2008

Profiles and correlatable humans

Mireille Hildebrandt, *Radboud University Nijmegen*

Profiles and correlatable humans

Mireille Hildebrandt[i]

1 Introduction: knowledge and information societies

This contribution originated from a presentation at the March 2005 international conference *Can knowledge be made just?* The title of this volume, *Who owns knowledge?*, suggests that ownership of knowledge may be an interesting way to achieve just knowledge (or a just distribution of knowledge?). In the field of profiling technologies Lawrence Lessig has argued that the creation of ownership of personal data could initiate a new technological framework that will empower individual citizens to regain control over their personal information (Lessig, 1999b:518-521; idem 1999a: 159-162). Others, however, have expressed serious doubts regarding this so-called commodification of information. For instance (Prins, 2004:7) argues that attributing property rights to individuals regarding their personal data misses the point, precisely because it is not their data in themselves that are of interest, but the knowledge constructed on the basis of these data, and the impact of this type of knowledge on ' position, social ordering, roles, individual status and freedom'. In the following I will develop a similar line of thought, leaving the 'ownership' of knowledge to other authors, and I will plead a different approach to the nexus of knowledge and information in our knowledge society.

The introduction to the conference that initiated this volume referred to the fact that we live in a 'knowledge society'. This concept – in counterpoint to the concept of an information society – raises the question whether there is a crucial difference between an information society and a knowledge society, for instance in the sense that they describe the same world from a different perspective. If this is the case, one may also ask whether and how the two are related. Based on the findings of the life sciences and information theory I will presume that our knowledge depends on information, while at the same time the question what counts as information depends on our knowledge (this is a circle, but not a vicious one). In that light it becomes interesting to describe nonhumans and their relationships in terms of knowledge and information (a gene *knows* how to produce a specific type of protein, depending on the input of specific information (epigenetics); a bird *knows* how to whistle to seduce a possible mate, depending on the input of specific information, like for instance a change in temperature or sunlight) (Van Brakel, 1999: 7/15). For centuries we have been used to think of both information and knowledge as produced by reflection or at least consciousness. *As if we know only what we know to know.* As if information only counts as information if we are aware of it as such (criticized by Mead, 1959/1934; Merleau-Ponty, 1945; Polanyi, 1966; Ryle, 1949). If we allow ourselves to recognize knowledge and information in the world of nonhuman organisms, our perception of an information or knowledge society may change and this change could shed some light on the advanced identification technologies that will soon impact our sense of self and our understanding of the validity, relevance and legitimation of knowledge. In what follows I will focus on profiling as the most advanced and comprehensive technology of identification. Profiling is understood here as the *automated* generation of a multiplicity of evershifting profiles that are preconditional for Ambient Intelligence and advanced risk-assessment. As such, profiling constitutes a very interesting nexus of information, knowledge and risk-assessment.

Hereunder I will discuss some implications of profiling practices for identity, legal subjectivity, democracy and rule of law. To this end I will attempt to answer (or at least raise) the question if and how the information and/or knowledge generated by these technologies can be made justiciable. First I will explain what is meant by the term profiling practices (section 2). Second I will discuss the purposes and some of the effects of the type of knowledge these practices produce (section 3). Third I will indicate in what ways profiling technologies produce a new type of

knowledge (section 4). Fourth I will explore the way the law tries to deal with this type of knowledge, raising some questions about the effectiveness of the law as it tries to fit this kind of knowledge into the present legal framework (section 5). Fifth I will discuss the crucial difference between a profile as a correlated data subject and the person of flesh and bones as a correlatable human (section 6). Sixth I will explore the possibilities for the law to make profiling practices justiciable. I will claim that for data protection legislation to take effect – that is, for profiles to become justiciable - both specific technological design and an effective fair trial are preconditional. Finally I will indicate how the law of a constitutional democracy in fact both presumes and – for this reason - should create a position from where *humans as linkable data subjects with a sense of self* can make knowledge-constructs like profiles justiciable (section 7).

2 Profiling practices

Profiling practices are a way to generate knowledge from data (Custers, 2004: 17-20). This knowledge consists of patterns or correlations between data(sets). To give a crude example: if we have a data base with data on the color of your eyes and a series of transactions at your local grocery, a correlation may be found between the color of your eyes and your preference for certain products (eye shadow would be predictable, but other, more unexpected correlations might turn up). In marketing this way of generating correlations (exploratory research) has been used since the '70-ties (called regression analysis). The tools for discovering such correlations have improved exponentially, while the transaction costs for searches in huge databases has decreased enormously. The whole process of profiling is often described in 4 steps: data collection, data preparation, data mining and interpretation (KDD-model: knowledge discovery in data bases) (Custers, 2004): 17-20). These are not obvious steps, as they all require a certain infrastructure and the restructuring of information in a way that fits a database. As Lyotard remarked in 1984:

> 'Along with the hegemony of computers comes a certain logic, and therefore a certain set of prescriptions determining which statements are accepted as 'knowledge' statements' (quoted by (Van Brakel, 1999:3/15).

