

Vrije Universiteit Brussel

From the Selected Works of Mireille Hildebrandt

2011

Autonomic and autonomous 'thinking': preconditions for criminal accountability

Mireille Hildebrandt



Available at: https://works.bepress.com/mireille_hildebrandt/80/

Autonomic and autonomous ‘thinking’: preconditions for criminal accountability

Mireille Hildebrandt

Abstract

Cognitive psychology suggests that unconscious ‘thought’ is capable of complex analyses, way beyond the capability of the conscious mind. In fact, many cognitive scientists claim that most – if not all – of our behaviour is the result of ‘the adaptive unconscious’ without which we would not be able to function at all. Explaining or justifying our actions depends on being consciously aware of what motivated these actions. If, however, the ‘causes’ of our behaviour are not accessible to the conscious mind, the ‘reasons’ we give may be qualified as a comfortable illusion. The difference between autonomic computing and human behaviour, in that case, is not obvious. This would either mean that attributing criminal liability to autonomic computing systems should not be a problem or it would mean that attributing criminal liability to human beings makes no sense anyway. In this paper I will challenge such a position by elaborating the distinction between autonomic and autonomous ‘thought’, behaviour and action, claiming its relevance for criminal accountability. I will build on recent findings of cognitive psychology about ‘The New Unconscious’, relating this to Judith Butler’s exploration of the constitutive opacity of the self, explored in the traditions of Hegel, Nietzsche, Freud and Foucault. The argument will be that the interplay between autonomic and conscious thought is the precondition for autonomous action, constituting the possibility to blame a person for wrongful action. In other work I hope to build on this, arguing that autonomic computing environments can manipulate our autonomic behaviours in ways that do not reach the threshold of consciousness, thus easily manipulating us into being nice and decent (or horrible and dangerous) individuals without us having a clue as to why we act as we do.

1 Inter-esse: Philosophy of Law and of Technology

I will start with a brief description of what is meant with autonomic computing and its relation to profiling and pattern-recognition (section 2), ending with claims made within the cognitive sciences that pattern-recognition is ‘the most powerful mechanism of successful cognition’ (Goldberg 2006:103). In section 3, I will approach the nexus between law and the cognitive sciences. This nexus is the middle ground on which I will explore the implications of autonomic computing for the notion of autonomous action that underlies the possibility to blame a person for violating the criminal law. After discussing the ‘new unconscious’ of the cognitive sciences, interpreted by some as exposing ‘the illusion of free will’, I will move onto a discussion of Judith Butler’s *Giving an Account of Oneself*. Butler discusses the inaccessible dimensions of the self, building on Hegel, Nietzsche, Freud, Adorno and

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

Foucault. She traces an intractable alterity and opacity *in* and *of* the self, that seems to be confirmed by recent findings in cognitive psychology. Section 4 will trace the implications of this constitutive opacity of the self for the attribution of criminal liability.

In my move from a cognitive psychology that excels in analytical and empiricist methodologies to a philosophy of the self inspired by continental philosophy the reader may encounter a transition of style, if not genre (Bakhtin 1993). The reader might expect me to join the current discussion on the compatibility of determinist claims of cognitive science with the voluntarist assumptions of Anglo-American legal philosophy, which is deeply entangled with moral philosophy. Though this discussion is very interesting, it builds on a curious melange of physicalism and mentalism that seems incapable of detaching itself from what has been called the folk psychology of human agency. This folk psychology is either attacked (suggesting that free will is an illusion) or rescued (suggesting that causal determinism does not threaten our concept of human agency), but in both cases the proponents speaks of agents as persons capable of acting for a reason, based on desires and beliefs about causal implications.ⁱ Rather than taking this definition of human agency for granted, I will change gear and move into another type of discourse. This change of perspective entails a shift from an analytical to a more imaginative articulation of what is at stake. I believe that both types of discourse are of equal importance and that their claims cannot be reduced to one another in an attempt to monopolise the discussions of the elusive opacity of the self.ⁱⁱ

2 Autonomic computing and the autonomic nervous systemⁱⁱⁱ

2.1 Autonomic computing

In 2001 IBM launched what it called a project on 'autonomic computing'. IBM claims autonomic computing to be the only answer to the increased complexity that will arise from the expansion of interconnected networked environments. If we imagine the realisation of Embedded Wireless Sensor Networks,^{iv} integrated into an RFID Internet of Things (ITU 2005) we have what the Philips and the European Commission have coined a vision of Ambient Intelligence (AmI) (Aarts and Manzano 2003, ISTAG 2001). According to IBM the ensuing complexity will soon be beyond repair (Kephart and Chess 2003: 41), requiring a new focus for research that should target the development of computer systems capable of self management, self configuration, self optimisation, self healing and self protection. The claim is that an environment that is capable of anticipating our inferred wishes in real time, always on the alert for changing circumstances, cannot be realised without a measure of autonomic computing.

On the IBM website, dedicated to autonomic computing, we find the explanation for the use of the concept 'autonomic':

The most direct inspiration for this functionality [real time adaptation of the environment, MH] that exists today is the autonomic function of the human central nervous system. Autonomic controls use motor neurons to send indirect messages to organs at a sub-conscious level. These messages regulate

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

temperature, breathing, and heart rate without conscious thought. The implications for computing are immediately evident; a network of organized, "smart" computing components that give us what we need, when we need it, without a conscious mental or even physical effort.

This is an interesting proposition, using a salient metaphor: the decisions made by our autonomic nervous system are subconscious in a non-Freudian sense. We simply have no access to them at all and even if we did, the language in which the various parts of the system communicate is not symbolic or meaningful in a way that our conscious mind would be capable of understanding. This is not to say that the unconscious is not a part of our self: if the system breaks down this is the end of us. We do not only *have* an autonomous nervous system but we actually *are* our autonomous nervous system. This does not imply that this is all there is to us, but realising its constitutive role should bring home some of the challenges presented by visions like Aml and by autonomic computing as their precondition.

2.2 Profiling and pattern recognition

The idea of autonomic computing is that our external environment begins to cater to our needs in a similar vein as our internal environment. This requires continuous profiling of our every movement, biological states, interactions, moods and responses to what happens in the environment. Instead of waiting for our deliberate input on how we want our coffee, what room temperature we like, which music fits our mood, which is the best moment to wake up, how we prioritize and respond to incoming information, the environment profiles our keystroke behaviour and correlates the patterns it exhibits to our health, our mood, our productivity and – for instance to the moment we need a shot of strong black or mild and milky coffee. This is just one example. The range of patterns that can be disclosed by profiling technologies seems unlimited. A combination of facial expressions, like the blink of an eye, yawning and the size of the iris, can correlate with a driver's fatigue (Jin et al. 2007), enabling a smart car to decide when to prohibit speeding or driving per se. Ambient Intelligence implies persistent, pervasive, ubiquitous and seamless monitoring and this has given rise to a state of alarm with privacy activists. The fear is that this monitoring provides extensive traces of what we did when where how and with whom. I think this fear is partly unjustified, because the sheer amount of trivial data collected and stored will turn the information they may present into noise. Once collected and stored these data become decontextualised, devoid of meaning, empty digital signs. My own fear, however, is a different one. Either all this data is lost because we have no means to retrieve it, to recontextualise it, to provide meaning and use it, in which case the whole enterprise is a waste of time, money and energy, *or* we develop techniques to trace and correlate the data, to detect patterns invisible to the naked eye, to recombine discrete data into new contexts and to anticipate futures on the basis of patterns found in past behaviours (Fayyad et al. 1996, Custers 2004, Hildebrandt 2008a, Hildebrandt and Gutwirth 2008, Kallinikos 2006 and 2008). Wasting time, money and energy is a serious matter, but what concerns me here is the other possibility. Imagine a smart energy meter that provides information on how much energy you are using at any moment in time in pence per hour, tonnes of CO₂ or kWh.^v This could allow you to change your energy spending habits in simple ways, since you have the relevant feedback to make effective decisions. However, if this information is collected and

