2013

Profile Transparency by Design? Re-enabling Double Contingency

Mireille Hildebrandt, *Radboud University Nijmegen*

Draft, please cite the published version in: M. Hildebrandt, K. de Vries (eds.), *Privacy, Due Process and the Computational Turn*, Routlegde 2013.

Chapter 9

Profile transparency by design?

Re-enabling double contingency

Mireille Hildebrandt

Introduction

The technologies of machine-learning render us transparent in a rather counterintuitive manner. We become *transparent* in the sense that the profiling software looks straight *through us* to 'what we are like', instead of making transparent 'what or who we are'. This reminds me of a cartoon that shows a couple, seated in bed – after the act – confronted with a voice-over that proclaims: 'I'm glad you enjoyed that. People who like that technique also enjoyed these other sexual techniques:…'.[1] It is interesting to note that the couple - who may have felt they just had a unique experience - is brought down to earth with a reminder of the repetitive nature of human interaction. They are reduced to being like many others and invited to explore the consolidated repertoire of those who are like them. In machine learning jargon the couple is mapped to its 'nearest neighbours' and even if their 'k-anonymity' prevents their unique identification, they are machine-readable in terms of their likeness to a 'cluster' of other couples.

Privacy advocates often focus on unique identification as the main attack on our privacy. Data minimisation and anonymity are often depicted as the prevalent strategies to protect what is understood to be the core of privacy: the need not to be singled out, not to be recognized as the unique individual person we hope to be. So-called user-centric identity management systems are developed to allow people to maintain contextual integrity, to restrict information flows within specific contexts and to manage the set of different roles they play in different environments such as work, leisure, home, school, sports, entertainment, shopping, healthcare, and the more. The holy grail of this version of contextual integrity is unlinkability, a notion that refers to techniques that should disable cross-contextual aggregation of individual profiles. In line with this, user-centric identity management refers to the use of credentials or attributes instead of full identification, which means that access to specific services is gained by merely disclosing the relevant attribute of a person (e.g. being over 18 years old, being female, being an employee, having paid for the service).

In the meantime reality presents us with an incentive structure that encourages business models that thrive on *consent* to override the purpose limitation principle,[2] or on *anonymisation* that renders most of data protection legislation inapplicable.[3]

Together, uninformed consent and anonymisation facilitate continuous and persistent interpenetration of contexts, enabling smart infrastructures to develop fascinating cross-contextual profiles of consumers and citizens, often creating new types of knowledge of cross-contextual behaviour. In an elucidating text (Massiello and Whitten, 2010) consider the high potential of function creep; they celebrate the re-use of data for unforeseen purposes that creates unexpected added value. This is assumed to create a win-win situation. Consumers find their desires satisfied before they become aware of them and advertising networks charge advertisers for the most precious scarce good of the information-driven society: human attention.

As indicated in the title of this volume, this volume seeks to flesh out the impact of data science on privacy and due process. In this chapter I will focus on due process, understood in the broad sense of the effective capability to contest decisions that have a significant impact on one's life. To contest such decisions, a person must be aware of them and be able to foresee their impact. In that sense due process requires transparency and/or knowledge symmetry. The research questions for this chapter are, first, how the application of data science challenges such transparency and, second, how we can reinvent it with regard to the proliferating machine-generated profiles that have an increasing influence on our life.

In section 2, I will introduce the Deleuzian concepts of de-realisation and virtualisation to elucidate what it is that profilers construct when they create large 'populations' of anonymous profiles that can be applied to large populations of individual human beings. In section 3, I will continue this line of thought and add the Deleuzian concept of the dividual, aligned with terms from the domain of computer science: data, data models, attributes, characteristics and properties. This should help to prepare the ground for an answer to the question of whether data science practices in the field of commerce and law enforcement afford virtualisation or merely de-realisation. In section 4, I explore the notions of transparency and enlightenment, connecting them to Parsons's and Luhmann's concept of double contingency. This regards the double and mutual anticipation that is constitutive for self, mind and society. The concept refers to the fundamental uncertainty that rules human communication, creating the possibility for new meaning amidst inevitable but productive misunderstandings. It also refers to the need for socio-technical infrastructures that stabilise meaning amongst individual minds, notably language, writing, the printing press, hyperlinked digitisation and finally the hidden complexity of computational decision-systems. In section 5, I will engage with Stiegler's notion of tertiary retention and the need to reinvent what he terms 'introjection' in the digital age. He introduces these notions in a plea for a new Enlightenment that should inform a new Rule of Law. Finally, in section 6, I will argue that renegotiating a novel double contingency will require profile transparency at the level of the digital infrastructure. I conclude with a brief sketch of what this could mean.

Through the looking glass: de-realization or virtualization?

Data science can be seen as a derivative of *Artificial Intelligence the Modern Approach* (AIMO, see Russell and Norvig, 2009)). It generates non-trivial information on the basis of statistical inferences mined from what has been called 'Big Data'. Some would claim that machines have 'come off age', generating types of pattern-recognition way beyond 'good old fashioned artificial intelligence' (GOFAI).[4]

Inductive learning, bottom-up algorithms, contextual awareness and feedback mechanisms all contribute to a novel type of transparency, an invisible kind of visibility (Hildebrandt, 2009), based on a continuous, pervasive, seamless stream of comparisons. Individuals are thereby represented as an assemblage of different roles that cut across large 'populations' of similar roles. The term population refers to a concept within the domain of statistics, where it denotes the complete set of phenomena that is under investigation, of which only a sample can be examined in detail. Statistics contains the rules along which the outcome of the study of the sample can be extrapolated or generalised to the entire set, the population. Data science seems to allow operations on a subset of a population that is far more extensive than a sample. Together with the mathematical complexity of the operations that can be performed by current computing systems, data science is capable of generating types of pattern recognition far beyond the reach of earlier statistical inference. In fact, unsupervised learning algorithms can generate, test and adapt hypotheses instead of merely confirming or refuting them. This is why Anderson (2008) spoke of the end of theory: he asserted that machine learning will soon be better at constructing and fine-tuning hypotheses than human beings will ever be. This is made possible by the fact that data mining operations are capable of working on unprecedented populations of data, creating what Amoore (2011) has coined 'data derivatives' that easily turn into new populations: resources for further research. Note, that we are now speaking of generations of populations: starting from the original flux of lifeworld phenomena, followed by their translation into machine-readable data models, followed by the inferred profiles (data derivatives), followed by the inferences that build on these first generation derivatives (constituting second generation derivatives). The original phenomena as framed by the mind of whoever design the research is the first generation population, their translation into discrete data models is the second generation population, the first set of inferences is the third generation population and so forth. The interesting and pertinent question is how fourth or fifth generation 'populations' connect with the first generation population. And, of course, how all this connects with the population that constitutes human society.

In the present infosphere, the plethora of machine-readable profiles do not offer the individuals to whom such profiles are applied a looking glass (mirror), where they can see how they are being matched against inferred profiles. Instead these profiles provide the company or authority who paid for the software with a way to reach out behind *their* looking glass, gazing straight through our condensed selves into the disentangled sets of 'similars', showing a maze of association rules that link us with – statistically – relevant lifestyles, demographic or geographic types, health risks or earning capacities. Just like in the famous story of *Through the Looking Glass, and What Alice Found There* (Carroll, 2000), 'profiling machines' (Elmer, 2003) open up an alternative world, basically consisting of simulations of our future states or behaviours. Profiling machines or 'inference engines'[5] thus function as a looking glass that provides an opportunity to peak into this alternative world. This enables the industry to calculate what profits can be gained from catering to consumers' inferred preferences, and similarly enables public authorities to calculate what types of offences may be committed where, when and possibly by whom, e.g. social security fraud.