Data collection requires that certain types of behavior (buying things, visiting places, surfing on internet, speaking with colleagues, friends, clients, suppliers) are not only observed but also recorded and stored. This is a major difference with previous observation of behavior in public space as this was not often recorded and stored (Lessig, 1999a: 143 and 150-151). For instance, if you buy bread around the corner and pay with cash, this will be observed by others, but your 'transaction' will not be recorded, stored and linked with other data that were recorded, stored and aggregated; money is still mostly anonymous today, even if this may be otherwise once banknotes are equipped with RFID-tags. Second, these data have to be aggregated into databases. To profile clients of a bank, all transaction data that come into different departments of the bank, have to be stored as transactions of the same client, because, to profile customers as high or low spenders it is necessary to locate as many different transactions possible as transactions of this one person. This is why integration of different databases becomes interesting (and profitable), leading to a market for data and databases. Third, data mining - the process of generating correlations or checking on the outliers of established correlations - requires creativity (to select data to be correlated and the algorithms that may produce correlations) and professional expertise (to detect spurious correlations that do not have any independent explanatory power).

The fourth step is the central issue of profiling: the emergence and interpretation of correlations. The point is that a correlation does not mean anything until it is interpreted: it does not necessarily indicate reasons nor causes. Of course it can be used as a hypothesis, claiming a causal link. The interesting thing about profiling is, however, that it does not start from a hypothesis that is than rigorously tested/falsified/verified, but that it generates correlations *without necessarily even being interested in causality or human reason*. For instance, it is possible to observe online behavior like the relative speed with which certain keys are touched on the keyboard. It is also possible to profile this behavior to such an extent that a person can be recognized as the same person on the basis of her typing behavior. Those who use this kind of identification technology are not interested in the causes of your typing behavior, nor in your reason to wait between the *a* and the *i*. They are interested in identifying you as the same person over a period of time (personalizing), and perhaps, linking the information with that of others,

being able to predict certain traits or behavior as correlating with your typing behavior (group profiling).

Profiling thus does not necessarily build on the traditional methodologies of for instance the social or natural sciences (Custers, 2004: 56-58) that aim at explanations in terms of causes or understanding in terms of meaning. Interpreting the correlations is not usually done to construct theoretical knowledge about society or persons (e.g. in sociology or psychology), but to decide on the next step to be taken. Data mining technologies are instrumental for decision-making processes. For instance in the case of marketing the question will be: which type of persons will be identified as potential customers, and to which approach will they yield; in the case of crime control: which type of persons will be targeted as potential suspects to further investigate; and, in the case of anti-terrorism or immigration: which type of persons will be identified as potential terrorists or illegal immigrants. Actually correlations seem to create new meaning. Insignificant personal data may thus turn out to be highly significant. Also they may correlate with sensitive personal information that is protected by privacy legislation. To avoid the constraints of data protection one could use seemingly insignificant data that correlates with sensitive personal data. This is called masking, (Custers, 2004: 57). Traditional protections may not work here, and even fighting traditional conceptions of science may be an irrelevant exercise. Profiling is about generating a new type of knowledge, consisting of patterns in data that are interesting and certain enough for a user; it is not necessarily about science. And it could very well be that the impact of these technologies and practices is more fuzzy, more precise, more implicit and more worrying that the impact of scientific knowledge.

3 Purpose and effects of profiling practices

Western societies are thought of as information and knowledge societies, but also as risk societies. Not because other types of societies are not prone to risk, but because while manipulating some of the risks of 'nature', we seem to fabricate others (Beck, 1992, Shklar, 1990). Looking into the relationship between risk, information and knowledge at the level of nonhumans again, it should be clear that information can indicate certain risks if we have the knowledge to read it like that. At the same time it is obvious that the use of explicit knowledge and information can itself be a risk, as our modern risk society demonstrates. Profiling technologies, which are focused on processing information to produce knowledge, may also create specific types of risk. To detect such risks we will investigate the purposes and effects of profiling practices (Custers, 2004: 74-78).

The purpose of profiling is *selection*, implying the wish to include certain objects or persons and to exclude others. This in itself is everyday business, life depends on it. Organisms select certain traits to fit in their environment and sometimes to fit their environment to their needs. While neo-Darwinists understand the survival of the fittest as justification for power play, one could say that Darwin demonstrated that evolution in the end rewards those organisms that fabricate the best fit with their environment (Sonigo & Stengers, 2003). In that sense evolution is a constant process of profiling: taking in and processing information to decide the next step (profiling as risk-assessment and screening for opportunities).

Moving to the production of scientific knowledge we can look at genetic profiling, not *by* genes, but *of* genes as practiced *by* molecular biologists and epidemiologists that aim to select genes that correlate with disease. Having detected such correlations they may hope in the end to find therapies to prevent or cure the correlated disease and/or to be able to target certain medication or surgery to the type of patient that will most likely respond. So far, however, the causal chains between genotypes and phenotypes have by far not been discovered – due to the marvelous complexities of the gene and its context. More importantly, in human society, selection is not only effective or ineffective but also legitimate or illegitimate; it can for instance exclude people from equal opportunities on grounds that we find unjust. As Lawrence Lessig describes, social hierarchies require information to discriminate between different social ranks. With increased mobility the costs of acquiring such information rose, because of the difficulties in tracing people. Thus many old hierarchies broke down. Profiling changes this, empowering profile users to re-establish inequalities (Lessig, 1999a:155). Selection, like technology, is neither good nor bad, but it is never neutral (Kranzberg, 1986). It impacts the lives of those that are selected and thus calls for justification.