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

stored with the owner of the infrastructure and disclosed to government officials or sold to commercial enterprises the data could be mined for significant patterns that reveal intimate details of your lifestyle. More interestingly, such data can be aggregated on a large scale, enabling data mining operations on the data of an entire population. In that case, if data on energy usage per quarter of an hour are correlated with data on health, employment, travels, shopping, political and religious affiliation, refined group profiles will emerge that correlate specific patterns of energy usage with a specific earning capacity, health status, shopping preferences etc. Such profiles, even though they are based on statistical inferences and merely present probabilities, can be applied if they match with your data, and this application will impact on your life.^{vi} Autonomic profiling, as a precondition for smart environments, means that unprecedented amounts of data will be mined for relevant patterns while these patterns will be used to anticipate your preferences and to proactively adapt the environment to fit your priorities. This is what will give smart environments a competitive advantage over others that have not succeeded in accommodating to its users. This is how smart environments may develop a new cognitive power that could eventually match the cognitive power of living creatures, whose continued existence in a changing environment depends on adequate pattern recognition. The idea that profiling or pattern recognition is the hallmark of life, of keeping in touch with and attuning one's responses to one's ecological niche, can be traced in the work of Maturana and Varela (1991, where it could be defined as structural coupling between an autopoietic system and its environment), James Gibson (1976, where it could be defined as locating the affordances of the environment by an organism) but also of Herbert Simon who defined intuition as subconscious pattern recognition that can – in principle – be performed by intelligent machines (Franz 2003). In this contribution I want to prepare the comparison of unconscious pattern recognition by individual humans with machine profiling, in order to investigate to what extent autonomic computing could impact human autonomy. This is highly relevant for the legal framework of constitutional democracy in as far as this both assumes and produces a measure of human autonomy (Hildebrandt 2008b).

3 The opacity of the self: autonomic and autonomous thought

3.1 *The wrong question: Do reasons causes actions?*

In *Ethical Know-How. Action, Wisdom, and Cognition* Francisco Varela (1999:ix) undertakes 'an understanding of ethics in a nonmoralistic framework'. He opens his first chapter with the proposition that 'Ethics is closer to wisdom than to reason, closer to understanding what is good than to correctly adjudicating particular situations' (Varela 1999:3). He refers to Charles Taylor's distinction between a moral philosophy that builds on the Kantian tradition of moral judgement and an ethics of situatedness that builds on the tradition of Hegel. He explains the difference by stating that (Varela 1999:4):

As a first approximation, let me say that a wise (or virtuous) person is one who knows what is good and spontaneously does it. It is this immediacy of perception and action which we want to examine critically. This approach stands in stark contrast to the usual way of investigating ethical behavior, which begins by analyzing the intentional content of an act and ends by

© Hildebrandt, chapter 9 in:

'Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology', M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

evaluating the rationality of particular moral judgements.

The question that comes to mind is how we can be sure that a person is right about what she thinks to be good and a moral philosopher would probably retort that this is exactly what moral philosophy is about. This may indicate a division of labour: those interested in embodied and situated action investigate how a person implements what she knows to be good, while those interested in criteria to decide which behaviour is right investigate how we can develop such criteria. However, such a division of labour implies that knowing precedes acting, reducing action to the implementation of representational knowledge about the right way to act.^{vii} In speaking of ‘the immediacy of perception and action’ and ‘wisdom rather than reason’ Varela seems to suggest that ethical knowledge is a matter of perception instead of the result of correct reasoning. His usage of the term ‘immediacy’ of both perception and action seems to indicate that in perceiving a situation we intuit how to act while performing the action, and the usage of the term ‘spontaneously’ indicates that the action is not premeditated by a conscious effort. The action is not ‘caused’ by our conscious will. This is a controversial proposition that flies in the face of the idea that an action is ‘good’ or ‘right’ if we can give good reasons for it and have performed it with those reasons in mind. Interestingly Varela’s position has some affinity with recent findings in neuroscience and cognitive psychology ‘demonstrating’ that the whole idea that our intentions ‘cause’ our behaviour is an illusion. Following Libet’s experiments in the beginnings of the 80s several authors have defended the position that conscious intention *does not precede but actually follows* motor action (Haggard and Libet 2001; Wegner 2002). Since to be a cause something must precede its consequence, conscious thought cannot be the cause of our actions. It rather seems to be a by-product of actions, providing us with a sense of control, while in fact the causes reside in complex brain mechanics to which we have no conscious access whatsoever. The problem with Libet’s experiments is that they concern simple motor actions, like the movement of a hand. It is unclear what conclusions can be drawn from this kind of controlled laboratory experiments for more complex behaviours, for instance for those that are the result of formal learning processes. Another question raised by the experiments is how the subconscious brain mechanics that are supposed to ‘cause’ the behaviour have been inscribed in the brain. If this was not part of our genetic make-up, this inscription must be the result of a learning process that may have involved conscious thought, especially in the case of more complex behaviours that depend on linguistic interaction. The fact that much of our behaviour is automated does not imply that we have not consciously initiated this automation at some point in the past. To complicate the issue further, Dijksterhuis (2006) discusses experiments that demonstrate the capacity of *unconscious thoughts* to deal with complex decision-making. His experiments ‘demonstrate’ that allowing for unconscious thought processes provides for markedly better decisions - in the case of complex problems - than either conscious thought or immediate decision-making. This seems to require qualification of Varela’s idea of immediate spontaneous action as a superior type of behaviour, reflecting knowledge of what is good. Well-tuned spontaneous action may require a measure of time during which the action is prepared at a level of subconscious thought processes, capable of more refined and complex ‘processing’ of information than conscious thought will ever be. Varela’s position can also be linked to the writings of Goldberg (2006) on wisdom as sedimented pattern recognition that is typical for the brain of an experienced person, providing for a reliable resource of

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

automated (re)actions. The subtitle of his *The Wisdom Paradox* (2006) is *How Your Mind Can Grow Stronger As Your Brains Grow Older*. Pattern recognition is supposedly sedimented in the left hemisphere of the brain, being the physical correlate of what we tend to call our intuitions about how to understand a situation and of how to act in it. According to Goldberg this intuition is not pre- or nonanalytic but postanalytic, because it is the result of a learning process that can be traced in the brain as a matter of inscribed pattern recognition. This concurs with Gigerenzer's nuanced discussion of gut feelings as a form of unconscious intelligence (Gigerenzer 2007), providing the background of Gladwell's popular exaltation of gut feelings in his bestselling *Blink* (2005). What makes Gigerenzer more interesting is that he aims to discuss *when* we can trust our gut feelings, meaning that we cannot take our intuitions for granted as being right or good. Indeed, Varela did not write that a wise person is one who spontaneously acts in accordance with whatever she feels is good, he specifically wrote that a wise person is one who knows what is good and does it.

We conclude that though a Cartesian perspective invites the idea that thoughts are the causes of actions, cognitive science suggests that in fact actions 'cause' our thoughts even while having us believe that we are in control. Actions themselves are 'caused' by subconscious brain processes to which we have no conceptual access. So, we have thoughts, unconscious thoughts, intentions, behaviours and actions and we have the personal experience of control over our actions and – just like in the case of our memory (Loftus and Ketcham 1994)– psychology is demonstrating that the control we think we have over our behaviour may be an illusion.

