

Chapter 3

Remote control

Human autonomy in the age of computer-mediated agency

Jos de Mul and Bibi van den Berg

‘Je est un autre’

(Rimbaud 1871)

Introduction

Human beings have always used instruments, media and machines to strengthen and expand their agency. These technologies enable them to have ‘remote control’ over both the natural and human world. Technological extensions serve to increase the ‘action radius’ of human autonomy. They enable us to do things we couldn’t do without them: writing makes it possible for us to delegate our memories to clay tablets, papyrus or paper. Pulleys facilitate lifting things that are far too heavy for our human bodily constitution. Telephones and e-mail enable us to be socially present in places while being physically absent from them (cf. Gergen 2002). Gamma knives allow us to target brain tumors with high doses of radiation therapy without affecting (much of) the surrounding tissue. And the Mars Exploration Rover enables us to gain insight into the geological history of Mars under circumstances that are physically impossible to survive for humans.

However, as the human life world transformed from a ‘biotope’ into a ‘technotope’ in modern culture, a fear emerged that human beings would become dependent on, or even slaves of technology (cf. Ellul 1988, Heidegger 1962). This dystopian perspective of the technological world is all the more worrying, to its adherents’ mind, because the *responsibility* for that world and what happens in it is still in the hands of human beings and not in the hands of the technologies. After all, human beings are the architects, designers and users of technologies, and for that reason they are responsible for their creations and their creations’ output.

With the advent of ‘autonomic computing’ – ubiquitous computing, Ambient Intelligence, pervasive computing, expert systems, artificial intelligence, artificial life, converging technologies, etc. – it seems that we can no longer understand these matters in a merely metaphorical sense. Autonomic computing appears to mark the transition into a phase in which technologies actually gain *agency* and become a potential threat to human autonomy.

In this chapter we will argue that this fear is excessive, because it starts from a misleading opposition of human agency and technical artefacts. Discussing the intimate relationship of man and technology, we will develop a notion of autonomy that focuses on the concept of ‘*remote control*’. We will argue that autonomic computing does not necessarily form a threat to our agency, but that, quite to the contrary, it may *strengthen* it. Note that we do not claim that autonomic computing *necessarily* strengthens human agency and autonomy. The most pressing question, we argue, is not whether autonomic computing strengthens human agency or not, but rather *under which circumstances* it does, and under which circumstances it threatens human agency. We will investigate this question by discussing a number of real and fictional cases dealing with increasingly more radical instances of ‘autonomic voting’.

Electoral compass(ion)

Perhaps one of the social phenomena in which we express our human autonomy most explicitly is that of democratic elections. In elections our choices, made freely and on the basis of (rational) arguments, may contribute to the maintenance and management of our society. During elections we must use (explicit) reasoning to choose which political programme we approve of most, which concrete policies we endorse, and which political ideals we would like to see realized.

As is the case in many Western countries, in the Netherlands this is no easy feat: there are numerous political parties and whoever makes it his explicit goal to choose responsibly, must have access to the right information (both in terms of channels and in terms of content) with regard to the political agendas of all these parties. Thankfully, there are many ways to go about getting this information: electoral meetings, paper and online party programmes, flyers, websites, commercials on television and message boards on the street, election debates on television, etc. However, despite all these sources of information, research shows that there are relatively few Dutch citizens who inform themselves elaborately and take the time to study all party programmes.¹ Of course many reasons can be cited for this behaviour, ranging from lack of interest to laziness, and from feeling politically underrepresented to downright rejection of the democratic system as a whole. Another reason, which is less often cited, may be information overload. Who has the time to read all of these party programmes, to watch every election debate, to go to every political rally? This is the reason why many voters vote based on their gut feelings, or don’t vote at all.

This phenomenon was clearly visible in the democratic elections held in the Netherlands for the ‘Water Boards’ in the fall of 2008. In a country of which an important part of its landmass is below sea level the Water Boards are *literally* of vital importance. The Boards’ task is to maintain dams and dykes and protect the country from both seawater and river flooding. The Water Boards are one of the oldest institutions in the democratic system of governance in the Netherlands – the first Water Board was created in 1122 AD, so the Boards have a long history

indeed. The Water Boards literally form the backbone of the Dutch ‘polder model’: traditionally each Board regulates the water maintenance for a specific region of the country. This means that their main responsibility is to guarantee three goals: flood defenses, preserving water quality, and managing the general water economy of the region.

Until recently the Water Boards focused predominantly on technical management. Lately they have politicized to some extent. As a political institution the Boards used to consist of *individual* elects, who could be chosen as members of the Board for a four-year term. In the elections of November 2008 political parties were introduced for the first time in all 26 Boards nationwide. Partially, these consisted of the traditional political parties that also populate the Dutch parliament, such as the Christian democrats (CDA), the (neo)liberals (VVD) and the labor party (PvdA). But there were also two new national parties, which can only be elected in this specific election: Water Natural (‘Water Natuurlijk’) and the General Water Board Party (‘Algemene Waterschapspartij’). The central themes in this election were the projected rising sea level as a result of global warming on the one hand, and environmental conservation on the other hand – two themes that do not necessarily go well together, and that even stand in serious opposition in many cases.

Despite the fact that the political importance of an institution such as the Water Boards has increased considerably in light of both of these themes, the number of voters for this election turned out to be dramatically low. Only 24% of the adult population voted (of which, notably, a marked number turned out to be unlawful, because the voting bill was too complex and voters did not fill out the form correctly). The authors of this chapter also struggled with the question of what to vote, but both felt it was their democratic duty to take part in the elections. They turned to ICTs for help on the matter.