In principle profiling systems could function as virtualisation machines in the sense of Deleuze's (Deleuze, 1994) conceptualization of the virtual.[6] To clarify what he means with the virtual I will build on the work of cyber-philosopher Lévy (1998 and 2005) who elaborated Deleuze's notion of virtualisation for the digital age. Deleuze understands the *virtual* in relation to what he calls the *actual*, and opposes this pair to that of what he calls the *possible* and the *real*. For Deleuze what is real is what exists. However, the real has two modes of existence: the virtual and the actual. His use of the terms of virtual/actual and possible/real may not be congruent with our common sense, but he derives them from an imaginative reconstruction of our philosophical tradition, thus shedding light on phenomena that our current common sense may not grasp. While drawing on medieval philosophy Deleuze re-engineers philosophical concepts, to provide the conceptual tools needed in the era of data derivatives (Lévy, 1998: 23):

> The word 'virtual' is derived from the Medieval Latin *virtualis*, itself derived from *virtus*, meaning strength or power. In scholastic philosophy the virtual is that which has potential rather than actual existence. The virtual *tends* towards actualization, without undergoing any form of effective or formal concretization. The tree is virtually present in the seed. Strictly speaking, the virtual should not be compared with the real but the actual, for virtuality and actuality are merely two different ways of being.

There is – according to this particular understanding of the virtual – a crucial difference between the possible and the virtual logical (Lévy, 1998: 24):

> The possible is already fully constituted, but exists in a state of limbo. It can be realized without any change occurring either in its determination or nature. It is a phantom reality, something latent. The possibility is exactly like the real, the only thing missing being existence. The realization of a possible is not an act of creation, in the fullest sense of the word, for creation implies the innovative production of an idea or a form. The difference between possible and real is thus purely.

The virtual, therefor, should not be compared to the real (since it is already real), but to the actual. Back to the seed (ibid):

> The seed's problem, for example, is the growth of the tree. The seed is this problem, even if it is also something more than that. This does not signify that the seed knows exactly what the shape of the tree will be, which will one day burst into bloom and spread it leaves above it. Based on its internal limitations, the seed will have to invent the tree, coproduce it together with the circumstances it encounters.

This implies that actualisation can be understood as the solution to a problem, a solution, however, that is never entirely determined by that problem since it requires a co-creation with its not entirely predictable environment. Whereas realisation is the *predetermined* concretisation of a possible (e.g. execution of a computer program), actualisation is the production of a solution to a problem that entails a measure of uncertainty (neural networks hosting unsupervised algorithms?). In respect of data derivatives the more interesting transition is the reversal of actualisation: virtualisation. Whereas 'actualization proceeds from problem to solution, virtualization proceeds from a given solution to a (different) problem' (Lévy, 1998: 27). This move generates the generic set of problems that gave rise to the particular

solution, and creates room for alternative solutions. Note that for Deleuze, virtualisation is not a matter of de-realisation. The art of virtualisation is to stick to the realm of the real, to resist moving back to a predefined set of possibilities that merely lack reality. Derealisation severely restricts the kinds of solutions that can be generated, because it remains in the realm of necessity and mechanical application. On the other hand 'virtualization is one of the principle vectors in the creation of reality' (Lévy, 1998: 27): by shifting from concrete solutions to virtual problems we create the precondition for novel acts of creation, generated in the course of novel types of actualization.

The question at stake in this chapter is how we should understand the virtual machines,[7] inference engines or profiling technologies that 'look through' our selves at the myriad of potentially similar states or entities. Are they machines of virtualisation in the sense of Deleuze or machines of what he terms de-realization? Do they provide a range of overdetermined possibilities (in Deleuze's sense) or do they provide sets of virtuals that allow for novel, underdetermined actualizations? Is data science a science of de-realization or an art of virtualization? If these machineries merely present us with endless variations of what is already present (de-realization) we may be strangled in the golden cage of our inferred preferences. If data science, however, triggers pools of unexpected similarities that evoke and provoke unprecedented articulations of self and other – we may be the lucky heirs to an extended domain of co-creation. In that case the question becomes who are cultivating this extended domain and who get to pick its fruits?

Dividuals and attributes: possibles or virtuals?

There is a fascinating indifference with regard to individual persons in machine learning and other profiling technologies, summed up by Rouvroy (2011) under the heading of the 'statistical governance of the real'. Individuals seem to count only as a resource of data or as a locus of application; in a sense the individual has finally become what (Deleuze, 1992) famously coined an assemblage of 'dividuals'. As explained when discussing the concept of population, the aggregate that forms the basis of knowledge construction is not a mass of people but a 'bank' of assorted data, correlated in numerous ways by a variety of techniques. These data do not concern *in*dividuals, they are not necessarily meant to identify a unique person. Rather, they allow their masters to connect the dots, generating a plethora of permutations and combinations (Deleuze, 1992):

> We no longer find ourselves dealing with the mass/individual pair. Individuals have become 'dividuals,' and masses, samples, data, markets, or 'banks.'

The focal point of data science is not the indivisible person as the smallest unit in various types of populations: the individable *in*dividual. The focal point is the multiplicity of machine-readable attributes used to assemble the various types of units that compose a variety of populations that are not necessarily 'made up' of people. Populations may for instance be populated with hair colour types, employment segments, health risks, security threats or other units of comparison. This, however,

does not mean that the profiles mined from such data aggregates will not be applied to individual human persons.

In the Netherlands a producer of adult diapers phoned people on behalf of their pharmacy to inquire after their urine loss. Based on these inquiries people were categorized as fitting various 'user-profiles', which would be employed to decide on the compensation paid for incontinence diapers by their insurance company, who had asked the pharmacies for categorisation. The uproar this caused led to an immediate termination of the policy, with the Minister for Healthcare speaking out against commercial companies thus gaining access to sensitive data. The insurance company quoted the need to reduce costs as a reason for the construction of user profiles that determine the attribution of compensation (Pinedo, 2012). We may guess that at some point the quantifiable level of incontinence can be inferred from the fusion of various databases, or from information gleaned from medically prescribed apps even if these were originally dedicated to other purposes, [8] creating dividuals depicting incontinence probability. By then nobody has to phone patients or clients to remind them of their problem, and some would claim that the automation of profiling is therefor less invasive, while also more objective than human assessment. To the extent that such granular profiling informs automated or semi-automated decision-systems we may have to learn to live with an extended family of dividuals that co-determine how government agencies, companies, health insurance and public utility providers 'read' us. These dividuals are not of our own making, we do not choose them and are hardly aware of their 'existence'. They are virtual in the common sense of not-physical or abstract; they depict the kind of profiles we fit at the level of statistical inferences and to the extent that they involve mechanical application dividuals seem to stand for de-realization in the sense of Deleuze's genealogy of the real. They form abstractions of an individual, based on the match between some of her attributes with profiles inferred from masses of attributes from masses of individuals. Profiling thus transforms the original 'mass' that is composed of a mass of 'individuals', into an aggregate of attributes that cut across and divide the *in*dividuals into their elements, characteristics, properties or attributes. As explained, these elements or properties are not given, they are *attributed* by whoever write the algorithms of data analytics, taking into account that whatever dividuals are sought after they must be inferred from machine-readable data by machine-readable algorithms.