From a non-Cartesian perspective this is not necessarily a surprise. The idea that autonomous thought presumes an individual will that is the cause of movements and more complex behaviour but has not been caused itself because it falls within the domain of the untouchable *res cogitans*, is a problematic idea if not the engine of a series of illusions about material passivity and causality on the one hand, and mental activity and freedom on the other hand. In trying to solve the problems generated by the separation of a *res extensa* and a *res cogitans* empiricist and rationalist traditions end up with the mentalistic picture of an uncaused will that causes our behaviour. That science now proves this to be an utterly inadequate model for understanding the complex relations between thinking and acting, actually confirms longstanding suspicions of the self as an independent subject whose inner freedom can be taken for granted. Nietzsche, Freud and Foucault seem to line up here with the latest findings of the cognitive sciences – *bien étonnés de se trouver ensemble*. This, however, does not imply that these findings cannot clarify the relationship between automated behaviours, intentional action and the meaning of human freedom. In what follows I will first explore the nature of autonomic thought, by discussing some salient experiments performed within the realm of the cognitive sciences. After that I will try to come to terms with the idea of autonomous action as the hallmark of human freedom, safely rooted in a constitutive opacity that grounds the self in the abyss of its own inaccessible beginnings. Returning to Varela I will argue for an enacted view of ethical perception, which allows me to understand autonomic thought as a preconditional dimension of autonomous thought. This paves the way for the articulation of the right question: to what extent is autonomic thought compatible with autonomous thought and intentional action? This question is of crucial importance for the attribution of criminal liability.

© Hildebrandt, chapter 9 in:

'Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology', M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

3.2 *Autonomic thought: pre-, non- or postanalytic?*

Referring to Herbert Simon's work on pattern recognition Goldberg (2006:104) understands intuition as a postanalytic competence, the result of a learning process that enables automatic pattern recognition and thus affords what he calls a 'mental economy'.^{viii} Instead of reinventing the wheel by consciously deliberating on how to respond to a specific situation, the mind recognises types of situations and automatically fits them with previous responses. It should be obvious that this process is not mechanical in any simple way and we may readily assume that this process of pattern recognition is in fact too complex to be represented in the conscious mind, since it has to take into account a myriad of interrelated features of an environment that is never the same as when the pattern (the type of situation) was first encountered. Also, the response to a similar type of pattern cannot be a simply copy of a previous response, so below the threshold of consciousness an unconceivable amount of work is done to attune our responses to whatever challenges we face. Gigerenzer (2007:18-19) explains that this does not mean that the brain solves a complex problem with a complex strategy, following logical principles. The unconscious brain does not specify all consequences for each action, weighing them carefully and adding up the numbers in order to choose the one with the highest value or utility. This kind of deliberation would be more typical for the conscious mind, at least for the conscious mind of someone trying to be rational about things. The unconscious brain – according to Gigerenzer – does not function like a calculating machine, 'rather, nature gives humans a capability, and extended practice turns it into a capacity' (idem:18). He speaks of the unconscious as '*knowing, without thinking, which rule is likely to work in which situation*' (idem:19, my emphasis).

Other authors, however, claim that *the unconscious does think, but we just don't know it*. In his 'A Theory of Unconscious Thought' Dijksterhuis (2006) describes an interesting set of experiments. For instance, participants were given information about 4 apartments, described in terms of 12 different features. One of the apartments was significantly more attractive than the others, while another one was significantly less attractive than the others. The 48 bits of information about the apartments were presented to the participants in a random manner, confronting them with a 'daunting amount of information' (idem:96). After that they were asked to evaluate the apartments. One third of the group had to perform this evaluation immediately after the information was provided, leaving no time to think things through. Another third of the group were given 3 minutes to think about the information before their evaluation and the last third were told they would make be asked about the evaluation later, this group was distracted for 3 minutes before their evaluation. Surprisingly, perhaps, those who had time to unconsciously 'think' about the assignment did best in terms of coining the most and least attractive apartments as such. In another experiment Dijksterhuis and Van Olden (2006) gave participants a choice between 5 different art posters. One third of the group were presented with all 5 posters at the same time and had to decide immediately; one third was presented with each poster separately for 90 seconds and was asked to review each of them and to write down their reasons for liking or not liking a poster and to carefully analyze their preferences; a last group was presented with all 5 posters and told that at a later moment they would be asked to choose the one they liked best (they were distracted

with a task of solving anagrams during 450 seconds). After the participants made their choice they answered some questions and – to their surprise – were given the poster of their preference. This time the researchers had no *ex ante* qualification of which was the best option. Instead, they asked participants at a later point in time (3-5 weeks after the experiment) to indicate their satisfaction with the choice they had made, the regrets they might have developed about their choice and the amount of Euros for which they were willing to sell the poster. The final outcome of the experiment was that those who had time for unconscious thinking were more satisfied with their choice, had less regrets and asked a higher price for an eventual sale of their poster. Dijksterhuis and Van Olden (2006:630) conclude that:

People who were given the opportunity to think unconsciously about choices made superior decisions relative to those who thought consciously or who did not think at all.

It is not very difficult to criticize this conclusion on the basis that it assumes that - what the authors call - a higher 'post-choice satisfaction' is an indication of a superior decision. There may be other reasons or causes that explain the correlation between having some time between seeing the options and deciding while being distracted and being more happy with one's choice after a number of weeks. For instance, one could counter that not having consciously scrutinised the posters makes the choice less vulnerable to doubt, assuming that participants have not given their choice much thought after the experiment was over. In that case the higher 'post-choice satisfaction' doesn't mean very much. However, this does not explain the difference between those who chose immediately and those that had some time to unconsciously prepare the choice. I will not continue to question the conclusions at this point, because I want to take up the challenge posed by similar findings about the workings of the unconscious mind.

It seems evident that the conscious mind does well in solving relatively simple problems that do not require taking into account too many interrelated factors, which would require exponential processing powers compared to those available to the conscious mind. Such processing powers seem to be a capability of the unconscious mind but we have no conscious access to them. 'They' (these brain processes) 'think for us' and we are bound to trust their outcome up to the point where we begin to reflect upon these outcomes and decide against them. This reflection may lead us to change our courses of action and when this implies changing our habits it will involve ingraining or inscribing new types of pattern recognition into the unconscious mind. Once these patterns are wired into the wetware of our mind, autonomic thinking – and acting – can take over and we can direct our conscious attention elsewhere, trusting that our unconscious awareness takes care of the rest. The conscious mind thus allows for postanalytic autonomic thinking,^{ix} but we can assume that our unconscious mind develops most of its autonomic behaviours without conscious input, meaning that much autonomic behaviour is learned but pre- or nonanalytic. Indeed, most of our self seems to evolve below the threshold of consciousness, taking care of our physical and emotional equilibria in the flux of events amidst which we live and sustain our self. The opportunities for conscious reflection are limited and their function seems to be just this: directing conscious attention to an issue that has emerged into consciousness, allowing for some direction over how to cope with it, providing an occasion to attune one's autonomic responses to goals that one can consciously

© Hildebrandt, chapter 9 in:

'Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology', M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

articulate. Retrieving issues from below the threshold of consciousness seems largely out of bounds as a process over which one has conscious control, if only because of the limited capabilities of the conscious mind as compared to the unconscious self.

Autonomic ‘thinking’ then, must be a dynamic *mélange* of pre-, non- and postanalytic ‘thinking’, with most of these processes being inaccessible for conscious reflection (nonanalytic) even if some are the result of conscious interventions (postanalytic) and still others are on the verge of emerging in the conscious mind, asking for our focused attention (preanalytic).