Another way to look at the purpose of profiling is to describe it as *prototyping*: enabling a data controller (business enterprise, police, immigration policy makers, doctor) to make decisions on the basis of a prototype, a knowledge-construct that filters our perceptions and expectations. Prototyping can be described as a psychological process (Canhoto & Backhouse, 2004), but also, at an epistemological level, as a *Vorurteil* in the sense of Gadamer: without some form of prototyping we would be flooded by meaningless information. From this perspective Schauer's (2003) defense of reasonable generalizations (profiles, probabilities and stereotypes) against postmodern 'context is all' particularisms seems an innocent and rather common sense reference to the unavoidable fact that we need some kind of profiling to survive the mass of detailed information that would otherwise paralyze both our personal life and organizational decision making processes. Two problems must be faced however. The first is that prototyping is close to stigmatization (Goffman, 1963;Hudson, 2005a), and it is not so easy to draw the line here. This means that even if it were effective to generalize at a certain point, it may still be illegitimate. For instance, it may be the case that profiling of persons as possible terrorists leads to unacceptable legal consequence, violating the presumption of innocence and the right to defend oneself against such legal consequence. The second point is that Schauer may simply be wrong in opposing a particularistic argument to the logic of generalization. Multivalent – fuzzy - logic demonstrates that categorization (generalization) itself can be made more precise and deliver better results when combined with contextual reference and used in a dynamic way (meaning that one has to be ready to readjust one's categories at any point in time). In fact, *automated* profiling reaches a level of sophistication that turns it into a very precise specification or personalization – even if still on the basis of statistically sound generalizations (group profiling). This may reduce wrongful categorization, but if that is the case the accuracy of the intimate knowledge it produces may reinforce stigmatization back to the level of social control in a village community.

A third way of looking at the purpose of profiling is warning persons of the risk they run and thus *confronting* them with knowledge about themselves they had no access to (genetically determined disease, for instance). This confrontation will enable them to take measures, but it may also impact their sense of self in an existential way. Profiling may offer them choices they would have lacked otherwise, but profiling may also reveal secrets of the self that force a person to reconstruct her identity (Hudson, 2005a).

A fourth function profiling may take on is *customization*. As a result of *targeted advertising* a person may be confronted only with those advertisements that should interest him, considering his past behavior. If we move on to the advance of targeted servicing and Ambient Intelligence this perspective becomes pervasive. In the case of Ambient Intelligence, the combination of ubiquitous computing and intelligent devices enables your environment to respond to your wishes before you become aware of them and to restructure itself in tune with the anticipations you carry 'under the skin'. This sounds like heaven – and like hell. It can reduce the feedback you get from your environment, as you begin to live in a world of your own making. The diversity of unexpected and unwanted confrontations with others, whether human or nonhuman, could be diverted by your intelligent agent that 'knows' - on the bases of your past behavior - what you would probably want and expect. We should add that the line between customization and manipulation is a thin one. To quote Lawrence Lessig (1999a:154):

> 'When the system seems to know what you want better and earlier than you do, how can you know where these desires really come from. (…) profiles will begin to normalize the population from which the norm is drawn. The observing will affect the observed'.

This brings us to the last paradoxical effect of profiling. While profiling seems to **individualize** (customize) your environment, it may in fact **de-individualize** your way of life. Both group profiling and personalization judge your needs, expectations and desires on the basis of past behavior, building a well-fitted and unusually comfortable cage from which escape will be nearly impossible, precisely when profiling becomes ever more ubiquitous and intelligent.

4 What is new about profiling?

As indicated above, profiling is as old as life on earth. With Varela one could identify living organisms by their capacity for self-organization, which enables them to profile their environment in order to match or fit the environment (co-producing themselves and their environment) (Maturana & Varela, 1991 and Sonigo & Stengers, 2003). Speaking of knowledge and information at the generic level of humans and non-humans alike makes it possible to subsequently detect what is special about human knowledge and information. Reserving the terms knowledge and information for human cognition may leave us ignorant of cognition in the world of nonhuman organisms. The automated profiling technologies that have been developed with the advance of computer technologies are different from the more traditional scientific knowledge construction we are familiar with. Even though profiling has been going on since the beginning of life on earth, the externalization and objectification *of* and conscious reflection *on* the knowledge it produces has been a recent phenomenon in our evolution, typical for human society that uses language to articulate its dealings within an environment. In terms of Helmuth Plessner human beings suffer and enjoy a sense of externality or bi-aspectivity that is not present in other conscious beings (De Mul, 2003): we are able to look back at ourselves and reflect on alternative choices of action. The capacity to reflect on one's knowledge and information – to be both conscious and *self*-conscious – seems crucial for human beings, and in a way, profiling can be understood as a new type of implicit knowledge that hits us 'under the skin' as it is not based on causes or reasons but on inference of correlated data that allow anticipation of future behavior. It compares well to the implicit types of knowledge and the implicit information exchange typical for all organisms that find themselves in an environment they have to cope with (anticipating risk and opportunity). In other words, profiling as knowledge construction is new in that it seems to revert back – in a very sophisticated way - to very old and very successful mechanisms of being in the world.

Before proceeding I will now briefly summarize what is new about profiling, as compared to traditional scientific knowledge construction. First, the scope of the data that can be recorded, aggregated and researched at a reasonably low cost is enormous; second, the low transaction costs have as a consequence that profiling practices often do not involve extrapolation of samples to populations, but pretend searching an entire field (Custers, 2004:56-58); third, profiling is typically ubiquitous and unobtrusive, which implies that the invasive character of profiling seems absent (Lessig, 1999a: 144); fourth, the low cost of searching an entire data base leads to exploratory research, generating correlations instead of starting with a theory and then deducting hypotheses that can be falsified/verified (Custers, 2004; Scott Armstrong, 1970); fifth, profiles can impact our lives in a number of ways without us ever being aware of the fact that we were included in or excluded from certain opportunities or risks on the basis of a profile; sixth, profilers may 'know' things about us we don't know about ourselves (Hudson, 2005a; Rose, 2003:86-87).