To investigate whether – or in which types of cases - the computerized gaze through the looking glass is a matter of de-realization or an act of virtualization, we need to look into the subtle negotiations that determine the attributes filling the databases. We must take into account that such databases are often seen as equivalent with the population of statistical inferences. It is tempting to assume that Big Data mining does not work with samples but with – nearly – complete populations, and this reinforces the assumption that inferred predictions are accurate, precisely because all data have been taken into account. This, however, is an illusion, as any machine-learning expert can explain. The term 'attribute' is salient here, because it highlights the performative act of negotiating the types of data that will fill the databases, data servers and cloud computing systems on which data mining operations run. To demonstrate this point it is instructive to check the Wikipedia entries for the term 'attribute'. Wikipedia distinguishes between an attribute in research, philosophy, art, linguistics and computer science (Wikipedia contributors, 2012). In research, it qualifies an attribute as a characteristic of an object that can be operationalized by giving it a certain value (e.g. yes/no, or blue/red/green/yellow, or 1, 2, 3, 4) to allow

for further data processing. Similarly, Wikipedia qualifies a property in modern philosophy, logic and mathematics a property as an attribute (a quality) of an object, while this attribute may be considered an object in its own right, having various properties of its own. Numerous highly refined philosophical debates have assembled around the notion of property (essential or accidental, determinate or determinable, ontological or epistemological), which we need not enter here (Swoyer, Orilia, 2011). What is relevant is the fact that attributes predicate a noun, qualifying and thus limiting its denotation. This implies that attributes are always attributes of something, even if that something is another attribute. In relational databases, attributes are used to define a property of an object or a class; to work with such attributes they are associated with a set of rules called operations, which define how they are computed within the relevant database. In a particular instance an attribute has a particular value. The class could for instance be 'woman', the attribute could be '> 40 ' (value: yes or no). Evidently we can imagine a class 'human being', with attributes 'sex' (value: man or woman) and 'age' (value: any number between 0 and 150). So, whether something is a class or an attribute depends on the structure of the database. We can easily think of attributes of attributes and various types of relationships between classes and objects that are defined in terms of required attributes. The crucial point here is the fact that attributes, characteristics or properties are not given, but attributed. The decision on which attributes define what classes or what objects as members of classes has far reaching implications. It determines the scope and the structure of the collective that populates the database and this has consequences for the output of the operations that are performed on the database. For instance, to the extent that the output feeds into an assessment-system or a decision-system, the attributes and the way they structure the database co-define the output. The chosen data models thus co-define how the user of the system perceives and cognizes her environment and how she will act, based on such perceptions and cognition. Especially in complex technological environments that integrate pervasive or even ubiquitous machine-to-machine communications that continuously assess the environment, machine learning will be based to a large extent on feedback loops between computing systems, potentially also grounding their key performance indicators on the output of a string of interacting inference machines. In that sense computing systems may become increasingly self-referential, deferring to subsystems or collaborating in the context of multi-agent systems capable of generating emergent behaviour. Much will depend, then, on the manner in which the flux of real life is translated into machine-readable terms. Referring to the effects of the proliferation of information-processing machines on learning processes, Lyotard wrote – back in the '70s of the last century (Lyotard, 1984: 4):

> The nature of knowledge cannot survive unchanged within this context of general transformation. It can fit into the new channels, and become operational, only if learning is translated into quantities of information. We can predict that anything in the constituted body of knowledge that is not translatable in this way will be abandoned and that the direction of new research will be dictated by the possibility of its eventual results being translatable into computer language.

My point here is not that some types of knowledge are not translatable into computer language. This seems an obvious, though somewhat trivial observation. The same point can indeed be made for the script and the printing press, which require their own translations and – just like data-driven environments - require and produce a novel mind-set as compared to a previous or later ICT infrastructure.[9] My point is that the

flux of real life events can be translated into computational formats *in different ways* and that what matters is to what extent alternative translations produce *alternative outcomes.* To come to terms with this we will have to find ways to play around with the multiplicity of dividuals that are used to profile us. Whereas we have learned to play around with written language, we may have more difficulty in achieving similar standards of fluency under the computational paradigm of proactive technological environments. This seems far more challenging, precisely because these environments outsource major parts of knowledge production and decision-making to complex interacting computing systems. Cognitive scientists claim that we are not hard-wired to understand statistics (Gigerenzer, 1991), let alone to absorb the complexities of knowledge discovery in databases. Our bounded rationality seems to require hidden complexity and intuitive interfaces to come to terms with an environment that seamlessly adapts to our inferred preferences (Weiser, 1991). This may, however, be a relief as well as a problem, depending on the extent to which we can guess what dividuality is attributed to us and how that may impact our life. The problem may seem to relate to the use of pseudonyms to separate contexts and roles, but the dividuals created by automated decision systems are not of our choice. They stand for the new stereotypes generated by our smart environments and could thus be termed artificial stereotypes, created by the unbounded computational irrationality of our environments – instead of being the result of the bounded rationality of human cognition.[10]

Deleuze decribed the transformation of societies structured by practices of discipline into societies organized by practices of control. Inspired by Foucault, Deleuze's disciplinary societies are characterized by enclosed spaces (monasteries, prisons, schools, hospitals, factories) and regulated temporalities, organised in a manner that renders individual subjects observable and predictable, thus inducing a process of self-discipline that aligns their behaviour with the regularity of the average monk, inmate, student, patient or employee of the relevant rank. Control societies differ because their regulatory regime no longer depends on a stable separation of spaces nor a predictable regulation of temporalities. Home, school, work and leisure increasingly overlap, both in space and in time. It is no longer the creation of the average individual subject that is at the heart of the mechanisms that produce modern or postmodern society. Instead, the individual is divided, mixed and mocked-up into a range of dividuals that are controlled by the invisible manipulation of complex data models. Deleuze in fact relates this to the further virtualization of financial markets (ibid):

> Perhaps it is money that expresses the distinction between the two societies best, since discipline always referred back to minted money that locks gold as numerical standard, while control relates to floating rates of exchange, modulated according to a rate established by a set of standard currencies.

Floating exchange rates were of course just the beginning. By now we know that flows of money, interest, options, derivatives and futures are increasingly determined by the automation of machine-readable inferences. And we are rapidly becoming aware of the complex feedback loops this entails between what Esposito (2011) has called the impact of the present futures on the future present. This may suggest that human subjects are progressively left out of the equation, but it is important to note that the automation ultimately concerns *inferences from* and *associations of* data

points that are traces from human behaviours. The use of the plural in the term 'behaviours' is telling. It refers to the machine-observable behaviour of persons, cut up into discrete data points that can be processed to compare and reconstruct dividuals, displaying a variety of probable future behaviours. The shift from a singular behaviour to a plurality of behaviours is significant; it alludes to the fragmentation and recombination that is typical for data mining operations and builds on the de- and re-contextualisation that is pivotal for pattern recognition in databases that have been fused (Kallinikos, 2006).

Summing up, the architecture of the data models that forms the basis for mining operations is decisive for whatever outcome the process produces. On top of that, Sculley and Pasanek (2008) have suggested that data mining builds on five assumption that are not necessarily valid.[11] In a salient article on 'Meaning and Mining: the Impact of Implicit Assumptions in Data Mining for the Humanities' - they point out that (1) machine learning assumes that the distribution of the probabilistic behaviour of a data set does not change over time, whereas much of the work done in the humanities is based on small samples that do not pretend such fixed distribution (focusing on change and ambiguity rather than invariance over the course of time), (2) machine learning assumes a well defined hypothesis space because otherwise generalization to novel data would not work, (3) for machine learning to come up with valid predictions or discoveries the data that are being mined must be well represented, avoiding inadequate simplifications, distortions or procedural artifacts, (4) machine learning may assume that there is one best algorithm to achieve the one best interpretation of the data, but this is never the case in practice, as demonstrated by the 'No Free Lunch Theorem' (which says there is no data mining method without an experimenter bias).[12] To illustrate their point they develop a series of data mining strategies to test Lakoff's claim about there being a correlation between the use of metaphor and political affiliation, both via various types of hypothesis testing (supervised learning methods) and via various types of clustering (unsupervised learning methods). They conclude that (ibid 2008:12):

> Where we had hoped to explain or understand those larger structures within which an individual text has meaning in the first place, we find ourselves acting once again as interpreters. The confusion matrix, authored in part by the classifier, is a new text, albeit a strange sort of text, one that sends us back to those text it purports to be about.