3.3 *Autonomous thought and the constitutive opacity of the self*

The question that surges into my consciousness at this point, is how to understand the idea of autonomous thought in the light of the constitutive opacity of the self under discussion, and the related question of what room is left for what we like to think of as human freedom. To explore this issue we need to leave the realm of motor action and relatively simple consumer choice, taking into account the complexity of contemporary society and of the technological infrastructure that affords such complexity.^x I will initiate this further exploration with a substantive quotation taken from Judith Butler’s *Giving an Account of Oneself* (2005:65):

To say, as some do, that the self must be narrated, that only the narrated self can be intelligible and survive, is to say that we cannot survive with an unconscious. It is to say, in effect, that the unconscious threatens us with an insupportable unintelligibility, and for that reason we must oppose it. The ‘I’ who makes such an utterance will surely, in one form or another, be besieged by what it disavows. An ‘I’ who takes this stand – and it is a stand, it must be a stand, an upright, wakeful, knowing stand – believes that it survives without the unconscious. Or, if it accepts the unconscious, this ‘I’ accepts it as a possession, in the belief that the unconscious can be fully and exhaustively translated into what is conscious. It is easy to see that this is a defended stance, but it remains to be seen in what this particular defense consists. It is, after all, the stand that many make against psychoanalysis itself. In the language that articulates opposition to a non-narrativizable beginning resides the fear that the absence of narrative will spell a certain threat, a threat to life, and will pose the risk, if not the certainty, of a certain kind of death, the death of a subject who cannot, who can never, fully recuperate the conditions of its own emergence.

But his death, if it is a death, is only the death of a certain kind of subject, one that was never possible to begin with, the death of a fantasy of impossible mastery, and so a loss of what one never had. In other words, it is a necessary grief.

This quotation recounts in a nutshell Butler’s salient argument for an ethical responsibility based on ‘an apprehension of epistemic limits’ (Butler 2005:43) concerning knowledge of the self. Instead of basing the ethical response on our capacity to provide a full account of oneself, Butler suggests that we must find a way to accept the fact that as we are not entirely transparent to ourselves we can not provide the definite reasons for our actions.^{xi} This acceptance grounds a shared way

of being in this world, and demands of us that we suspend our judgement of the other in the face of the opacity that constitutes not only my self but her self too. It entails that we recognise the extent to which we are not the authors of our actions, which are ultimately constituted in the unconscious and inaccessible limbo of the self, as it ‘emerges’ from the complex entanglements of linguistic and technological norms that constrain and make possible that we ‘come about’.^{xii} Butler’s briefly refers to psychoanalysis when asking us to attend to the opacity of an unconscious we cannot entirely retrieve for conscious introspection. Her discourse on what she calls ‘the prehistory of the subject’ (Butler 2005:79) is an extended dialogue with Adorno, Nietzsche, Foucault, Hegel, Cavarero, Levinas, Kafka, Freud, Bollas, Klein, Laplanche, rather than an analysis of the findings and an exploration of the implications of what has been coined as ‘The New Unconscious’ within cognitive science.^{xiii} Her dialogical monologue on the vulnerabilities of an exposed subject that stumbles upon its indeterminate, unfinished nature seems to turn Sartre upside down: the end of our life will not undo the enigma of who we are, finally or definitively, because the beginning that precedes us and constitutes us cannot be uncovered. Self-identity and a coherent account of who we are ‘will have to fail in order to approach being true’ (idem:42),^{xiv} ‘the ‘I’ cannot give a final or adequate account of itself because it cannot return to the scene of address by which it is inaugurated and it cannot narrate all of the rhetorical dimensions of the structure of address in which the account itself takes place’ (idem:67). However different the style of writing may be between cognitive scientists who wish to speak a clear and distinct language in which transparency reigns and philosophers who wish to speak a more imaginative language that discloses the ambiguity that springs from the opacity of the self, I think they are both – in their different idiom – referring to the same abyss from which our conscious experience wells up time and again. Butler indeed acknowledges that the opacity of our beginnings concerns a *reiterant event* (‘prehistory does not stop happening’, idem:78), as we are continuously ‘impinged’ by others and by ‘the way social force takes up residence within us, making it impossible to define ourselves in terms of free will’ (idem:106). What Butler accomplishes in her account of the impossibility of a full account of oneself is that she – probably unintentionally – provides a link between the detailed experiments of cognitive science that refute our sense of being sovereigns of our own self and philosophers and psychoanalysts who have been arguing for a relational conception of the self, while struggling with the implications for human agency, intentional action and individual responsibility.

If autonomous action was never a matter of undiluted control of a person over her own actions, how can we steer free of overconfident accounts of scientists claiming that free will is nothing more than an illusion? Must we agree with Nietzsche in lamenting the human condition, as he wrote more than a hundred years ago (1887) that people:

felt unable to cope with the simplest undertakings; in this new world they no longer possessed their former guides, their regulating, unconscious, and infallible drives: they were reduced to thinking, inferring, reckoning, co-ordinating cause and effect, these unfortunate creatures; they were reduced to their ‘consciousness’, their weakest and most fallible organ!^{xv}

Or, alternatively, is there good reason to celebrate the complexities of a conscious mind that is capable of self-reflection while being deeply – and to some extent safely

– immersed in the inaccessible processes of autonomic behaviours that attune the organism to its internal and external environments? And could it be that to understand how this reflection came about and developed to its present form of reflexivity, we must turn to language and technology as constitutive for our capacity to externalise our thoughts, to create a distance from our self for the self and – in doing this – witness our birth as self-conscious beings (Ambrose 2001)? If I tell a small child that she is Carol, whereas I am Mireille, by addressing her with: “you are Carol” (pointing to her), and “I am Mireille” (pointing to myself), she will initially answer me by confirming what I just said, pointing to herself and saying: “you Carol”, and pointing to me while saying: “I Mireille”. I will retort, claiming “no, I am Mireille and you are Carol”. At some point in time children learn to change perspective and suddenly realise that to others they are the ‘you’ that others are to them, while for themselves they are the ‘I’ that others claim to be themselves. This looking at the self from the perspective of the other is the birth – in language – of the subject, the first person singular.^{xvi} It is indeed an affordance of language. In that sense the subject is the consequence, though not in causal terms, of being addressed by the other(s), while this address is itself grounded in something that extends beyond the singularity of a particular address (language, social norms). To be born as *another* and experience a first sense of self, however, is only the beginning of an adventure that has taken us from small scale face-to-face communities to large scale societies of strangers who are nevertheless constrained by shared social norms due to the jurisdictions to which they are subject(ed). We need to remember that the distantiation in space and time that conditions the appearance of large scale societies has been an affordance of the technology of the script, while the introduction of the printing press allowed for even greater distantiation, thus giving birth to systemisation, indexation, rationalisation, visualisation and a concurrent turn to empirical science with its emphasis on visual representation (Ong 1982, Eisenstein 2005, Goody and Watt 1963, Ricoeur 1986, Lévy 1990, Ihde 1990a, 1990b and 2002). It seems that the script and the printing press afford an intensification of the externalisation and fixation of thoughts, making them available for reflection. This externalisation and fixation ‘saves’ them from the ephemeral quality of autonomic ‘thoughts’ as they appear in the conscious mind, thus reinforcing the process of translating autonomic thoughts into what I would call autonomous (reflective) thought. If this is a fact, then the move to writing and printing must have triggered a shift to autonomy, to the possibility of giving an account of myself ‘that not only explains what I do but allows me to assume greater agency in deciding what to do’ (Butler 2005:79). Indeed the move from speaking to writing and printing may be the ‘cause’ of our illusion of control, of willing our actions. Intentional action depends on conscious anticipation of the context in which we will act and of the consequences that our action may trigger. This anticipation is the fruit of deliberate reflection on past actions, allowing us to appropriate these actions and their consequences as *our* actions and as effects that *we* caused. Thus, in hindsight, we *become* the authors of these actions. However, in becoming the authors of our actions we create the conditions for anticipating future actions and their consequences, thus creating the conditions for intentional action. This is not only related to the use of written and printed language, but already initiated by the interplay of remembering and anticipation made possible by language as such. In fact Ambrose (2001) describes, from an anthropological and archaeological perspective, how the use of tools presumes and produces some anticipation in terms of cause and effects, indicating that language may well be a necessary corollary of the instrumental usage

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

of technologies that have turned us into animals that can take on – and at some point in time can also articulate – an intentional stance in this world.