5 Data Protection and informational privacy

5.1 Good practice guidelines, technological design and legislation

If the knowledge produced by profiling practices entails exclusion, stigmatization, confrontation, customization and even de-individualization, the question is how to constrain these practices in order to make the knowledge they produce just. The traditional means to constrain application of new technologies is to provide good practice guidelines for industry and/or legislate on the matter. In the last decades alternative instruments have been elaborated, of which technological design is perhaps the most interesting. In the field of data mining so-called PET's (privacy enhancing technologies) have been developed, that combine a measure of linkability (a necessary precondition for profiling) with anonymity and/or pseudonymity. To fine-tune one's level of privacy and security (in terms of linkability) per contact would be unthinkable, so to make these PET's work one needs a digital agent (or identity management device, IMD) that is programmed to choose the desired level of anonymization for you (Agre, 2001, Clarke, 1994, Lessig, 1999a).

As to the legal constraints, from the '70-ties onward attempts have been made to regulate the collection, storage, exchange and use of personal data (Bennett, 2001). The purpose of this regulation is of course not the protection of data in itself, but the protection of the persons that can

be harmed by the use of those data. Data protection legislation is a tool to protect the informational privacy of persons or groups. This legislation is generally based on a set of principles, first developed in the 1974 US Privacy Act, later expressed in the (non-binding) guidelines of the OECD and numerous national statutes on data protection (see e.g. the EC Directive 95/46/EC). The principles can be summarized as (1) the collection limitation principle, stating that collection of personal data should not be unlimited; (2) the data quality principle, stating that personal data should be correct, complete and up-to-date; (3) the purpose specification principle, stating that the purpose for which personal data are collected must be specified, and that they may only be used for that purpose; (4) the use limitation principles, stating that disclosure or use for other purposes is only allowed in case of consent of the data subject or on the basis of the authority of the law; (5) the transparency principle, stating that the data subject should be able to know about the collection and storage of personal data, their purpose and the identity of the data controller; (6) the individual participation principle, stating that a data subject has the right to erase, rectify, complete or amend her data; and finally (7) the accountability principle, stating that the data controller should be accountable for complying with these principles. In more generic terms we could sum up these principles under the heading of a *principle of minimum asymmetry*, combining demands of opacity for personal information with demands for transparency in the case of lawful monitoring of such data. The combination of opacity and transparency, argued by Gutwirth and De Hert (2005), offers a more balanced regime of informational freedom than David Brin's breath-taking thought-experiment. Brin (1998) claims that attempts to achieve opacity for individuals is outdated and should be substituted for attempts to provide total transparency of all information everywhere (implying that even the White House will be a Glass House, Rosen 2004:195). As such, the principle of minimum asymmetry has been described by Jiang (2002: 4) in terms of a privacy-aware system that 'should minimize the asymmetry of information held between data owners and data collectors and data users, by: (1) decreasing the flow of information from data owners to data collectors and users, and (2) increasing the flow of information from data collectors and users back to data owners'. The first part of the principle is equivalent with the data minimization principle, demanding maximum opacity of personal information, heralded by privacy propagators; the second part of the principle adds the principle of transparency that is constitutive for data protection regimes. Other than Jiang, data protection legislation legislation so far does not think in terms of data owners, especially in Europe data protection is considered a personality right and/or human right that cannot be traded with (Gutwirth & De Hert, 2005; Prins, 2004).

5.2 Effectiveness of data protection legislation

This all sounds very fair and very just. However, as we have seen, the essence of profiling is the *ubiquitous process* of collection, storage, aggregation and processing of data in databases. If such ubiquitous processes take place and develop into forms of ambient intelligence whereby our personal digital agents interact with an animated environment on a real time basis, data protection with its dependence on traditional legal tools seems to become totally inadequate. The sheer amount of decisions taken by intelligent devices, software programs and personal digital agents seems to invalidate traditional legal concepts like (1) liability for individual actions, (2) transparency of and access to personal information, (3) limitation of the use of data for specific purposes and, especially, (4) consent as the basis for collection, storage and processing of personal data. This is the case because (1) the actions of electronic devices that (will) impact our lives may at some point resist reduction to actions of a specific human agent; (2) the amount of data being collected and the low transaction costs for the data controller make transparency and access virtually impossible; (3) as the essence of profiling is linking data and discovering for what purpose the emerging profiles could be used, not much can be expected from attempts to prohibit linking and using data for other purposes; and (4) this is even more pertinent as so many daily transactions require consent of a kind that can hardly be taken serious, considering the consequences of refusing consent and the impossibility to go over all the different conditions on which consent is given.

We may conclude that the logic of profiling (ubiquitous linkability, unobtrusive correlatability) is at odds with the logic of data protection (providing citizens with the means to refuse and/or direct their linkability). The one builds on invisibility, the other on transparency. This means that unless the principles of informational privacy can – for instance - be built into the personal digital