In fact they continue (ibid:17):

> Machine learning delivers new texts – trees, graphs, and scatter-grams – that are not any easier to make sense of than the original texts used to make them. The critic who is not concerned to establish the deep structure of a genre or validate a correlation between metaphor and ideology, will delight in the proliferation of unstable, ambiguous texts. The referral of meaning from one computer-generated instance to the next is fully Derridean.

Under the heading of section 6 I will return to this point and briefly discuss a set of recommendations, provided by Sculley and Pasanek, that should mitigate the risks generated by these assumptions.

Here I conclude that a closer look at the construction work needed to produce dividuals or data derivatives gives us some insight into the question of whether dividuals are virtuals or possibles. In management speak: tools that empower the inhabitants to play around with their smart environments or tools that manipulate them as mere resources for the computing systems that run the infrastructure, keeping them hostage to their inferred preferences. The answer is that this will depend on whether the inhabitants of these new lifeworlds will be capable of figuring out how their dividuals determine what and how they can act upon that. As with all capabilities,[13] this does not merely depend on their intelligence. To bring them in a position from where they can interact with their own dividuals (deleting, modifying or enhancing them) they will need a legal and technical infrastructure that affords such reconfigurations. I will return to this in section 6, after investigating how pre-emptive computing upsets and disrupts one of the core assumptions of human intercourse, siding with Stiegler's call for a new Enlightenment and a new Rule of Law.

Double contingency in the era of pre-emptive computing

In *Looking Awry*, Zizek (1991: 30) suggests that 'communication is a successful misunderstanding'. This may be the most salient summary of what is known as the concept of double contingency in sociology and philosophy, which denotes the most fundamental level of analysis concerning the coordination of human action.[14] Before fleshing out how data science practices may alter this 'primitive' of self, mind and society I will present a brief overview of the concept.[15]

The theorem of double contingency was first coined by Parsons (Parsons and Shils, 1951; Parsons, 1991), depicting the fundamental uncertainty that holds between interacting subjects who develop mutual expectations regarding each other in a way that objects do not. Pivotal here is that subjects must develop expectations about what the alter (the interacting subject) expects from them, to be able to 'read' their own interactions, and the same goes for their alter. The temporal dimension of interaction introduces a contingent and inevitable uncertainty about how the alter will understand one's action, knowing that the same goes for all interacting individuals. Parsons named this condition of fundamental uncertainty the double contingency or social interaction, highlighting the interdependence of the mutual expectations that provide a virtuous or vicious circle of iterative interpretation. Parsons basically builds on Mead's notion of the 'generalized other', that depicts the need to anticipate how others will understand us and how they will act upon our gestures, speech and actions. (Mead and Morris, 1962) explains this 'generalized other' with the example of a ball-game that requires a player to internalise the positions of the other players with regard to each other, to the rules of the game and to herself, to be able to interact fluently and successfully as a player of that game.

In other work we have coined this mutual and co-constitutive set of anticipated expectations 'double anticipation' (Hildebrandt, 2009). We have argued that this is what enables and constrains human interaction, and elaborated how it co-constitutes individual identity. We draw on Ricoeur's (1992) *Oneself as Another*, that provides a penetrating analysis of human identity that seems pertinent for the issue of transparency in computationally enhanced environments. If the construction of identity depends on our capability to anticipate how others anticipate us – we must

learn how to figure out the way our computational environment figures us out. Ricoeur discusses identity in terms of a relational self that must be situated on the nexus of the pair of continuity and discontinuity (diachronic perspective) and that of sameness and otherness (synchronic perspective). The most intriguing part of Ricoeur's analysis of human identity consists in his introduction of the concepts of *idem* (identical and identity, similarity and sameness, third person perspective) and *ipse* (selfhood, first-person perspective). In his account of personal identity Ricoeur demonstrates how our understanding of self-identity is contingent upon our taking the role of the other (the second person perspective) that eventually provides us with something like a third person perspective on the self (cf. also Mead's generalised other), which is constitutive for our developing 'sense of self' (first person perspective).

Parsons was less interested in personal identity than in the construction of social institutions as proxies for the coordination of human interaction. His point is that the uncertainty that is inherent in the double contingency requires the emergence of social structures that develop a certain autonomy and thus provide a more stable object for the coordination of human interaction. The circularity that comes with the double contingency is thus resolved in the consensus that is consolidated in sociological institutions that are typical for a particular culture. Consensus on the norms and values that regulate human interaction is Parsons' solution to the problem of double contingency and thus also explains the existence of social institutions. As could be expected, Parsons' focus on consensus and his urge to resolve the contingency have been criticised for its 'past-oriented, objectivist and reified concept of culture' and for its implicitly negative understanding of the double contingency.

A more productive understanding of the double contingency may come from Luhmann, who takes a broader view of contingency; instead of merely defining it in terms of dependency he points to the different options open to subjects who can never be sure how their actions will be interpreted. The uncertainty presents not merely a problem but also a chance, not merely a constraint but also a measure of freedom. The freedom act meaningfully is constraint by the earlier interactions, because they indicate how one's actions have been interpreted in the past and thus may be interpreted in the future. Earlier interactions weave into Luhmann's emergent social systems, gaining a measure of autonomy – or resistence - with regard to individual participants. Ultimately, however, social systems are still rooted in the double contingency of face-to-face communication.[16] The constraints presented by earlier interactions and their uptake in a social system can be rejected and renegotiated in the process of anticipation. By figuring out how one's actions are mapped by the other, or by the social systems in which one participates, room is created to falsify expectations and to disrupt anticipations. This will not necessarily breed anomy, chaos or anarchy, but may instead provide spaces for contestation, self-definition in defiance of labels provided by the expectations of others, and the beginnings of novel or transformed social institutions. As such, the uncertainty inherent in the double contingency defines human autonomy and human identity as relational and even ephemeral, always requiring vigilance and creative reinvention in the face of unexpected or unreasonably constraining expectations.

This is where Zizek's phrase comes in. By referring to communication as a misunderstanding, Zizek seems to acknowledge the inherent uncertainty that constitutes the meaning of our expressions. In a sense, we can never be sure whether what we meant to say is what the other understood. We can take the perspective of the

other to guess how they took what we uttered, but this switch of perspective is always an anticipation, or an interpretation. It will create interstitial shifts of meaning between one utterance and another, even if the words are the same. However, Zizek also acknowledges that the misunderstanding that grounds our attempt to communicate is productive. Insofar as the attempt succeeds and communication 'works', meaning is created in between the black boxes that we are – thus also allowing us to reinterpret our own intended meanings. Our self-understanding emerges *in* and *from* this process of meaning attribution, contributing to a sustained practice that constitutes self, mind and society.