This intentional stance is still rooted in the grounding opacity of the self, building on numerous autonomic processes and relying on inscribed (learned) pattern recognition and the related autonomic behavioural responses. Autonomy was not first, it cannot master its beginning, but this does not make it any less cause of celebration. It does not turn intentional action into an illusion; it only turns transparency and sovereignty of the self an illusion. Autonomy is relative as well as relational. The self, as a body, as a mind that for the most is an array of subconscious autonomic processes to which its conscious dimension has no access, is exposed to permanent and pertinent impingement by the other(s), without ever having a final grip on its own development. But as this self arises from the depths of its subconscious and the width of its sociotechnical constitution, its relative and relational autonomy springs up, its natality (Arendt 1998) takes root and begins to influence both itself and its environment.^{xvii} This is where, so far, autonomic computing is no match. Though the qualities of self management, self configuration, self optimisation, self healing and self protection of autonomic computing require a certain spontaneous, random, unpredictable to be programmed into the system, this does not compare to a natality that guides self conscious reflection. It may at some point approach the adaptive qualities of living organisms, but this does not equate to the conscious awareness that is preconditional for autonomous action. Autonomic computing may be vulnerable to destruction, to failure but it has no reflective awareness of this vulnerability and it can for this reason not be held to account for the harm it causes.

4 Potential implications for criminal liability

4.1 Law and the ethics of judgement

The differences between ethics and law are manifold. What interests me here is that since the criminal law calls to account an alleged offender, modern law requires more than ethical perception: it requires judgement. Varela's attempt to suspend the move towards judgement, which is structured in terms of the reasons that can be given for a particular action, is praiseworthy in as far as it concerns our capacity for *ethical* action. Butler's praise for an ethics of recognition that suspends judgement and takes into account the epistemic limits of our knowledge of the self and the other, is equally pertinent but seems similarly helpless when the *law* requires us to give an account of our actions in view of being accused of a violation of its norms. Modern law assumes and produces an ethics of judgement. The judge has been attributed the task of deciding whether we are to blame, whether we have failed to live up to the norms that exceed our particular situation, which we have invented and which were imposed upon us before we 'came about'. Legal norms that, to some extent, co-constitute our self. Or, if we were socialised outside the safe harbour of mainstream normativity, whether we violated legal norms that are foreign to us, that are not inscribed in our autonomic 'thoughts' even if we are consciously aware of their imperative nature. Can we reject the accusation with a story about the opacity of our self? Can we claim that we have no access to our autonomic mind and have no control over actions that may seem autonomous but are, in the end, 'caused' by inaccessible brain processes? Can the judge follow this logic and reply that her own autonomic program directs her

© Hildebrandt, chapter 9 in:

'Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology', M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

towards a conviction, because if we – who stand accused – are ‘caused’ by the unconscious, so is she? In the next subsection (4.2) I will discuss the idea that we were never capable of autonomous action anyway, implying that criminal liability in the end makes no sense. Equally pertinent for the subject of this colloquium, one could argue that the inevitable opacity of the unconscious dimensions of our mind dissolves the problem of autonomic computing environments having an invisible impact on our behaviour, since so much of our actions seem informed by invisible processes anyway? In section 4.3 I will indicate how this relates to the next question, which concerns the idea that autonomic computing does not really make a difference, since we are already out of control.

4.2 *We were never capable of autonomous action: Don’t blame us*

From our discussion of the unconscious brain processes that inform the interactions with our environment we have concluded that there is a fundamental opacity of the self that grounds us while we have no unmediated access to it. Nietzsche, Freud and Foucault have described the alien aspects of and within the self long before cognitive psychology disclosed the untenable claims of the Cartesian ego, untenable at least in as far as it posits that when all distractions are peeled away we encounter a *res cogitans* that is capable of transparent rational thought about itself. From an ecological, relational perspective the distractions that hit us under the threshold of consciousness co-constitute our self and form an intractable layer of pattern recognition wired into the ‘plastic’ brain, perhaps reducing its plasticity while providing us with numerous sets of sophisticated well-attuned autonomic responses to similar events.

This, however, does not refute the possibility of autonomous action. Even if autonomous action depends on and produces autonomic action, it can be distinguished – though not separated - from the subconscious resources on which it nourishes. Although we cannot account for our own emergence as a human being in a particular time and place, the conscious self-awareness that is triggered by being addressed as the author of our action paves the way for intentional action. In that sense we can agree with Nietzsche when he claims that being called to account institutes self-reflection, delivering us to the task of providing reasons for our actions, even if we have little access to the correlated brain processes that are the precondition for any type of action. Being called to account thus creates what it assumes. There is a problem here, which lies at the core of criminal liability. I would nevertheless argue that the circularity we encounter with regard to the institution of accountability is not vicious but virtuous, as long as we remember to incorporate the ethics of recognition that should precede an ethics of judgement. Keeping in mind that our roots are mostly invisible, untraceable and intractable we should be careful not to condemn a person beyond what she could (and should) have avoided. This care is not required because the accused was badly programmed but because we recognise a common vulnerability to primary impingement by caretakers we have not chosen. This care does not imply that we stop calling to account those who misbehave, because we think their behaviour is caused by wrongly inscribed autonomic thought. On the contrary, this care implies that we acknowledge the importance of being called to account, as this is the most salient way to appropriate our actions as *our* actions and to generate autonomous action from the wells of autonomic thought.

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

The care that should inform legal judgement is inscribed in the procedural safeguards of the legal process. The fair trial that precedes the legally valid identification of a defendant as an offender and – if the defended is found guilty as charged - the trial enables calling to account the offender. It provides for a delay, a hesitation, an extended period of time during which the charges are investigated in a court of law, instituting the possibility to contest the charges, claiming that one did not commit the incriminated act; that the action one did perform does not qualify as a criminal action; that the action falls within the scope of the criminal law but can be justified or excused. The contestation of the charge takes the defendant serious as a legal subject, capable of acting in law, having standing in court. Despite the temptation of treating the defendant as an object of investigation and punishment, the trial institutes legal constraints that force the judge to face the defendant as a person who, like the judge, is vulnerable to the inscriptions of her environment over which she had no little control. The judge should suspend her judgement until she is lawfully convinced of the guilt of the defendant, or – if this point is not achieved – the judge must acquit. And in facing the defendant the judge may develop an understanding of the other that clarifies what is opaque for this other, lost in the mist of early impingement by primary caretakers or by social norms that have imposed dangerous habits of the mind on the defendant, explaining her behaviour without necessarily excusing it. In convicting the offender the judge qualifies the other as an autonomous person, capable of resisting the imposition of social norms that invite behaviour that hurts or damages others to an extent that it violates the criminal law. But this is never an easy task, as it requires of a judge to be in touch with her own opacity, acknowledging that resisting autonomic behaviours can be extremely hard at some point in life.

The fact that our actions are grounded in autonomic behaviours over which we have limited control is not a reason to accept an appeal to being determined by autonomic brain processes. On the contrary, being called to account is preconditional for developing autonomous thought and for this reason the criminal trial is a pertinent occasion to generate autonomous action. In that sense criminal liability is productive of agency, to the extent that its attribution is organised in a fair trial that reminds the judge of the opacity she shares with the person she has to judge. In the mean time, the delay that is preconditional for an effective attribution of liability also relates to the capability of the judge to generate autonomous thought and to judge the other without being determined by her own autonomic thoughts. The findings of cognitive science confirm that the optimal decision is not the one taken ‘on the spot’ without any contemplation. However, these findings also suggest that the optimal decision may not be the one that is dictated by conscious deliberation, followed immediately by the decision. It would be interesting to develop experiments that test the excellence of more complex decisions that have normative implications and impact the life and well-being of others, to be taken after conscious deliberation, followed by an extended period of distraction, thus involving the unconscious brain in the process. Experiments that test the capacity of the unconscious brain to resist stereotyping (Glazer and Kilhstrom 2005:186-189) indicate that autonomic thought is capable of applying normative constraints – somehow confirming Varela’s proposition ‘that a wise (or virtuous) person is one who knows what is good and spontaneously does it’. We must however qualify his statement by adding that in specific situations, such as the taking of decisions that may impact one or more others, spontaneity will only lead to ‘good’ actions if it is the result of conscious deliberation followed by a delay that

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

allows for the unconscious mind to work things out, perhaps followed by a renewed deliberation, until a 'reflective equilibrium' is reached.