agents (PDA) that manage the exchange of data, and unless these privacy-enhanced PDA's are widely used by human data subjects, profiling technologies will simply disable data protection legislation. This confronts us with a new problem. The risks of illegitimate (unjust?) profiling *could* be managed by developing a technological infrastructure that can prevent ubiquitous transparency of citizens, while at the same time promoting transparency of data controllers (principle of minimum asymmetry). However, such a technological infrastructure would need a legal, economic and social infrastructure that is conducive to the installment and real time maintenance of the balance between personal opacity and organizational transparency. We may in fact need new ways of interdisciplinary thinking to develop the adequate business-models, engineering design and legislation to allow the emergence of such technologically embodied law. In the following I will focus on the legal preconditions for such a system of checks and balances, notwithstanding the fact that the legal exploration will need adequate integration with the domains of what Lawrence Lessig calls the market, social norms and the architecture of our environment (Lessig, 1999a, see also Koops & Leenes 2005). I would add that in propagating technologically embodied law I do not plead an instrumentalist attitude to law and technology, whereby technology is used to enforce social norms without public debate and without any possibility to choose alternative actions. In reaction to (Lessig, 1999a; Reidenberg, 1998) several writers have indicated that such instrumentalism would in fact treat citizens as objects to be manipulated into preferred behavior, instead of appealing to them as rational agents or reasonable subjects (Brownsword, 2005; Tien, 2004).This sounds like the Pavlov reaction of a lawyer, who can only think in terms of law as the democratically legitimated commands of a sovereign, in opposition to technology as an implicit form of regulation that forces your hand without a possibility for contestation either in the political or the judicial forum. The point I am making – and that I believe Lessig to be advocating – is rather that technology regulates our lives anyway, and that it is up to us to make an informed choice about the technological infrastructure that facilitates the type of society we want to live in.

6 Correlated data subjects and correlatable humans

6.1 Two types of questions

The simple fact that profiles can, do and will affect our lives in both positive and negative ways raises many questions. On the one hand these questions concern the impact of false positives and false negatives (wrong categorization): how to organize the possibility of resistance against knowledge-claims regarding an individual based on the group-profile that is applied to her? If the profile is non-distributive, meaning that not all members of the group or category share all the attributes of the profile, it cannot be concluded that the profile applies to an individual member of that group or category. Applying the profile may thus lead to selection or exclusion on false grounds. On the other hand the questions raised concern the impact of knowledge about a person that stigmatizes this person or an entire group of persons; knowledge that confronts a person with information about herself she may not want to know; knowledge that is used for targeted servicing leading to customization and de-individualization. Can the law make these knowledge-constructs justiciable in relation to both types of questions: (1) questions about fitting people into wrong categories and (2) questions about effects of fitting people into categories as such? The first type of question is important but seems obvious: if I can give evidence that the profile does not apply to me, I have a case. The second type of question is less obvious: if the profile applies to me and is not abused for unjustified discrimination, what could be the problem?

6.2 The correlatable human and the correlated data subject

To explain what could be the problem we must look into the proliferation of profiles that are emerging. These profiles can be understood as *correlated data subjects*: a data subject is the subject (human or nonhuman, group or individual) the data refer to. If my online behavior is monitored and stored in databases together with the online behavior of many other data subjects, the reservoir of searchable data will be enormous and ever expanding. By means of data mining techniques new patterns can be discovered at any point in time, delivering new profiles consisting of correlated data that constitute correlated data subjects. What should interest us here is that nobody can predict which or how many different profiles will emerge and appear useful for a data controller at any point in time. In fact this would mean that in the end any trivial data can become

personally identifiable information (PII) which would make the EC Data Protection Directive 95/46 applicable on any trivial data (Schreurs and Hildebrandt, 2005: 38). The set of correlated data subjects that may be constructed and applied is unlimited as long as we continue interacting with our environment. This means that the human of flesh and bones, whose data are being registered is - in terms of the correlations that are claimed to represent her – a *correlatable* human and will never be congruent with any of the profiles that aim to define her for any specific purpose. This is a very crucial aspect of profiling that reveals the beginning of an answer to the question what could be the problem of profiling as such.

To explore this further I will use Gilles Deleuze's opposition of the *virtual* and the *actual*, as compared to his opposition of the *possible* and the *real* (Sasso & Villani, 2003:22-30). For Deleuze the virtual is not opposed to the real but to the actual. He relates the possible to the real, in the sense that the difference between them is merely that the one exists while the other does not. When the possible becomes real it does not change, only becomes existent (it is already determined). The virtual however is already real, it already exists. Whether, how and when the virtual actualizes is, however, not entirely determined. The virtual – present in the actual – is underdetermined. Thus being correlatable implies being virtual. It means that the correlations that will proliferate cannot all be determined in advance. As Pierre Lévy writes about the actual seed that contains the virtual tree:

> 'starting from the constraints that constitute the seed as this particular seed, it has to invent the tree, to coproduce it together with the circumstances it will meet' [my translation, mh] (Lévy, 1997).

In that line *realization* must be understood as the occurrence of a predefined possibility (or probability (Stengers, 1997: 27, footnote 10), while *actualization* is the invention of a solution required by a complex problematic. Pierre Lévy goes on to ask the question how we should understand *virtualization* (the inverse of actualization). He answers that question by saying that

> 'virtualization of an entity consists in discovering a general question to which this entity relates; forcing the entity to move in the direction of this interrogation and pushing it to redefine the actuality of its starting point as an answer to a particular question [my translation, mh]' (Lévy, 1997).

Actualization goes from problem to solution; virtualization passes from a given (actualized) solution back to the problem,

> 'thus making the instituted distinctions fluid, augmenting degrees of liberty, creating a productive vacuum [my translation, mh]' (Lévy, 1997).

In fact, Lévy writes,

> 'Virtualization is one of the main vectors of the creation of reality [my translation, mh]' (Lévy, 1997).