Kieskompas enterprise, a Dutch private enterprise in which entrepreneurs and social scientists from the VU University of Amsterdam collaborate, designed a unique Electoral Compass for every Water Board. The Compass consisted of 36 theses in 12 categories: taxes, governmental innovation, democracy, dykes and roads, economy and environment, energy and climate change, nature and recreation, plants and animals, polders and landscape, water, floods, and living and working. These theses had to be valued on a five-point scale (completely agree, tend to agree, neutral, tend to disagree, completely disagree – alternatively, users could choose ‘no opinion’). After completing the questionnaire, one could see one’s position in the political landscape as it aligned with the parties participating in one’s own Water Board, and thus find out which party would represent one’s preferences best. The political landscape was represented along two axes, one ranging from ecology to economy, the other from a broad to a small range of responsibilities for the Water Boards. The Electoral Compass also provided information about the history and tasks of the Water Boards, and voters were given information regarding the policy proposals of the participating parties. It turned out to be very popular, not only in this election but in any of the democratic elections for which it was developed (elections for the national parliament, local

elections, and even the US presidential election). It did, in fact, help us to make a choice for a party in the Water Board elections.

What is interesting about the case of the Electoral Compass as a technology is this: in its current form we delegate the task of delving through all the information involved in a single election to the technology, but the individual voter remains in control of the final casting of the vote. The technology sifts through all the information to provide us with an easier choice, which means that it has taken over the *process* of collecting and weighing all the information, yet the voter himself makes the final decision – the goal of voting for an election is still in the hands of the human agent.

The fact that the Electoral Compass leaves the final decision up to the voter is one of the main differences with another version of this type of technology in the Netherlands, called VoteMatch (StemWijzer). The latter, which has been developed by the Dutch Centre for Political Participation (Instituut voor Publiek en Politiek, IPP), gives a clear voting advice, whereas the Electoral Compass does not. The Electoral Compass only aims at providing the user with an easy overview of his personal alignment in relation to various parties in one election. Another difference is that VoteMatch only offers three options to the theses presented (agree, disagree, don't know). On the other hand, VoteMatch provides the possibility to indicate the relative weight of each topic whereas the Electoral Compass does not. Yet another voting aid, the ChoiceAdviser (KiesWijzer), does not present theses, but questions to the user. There are only ten of them, and the user has to choose between three different answers (and 'no opinion'). After answering the questions, the user is shown the amount of affinity he has with each of the political parties, and with a table with the answers of all parties to these questions.

After the introduction of these technological voting aides each of them received critiques regarding the type of questions they pose, their objectivity and transparency, the way they present results, and the type of advice they provide. The Electoral Compass has been accused of being one-sided and even strongly biased, because it favours parties in the centre of the political landscape. We will come back to this below. What became clear is that technological voting aides are *never fully neutral*, but always reflect, to a certain extent, the technological, methodological and political decisions that have been made by the designers of these aides.

Such biases have been researched extensively in relation to various technologies under the name of 'scripts'. Three different meanings of the term script can be distinguished in relation to technologies (Van den Berg 2009). First, there is the idea that designers implement ideas about users and prospected contexts of use into the technologies they design. For instance, conceptions of users in terms of gender may result in product designs that are not only very different for men and women, but also *express* this difference in their materiality. Van Oost has shown that shavers for women and for men are different in shape, in the buttons they have, and in the ways they can and cannot be used. She concludes that the designers and manufacturers of these products not only sell shavers for men and women, but also affirm and reify gender (Van Oost 2003: 207). This line of research has come to

be known as ‘script analysis’ and over the years has come to play an important role in Science & Technology Studies (cf. Akrich 1992, 1995, Berg 1999, Gjøen and Hård 2002, Latour 1992, Van Oost 2003).

Second, in artificial intelligence the term script refers to the human ability to quickly and easily come to understand a wide array of everyday, recurring and ritualistic ‘scenes’ and to know how to act in them (Schank and Abelson 1977). For instance, when entering a restaurant, we instantly know that a cycle of actions such as ‘finding a seat’, ‘reading the menu’, ‘ordering food’, ‘eating the food’ and so on can be expected. Human agents apparently possess a type of knowledge, called scripts, to deal with such standardized, regularly occurring sequences of action related to specific scenes. Artificial intelligence research aims at making such scripts explicit, so that they can be mimicked by computer technologies, which as a result, the argument goes, will lead to smarter, more life-like machines.

Third, artefacts *themselves* may act as scripts, sometimes as intended by their designers, and sometimes unintentionally. This is evident for instance in the way in which a groom on a door, that makes the door harder to open, discriminates against specific groups of users, most notably older people, children, and people carrying things in their hands (Latour 1992).

Technological artefacts have the interesting characteristic that they often *contain* scripts in the first two senses discussed here, but *are* scripts themselves (the third sense) as well (Van den Berg 2009). In the example of voting aides discussed here this is evident. First, as we have seen one of the critiques against voting aides is that they contain the designers’ ideas – both political opinions and conceptions of users – that may influence the opinions and voting behaviours of users. For instance, the Electoral Compass, developed by the VU University in Amsterdam, a protestant university, we have seen above, was accused of favouring parties in the political centre, and most notably the Christian democrats (CDA). The designers’ political preferences are thus scripted into the Compass’s design. These scripts embedded into the software may (implicitly) steer the voters’ eventual choices. This is the meaning of scripts in Science & Technology Studies. Second, the voting aide uses ideas about the way in which people generally tend to choose between different bits of information given to them and how they value these various offerings. These ideas are translated into algorithms that thus represent scripts in the second meaning of the term, as presented by artificial intelligence. Third, the voting aide *itself* has a script force in the sense that it steers people in certain directions. When the user completes the answers to the questions, he will feel he has every reason to believe that the choice presented to him by the voting aide represents, or even *is* in fact, his own choice. Particularly this last point is interesting in light of developments in the direction of autonomic computing, and therefore we will come back to it more extensively below.