4.3 The next question: does autonomic computing make a difference?

Careful decision-making requires combining conscious deliberation with autonomous elaboration, trusting the unconscious mind to come up with relevant solutions, while taking the responsibility for the decision by returning to conscious deliberation. This return to conscious deliberation is triggered by the anticipation (the possibility) of being addressed as an autonomous agent, by being called to account. Though many decisions can and must be made in the flux of life, with no need and no time for such reflective interventions, autonomous action implies that at some point we must be open to debate on which course of action is either good or bad. We must, however, remember that we have no unmediated or complete access to the self, and need to be sensitive to the opacity of the roots of our thoughts. This opacity does not preclude us from opening our mind to what others may tell us about our self, providing us with a new view of our self and lifting the veil of our invisible motivations. Though we have no direct access to unconscious processes, the gaze of the other may inform us about our self, thus also co-constituting our sense of self, or self-consciousness. If this openness and the capacity to move back and forth between conscious and unconscious thought are the hallmark of autonomous action the question becomes relevant in which way autonomic computing systems make a difference to human identity and criminal liability.

Bateson once wrote that what matters is to detect 'the difference that makes a difference' (Bateson 1972:315). Obviously many differences can be discerned between the human mind, building on her own autonomic thought and the human mind, building on processes of autonomic computing of an environment that proactively caters to her inferred preferences. In this text I have tried to describe the emergence of autonomous thought in the midst of our autonomic mind, thus preparing the ground for an investigation of 'the difference that makes a difference'. My guess would be that two differences may be relevant here. First, there is the matter of access to the way the autonomic environment profiles us: which opacity is exposed, correlated and disclosed by the profiling machines that mine our data? Second, the environment does not share this exposure with us. It does not share a fundamental opacity, because the environment is not (yet?) conscious of its own being. It does not expose itself, because it does not care to be exposed. These two issues need further exploration, possibly providing new insights into our own vulnerabilities and opacity, while potentially also providing new ways to co-constitute the self.

5 Concluding remarks

Technological innovations make a difference to our life world. There is nothing inherently disturbing or bad about this. However, if the socio-technical infrastructure of knowledge production and communication changes radically, this may entail far reaching changes in our perception of time, space and of the others we encounter. These changes may increase our freedom or bondage, depending on the specific

© Hildebrandt, chapter 9 in:

'Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology', M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

affordances of the technologies involved and on the way we incorporate them in our reconstituted life world. In this contribution I am concerned with the impact of autonomic computing systems on the constitution of the self and with the implications this may have for criminal liability. To tackle this concern I have decided to make a preliminary and necessary detour to investigate to what extent our self is constituted by autonomic thoughts and actions. As this detour requires an exploration in its own right, I have restricted myself to precisely this issue, only briefly touching upon the subsequent question of how autonomic computing will impact our capacity for autonomous action.

I have taken up some of the challenges posed by findings in the field of cognitive psychology, which suggest that the sense of control we have with regard to our own actions is an illusion, produced after the fact. My point was not to deny the importance of these findings, but to invite another way of reading them. From a postphenomenological point of view the idea that our reasons are not the causes of our actions is not a surprise. A hermeneutics of suspicion, as practiced by Marx, Nietzsche and Freud, has preceded the experimental evidence now available for the fundamental opacity of what induces our interactions. This has allowed me to discuss several experimental findings within the domain of cognitive psychology in relation to Judith Butler's discussion of what it means to 'give an account of oneself'. This is particularly interesting because Butler is concerned with an ethics of recognition that is inspired by the constitutive opacity of the self. In fact she situates human responsibility in a shared vulnerability that denies us full mastery over (the consequences as well as the motivations of) our own actions. Autonomic thought processes seem to form a core dimension of our self, meaning that most of what we think and do is prepared below the threshold of consciousness. We have no conceptual access to this level, but this does not imply that our conscious attention does not influence the patterns and strategies developed at the subterranean strands of our mind. One of the implications of this reading of what has been called 'The New Unconscious' is that we have to be more precise and more modest about our notion of autonomous action. We must admit that autonomous action is not as straightforward and transparent as Cartesian separatists of the *res cogitans* would have it. Autonomous action is rooted in and constituted by complex autonomic mechanisms, many of which are the result of learning processes that ingrain specific types of pattern recognition and equally patterned responses into the wetware of our brains. Such patterns allow us to seamlessly anticipate familiar contexts without the need to pay conscious attention to all that requires an adequate response. We must, however, also admit that conscious intervention is part of the human condition, allowing for the instigation of new habits that are conditioned by novel autonomic processes at the level of the brain and the body.

Though reasons do not cause actions we have a capability for autonomous thought and action even if rooted in autonomic actions. This is the case, because autonomic actions can become the object of reflection of the conscious mind, thus creating room for autonomous thought and action. One of the crucial events that evoke this reflection is calling a person to account for an action that has caused harm to another. In accusing a person of having acted in a way that unduly hurts another person the accused is challenged to see the incriminated action as her own action, for which she bears responsibility. This will invite her to anticipate the consequences of her actions and to redirect them if they are met with reproach. This actually initiates the birth of

© Hildebrandt, chapter 9 in:

'Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology', M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

intentional action, even when we decide to resist the accusation. Criminal liability thus not only assumes human agency but also creates a sense of agency that is constitutive of autonomous action. Another important insight to be gained from Butler's ethics of recognition is that it should precede the ethics of judgement that informs the judicial process. Though we can easily judge a person we claim not to recognise, this would in fact be unethical and unjustifiable. The constitutive opacity in which autonomous action takes root is something we share with the other. The judge who forgets that her judgement is informed by her own autonomic thought processes may take a shortcut to conviction, ignoring the need to first suspend her judgement. Even if the position of the judge is one of authority, calling the defendant to account for his incriminated behaviour, she should remember the vulnerability of autonomous action. At the same time we should not defer from judging another if he has wrongfully and culpably violated shared norms, because since this is how we learn to become responsible persons, even if we decide to denounce the norms we find to imposed upon us. Autonomous thought and action is relative, relational and grounded in a constitutive opacity. Nevertheless our capability to distance ourselves from the autonomic sources of the self through language and technologies of perception and communication is cause for celebration. It has turned us into vulnerable and open beings, suffering from many illusions that sometimes create their own reality, but capable of retracing our steps and reinventing *a* future.

Building on the recognition of this vulnerability we need to research how proactive environments 'wired' with autonomic computing systems will impact our capability for autonomous thought. To what extent will digital technologies overrule the distantiation afforded by the written and printed script and what does this mean for reflective consciousness and the constitution of the self? I hope that this chapter will contribute to an exploration of these issues, in urgent need of further interdisciplinary research.