The question is what this means for self, mind and society in the era of pre-emptive computing. In his description of behavioural advertising McStay (2011: 3) speaks of the pre-emption of intention as a crucial characteristic of targeted advertising. More generally one can see that the original idea of ubiquitous computing, Ambient Intelligence and the Internet of Things relies on the same notion: we are being serviced before we have become aware of our need for such service. In other words, before we have formed an explicit or conscious intention, the computational layer that mediates our access to products or services acts upon the inferred intention. That includes, for instance, the personalisation of search engine results or the 'auto-complete' functions in mail programs. Some speak of digital butlers (Andrejevic, 2002), who pre-empt the idiosyncratic urges of their masters without making a point of it. Jeeves revisited after the computational turn. Negroponte (1996: 149) explained the need for an *i*Jeeves in 1996:

> The idea is to build computer surrogates that possess a body of knowledge both about something (a process, a field of interest, a way of doing) and about you in relation to that something (your taste, your inclinations, your acquaintances). Namely, the computer should have dual expertise, like a cook, gardener, and chauffeur using their skills to fit your tastes and needs in food, planting, and driving. When you delegate those tasks it does not mean you do not like to prepare food, grow plants, or drive cars. It means you have the option to do those things when you wish, because you want to, not because you have to.
>
> Likewise with a computer. I really have no interest whatsoever in logging into a system, going through protocols, and figuring out your Internet address. I just want to get my message through to you. Similarly, I do not want to be required to read thousands of bulletin boards to be sure I am not missing something. I want my interface agent to do those things.
>
> Digital butlers will be numerous, living both in the network and by your side, both in the center and at the periphery of your own organization (large or small).

It is important to acknowledge that we are already surrounded by cohorts of digital butlers and I dare say we are better off with them than without. My argument in this chapter is not one of techno-pessimism and I do not believe in a romantic offline past where all was better, more authentic or less shallow. However, to the extent that our computational environment provides for an external artificial autonomic nervous system we must come to terms with the implications. This 'digital unconscious' thrives on 'subliminal strategies' to cater to our inferred preferences,[17] as long as whoever is paying for the hardware and the software can make a profit. The element of pre-emption that is hardwired and softwired into the computational layers that surround us may be a good thing, but we must find ways to guess how they are

guessing us. We must learn how to anticipate how these machineries are anticipating us. We must – in other words – reinvent a double contingency that reintroduces a successful misunderstanding between us and our computational butlers. If the misunderstanding fails, we may end up as their cognitive resources.[18]

1. A new Enlightenment: tertiary retention and introjection

In a presentation at the World Wide Web Consortium (W3C), Stiegler (2012) has called for a new Enlightenment, under the title of *Die Aufklärung in the Age of Philosophical Engineering*. The term philosophical engineering comes from an email by Berners-Lee, one of the founding fathers of the world wide web:[19]

> Pat, we are not analyzing a world, we are building it. We are not experimental philosophers, we are philosophical engineers. We declare 'this is the protocol'. When people break the protocol, we lament, sue, and so on. But they tend to stick to it because we show that the system has very interesting and useful properties.

Philosophy was done with words, mediated by handwritten manuscripts and later by the printing press. Now, Berners-Lee writes, we do it by means of protocols that regulate online behaviours. He paraphrases Marx's famous Thesis XI on Feuerbach (Marx and Engels, 1998: 571): 'The philosophers have only interpreted the world in various ways; the point is to change it'. The point of Berners-Lee is that this is precisely what engineers do, whether they like it or not. He is calling on them to acknowledge their impact and – if I may summarise his position – to engineer for a better world, or at least to refrain from engineering a bad one.

In his presentation, Stiegler goes even further. He traces the role of technics (the alphabet, the printing press, the digital) in the construction of thinking, relating Enlightenment thought to the workings of the printing press. His point is that the digital brain – constituted by what he calls the tertiary retention of the digital era, will not necessarily have the same affordances as the reading brain. If we want to preserve some of the affordances of 18th century Enlightenment that still inform our self-understanding, we must take care to engineer these affordances into the digital infrastructure. His main worry is that (ibid: 1/12):

> The spread of digital traceability seems to be used primarily to increase the heteronomy of individuals through behaviour-profiling rather than their autonomy.

Before explaining the notion of tertiary retention let me briefly reiterate that Stiegler emphasises that 'the web is a function of a technical system which could be otherwise', highlighting that whatever the present web affords, may be lost if its basic structure is amended. This seems to accord with the idea that 'technology is neither good not bad, but never neutral' (Kranzberg, 1986). Though a technological infrastructure such as the printing press, electricity or the Internet is not good or bad 'in itself' – it has normative consequences for those whose lifeworld is mediated. It changes the constraints and the enablers of our environments, opening up new paths but inevitably closing down other. Whether that is a good thing, is a matter of

evaluation – and this will depend on what is gained and what is lost compared to a previous or alternative infrastructure. And, for whom.

The concept of tertiary retention builds on Husserl's understanding of memory. The first retention is that of perception, unifying the flux of impressions generated by one's environment into the experience of one's own perception. This first retention is entirely ephemeral: 'the perceived object only appears in disappearing' (ibid: 5/12). Secondary retention is the imprint 'in the memory of the one who had the experience, and from which is may be reactivated' (ibid: 6/12). Note that Stiegler does not speak of information retrieval, since we know that secondary retention is an ongoing process whereby each novel secondary retention and each reactivation transforms – however little – the initial secondary retention.[20] According to Stiegler a tertiary retention is 'a spatialisation of time', meaning a transformation of 'the temporal flow of a speech such as the one I am delivering to you here and now into a textual space' (ibid: 4/12). In a way, this is a materialisation of the seemingly immaterial matter of time, though speech itself is obviously not immaterial (being produced by vocal organs disseminating sound waves etc.). A tertiary retention, such as writing, printing on paper or on silicon chips externalises, spatialises and materialises the flux and the imprint of primary and secondary retention (ibid: 5/12):

> One can speak of a visibly spatialising materialisation to the extent that there is a passage from an invisible, and as such in-discernable and unthinkable material state, to another state, a state that can be analysed, critiqued and manipulated – in both senses that can be given to this verb, that is:
>
> 1. on which analytical operations can be performed, and intelligibility can be produced; and
> 2. with which one can manipulate minds – for which Socrates reproached the sophists in the case of writing, writing being the spatialisation of time of what he called 'living speech'.[21]

Such tertiary retention is called by the name of grammatisation, which, according to Stiegler (ibid: 4/12):

> describes all technical processes that enable behavioural fluxes or flows to be made discrete (in the mathematical sense) and to be reproduced, those behavioural flows thought which are expressed or imprinted the experiences of human beings (speaking, working, perceiving, interacting and so on). If grammatisation is understood this way, then the digital is the most recent stage of grammatisation, a stage in which all behavioural models can now be grammatised and integrated through a planetary-wide industry of the production, collection, exploitation and distribution of digital traces.

Tertiary retention or grammatisation enables the sharing of content, of thoughts, of externalisations across time and space (Ricoeur, 1973; Geisler, 1985), across generational and geographical distances. It constitutes a transindividual retention that can survive the death of its author, but – as Stiegler emphasises – to empower individual persons it must be re-interiorised, re-individuated, reinforcing the capability 'to think for oneself' (ibid: 11/12). Here Stiegler paraphrases Kant's famous essay *Beantwortung der Frage: Was ist Aufklärung*? in which Kant calls on his reader to 'dare to think for themselves': *sapere aude!* Kant writes: 'have the

courage to use your own mind' (Kant, 1784: 481). This is what Stiegler is up to: we must preserve the particular dimensions of the bookish mind capable of arriving at such a thought. Instead of taking for granted that such thinking is the achievement of a disembodied transcendental ego (as Kant himself did propose), we need to investigate the grammatisation that is a condition for this type of thinking. This means, above all, that any tertiary retention that becomes entirely self-referential will be dead, and remain so 'if it does not trans-form, through a reverse effect, the secondary retentions of the psychical individual affected by this tertiary retention' (Stiegler 2012: 10/12). Today, neuroscience is capable of experimentally testing the 'constitution of the mind through the introjection of tertiary retentions' (ibid: 10/12), tracing the implications of reading and writing for the morphology and the behaviours of our brains (Wolf, 2008). This way we can localise the correlates of capabilities such as reflection, consideration, deliberation and intentional action in what Wolf has called 'the reading brain'. If pre-emptive computational layers shortcut the introjection of tertiary retention, the point is reached that these layers are not merely our digital butlers but that we become their cognitive resource, part of their extended mind. Stiegler therefor concludes that the digital is like a *pharmakon*:[22] depending on its usage it may reinvent or destroy us. To prevent the digital brain from being short-circuited by automata, we must make sure that individual users of the Internet have the capability to 'think for themselves' and know how to get their finger behind attempts to bypass the neo-cortex. Thus, a new Enlightenment, a new transparency must be engineered. Words are not enough here.