We may conclude that all voting aides ultimately raise the same question: do such technologies threaten or undermine our autonomy? Ought we not to decide for ourselves, as individual human agents, what party we want to vote for? After all, as such voting aides contain hosts of scripts and choices, they affect our

behaviour and therefore guide us in directions that we may not have chosen, had we taken the time and put forth the effort to gather all the information required to vote responsibly for ourselves. In short, *voting aides seem to affect our (political) autonomy*. Surprisingly, this effect on our autonomy applies to the very domain in which human beings claim to express her most strongly: the design and use of technological artefacts.

Autonomy and distributed agency

Our modern Western (liberal, secular, scientific) culture values human autonomy as one of the pillars – if not the most important pillar – of agency.² Some argue that this emphasis can be traced back to the philosophy of René Descartes (Gontier 2005). In Descartes' philosophy the rationality of human agency is not only underlined, but even *assumed* as the starting point from which we ought to understand the reasons that motivate human action. Over time this assumption in Descartes' work has been criticized with various arguments. One line of criticism points towards the fact that there are numerous *other* factors that motivate human agency apart from rationality, and that motivate human agency to such an extent that our rationality only has a fraction of the importance that we generally ascribe it. Human agency, some of these critics say, is predominantly the result of our genetic makeup, or of our upbringing, as others claim, or of our social class, or our gender, as yet others argue (Rosenberg 2008). These lines of criticism all lead to harder or softer forms of determinism – human agency, in their view, is largely the result of forces that are either biologically or socially determined, or a mixture thereof, and that, in any case, fall outside the sphere of complete control of the individual agent. Contra Descartes it is generally held nowadays that human agency is the result of a *complex* of various factors, which includes both our rational deliberation, and a set of 'natural' forces such as our genes and passions, and of 'nurture' forces such as our upbringing, social class, gender, etc.

A second line of criticism confronting the Cartesian emphasis on the rationality of human agency starts from the factual claim that in the vast majority of our actions rational deliberations are at best implicit, but more often than not wholly absent. When rational deliberations *do* play a role in our actions, most of the time these deliberations emerge retrospectively – one formulates reasons for one's actions *post actio* rather than *pre actio*, or even *in actio*. Apart from a longstanding tradition in philosophy that debates the alleged freedom of choice that human agents have, research in neurosciences has empirically shown that there are many instances in which our brain have already 'made a decision'³ before we even become *aware* of what we want to do (Burns and Bechara 2007, Libet 1985).

These two lines of critique have given rise to a postmodern denial of the existence of autonomy and free will, both in the natural sciences (cf. Dawkins 2006) and in the humanities (Nietzsche and his postmodern heirs). However, we argue that this extreme position takes matters too far. In our everyday lives we experience ourselves *as agents* with *some level of control and freedom of choice*, even

if we are willing to grant the postmodern suspicion that our levels of control are far from complete, and that our freedom of choice may often be informed by motives and processes that are either unknown to us or principally not (completely) insightful for us. Our *experience* of ourselves *as agents with some level of control* is ignored or denied by both the natural scientists and the postmoderns in the humanities. However, we feel that this first-person perspective needs to be taken seriously, as it seems real to us in everyday life, despite the limitations of its reality as pointed out by these scientists. Obviously, having *incomplete* control does not imply having *no control at all*.

As we have seen above, human agency is the result of a number of factors combined, including biological, social, and cultural ones, and including our rational facilities for deliberation. What we want to emphasize is the fact that, although we have limited control over the forces that motivate our action and the elements that become part of our identities, human agency has a distinctive ability to *affirm* its actions in a unique manner: via a ‘*reflexive loop*’. This reflexive loop enables us to view and judge the internal and external forces that motivate us *as forces that motivate us* to act, and to affirm and embrace those actions *as our own* (or, alternatively, to distance ourselves from them). One could argue that what happens in the reflexive loop is that we **leave our** first-person perspective and take a third-person, a *remote* perspective towards ourselves and our own actions. In this remote stance we gain a certain amount of freedom towards the forces that drive us.

The human constitution varies from that of **a plant** in the sense that a human being not only *is* a body (as is a plant), nor *is and has* a body (as does an animal), but *is, has,* and simultaneously can always *relate to* its body from an external position. Or, to phrase it in experiential terms: ‘Man not only lives (*lebt*), and experiences his life (*erlebt*), but he also experiences this experience of life’ (Plessner 1981: 364, also see De Mul 2003). This **fact** is precisely the reason why we are always engaged in a reflexive loop: we **can** view and judge our actions from a distance – though we need not always do so.

A classical example of the workings of this reflexive loop, and of autonomy in relation to actions, can be found in Euripides’ tragedy *Medea* (Euripides 2006). In this Greek tragedy we encounter Medea, who has been left by her husband Jason for a younger woman. She is furious, feels utterly humiliated and therefore seeks revenge. Tormented by conflicting emotions she struggles to weigh up options for vengeance and in the end she chooses to kill their children to get back at Jason. *Medea* is often cited as the first example of the expression of free will. In fact, this interpretation is at best one-sided. Medea does not express free (Cartesian) agency – rather, she is motivated by various forces, and is ripped apart between clashing emotions: on the one hand the hatred she experiences with respect to her (ex)husband Jason, who has left her in the most humiliating and abominable way, and on the other hand the love of her children. Both of these emotions battle for dominance within the person of Medea. Medea, therefore, is not ‘free’ in the sense that she can make a decision based on pure rational deliberation. One could even argue quite the opposite, with the postmoderns above: that Medea does not have

agency at all, because she is ruled by her passions (or ‘daemons’, in the language of Euripides).