If we focus on the importance of the difference between a predefined possibility that comes into existence in a mechanical, predictable way and an underdetermined virtuality that is actualized in a specific way by co-constructing its environment as a solution to a concrete problem, we can see the importance of differentiating between a correlatable human of flesh and bones and the correlated data subjects that can be constructed on the basis of her past behavior. If these correlated data subjects are understood as realization of a predefined probability we deny the human that gave rise to these representations his indeterminate nature, we lock him up in the solution he chose. However, if we take the correlated data subjects for what they are we will keep in mind that they can always surprise us by inventing new solutions to new problems. Keeping this in mind is a kind a virtualization and implies the freedom that comes with indeterminate organisms. We should of course not be naïve about this freedom: once the tree has actualized we can dream back to the seed that contained an infinite set of possible trees – but this does not mean that the three can move back into the seed and choose another way of life. Human indeterminacy seems to move one step further, however, because the person of flesh and bones can sit down and reflect on her past behavior. She can become aware of her virtual self and can indeed attempt other solutions to new problems – even if this does not mean she can erase her actual self entirely to become a clean slate: we cannot become embryo's after growing up. The human capacity to be aware of oneself as a self is closely related to the development of the *sense of self* that constitutes

our identity (Hildebrandt, 2006a, Ricoeur 1992, Mead 1959/1934). The danger of profiling humans as data subjects lies in confusing the virtual with the possible: in the attempt to *define* the human person of flesh and bones by means of a profile that may provide intimate knowledge about this human. Such attempts to define a person – possibly with legal or other real consequences – may impact the sense of self of a person to an extent that destroys the freedom to (re)construct his identity. This is not to say that profiling in itself is dangerous, or that the reconstruction of our identity is a voluntaristic affair. Rather on the contrary: stereotypes provide us with the raw material to (re)construct our identity, they indicate how others perceive us and thus empower us to resist, amend or comply with the image that is projected upon us. This is what identification is all about. But if we are not able to reflect upon the prototypes that direct the way others deal with us, because only the profiler has access to them, these images of the self determine us beyond recognition. The thin line between monitoring and manipulation will loose its significance if we don't know what profiles are at work; we will find ourselves at the mercy of those that are so eager to service us with anything that brings them a profit.

The attempts to define a person can thus have as a consequence that targeted servicing creates an environment for me that responds to desires I did not know I cherished, thus pinning me down to the inferences made on the basis of past transactions. How can I challenge these inferences, and counter the future laid out for me? What happens if a suspect is faced with a profile that defines him as a psychopath, based on strings of data that have been correlated and prove a high probability of repeated criminal actions? Should he accept the profile and conform to the expectations thrown in his face, or could he accept the profile and decide to change his lifestyle? Should the law think of him as a lost case or should the law provide the means to counter both the applicability of the profile ànd the deterministic implications of the relevant profile, even if it is accepted as applicable?

7 Can Profiles be Made Just; Are Profiles Justiciable?

7.1 Making knowledge just or just justiciable?

What could it mean to make profiles, as knowledge-constructs, justiciable? And is making them justiciable the same as making them just? As a lawyer I feel rather attracted to the idea of making knowledge justiciable, because it is a more modest claim than making knowledge just. Justice is something to be strived for, certainly by the law, but to *claim* justice is – perhaps paradoxically – easily a bridge too far (Derrida, 1994). The difference between law and morality is that the first can settle disputes, while the last has probably caused them since we so often differ on what morality demands. At the same time, however, morality is also a part of the law: the disputes it causes at the heart of the law nourish the vitality, complexity and responsiveness of the law. Above that the law can even be said to be embody a specific morality, precisely because it makes things disputable (justiciable). To make knowledge just would be claiming the archimedic foothold from where to dictate true knowledge, while to make knowledge justiciable is rather the opposite. It means that we bring competing knowledge claims within the jurisdiction of a fair trial that allows their proponents to plead their case. I will thus restrict myself to raising the question what it takes to make the knowledge claims of profiling practices justiciable.

7.2 Three legal tools to make knowledge-claims justiciable

From the perspective of democracy and rule of law, the answer is three-fold: first, we need *legislation* that constrains technological design and its applications to fit the demands of a democratic constitutional state. This means that I will at least have access to the profiles that are applied to me and have the legal tools to contest the knowledge claim they present. Second, the most important legal tool to accomplish this is a *fair trial* that makes the knowledge claims of profiles justiciable whenever they have legal consequence. As discussed in other work the fair trial offers an interesting 'ideal type' or 'good practice' for the testing of knowledge claims (Hildebrandt, 2006b). The combination of the interrelated principles of an independent and impartial judge, a public hearing, equality of arms, presumption of innocence, contradictory proceedings and the principle of immediacy provides a setting that allows lay persons to have the last word on competing expert knowledge claims. In fact, some social research claims that jury

trials contain an interesting setting for rethinking both the construction of knowledge and the issue of representation (participative instead of aggregative, (Wakeford, 2002). As indicated before we need to develop a technological infrastructure that detects when and where profiles are constructed and anticipates the use made of them. Otherwise we have no way of knowing which and how knowledge claims are impacting our lives. Third, we need a concept of the *legal person* that empowers a person to question the construction and application of profiles as they impact humans as correlatable data subjects and as persons with a sense of self

7.3 The concept of the legal persona

In this last section I will focus on the concept of the legal person that is preconditional for the democratic constitutional state and for the fair trial that embodies the constitutive principles of the rule of law. The concept of the legal person affects two things. First it *provides a position from which* actions – like knowledge construction - can be made justiciable if they are claimed to cause harm. This position is effectively actualized *in* the 'fair trial'. The formal equality attributed to the legal persona in court empowers humans of flesh and bones to contest unjustified discrimination on the basis of inferred profiles – provided the technological infrastructure exists that can detect the use of profiles.