A new Rule of Law: profile transparency by design

The computational turn is to be seen as a *pharmakon*. This will allow us to carefully distinguish between its empowering and destructive affordances. To the extent that the plethora of dividuals, data derivatives and other computational models create room for new actualisations, they are in the domain of virtualisation. But to the extent that the mass of inferences, profiles and automated decision-systems pre-empt our intention they reduce to de-realizations that stifle innovation and present us with nothing else than a sophisticated recalculation of past inclinations. To prevent the last and to preserve the first may require hard work and nothing can be taken for granted.

In this section I will investigate what is required to reinstate the double contingency that constitutes self, mind and society – enabling us to guess how our computational environment anticipates our states and behaviours. This can be framed as a transparency requirement, but – taking note of the previous section – we must be cautious not to reduce transparency to a simple information symmetry. Such a symmetry will easily cause a buffer overflow:[23] the amount of information it would involve will flood our bounded rationality and this itself will enable manipulation by what escapes our attention. Though some authors may applaud the Enlightenment of Descartes' *idées claires et distinctes*, others may point out that this generates overexposure, wrongly suggesting the possibility of light without shadows. The metaphor of the buffer overflow actually suggests that we may require selective enlightenment, and are in dire need of shadows. The more interesting question, therefor, will be what should be in the limelight and where we need darkness. In renaissance painting the techniques of the *claire-obscure*, the *chiaroscuro*, the *Helldunkel* were invented and applied to suggest depth, and to illuminate what was meant to stand out. By playing with light and shadow the painting could draw the

attention of the onlooker, creating the peculiar experience of being drawn into the painting – as if one is standing in the dark, attracted by the light.

The computational turn invites us to reinvent something like a claire-obscure, a measure of transparency that enables us to foresee what we are 'in for'. This should enable us to contest how we are being clustered, correlated, framed and read, thus providing the prerequisites for due process. Therefor the question is how we can present the plurality of matching dividuals to the bounded rationality that constitutes our individuality. This should enable us to play around with our digital shadows, acquiring the level of fluency that we have learned to achieve in language and writing.

There are two ways of achieving such transparency. Neither can do without the other. The first involves intuitive interfaces that develop the *chiaroscuro* we need to frame the complexity that frames us (reinstating a new type of double contingency). In terms of computer engineering this involves the front-end of the system. To make sure that the front-end does not obscure what requires our attention we need a second way of achieving transparency. This involves the possibility to check, test and contest the grammatisation that defines the outcome of computational decision-systems. It concerns the back-end of the system and this may not fare well with current day business models, which thrive on trade-secrets and intellectual property rights on data mining algorithms. This transparency, however, is our only option to regain the introjection of tertiary retention, i.e. to re-individuate the sets of proliferating dividuals that are used to target, trade with, or to circumvent our attention. To make sure that the information that derives from testing computational mediations does not cause a buffer-overflow with our bounded rationality, we must design a front-end that manages our clair-obscure. And somehow, we will have to become grammatised in the language of computational retention, to play our bit on the nexus of the front-end and the back-end. If we don't, we will lose the precious capability for which Kant called on us: to think for ourselves.

To achieve transparency about the back-end of computational systems that profile us I return to the five recommendations provided by Sculley and Pasanek, as promised above. I adept their recommendations, that were constructed for the domain of the digital humanities, to better fit the broader scope of marketing, law enforcement and the whole plethora of automated decision-making systems that co-define our lifeworld. For the original recommendations see Sculley and Pasanek (2008).

First, a collaborative effort is required between the engineers, designers and users of the relevant computing systems and those whose capabilities will be affected (for instance consumer organisations, citizens juries, NGOs). All stakeholders should make the effort of clarifying their assumptions about the scope, function and meaning of systems, as they are developed. After all, the construction of these systems concerns the architecture of the polis, it requires as much political participation as any other interaction with significant impact on third parties (Dewey, 1927; Hildebrandt, Gutwirth, 2007; Marres, 2005).

Second, those using the systems for data mining operations should employ multiple representations and methodologies, thus providing for a plurality of mining strategies that will most probably result in destabilizing any monopoly on the interpretation of what these systems actually do. This will clear the ground for contestation, if needed. Without an evidence-based awareness of the alternative outcomes generated by alternative machine learning techniques, data mining may easily result in holding people hostage to inferences drawn from their past behaviours.

This would amount to de-realisation instead of virtualisation. We can think of software verification, sousveillance or counterprofiling as means to prevent this. Searls' (2012) notion of vendor relationship management may be of help here, turning the tables on the mantra of customer relations management.

Third, all trials should be reported, instead of 'cherry-picking' those results that confirm the experimenters' bias. At some point failed experiments can reveal more than supposedly successful ones. This is particularly important in a setting that generates automated decisions that impact the capabilities of groups as well as individuals. Especially with regard to prohibited or undesirable discrimination this seems important (Pedreshi et al., 2008). The imposition of documented auditability obligations on data controllers under the proposed EU General Data Protection Regulation (GDPR) confirm the import of situating the experimenter's bias. As indicated above, such bias is inevitable but this does not imply that any bias will do.

Fourth, whenever such impact is to be expected the public interest requires transparency about the data and the methods used, to make the data mining operations verifiable by joint ventures of e.g. lawyers and software engineers. This connects to the fifth recommendation, regarding the peer review of the methodologies used. The historical artifact of constitutional democracy, nourishes on a detailed and agonistic scrutiny of the results of data mining operations that can sustain the fragile negotiations of the rule of law. Without such a reappropriation or introjection of the tertiary retention of computational grammatisation, civil society will cease to exist.

Having started with a brief exploration of transparency of the back-end, we now turn back to the front-end. The recommendations summed up above focus on transparency about assumptions inscribed into the system, methods used and results obtained. Combinations of requirements engineering, software verification and impact assessments regarding potential violations of fundamental rights should do at least part of the job here. Just like in the case of the front-end total transparency is neither possible nor desirable. The challenge will be how to monitor compliance with, for instance, data protection legislation without posing new privacy risks or how to prevent compliance models that create an illusion of compliance instead of the substance. This brings us to the front-end. How to feed the results of critical and constructive discussions on the back-end into the front-end; how to provide consumers and citizens with the kind of anticipations that nourish their capability to play with the system; how to engage the industry in a way that empowers it to invest in the kind of interfaces that take customers serious as players instead of merely as cognitive resources for data mining operations? Do we need more icons instead of text, must we design intelligent agents that are programmed on our behalf, must we help customers to stop 'using' technology and start 'interacting' with it? Who is we? These are the hard questions, requiring the hard work.