However, this is not the case either. Medea *is* in fact an agent, because in the process of struggling with these clashing forces inside her, forces that pull her to this side and that, she makes a decision and affirms one force at the cost of another. She ends up embracing her hatred for Jason as the main motivator for her actions and thus decides to kill her children. While this horrible death may make her an unsympathetic character, we have to grant Medea the fact that she does take responsibility for her *daemon*. She embraces the action she completes. She identifies herself with her action, despite the fact that it originated in an overwhelming force over which she had little control. Therefore, Medea is in fact a good example of what one could call ‘responsibility without freedom’ (Alford 1992). And this, we argue, is precisely the minimal requirement of what it means to have human agency (cf. De Mul 2009: 179–244).

Human agency, we argue, is based on *reflexive remote control*: our actions are remote(ly) controlled, that is, they are motivated, stimulated, challenged, and shaped, by countless internal and external forces, but as reflexive beings we simultaneously exert remote (self-)control *via* the forces that motivate us. This is comparable to the zapper handling a remote control in front of the television: although he has a rather limited influence on the choice of television shows that are on TV, nevertheless handling the remote control enables him not only to make choices with regard to which shows and channels he wants to watch, but more importantly: he is responsible for his choices, for the self-reflexive cycles they engage, and therefore for the ‘bricolaged’ identities he zaps together.

What the example of Medea shows is that external and/or internal forces outside our deliberative faculties may limit our human autonomy. In extreme cases, such as blind anger or senseless panic this is indeed the case. In many other cases, however, we find forms of externalization that don’t undermine, but rather enhance human autonomy instead. Dilthey has shown that, contrary to the rationalistic, introspective tradition instigated by Descartes, our *lived experiences* (*Erlebnisse*) more often than not are far from transparent. Our thoughts, motives and feelings often remain implicit and we only get to know them or gain insight into them in the process of *expressing* them (*Ausdruck*), that is: in speaking, in the language we use, in our actions, in the clothes we wear, in the laws we write and adhere to, in the institutions we construct and embrace, etc. Implicit meanings, ideas and feelings are articulated in our expressions, and thus instigate an *understanding* (*Verstehen*) of ourselves, of our motives, and our drives. Dilthey explains this autonomy-enhancing reflexive loop as follows:

An expression of lived experience can contain more of the nexus of psychic life than any introspection can catch sight of. [. . .] In lived experience we grasp the self neither in the form of its full course nor in the depths of what it encompasses. For the scope of conscious life rises like a small island from inaccessible depths. But an expression can tap these very depths. It is creative.

[Finally] it is the process of *understanding* through which life obtains clarity about itself in its depths [. . .] At every point it is understanding that opens up a world.

(Dilthey 1914–2005: 206, 220, 87, 205, also see De Mul 2004: 225–56)

Technologically mediated agency

In the evolution of the human life form cognitive artefacts have played a crucial role in the reflexive loop of lived experience, expression, and understanding. The act of writing is a good example. Since the so-called ‘mediatic turn’ in the humanities – initiated by McLuhan and his Toronto school – much attention has been paid to the fact that writing is not just a neutral instrument to express thoughts, but structures and enhances human thought in specific ways. In his book *Orality and Literacy* Walter Ong has argued that the transformation of oral cultures into writing cultures opened a whole new domain of human agency and culture:

Without writing, words as such have no conceivable meaning, even when the objects they represent are visual. They are sounds. You might ‘call’ them back – ‘recall’ them. But there is nowhere to ‘look’ for them. They have no focus and no trace (a visual metaphor, showing dependency on writing), not even a trajectory. They are occurrences, events. [. . .] By separating the knower from the known, writing makes possible increasingly articulate introspectivity, opening the psyche as never before not only to the external objective world quite distinct from itself but also to the interior self against whom the objective world is set.

(Ong 1982: 31, 105)

Without the use of these very fundamental artefacts, which we may call ‘external devices of reflection’, humans as we know them would not exist. The reverse, of course, is true as well: it is human beings that create and develop artefacts, and that interpret them *as artefacts* in their use, hence without them these artefacts would not exist, neither practically nor ontologically.

Interestingly, initially writing met severe criticism, because it was not so much understood as an autonomy-enhancing technology, but rather as a threat to human autonomy. For example, in *Phaedrus* Plato critically discusses the invention of writing and what he conceives to be the downfall of both oral culture and human memory (Plato 1914, *Phaedrus* 275A). In this dialogue the Egyptian king Thamus argues that writing will eliminate the human capacity to remember, because humans will forget to practise their memories. By delegating human abilities to technological artefacts, Thamus reasons, humans will lose powers, capabilities, sources of agency. The underlying theme voiced by Plato in this dialogue is a perspective regarding human *autonomy* in relation to technological artefacts – one that has been echoed many times over in recent decades with regard to the delegation of cognitive tasks to computer technologies. In these modern variants

of Plato's argument the central line of reasoning is not that the *products* of our thinking are delegated to technological artefacts, as was the case in *Phaedrus* discussion on writing and memory, but even worse: important parts of the rational and moral *process* of thinking itself is delegated to computers. According to these critics, this will lead to an undermining of human autonomy.