6 References

- Arendt, H. (1998), *The Human Condition*, Chicago, IL: The University of Chicago Press
- Ambrose, S.H. (2001), Paeleolithic Technology and Human Evolution, (291) *Science*, p. 1748-1753
- Aarts, H. (2007), On the Emergence of Human Goal Pursuit: The Nonconscious Regulation and Motivation of Goals, *Social and Personality Compass*, p. 183-201
- Aarts, E. and S. Marzano (2003), *The New Everyday. Views on Ambient Intelligence*, Rotterdam: 010
- Bakhtin, M. (1993), The Problem of Speech Genres, in: C. Emerson and M. Holquist, *Speech Genres and Other Late Essays*, Austin: University of Texas Press
- Bargh, J.A. (2005), Bypassing the Will: Toward Demystifying the Nonconscious Control of Social Behavior, in: Hassin, R.R., J.S. Uleman, J. A. Bargh (eds.) (2005), *The New Unconsciousness*, Oxford University Press US, p. 37-58
- Bateson, G. (1972). *Steps to an Ecology of Mind*. New York, Ballantine
- Butler, J. (2005), *Giving an Account of Oneself*, New York, Fordham University Press
- Custers, B. (2004). *The Power of Knowledge. Ethical, Legal, and Technological Aspects of Data Mining and Group Profiling in Epidemiology*. Nijmegen, Wolf Legal Publishers
- De Hert, P. and S. Gutwirth (2006). Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power. *Privacy and the Criminal Law*. E. Claes, A. Duff and S. Gutwirth. Antwerpen Oxford, Intersentia
- Derrida, J. (1994), *Force de Loi*, Paris: Gallilée
- Dijksterhuis, A. and L.F. Nordgren (2002), A Theory of Unconscious Thought, 1 *Perspectives on Psychological Science* (2), p. 95-109
- Dijksterhuis, A. and Z. Van Olden (2006), On the benefits of thinking unconsciously: Unconscious thought can increase post-choice satisfaction, (42) *Journal of Experimental Social Psychology*, p. 627-631
- Eisenstein, E (2005), *The Printing Revolution in Early Modern Europe* (2nd edn), Cambridge and New York, Cambridge University Press
- Fayyad, U. M., G. Piatetsky-Shapiro, et al., Eds. (1996). *Advances in Knowledge Discovery and Data Mining*. Meno Park, California - Cambridge, Mass. - London England, AAAI Press / MIT Press
- Franz, R., Herbert Simon. Artificial intelligence as a framework for understanding

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

intuition, (24) *Journal of Economic Psychology*, p. 265-277

Gibson, J. (1986), *The Ecological Approach to Visual Perception*, New Jersey: Lawrence Erlbaum

Gigerenzer, G. (2007), *Gut Feelings. The Intelligence of the Unconscious*, New York: Penquin

Gallagher, D. and D. Zahavi (2008), *The Phenomenological Mind. An Introduction to Philosophy of Mind and Cognitive Science*, Routledge: London and New York

Glaser, J. and J.F. Kihlstrom, Compensatory Automaticity: Unconscious Volition Is Not and Oxymoron, in: Hassin, R.R., J.S. Uleman, J. A. Bargh (eds.) (2005), *The New Unconsciousness*, Oxford University Press US, p. 171-195

Goody, J and I. Watt (1963), The Consequences of Literacy (5) *Comparative Studies in Society and History* 3, p. 304–45

Goldberg, E. (2005), *The Wisdom Paradox. How Your Mind Can Grow Stronger As Your Brain Grows Older*, Gotham: New York

Goldberg, E. (2006), The Wisdom Paradox, in: H.R. Moody, *Aging*, Thousands Oaks: California

Haggard, P. and B. Libet (2001), Conscious Intention and Brain Activity, (8) *Journal of Consciousness Studies* 11, p. 47-63

‘t Hart, A.C. (1995), Mensenwerk? Over rechtsbegrip en mensbeeld in het strafrecht van de democratische rechtsstaat, KNAW Mededelingen van de Afdeling Letterkunde, Nieuwe Reeks, Deel 58 no. 4 (Proceedings of the Dutch Royal Academy of Sciences)

Harcourt, B.E. (2007), *Against Prediction. Profiling, Policing and Punishing in an Actuarial Age*, Chicago and London: The University of Chicago Press

Hassin, R.R., J.S. Uleman, J. A. Bargh (eds.) (2005), *The New Unconsciousness*, Oxford University Press US

Hildebrandt, M. (2006a). Privacy and Identity. *Privacy and the Criminal Law*. E. Claes, A. Duff and S. Gutwirth. Antwerpen - Oxford, Intersentia: 43-58

Hildebrandt, M. and B.-J. Koops (2007). *A Vision of Ambient Law*. Brussels, FIDIS Deliverable 7.9, available via www.fidis.net

Hildebrandt, M. (2008a) 'Profiles and Correlatable Humans', in: Nico Stehr (ed.), *Who Owns Knowledge. Knowledge and the Law*, New Brunswick, New Jersey: Transaction Books, p. 265-285

Hildebrandt, M. (2008b), Profiling and the Identity of the European Citizen, in:

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

- Hildebrandt, M. and S. Gutwirth (eds.) (2008), *Profiling the European Citizens. Cross-Disciplinary Perspectives*, Springer: Dordrecht, p. 303-326
- Hildebrandt, M. (2008c), Ambient Intelligence, Criminal Liability and Democracy, (2) *Criminal Law and Philosophy* 2, p. 163-180
- Hildebrandt, M (2008d), A Vision of Ambient Law, in: R. Brownsword and K. Yeung (eds.) *Regulating Technologies*, Oxford: Hart, 2008, p. 175-191
- Hildebrandt, M. (2008e) Legal and technological normativity: more (and less) than twin sisters, (12) *TECHNÉ* (12) 3, p. 169-183
- Hildebrandt, M. and S. Gutwirth (eds.) (2008), *Profiling the European Citizens. Cross-Disciplinary Perspectives*, Springer: Dordrecht
- Hildebrandt, M. (2009), Technology and the End of Law, in: Erik Claes en Bert Keirsbilck, *Facing the Limits of Law*, Dordrecht Springer (forthcoming)
- Hudson, B. (2005). Secrets of Self: Punishment and the Right to Privacy. *Privacy and the Criminal Law*. E. Claes and A. Duff. Antwerp Oxford, Intersentia, p.
- Ihde, D. (1990a), *Technology and the Lifeworld*, Bloomington: Indiana University Press
- Ihde, D. (1990b), *Instrumental Realism: The Interface between Philosophy of Technology and Philosophy of Science*, Bloomington: Indiana University Press
- Ihde, D. (1993), *Postphenomenology: Essays in the Postmodern Context*, Evanston: Northwestern University Press
- Ihde, D. (2002), *Bodies in Technology*, Minneapolis: University of Minnesota Press
- ISTAG (2001) Scenarios for Ambient Intelligence in 2010, Information Society Technology Advisory Group; available at <<ftp://ftp.cordis.europa.eu/pub/ist/docs/istagscenarios2010.pdf>>
- ITU (2005), *The Internet of Things*, Geneva: International Telecommunications Union
- Jin, S., S-Y. Park, J.J. Lee (2007), Driver Fatigue Detection Using a Genetic Algorithm, (11) *Artificial Life and Robotics* 1, p. 87-90
- Kallinikos, J. (2006). *The Consequences of Information. Institutional Implications of Technological Change*. Cheltenham, UK Northampton, MA, USA, Edward Elgar
- Kallinikos, J. (2008), Living in Ephemeria. The Shortening Life Spans of Information, available at: <http://www.informationgrowth.info/docs/livinginephemeria.pdf>
- Kephart, J.O. and D. M. Chess (2003), The Vision of Autonomic Computing, *Computer*, January

© Hildebrandt, chapter 9 in:

‘Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology’, M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