Second, the fact that the law thinks in terms of the legal person (legal subjectivity) provides an artificial position that shields humans as correlatable data subjects and as a person with a sense of self from complete determination. The concept of the legal person refers to the Greek *persona*, which was the mask behind which actors hid themselves when they performed in a theatre performance. The mask indicated the role they played. When a subject takes the stand in a court of law, the construct of the legal person prevents confusion between the role the subject takes on the one hand and the indeterminate subject of flesh and bones on the other hand (Foqué, 1996; Foqué & 't Hart, 1990). One could articulate this in another way by saying that the concept or construct of the legal person allows the human person as a correlatable data subject with a sense of self to resist the profile (the correlated human) and/or the way it affects her life. The importance of the legal person is that it protects the essentially underdetermined nature of the human person against the desire of the state and the market to categorize its subjects in such a way that they fit the logic of state bureaucracy and/or market imperatives. By insisting on the correlatability of humans against knowledge claims concerning correlations, the legal person confirms and protects what Deleuze would perhaps have called the virtual character of the correlatable human. The law, by instituting the legal person, creates a distance between the correlatable human and the correlated data subject, thus creating a specific type of freedom. This freedom allows us to challenge the actual profile; by virtualizing it back to the questions it presumes, thus creating the possibility to claim for instance the irrelevance of the actual profile.

Imagine a group profile that attributes certain properties to members of the group. If the profile is non-distributive (as most non-trivial profiles are), categorizing members by means of the profile will produce false negatives and false positives. If a person takes action in a court of law claiming that the profile is not applicable in her case, she in fact challenges the move from the probable to the real. This is the more obvious reason why we want to know that, which and how profiles are used to infer things about our person. Imagine a person is profiled as a psychopath according to the checklist of the famous professor Hare (Edens, 2001), because she has the properties that define a psychopath. Though she may in fact fit the profile, she may want to question the validity of knowledge construct that lies at the basis of her profile (claiming in her case the profile is a false positive) and/or she may want to question the relevance of categorizing people on the basis of Hare's checklist. In that case she may claim that the profile of a psychopath as constructed by Hare answers the wrong question, looking only for the probabilities that define an ensuing reality, instead of virtualizing actual traits and thus opening the way for new actualities (Stengers, 1997: 147).

8. Concluding remarks

Profiling practices entail a specific form of knowledge construction, used to provide an assessment of risks and opportunities. On the basis of these knowledge constructs decisions are made that impact the life of the data subjects and their sense of self. This in itself is not a new fact; the nexus of the life and information sciences indicate that profiling is a typical way for organisms to meet the demands of their environments. However scope and scale of the collection, aggregation and processing of data that is made possible by pervasive, ubiquitous and intelligent computing impact power relations between those that are profiled and those that control or use profiles.

For democracy and rule of law it is important to understand the human person as a correlatable data subject with a sense of self, that cannot be defined entirely in terms of probabilities. This is precisely why it is important to make profiles justiciable and also indicates the importance of the concept of the legal person. The legal person empowers the actual person (the correlated human) to claim her virtual identity (the correlatable human), to resist determination by probabilities. In the end it will be the access to the fair trial and the access to information on the profiles built on our data that will open the possibility to challenge profiles and their implications.

References

Agre, P.E. (2001) Beyond the Mirror World: Privacy and the Representational Practices of Computing, in: P.E. Agre & M. Rotenberg (Eds) *Technology and Privacy: The New Landscape* (Cambridge, Massachusetts, MIT).

Beck, U. (1992) *Risk Society: Towards a New Modernity* (London, Sage).

Bennett, C.J. (2001) Convergence Revisited: Toward a Global Policy for the Protection of Personal Data?, in: P.E. Agre & G. Bramhall (Eds) *Technology and Privacy: The New Landscape* (Cambridge, Massachusetts, MIT).

Brin, D. (1998) *The Transparent Society. Will Technology Force Us to Choose Between Privacy and Freedom?* (Reading, Massachusetts, Perseus).

Brownsword, R. (2005) Code, control, and choice: why East is East and West is West, *Legal Studies*, 25(1), pp. 1-22.

Canhoto, A. & Backhouse, J. (2004) Constructing categories, Construing signs - analysing differences in Suspicious Transaction Reporting practiceInformation Systems Integrity Group, London School of Economics, London).

Clarke, R. (1994) Human Identification in Information Systems: Management Challenges and Public Policy Issues, *Information Technology & People*, 7(4), pp. 6-37.

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 Mul, J. (2003) Digitally mediated (dis)embodiement. Plessner's concept of excentric positionality explained for cyborgs, *Information, Communication & Society*, 6(2), pp. 247-266.

Derrida, J. (1994) *Force de loi* (Paris, Galilée).

Edens, J.R. (2001) Misuses of the Hare Psychopathy Checklist-Revised in Court, *Journal of Interpersonal Violence*, 16(10), pp. 1082-1094.

Foqué, R. (1996) Legal Subjectivity and Legal Relation. Language and Conceptualization in the Law*Festschrift for Jan M. Broekman "Law, Life and the Images of Man"* (Berlin,

Foqué, R. & 't Hart, A.C. (1990) *Instrumentaliteit en rechtsbescherming* (Arnhem Antwerpen, Gouda Quint Kluwer Rechtswetenschappen).

Goffman, E. (1963) *Stigma. Notes on the Management of Spoiled Identity* (Englewood Cliffs, NJ, Prentice-Hall).

