, in their practices and conversations. Such research can shed more light on the moral significance of technologies-in-design. Second, ethical reflection should be brought to the field of technology design. Design methods need to be expanded with tools for the anticipation, assessment and design of mediations. And, most importantly, design schools will need to teach their students that design is an intrinsically social activity, in which designers should learn to take responsibility for the ways in which they intervene in society.

References

- Achterhuis, H. (1995), 'De moralisering van de apparaten'. In: *Socialisme en Democratie* 52 nr. 1, 3-12.
- Berdichewsky, D. and E. Neuenschwander (1999), 'Toward an ethics of persuasive technology', *Communications of the ACM*, 42 - 5, p.51-58, May 1999
- Bouma, J.T., A.I.M Voorbij and W. A. Poelman (2009) 'The influence of changes in the physical and technical design on social interactions in a cohousing community'. In: E. Durmisevic (ed.), *Lifecycle Design of Buildings, Systems and Materials*. Enschede: International Council for Building Research Studies and Documentation (CIB) and the University of Twente, the Netherlands
- Dorrestijn, S. (2012) *The Design of our Own Lives*. Enschede: University of Twente (dissertation)
- Foucault, M. (1997). *Ethics: Subjectivity and Truth* (ed. P. Rabinow). New York: The New Press
- Foucault, M. (1997a). "What is Enlightenment?". In: M. Foucault, *Ethics: subjectivity and truth*, edited by Paul Rabinow. New York: The New Press
- Foucault, M. (2010). *The Government of Self and Others: Lectures at the Collège de France 1982-1983* (edited by Arnold I. Davidson, translated by Graham Burchell). New York: Palgrave Macmillan
- Heidegger, M. (1971). 'The Origin of the Work of Art', in: *Poetry, Language, Thought*, trans. J.M. Anderson and E.H. Freund. New York: Harper & Row.
- Hottois, G. (1996). *Entre symboles et technosciences. Un itinéraire philosophique*. Seyssel (Paris): Editions Champ Vallon
- Ihde, D. (1990). *Technology and the Lifeworld*. Bloomington/Minneapolis: Indiana University Press
- Latour, B. (1993). *We have never been modern* (trans. C. Porter). Cambridge, Harvard University Press. (translation of: *Nous n'avons jamais été modernes*, Parijs: La Découverte, 1991).
- R. Thaler and C. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (New Haven: Yale University Press, 2008).
- Tromp, N., P. Hekkert and P.P. Verbeek (2011). 'Design for Socially Responsible Behavior: A Classification of Influence Based on Intended User Experience'. *Design Issues* 27:3, pp. 3-19.

- Verbeek, P.P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Pennsylvania State University Press.
- Verbeek, P.P. (2009), 'The Moral Relevance of Technological Artifacts'. In: M. Duwell e.a. (red.), *Evaluating New Technologies*. Springer, Dordrecht, pp. 63-77
- Verbeek, P.P. (2010), 'Accompanying Technology: Philosophy of Technology after the Ethical Turn'. In: *Techné: Research in Philosophy and Technology* 14:1, pp. 49 - 54
- Verbeek, P.P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago and London: The University of Chicago Press.