However, Plato's argument in *Phaedrus* can easily be countered, and the same applies to its echoes in modern times. What both of these versions of the 'extension argument' overlook, is the fact that technological artefacts, though they are not part of our organic body, are an integral part of our distributed cognitive structure (cf. Magnani 2007: 5–6). They remain part of ourselves, as the artefacts are part of the conjoint network in which we operate and act with them. Moreover, what the extension argument misses is the fact that technological artefacts, in adopting and reconfiguring certain tasks from human beings, facilitate the development and blooming of all kinds of new 'typically human' capacities. By delegating the content of our memories to paper (in writing), our cognitive structure is less burdened with the task of remembering, and thus new roads are opened for the development of novel forms of rationality, structured by the medium-specific characteristics of writing. This same mechanism applies to delegating the process of rational thinking to computer technologies. Such delegation does not lead to a diminishment of human autonomy, but to an increase of human agency, and as such, to an expansion and strengthening of human autonomy. In a sense the more agency an artefact has, the more it potentially *enhances* human autonomy by inviting us to reach new goals and use new means.

The critique of Plato and his modern heirs starts from a dichotomous distinction between human agents and technological artefacts. This distinction is problematic, because human beings and artefacts have always and will always form *networks*, in which each mutually depends on the other (Latour 1993, 1999, 2005, Magnani 2007). Neither can exist without the other – a human being is not a human being without artefacts, nor is an artefact an artefact without human beings.

When we *do* distinguish between human beings on the one hand and artefacts on the other (either analytically or in practice), claiming that human beings are active whereas artefacts are passive is an obvious oversimplification. As we have argued in the preceding section, human beings have never been very autonomous. A considerable part of our actions is remote controlled by both internal and external factors that are outside our sphere of control. Our human agency is not a completely autonomous (*self-governing*) power, but rather a reflexive relation to that which motivates our actions – a relation, moreover, in which we can choose to affirm and absorb these motivating forces as *our* motives, drives, passions, ideas. Only in the interplay of our internal and external motivators on the one hand and our own reflexive appropriation on the other do we, as acting beings, as agents, emerge. Human agency, in this sense, has always been distributed agency.

Bruno Latour has raised a similar argument regarding the moral implications of technological mediation. Latour rejects the assumption that human ethics formulates moral goals, whereas technology merely supplies the means to realize these

goals. Technology always provides us with a *detour* towards the goals we want to reach:

If we fail to recognize how much the use of a technique, however simple, has displaced, translated, modified, or inflected the initial intention, it is simply because we have *changed the end in changing the means*, and because, through a slipping of the will, we have begun to wish something quite else from what we at first desired. [. . .] Without technological detours, the properly human cannot exist. [. . .] Technologies bombard human beings with a ceaseless offer of previously unheard-of positions – engagements, suggestions, allowances, interdictions, habits, positions, alienations, prescriptions, calculations, memories.

(Latour 2002: 252)

Beyond human agency

In the traditional conception of technologies, one could argue, we conceive of our own relationship towards these technologies as follows: we, as human beings, formulate one or more *goals* or outcomes we want to achieve, and we then proceed to create technologies to reach those goals. We are in charge of the outcomes of technologically mediated praxes, and we provide the technology with the *processes* to go about reaching the goal we have set. Technologies are thus viewed as simple instruments, with which we have a clear goal-means relationship.

In fact, our relationship with technologies is much more complex and diversified than that (cf. De Mul 2009: 245–61). While we do in fact create *some* technologies for which we define both the goals and the processes of reaching those goals, there are also examples of technologies for which we define *only* the outcomes. The process of reaching those outcomes is left to the artefact itself. For instance, in modern cars, when we press the brake to make it stop, the brake system, consisting of independently operating sub-systems, cleverly ‘decides’⁴ which systems it needs to engage in those specific circumstances to make the car stop. Moreover, in some cases not only the *process* of accomplishing certain goals is left to the technology, but the definition of the *outcome* itself as well. Both the goals and the process are thus delegated to the technology. This is the case, for instance, in the ‘power grid’, a network of power supply systems that manage the power distribution in Southern California. The grid decides how to distribute power optimally (process), but also defines what the best outcomes of distribution are (goal). It is clear that all three of these forms entail different relationships between human beings and technological artefacts, and have consequences for human autonomy.

At the beginning of this chapter we discussed our recent voting experiences surrounding the elections for the Water Boards in the Netherlands. We described the use of the Electoral Compass, designed and deployed to relieve us of the burden of having to muddle our way through large amounts of electoral information, ranging from flyers and websites to television debates and party programmes. We

concluded that what happens when we use a technology such as the Electoral Compass is a delegation of the *process* to the technology, whereas the *goal* – that is, the final decision of who to vote for and the actual casting of the vote – remains firmly in the hands of the human agent. It is the individual, autonomous agent, who uses the information provided by the technology, but weighs and decides for himself. In the case of the Electoral Compass, therefore, we would be hard pressed to argue that using this technology undermines our autonomy.

As we have seen, one could argue that matters are a little different in the case of one of the other voting aides we discussed. In the case of VoteMatch, the technology does not provide us with an overview in numbers of our alignment to each of the parties we can vote for, but with an explicit *advice* instead. The technology clearly points us in the direction of a specific vote, and this raises the question of influencing. How many of us would be recalcitrant and daring enough to ignore the advice and vote for an entirely different party? How many of us would be aware of the fact that they do not *have* to accept the advice? How many of us would think critically about either the content of the advice, or the ways in which it has been constructed? We have seen above how many faceted the scripting of this kind of technologies really is. One could argue, therefore, that in this second case the decision to vote for this party, rather than that one, is indeed delegated to the technological artefact in a way that diminishes (although it doesn't eradicate) our human autonomy.