Gutwirth, S. & De Hert, P. (2005) Privacy and Data Protection in a Democratic Constitutional State, in: M. Hildebrandt & S. Gutwirth (Eds) *Profiling: Implications for Democracy and Rule of Law, FIDIS deliverable 7.4* (Brussels, available at www.fidis.net,

Hildebrandt, M. (2006a) Privacy and Identity, in: E. Claes, A. Duff & S. Gutwirth (Eds) *Privacy and the Criminal Law* (Leuven, Intersentia).

Hildebrandt, M. (2006b) The Trial of the Expert: Èpreuve and Preuve, *The New Criminal Law Review*, 1(1 or 2).

Hudson, B. (2005a) Secrets of Self: Punishment and the Right to Privacy, in: E. Claes & A. Duff (Eds) *Privacy and the Criminal Law* (Antwerp Oxford, Intersentia).

Jiang, X. (2002) Safeguard Privacy in Ubiquitous Computing with Decentralized Information Spaces: Bridging the Technical and the Social*Privacy Workshop September 29, 2002, University of California, Berkeley* (Berkeley, available at: http://guir.berkeley.edu/pubs/ubicomp2002/privacyworkshop/papers/jiang-privacyworkshop.pdf.

Kranzberg, M. (1986) Technology and History: 'Kranzberg's Laws', *Technology and Culture*, 27, pp. 544-560.

Leenes, R. and Koops, B.-J., (2005) 'Code': Privacy's Death or Saviour?, International Review of Law Computers & Technology 19 (3), pp. 329-340.

Lessig, L. (1999a) *Code and other laws of cyberspace* (New York, Basic Books).

Lessig, L. (1999b) The Law of the Horse: What Cyberlaw Might Teach', *Harvard Law Review*, 113(501), pp. 501-547.

Lévy, P. (1997) Sur les chemins du virtuel

Maturana, H.R. & Varela, F.J. (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).

Polanyi, M. (1966) *The Tacit Dimension* (Garden City, New York, Anchor Books).

Prins, J.E.J. (2004) The Propertization of Personal Data and Identites, *Electronic Journal of Comparative Law, available at http://www.ejcl.org/*, 8(3).

Reidenberg, J.R. (1998) Lex Informatica: The Formulation of Information Policy Rules Through Technology, *Texas Law Review*, 76(3), pp. 553-585.

Ricoeur, P. *Oneself as Another*. Translated by Kathleen Blamey (Chicago and London, The University of Chicago Press)

Rose, H. (2003) The Commodification of Virtual Reality, in: A.H. Goodman, D. Heath & M.S. Undee (Eds) *Genetic Nature/Culture. Anthropology and Science beyond the two-culture divide* Universtity of California Press).

Rosen, J. 2004 *The Naked Crowd. Reclaiming Security and Freedom in an Anxious Age* (New York, Random House)

Ryle, G. (1949) *The Concept of Mind* (New York, Barnes & Noble).

Sasso, R. & Villani, A. (Eds.) (2003) *Le Vocabulaire de Gilles Deleuze* (Paris, Librairie Philosophique J. Vrin).

Schauer, F. (2003) *Profiles Probabilities and Stereotypes* (Cambridge, Massachusetts

London, England, Belknap Press of Harvard University Press).

Schreurs, W. & Hildebrandt, M. (2005) Legal Issues, in: W. Schreurs, M. Hildebrandt, M. Gasson & K. Warwick (Eds) *Report on the Actual and Possible Profiling Techniques in the Field of Ambient Intelligence* (Brussels, FIDIS deliverable 7.3, available at www.fidis.net).

Scott Armstrong, J. (1970) How to Avoid Exploratory Research, *Journal of Advertising Research*, (4), pp. 27-30.
Shklar, J. (1990) *The Faces of Injustice* (New Haven, Yale University Press).
Sonigo, P. & Stengers, I. (2003) *L'Evolution* (Paris, EDP Sciences).
Stengers, I. (1997) *Cosmopolitiques. Tome 7. Pour en finir avec la tolérance* (Paris, La Découverte / Les Empêcheurs de penser en rond).
Tien, L. (2004) Architectural Regulation and the Evolution of Social Norms, *International Journal of Communications Law & Policy*, (9).
Van Brakel, J. (1999) Telematic Life Forms, *Techné: Journal of the Society for Philosophy and Technology*, 4(3), pp. http://scholar.lib.vt.edu/ejournals/SPT/v4_n3html/VANBRAKE.html.
Wakeford, T. (2002) Citizens Juries: a radical alternative for social research, *Social Research Updat*, 37(summer).

[i] Mireille Hildebrandt teaches law and legal theory at the Law Faculty of Erasmus University Rotterdam and is seconded to the Law Science Technology and Society (LSTS) at the Vrije Universiteit Brussels as senior researcher on an interdisciplinary research project, coordinated by Serge Gutwirth, Bruno Latour and Isabelle Stengers, financed by the Belgium Science Policy Office, called: *The Loyalties of Knowledge. The positions and responsibilities of the sciences and of scientists in a democratic constitutional state*. See http://www.vub.ac.be/LSTS/people/Hildebrandt/index.shtml. This contribution has been inspired by the challenging research done within the project on the *Loyalties of Knowledge*. A first version of this text was presented at the international conference on 'Can Knowledge Be Made Just' at the Kulturwissenschaftliches Institut in Essen, Germany on 23rd March 2005; a second version was discussed at a seminar of the Information Systems Integrity Group of James Backhouse at the London School of Economics on 20th October 2005. I want to thank the participants of both meetings for their challenging and enthousiastic comments.