The proposed GDPR introduces a right to profile-transparency,[24] whenever automated decisions have a significant impact on the life of an individual, or legal effect. This right comprises the right to know about the existence of such a decision and the right to know the envisaged effects of the decision. Combined with the obligation to implement Data Protection by Design,[25] which requires those in charge of computational systems to implement appropriate technical and organisational measures and procedures to ensure that the processing will meet the requirements of profile transparency. The appropriateness is related to the technical state of the art and the economic feasibility. This seems a balanced and realistic challenge to engage in

the development of both back-end and front-end transparency tools. An optimistic note to end this chapter. I return, nevertheless to the enigma of the Sphinx on the cover of this book. Oedipus is depicted in the *clearing* of the *clair-obscure*. He stands out strong, wilful and looks somewhat impatient. The Sphinx stands in the shadow of a cave, potentially irritated that a trespasser has finally solved her riddle. However, Oedipus may have solved the riddle, he cannot evade the fundamental fragility it foretells. Transparency tools can invent a new version of the double contingency that constitutes our world – they cannot resolve the fundamental uncertainty it sustains. On the contrary, these tools should help to reinstate this uncertainty, rather than the over-determination that the computational turn could otherwise enable.

Notes

[1] Check: http://www.robcottingham.ca/cartoon/archive/2009-02-21-recommender/.
[2] Note that in the EU context consent cannot overrule the purpose limitation principle. Consent is a ground for legitimate processing (art. 7 D 95/46/EC), but such processing will have to comply with the norms for fair processing, of which purpose limitation is one (art. 6 D 95/46/EC). In the draft Data Protection Regulation it seems that data subjects can waive the right to purpose limitation with regard to 'further processing' (secondary use), cf. art. 6(4) of the Draft Regulation as presented on 25th January 2012. Note that privacy policies or service licence agreements often entail vaguely formulated indications of the purpose for which data may be used, which seems equivalent to obtaining consent to overrule purpose limitation.
[3] Data Protection legislation is built on the concept of personal data, i.e. data that relates to an identified or identifiable person (e.g. art. 2 D 95/46/EC). To the extent that data is successfully anonymised the legislation does not apply.
[4] I use the term 'machine' here in the broad sense of an artificial contraption that is used as a tool to achieve certain goals by means of leverage (the lever), transformation of energy (steam engine) or automation (the computer). In reference to machine learning as a branch of AI, I will include software programs under the heading of machine, but I will also assume that software must be articulated into matter to actually function as a machine.
[5] In computer science an inference engine is a computer program that derives answers from a knowledge base; it is based on pattern recogntion and can be qualified as data-driven because the rules that are applied to infer answers depend on the connections between the data.
[6] On the genealogy of Deleuze's quest for the virtual see (Smith, Protevi, 2011).
[7] A virtual machine achieves hardware virtualization by means of a software implementation of a machine, executing programs like a physical machine; it allows different operating systems to run entirely seperately on the same hardware, simulating different machines on the same platform.
[8] On apps for diabetes or heart disease see (Brustein, 2012). Though such function creep, especially in the case of health data, is prohibited by law we should not be surprised if such inferences will legitimated via explicit unambiguous consent packaged with insurance contracts that offer benefits for those who allow closer monitoring of their health-care related behaviours.
[9] On the impact of the ICT infrastructure of the script and the printing press on the brain and the mind, see (Wolf, 2008).
[10] If it is true that rational decision-making depends on the emotional fitness that allows us to make choices and to act with intention (e.g. (Damasio, 2000), then computational systems may not be capable of rational decision-making – unless guided by human intention. I leave aside the discussion of whether synthetic emotions will resolve this problem, but see (Velasquez, 1998).
[11] See (Hildebrandt, 2011b):
[12] See http://www.no-free-lunch.org/ for an overview of 'no free lunch theorems'. Cp. Giraud-Carrier and Provost 2005.
[13] Though the capability approaches of (Sen, 1999) and (Nussbaum, 2011) do not directly connect with notions like 'data protection by design', it may be important to elaborate this connection. In both cases human rights protection may impose imperfect duties on states and other actors to provide effective means of empowerment, without engaging in paternalism.

[14] Zizek actually refers to the French philosopher Lacan, whose theory is not equivalent with those of Luhmann and Parsons, who developed the notion of double contingency. Nevertheless, Zizek's phrase aptly describes the experience of the double contingency that is at stake in this chapter.
[15] In computer science a primitive is a basic building block, 'with which to model a domain of knowledge or discourse' (Gruber 2009: 1963)(Gruber, 2009).
[16] *Pace* Luhmann, a social system will reproduce the contingency at the level of the system, because it must interact with other systems that can reject, misunderstand, contest or renegotiate whatever a system does or communicates. Cf. (Vanderstraeten, 2007). It is important to note that the founding fathers of the concept of autopoiesis – on which Luhmann built his systems theory – reject the idea that social systems achieve the kind of autonomy that is characteristic of individual human beings. See e.g. (Maturana, Varela, 1998): 198.
[17] The term 'digital unconscious' was coined by Derick de Kerchove, Director of the McLuhan Program in Culture and Technology from 1983 until 2008. See http://www.mcluhanstudies.com/index.php?option=com_content&view=article&id=485:from-freud-to-digital-unconsciuos&catid=78&Itemid=472, last assessed 30 October 2012. On subliminal influences see (Hildebrandt, 2011a)(Hildebrandt, 2011c).
[18] This is Andy Clark's (Clark, 2003) idea of the extended mind 'inside-out'. Instead of machines being part of our mind, we become part of theirs. My aim is make sure that we are never merely an instrument for the information-driven cognition of these computing systems (tongue in cheek one could say I am requiring them to respect the Kantian moral imperative).
[19] See http://lists.w3.org/Archives/Public/www-tag/2003Jul/0158.html.
[20] Cognitive psychology challenges common sense intuitions about the accuracy of our memory, see e.g. (Stark u. a., 2010) on the complex differences between activation of true and false memories.
[21] A reference to Plato's critique of writing in the Phaedrus (Plato, 2012), which is all about the effects of tertiary retention on the capability for secundary retention.
[22] In the Phaedrus King Thamus offers 'writing' as a *pharmakon* (medicin) that can extend one's memory. The King refuses, suggesting that writing will generate forgetfulness, being a poison instead of a medicin. The notion of the pharmakon that can be medicin or poison, has been elaborated within French philosophy, e.g by (Derrida, 1983), (Stengers, 2010), and Stiegler (ibid).
[23] In digital security a buffer overflow is one of the most basic and prevailing vulnerabilities of computing systems. See e.g. (Leeuw, Bergstra, 2007): 639.
[24] Art. 20.4 of the GDPR, available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf, last accessed 30 October 2012.
[25] Art. 23.1 of the GDPR.

References

Amoore, L. (2011) 'Data Derivatives On the Emergence of a Security Risk Calculus for Our Times', in *Theory, Culture & Society*, 28 (6): 24–43.
Anderson, C. (2008). 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', in: *Wired Magazine*, 16 (7).
Andrejevic, M. (2002) 'The work of being watched: interactive media and the exploitation of self-disclosure', in *Critical Studies in Media Communication*, 19 (2): 230–248.
Brustein, J. (2012) 'Coming Next: Doctors Prescribing Apps to Patients, *The New York Times*, 19 August 2012.
Carroll, L. (2000) *Alice Through The Looking Glass*, Creation Books.
Clark, A. (2003) *Natural-Born Cyborgs. Minds, Technologies, and the Future of Human Intelligence*, Oxford: Oxford University Press.
Damasio, A.R. (2000) *The feeling of what happens: body and emotion in the making of consciousness*, New York: Harcourt.
Deleuze, G. (1994) *Difference and repetition*, New York: Columbia University Press.
Deleuze, G. (1992) 'Postscript on the societies of control', in: *October,* 59.
Derrida, J. (1983) *Dissemination*, Chicago: University Of Chicago Press.
Dewey, J. (1927) *The public & its problems*, Chicago: The Swallow Press.