Now, think of the following scenario. What if, in the near future, a voting aide would be able not only to provide us with an overview of our political alignment to various parties, or with a clear and concise, boxed-in, ready-made voting advice, but would also be able to take things one step further? What if the voting aide could also cast the vote for us? Imagine that we, as autonomous agents, would be too busy, too lazy, too cynical or otherwise engaged in any other form or shape, to put forth the effort of actually casting the vote on the designated election day. What would that entail for our autonomy? If the voting aide would consult us in a final decision on who to vote for before it actually casts the vote, this seems fair enough. After all, it doesn't actually matter (or does it?) who presses the button on the voting machine, or who handles the red pencil to fill out the form. What matters is that *my* decision as a voter (which, notably, I may have come by through my own rational or not-so-rational electoral choices, or with the more or less neutral help of a voting aide) is the final decision in this fundamental democratic practice. My human autonomy (or what is left of that after accepting VoteMatch's advice as *my* voting decision) is safeguarded as long as I have the final say.

But what if we take it yet another step further? Imagine a world in which the voting aide does all of the things discussed before, but on top of that it can also vote for us *in an independent way*. This means, for instance, that despite the fact that we have always voted for party X (and may even have told the aide explicitly to do so, this time, for us as well), it may decide that it has good grounds to ignore our voting history and our current voting preference, and to vote for an entirely different party *in our name*. Let's assume that it doesn't choose to do so out of

spite or confusion or any ‘irrational’ or non-benevolent motivation, but that it merely uses profiling to calculate that, even though we always thought that X was the party for us, in reality our ideas and our behaviours match Y much more closely, and therefore Y is the party we *should* vote for. In this scenario Y is not only the party we *should* vote for, but the one we effectively *do* vote for, because the voting aide will cast the vote on our behalf to its best judgement and without our explicit consultation. It is obvious that, in this scenario, our human agency is indeed seriously affected by the process of delegation to a technological artefact, and that our remote control is undeniably extremely remote in this case.

Now, it is easy to cast aside a scenario such as this with the argument that it is futuristic (ergo unrealistic) ‘what-if babble’. However, look at the following scenario, which we have clipped from one of the key documents presenting the European Commission’s perspective of the near technological future, called Ambient Intelligence:

It is four o’clock in the afternoon. Dimitrios, a 32 year-old employee of a major food-multinational, is taking a coffee at his office’s cafeteria, together with his boss and some colleagues. He doesn’t want to be excessively bothered during this pause. Nevertheless, all the time he is receiving and dealing with incoming calls and mails. [. . .] Dimitrios is wearing, embedded in his clothes [. . .], a voice activated ‘gateway’ or digital avatar of himself, familiarly known as ‘D-Me’ or ‘Digital Me’. A D-Me is both a learning device, learning about Dimitrios from his interactions with his environment, and an acting device offering communication, processing and decision-making functionality. Dimitrios has partly ‘programmed’ it himself, at a very initial stage. [. . .] He feels quite confident with his D-Me and relies upon its ‘intelligent’ reactions.

At 4:10 p.m., following many other calls of secondary importance – answered formally but smoothly in corresponding languages by Dimitrios’ D-Me with a nice reproduction of Dimitrios’ voice and typical accent, a call from his wife is further analysed by his D-Me. In a first attempt, Dimitrios’ ‘avatar-like’ voice runs a brief conversation with his wife, with the intention of negotiating a delay while explaining his current environment. [. . .] [However, when she calls back once more] his wife’s call is [. . .] interpreted by his D-Me as sufficiently pressing to mobilise Dimitrios. It ‘rings’ him using a pre-arranged call tone. Dimitrios takes up the call with one of the available Displayphones of the cafeteria. Since the growing penetration of D-Me, few people still bother to run around with mobile terminals: these functions are sufficiently available in most public and private spaces [. . .] The ‘emergency’ is about their child’s homework. While doing his homework their 9 year-old son is meant to offer some insights on everyday life in Egypt. In a brief 3-way telephone conference, Dimitrios offers to pass over the query to the D-Me to search for an available direct contact with a child in Egypt. Ten minutes later, his son is videoconferencing at home with a girl of his own age, and recording this real-time translated conversation as part of his homework. All

communicating facilities have been managed by Dimitrios' D-Me, even while it is still registering new data and managing other queries.

(Ducatel et al. 2001: 5)

This scenario is not about a voting aide, nor about our actions as autonomous voting agents, but it does portray a number of relevant parallels with the last stage of the voting aide we have discussed above. The man in the scenario has a personal technological aide that answers his incoming communications whenever he is otherwise occupied. Although this sounds quite appealing, and not even so uncommon at first – most of us use answer phones and automatic e-mail replies for precisely the same goal – there are two rather eerie elements to his aide's capacities and behaviours. First, the aide has been given the responsibility to decide whether incoming messages are important or not. Note that an answer phone or an automatic e-mail reply are entirely indiscriminate in this respect. The aide thus makes decisions based on *its* estimate of the importance of the content of the message and the nature of the relationship one has to the caller. This means that it *values* our communications for us *and acts* on the basis of these values. Second, what is eerie about the aide in this example is that it *mimics its owner*. It responds to incoming communications using an imitation of its owners' voice, including inflections and word choice. This means that we do not only delegate the process of valuing to the artefact, but also the *form and content of our* [sic] *response*. And these are precisely the same two issues that are at stake in the scenario we've sketched for the voting aide of the future.