- Lévy, P. (1990). *Les technologies de l'intelligence. L'avenir à l'ère informatique*, Paris, La Découverte
- Loftus, E. and K. Ketcham, *The Myth of Repressed Memory*, New York, St Martin's Press, 1994
- Maturana, H.R. and F.J. Varela (1991), *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht: Reidel
- Mead, G. H. (1959/1934). *Mind, Self & Society. From the standpoint of a social behaviorist*. Chicago - Illinois, The University of Chicago Press
- Merleau-Ponty, M. (1945), *Phénoménologie de la Perception*, Paris: Gallimard
- Mahlmann, Matthias (2007), 'Ethics, Law and the challenge of cognitive science', *German Law Journal*, 8 (6), 577-615.
- Morse, Stephen J. (2008), 'Determinism and the Death of Folk Psychology: Two Challenges To Responsibility from Neuroscience', *Minnesota Journal of Law Science & Technology*, 9 (1), 1-35.
- Ong, W. J. (1982), *Orality and literacy: The technologizing of the word*, London: Methuen
- Plessner, H. (1975), *Die Stufen des Organischen unter der Mensch. Einleitung in die philosophische in die philosophische Anthropologie*, Frankfurt: Suhrkamp
- Ricoeur, P. (1986), *Du texte à l'action. Essais d'herméneutique II*, Paris, Editions du Seuil
- Ricoeur, P. (1992), *Oneself as Another*, Chicago: The University of Chicago Press
- Rouvroy, A. (forthcoming), *L'art de l'oubli dans la société de l'information. Contribution aux Actes du Colloque Asphalès sur « La protection de l'individu numérisé »*, 22 et 23 novembre 2007 Paris, CNRS (forthcoming with L'Harmattan), available at: http://works.bepress.com/antoINETte_rouvroy/
- Ryle, G. (1949), *The Concept of Mind*, New York: Barnes and Nobles
- Selinger, E. (2006), *Postphenomenology. A Critical Companion to Ihde*, New York: State University of New York Press
- Stengers, I. (2000), *The Invention of Modern Science*, Minneapolis: University of Minnesota Press
- Strack, F., L. Werth and R. Deutsch (2006), *Reflective and Impulsive Determinants of Consumer Behavior*, (16) *Journal of Consumer Psychology* 3, p. 205-216

© Hildebrandt, chapter 9 in:

'Law, Human Agency and Autonomic Computing. Philosophers of Law meet Philosophers of Technology', M. Hildebrandt & A. Rouvroy (eds.), Routledge 2011

Varela, F. J. (1999). *Ethical Know-How. Action, Wisdom and Cognition*. Stanford: Stanford University Press

Varela, F.J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind. Cognitive Science and Human Experience*. Cambridge Massachusetts: MIT Press

Wegner, D.M. (2002), *The illusion of conscious will*, Cambridge, MA: MIT Press

Wegner, D.M. (2007), Who is the Controller of Controlled Processes? in: Hassin, R.R., J.S. Uleman, J. A. Bargh (eds.) (2005), *The New Unconsciousness*, Oxford University Press US, p. 19-36

ⁱ Cf. Morse 2008 and Mahlmann 2007. Defining human agency in terms of practical reason, based on beliefs and desires, corresponds with Donald Davidson's definition of agency.

ⁱⁱ Indeed, one of the problems of autonomic computing may be that its genre is limited to one only, which assumes and produces unambiguous 'ontologies'. In computer science an ontology is 'a formal representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain', see [http://en.wikipedia.org/wiki/Ontology_\(computer_science\)](http://en.wikipedia.org/wiki/Ontology_(computer_science)).

ⁱⁱⁱ This section draws on Hildebrandt 2008a, Hildebrandt and Gutwirth 2008, Hildebrandt 2007 and on exchange within the workpackage on profiling of the EU funded Future of Identity in Information Society (FIDIS) Network of Excellence (NoE), see FIDIS deliverables 7.2/3/4/5/7/9/12/14 at www.fidis.net. See also the research project on 'Law and Autonomic computing: Mutual Transformations' at <http://www.vub.ac.be/LSTS/research/lawandautonomiccomputing.pdf>.

^{iv} See the European Funded Coordinated Action (CA) on 'Cooperating Embedded Systems for Exploration and Control featuring Wireless Sensor Networks' at <http://www.embedded-wisents.org>.

^v See for instance: <http://www.energy-retail.org.uk/smartmeters.html>.

^{vi} The usage of profiling techniques is not restricted to marketing and smart homes; it can also be adopted by forensic intelligence or elsewhere in the criminal justice system, see Harcourt 2007.

^{vii} Representational in the sense that like certain ontological propositions are said to represent within the mind *the* reality as it 'is' outside the mind, certain deontological proposition can be thought to represent within the mind how actions should take place outside the mind (in the real world). This mentalistic picture is related to a correspondence theory of truth and depends on a strict separation of an inside and an outside, evoking the problematic of the homunculus. Profound criticism of this type of epistemology can be found, for instance, in Ryle (1949), Merleau-Ponty (1945) , Varela et al. (1991) and recently in Gallagher and Zahavi (2008).

^{viii} It is unclear whether Goldberg's use of the term 'analytic' refers to conscious analysis, or whether it also refers to information processing by the unconscious mind.

I will use the term in reference to conscious deliberation and analysis, as will become clear below.

^{ix} The idea of postanalytic autonomic thinking must not be taken to mean that we could ever consciously decide how to inscribe desired habits of thought and action into the neural pathways of the brain. For interesting descriptions of the impact of consciousness on habit formation see Bargh (2005:51-54), see also Strack and Werth (2006).

^x Let me be quick to note that previous or other societies must also be qualified as complex. However, with the advent of the script, the printing press and the digital era the scope of interhuman relationships has moved way beyond the face-to-face and the potential permutations and combinations of information, the ensuing de- and recontextualisation as well as de- and reterritorialisation has increased the complexity of relationships between humans and nonhumans exponentially. See Lévy 1990, Kallinikos 2006.

^{xi} This is an important clue to the importance of the right to privacy, which is what De Hert and Gutwirth (2006) have coined legal ‘opacity tools’. The fact that we may want to shield knowledge about our selves from the gaze of others does not imply that we are transparent to our selves. Cf. Hudson (2006) who discusses privacy as a tool to protect *the secrets of the self*.

^{xii} ‘Coming about’ is also a sailing term which implies that the sailboat is turned by passing the bow of the boat through the wind. In this text I use it equivalent with ‘emerging’ and one could connect the two by remarking that the eruption of a self-consciousness is like taking a turn that is both continuous and discontinuous with non-reflective consciousness.

^{xiii} Her relational understanding of the self comes close to that of Mead (1959/1934), Ricoeur (1992) and Plessner (1975), all stressing the alterity that resides at the core of the self as well as the ex-centric constitution of the self.

^{xiv} This is why we must cherish the possibility to provide different accounts of our self, resisting efforts by others to impose a kind of self-identity that demands transparency and forbids what has been called ‘the right to oblivion’. See saliently Rouvroy (2009): ‘Nous voulons soutenir que l’une des conditions nécessaires à l’épanouissement de l’autonomie individuelle est, pour l’individu, la possibilité d’envisager son existence non pas comme la confirmation ou la répétition de ses propres traces, mais comme la possibilité de changer de route, d’explorer des modes de vie et façons d’être nouveaux, en un mot, d’aller là où on ne l’attend pas.’

^{xv} Quoted by Butler (2005:11) from Friedrich Nietzsche, *On the Genealogy of Morals* (translated by Walter Kaufmann), New York: Random House 1969, at 84.

^{xvi} ‘t Hart (1995) has given a penetrating account of the fact that we are not born as subjects, referring amongst others to the work of the Swiss biologist Adolf Portmann (1887-1982), who emphasised that the biological birth of the child does not coincide with the birth of the person.

^{xvii} For Arendt human action entails an element of originality, spontaneity and freedom that cannot be reduced to an instrumental understanding of behaviours. This element, which she coined ‘natality’, seems to me to be on the verge of subconscious

autonomic processes (which should not necessarily be understood as mechanical) and conscious appropriation.