Elmer, G. (2003) *Profiling Machines: Mapping the Personal Information Economy*, The MIT Press.
Esposito, E. (2011) *The Future of Futures: The Time of Money in Financing and Society*, Cheltenham: Edward Elgar Publishing.
Geisler, D.M. (1985) 'Modern Interpretation Theory and Competitive Forensics: Understanding Hermeneutic Text', in *The National Forensic Journal*, III (Spring): 71–79.
Gigerenzer, G. (1991) 'How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases"', in W. Stroebe, M. Hewstone, (eds.) *European Review of Social Psychology*, Chichester: Wiley.
Gruber, T., Liu, L., Tamer, M. (eds.) (2009) 'Ontology', *Encyclopedia of Database Systems*, Berlin Heidelberg: Springer.
Hildebrandt, M. (2011a) 'Autonomic and autonomous "thinking": preconditions for criminal accountability', in M. Hildebrandt, A. Rouvroy (eds.) *Law, Human Agency and Autonomic Computing. The Philosophy of Law meets the Philosophy of Technology*, Abingdon: Routledge.
Hildebrandt, M. (2011b) 'The Meaning and Mining of Legal Texts', in D.M. Berry (ed.) *Understanding Digital Humanities: The Computational Turn and New Technology*, London: Palgrave Macmillan.
Hildebrandt, M. (2009) 'Who is profiling who? Invisible visibility', in: S. Gutwirth, Y. Poullet, P. De Hert (eds.) *Reinventing Data Protection?,* Dordrecht: Springer: 239–252.
Hildebrandt, M, Koops, B.J., De Vries, E. (2009) *Where idem-identity meets ipse-identity. Conceptual explorations*, Brussels: Future of Identity in the Information Society (FIDIS), available via http://www.fidis.net/resources/deliverables/profiling/#c2367, last accessed 30th October 2012.
Hildebrandt, M., Gutwirth, S. (2007) '(Re)presentation, pTA citizens' juries and the jury trial', *Utrecht Law Review*, 3 (1), available at http://www.utrechtlawreview.org/index.php/ulr/issue/view/5, last accessed 30 October 2012.
Hildebrandt, M. (2011c) 'Legal Protection by Design: Objections and Refutations', in: *Legisprudence*, 5 (2): 223–248.
Kallinikos, J. (2006) *The Consequences of Information. Institutional Implications of Technological Change*, Cheltenham, UK: Edward Elgar.
Kant, I. (1784) 'Beantwortung der Frage: Was ist Aufklarung?', in *Berlinische Monatsschrift*. Dezember-Heft: 481–494.
Kranzberg, M. (1986) 'Technology and History: "Kranzberg"s Laws', *Technology and Culture*, 27: 544–560.
Leeuw, K.M.M. de, Bergstra, J. (eds.) (2007) *The History of Information Security: A Comprehensive Handbook*, Elsevier Science.
Lévy, P. (1998) *Becoming Virtual. Reality in the Digital Age*, New York and London: Plenum Trade.
Lévy, P. (2005) 'Sur les chemins du virtuel', accessible at http://hypermedia.univ-paris8.fr/pierre/virtuel/virt0.htm, last accessed 30th October 2012.
Lyotard, J.-F. (1984) *The Postmodern Condition: A Report on Knowledge*, Manchester: Manchester University Press.
Marres, N. (2005) *No Issue, No Public. Democratic Deficits after the Displacement of Politics*, Amsterdam: published by the author, available via: http://dare.uva.nl.
Marx, K., Engels, F. (1998) *The German Ideology, including Theses on Feuerbach*, Prometheus Books.
Massiello, B., Whitten, A. (2010) 'Engineering Privacy in a Age of Information Abundance', in *Intelligent Information Privacy Management*, AAAI: 119–124.
Maturana, H.R.; Varela, F.J. (1998): *The Tree of Knowledge. The Biological Roots of Human Understanding*, Boston & London: Shambhala.
McStay, A. (2011) *The mood of information : a critique of online behavioural,* New York: Continuum.

Mead, G.H., Morris, C.W. (1962) *Mind, self, and society from the standpoint of a social behaviourist*, Chicago: University of Chicago Press.
Negroponte, N. (1996) *Being digital*, New York: Vintage Books.
Nussbaum, M.C. (2011) *Creating Capabilities: The Human Development Approach*, Belknap Press of Harvard University Press.
Parsons, T. (1991) *The Social System* (2nd ed.) Routledge.
Parsons, T., Shils, E. (1951) *Toward a General Theory of Action*, Cambridge, MA: Harvard University Press.
Pedreshi, D., Ruggieri, S., Turini, F. (2008) 'Discrimination-aware data mining', in: ACM Press: 560-
Pinedo, D. (2012) 'Niet alleen Tena belt incontinente patiënten', *NRC Next*, 10 August 2012.
Plato (2012) *Phaedrus*. o.V.
Ricoeur, P. (1992) *Oneself as Another*, Chicago: The University of Chicago Press.
Ricoeur, P. (1973) 'The Model of the Text: Meaningful Action Considered as a Text', *New Literary History*, 5 (1): 91–117.
Rouvroy, A. (2011) 'Technology, Virtuality and Utopia: Governmentality in an Age of Autonomic Computing', in M. Hildebrandt, A. Rouvroy (eds.) *Law, Human Agency and Autonomic Computing. The Philosophy of Law Meets the Philosophy of Technology,* Abingdon: Routledge.
Russell, S., Norvig, P. (2009), *Artificial Intelligence: A Modern Approach*, Prentice Hall.
Sculley, D, Pasanek, B.M (2008) 'Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities', *Literary and Linguistic Computing*, 23 (4): 409–424.
Searls, D. (2012) *The Intention Economy: When Customers Take Charge*, Harvard Business Review Press.
Sen, A. (1999) *Commodities and Capabilities*, Oxford University Press.
Smith, D., Protevi, J. (2011) 'Gilles Deleuze', in: E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2011.
Stark, C.E.L., Okado, Y., Loftus, E.F. (2010) 'Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms', *Learning & Memory*, 17 (10): 485–488.
Stengers, I. (2010) *Cosmopolitics I*, Univ Of Minnesota Press.
Stiegler, B. (2012) 'Die Aufklaerung in the Age of Philosophical Engineering', Lyon 20th April 2012.
Swoyer, C., Orilia, F, (2011) 'Properties', in: E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, Winter 2011.
Vanderstraeten, R. (2007) 'Parsons, Luhmann and the Theorem of Double Contingency', *Journal of Classical Sociology*, 2 (1): 77–92.
Velasquez, J.D. (1998) 'Modeling Emotion-Based Decision Making', in *Emotional and Intelligent: The Tangled Knot of Cognition*: 164–169.
Weiser, M. (1991) 'The computer for the 21st century', *Scientific American*, 265 (3): 94–104.
Wikipedia contributors (2012) 'Attribute', *Wikipedia, the free encyclopedia*, Wikimedia Foundation, Inc., last accessed 30th October 2012.
Wolf, M. (2008) *Proust and the Squid: The Story and Science of the Reading Brain*, Icon Books Ltd.
Zizek, S. (1991) *Looking awry: an introduction to Jacques Lacan through popular culture*, Cambridge, MA: MIT Press.