Delegating agency to artefacts is something human beings have been prone to do since the beginning of time. No harm is done in most of our delegations – quite the reverse. They enhance our abilities to act in the world and create new possibilities for action that would be impossible without such delegation. With the advent of autonomic computing and Ambient intelligence the delegation of agency reaches hitherto unimaginable levels, and our degree of 'competence', effectiveness and autonomy will thereby stretch to new limits. This is why these technological developments deserve our support and attention. But at the same time, we must always be vigilant of the turning point, at which the autonomy and agency of human agents are externalized to such a degree that they in fact undermined considerably. This means that the challenge for both designers and social scientists and philosophers is to find this turning point, to approach it as closely as possible, yet to never cross it.

We argue that the reflexive loop that we have discussed in this chapter is crucial in this respect. The danger is not so much in delegating cognitive tasks, but in distancing ourselves from – or in not knowing about – the nature and precise mechanisms of that delegation. As we noted in our discussion of the voting aides artefacts contain scripts on two different levels: they contain various (technological and political) ideas and norms from the designers who built them, and they influence users' thinking and actions. Awareness of, and insight into the 'scriptural character' of the artefact, and having the ability to influence that character, is crucial

for users in light of the delegation of their autonomy. If we lack awareness and insight with respect to the way a voting aide works, the ‘prejudices’ that it (unavoidably) contains, and the grounds on which ‘our’ choice is made, then our autonomy is threatened, even if this choice is in line with our political preferences and interests. If we do have this awareness and insight, and a reflexive loop enables us to toy with the aforementioned parameters and to confirm or reject certain values, then the knowledge and decision rules that are built into the voting aide will strengthen our autonomy instead. In that case distributed agency entails an enhancement of our power to act.

Of course, human awareness and knowledge are limited. As computer systems become more and more complex, it will be ever more difficult to open and understand the black box. It is likely, therefore, that the reflexive loop will gradually move from the organic to the artificial components of the network to an ever larger degree. Conceivably, such ‘intentional networks’ will be superior to networks in which the human ‘component’ is the final link. From an anthropocentric perspective that is quite something. Yet it would be unwise to follow Medea and kill our mind children, the technological artefacts, based on hurt pride. Instead, maybe we can find comfort in these words, uttered by Nietzsche’s Zarathustra:

Man is a rope, tied between beast and Overman – a rope over an abyss [. . .]
What is great in man is that he is a bridge and not an end: what can be loved
in man is that he is an *overture* and a *going under*. . .

(Nietzsche 1980, Vol. 4: 16)

Notes

- 1 For instance, in 2006, a year of national elections in the Netherlands, 64% of the voters expressed that they had used none of the sources of information discussed here to inform themselves before casting their vote (CBS 2006).
- 2 ‘Agency’ here is to be understood as the ‘capacity to act’, whereby we leave open the question of the precise necessary and/or sufficient conditions for such a capacity to arise. The ‘standard conception’ of agency summarizes the notion of agency in the following proposition: ‘X is an agent if and only if X can instantiate intentional mental states capable of directly causing a performance.’ (Himma 2008: 3). However, this entails a discussion of what intentionality is, and which beings qualify as ‘really’ intentional – as Daniel Dennett has remarked 15 years ago already: ‘. . .for the moment, let us accept [the] claim that no artifact of ours has real and original intentionality. But what of other creatures? Do dogs and cats and dolphins and chimps have real intentionality? Let us grant that they do; they have minds like ours, only simpler, and their beliefs and desires are as underivedly about things as ours. [. . .] What, though, about spiders or clams or amoebas? Do they have real intentionality? They process information. Is that enough? Apparently not, since computers – or at least robots – process information, and their intentionality (*ex hypothesi*) is only derived.’ (Dennett 1994: 100). We follow Dennett in his solution to the intentionality question: what matters is not so much whether an organism has intentionality or not, but whether it displays something that convinces us as being intentionally aimed – Dennett calls this ‘*as-if*’ intentionality’ (cf. Adam 2008, Dennett 1994). Moreover, in our conception of agency, we side with Floridi and

- Sanders, who formulate three criteria for agency: (1) interactivity, (2) autonomy, and (3) adaptability (Floridi and Sanders 2004: 349, 357–58).
- 3 We deliberately put the phrase ‘made a decision’ between quotation marks here to indicate that we should take this phrase as a *façon de parler*. Although many contemporary neuroscientists ascribe psychological attributes (such as making decisions) to the brain, this should be regarded as a category mistake if this is taken literary and not as a metaphor. After all, brains do not make decisions, only human beings do. Neuroscience can investigate the neural preconditions for the possibility of the exercise of distinctively human powers such as thought, reasoning and decision-making and discover correlations between neural phenomena and the possession (or lack) of these powers, but it cannot simply replace the wide range of psychological explanations with neurological explanations. When neuroscientists ascribe psychological attributes to brains instead of to the psychophysical unity that constitutes the human being, they remain victims of (an inverted version of) Cartesian dualism (cf. Bennett 2007: 6–7, 142ff., Bennett and Hacker 2003: introduction). The fact that neuroscientific investigations show that (in specific cases) neural processes that accompany bodily action precede conscious decision, does not prove that the brain makes the decision instead, but rather that in these cases the psychophysical unity decides unconsciously.
 - 4 Here, too, we have placed the word ‘decides’ between quotation marks, since the brake system does not decide in the ordinary sense of the word, but rather acts mechanically according to its programme. The point is, however, that the brake system functions independently of the driver. The more complicated an automated device, the more we will tend to ascribe intentionality and even rational decision-making to it.

References

- Adam, A. (2008) ‘Ethics for things’, *Ethics and Information Technology*, 10: 149–54.
- Akrich, M. (1992) ‘The de-scription of technical objects’, in Bijker, W. E. and Law, J. (eds) *Shaping technology/building society: Studies in sociotechnical change*. Cambridge, MA: MIT Press.
- (1995) ‘User representations: Practices, methods and sociology’, in Rip, A., Misa, T. J. and Schot, J. (eds) *Managing technology in society: The approach of constructive technology assessment*. London, New York: Pinter Publishers.
- Alford, C. F. (1992) ‘Responsibility without freedom. Must antihumanism be inhumane? Some implications of Greek tragedy for the post-modern subject’, *Theory and Society*, 21: 157–81.
- Bennett, M. R. (2007) *Neuroscience and philosophy: Brain, mind, and language*. New York: Columbia University Press.
- Bennett, M. R. & Hacker, P. M. S. (2003) *Philosophical foundations of neuroscience*. Malden, MA: Blackwell Pub.
- Berg, A.-J. (1999) ‘A gendered socio-technical construction: The smart house’, in MacKenzie, D. A. & Wajcman, J. (eds) *The social shaping of technology*. 2nd ed. Buckingham (UK), Philadelphia (PA): Open University Press.
- Burns, K. & Bechara, A. (2007) ‘Decision making and free will: A neuroscience perspective’, (25) *Behavioral Sciences & the Law*, 2: 263–80.
- CBS (2006) ‘Politieke en sociale participatie’. CBS.
- Dawkins, R. (2006) *The selfish gene*. Oxford: Oxford University Press.
- De Mul, J. (2003) ‘Digitally mediated (dis)embodiment: Plessner’s concept of excentric positionality explained for cyborgs’, (6) *Information, Communication & Society*, 2:

- 247–66.
- (2004) *The tragedy of finitude: Dilthey's hermeneutics of life*. New Haven: Yale University Press.
- (2009) [2006] *De domesticatie van het noodlot: De wedergeboorte van de tragedie uit de geest van de technologie*. Kampen (NL), Kapellen (Belgium): Klement/Pelckmans.
- Dennett, D. C. (1994) 'The myth of original intentionality', in Dietrich, E. (ed.) *Thinking computers and virtual persons: Essays on the intentionality of machine*. San Diego: Academic Press.
- Dilthey, W. (1914–2005) *Gesammelte Schriften (23 vols.)*. Stuttgart/Göttingen: B.G.Teubner, Vandenhoeck & Ruprecht.
- Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J. & Burgelman, J.-C. (2001) 'ISTAG: Scenarios for Ambient Intelligence in 2010'. Seville (Spain): IPTS (JRC).
- Ellul, J. (1988) *Le bluff technologique*. Paris: Hachette.
- Euripides, (2006) *Medea*. Oxford; New York: Oxford University Press.
- Floridi, L. & Sanders, J. W. (2004) 'On the morality of artificial agents', *Minds and Machines*, 14: 349–79.
- Gergen, K. J. (2002) 'The challenge of absent presence', in Katz, J. E. & Aakhus, M. A. (eds) *Perpetual contact: Mobile communication, private talk, public performance*. Cambridge (UK), New York (NY): Cambridge University Press.
- Gjøen, H. & Hård, M. (2002) 'Cultural politics in actions: Developing user scripts in relation to the electric vehicle', (27) *Science, Technology & Human Values*, 2: 262–81.
- Gontier, T. (2005) *Descartes et la causa sui: Autoproduction divine, autodétermination humaine*, Paris: J. Vrin.
- Heidegger, M. (1962) *Die Technik und die Kehre*, Pfullingen: Neske.
- Himma, K. E. (2008) 'Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent?', (11) *Ethics and Information Technology*, 1: 19–29.
- Latour, B. (1992) 'Where are the missing masses? The sociology of a few mundane artifacts', in Bijker, W. E. & Law, J. (eds) *Shaping technology/building society: Studies in sociotechnical change*. Cambridge, MA: MIT Press.
- (1993) *We have never been modern*. Cambridge, MA: Harvard University Press.
- (1999) *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
- (2002) 'Morality and technology: The end of the means', (19) *Theory, Culture & Society*, 5–6: 247–60.
- (2005) *Reassembling the social: An introduction to actor-network-theory*. Oxford (UK), New York (NY): Oxford University Press.
- Libet, B. (1985) 'Unconscious cerebral initiative and the role of conscious will in voluntary action', (8) *Behavioral and Brain Sciences*, 4: 529–66.
- Magnani, L. (2007) 'Distributed morality and technological artifacts', paper presented at Human Being in Contemporary Philosophy, Volgograd (Russia), 28–31 May.
- Nietzsche, F. (1980) *Sämtliche Werke (15 vols.)*. Berlin: De Gruyter.
- Ong, W. J. (1982) *Orality and literacy: The technologizing of the word*. London, New York: Methuen.
- Plato (1914) *Plato, with an English Translation* (trans. North, H., Fowler, L. & Maitland, W. R.). London, New York: W. Heinemann, The Macmillan Co.
- Plessner, H. (1981) *Die Stufen des Organischen und der Mensch: Einleitung in die philosophische Anthropologie*. Frankfurt am Main: Suhrkamp.

- Rimbaud, A. J. (1871) *Lettre du Voyant*, in a personal communication to Demeny, P., 15 May 1871.
- Rosenberg, A. (2008) *Philosophy of social science*. Boulder, CO: Westview Press.
- Schank, R. C. & Abelson, R. P. (1977) 'Scripts', *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale (NJ), New York (NY): L. Erlbaum Associates.
- Van den Berg, B. (2009) *The situated self: Identity in a world of Ambient Intelligence*. Rotterdam: Erasmus University.
- Van Oost, E. (2003) 'Materialized gender: How shavers configure the users' femininity and masculinity', in Oudshoorn, N. and Pinch, T. J. (eds) *How users matter: The co-construction of users and technologies*. Cambridge, MA, MIT Press.