Profiling into the future:  
An assessment of profiling technologies in the context of Ambient Intelligence

Mireille Hildebrandt

Vrije Universiteit Brussel, Erasmus Universiteit Rotterdam,  
Workpackage Leader of Profiling in FIDIS

Abstract

Both corporate and global governance seem to demand increasingly sophisticated means for identification. Against a justification in terms of dealing with security threats, fraud and abuse, citizens are screened, located, detected and their data stored, aggregated and analysed. At the same time potential customers are profiled to detect their habits and preferences in order to provide for targeted services. Both industry and the European Commission are investing huge sums of money into what they call Ambient Intelligence and the creation of an Internet of Things. Such intelligent networked environments will entirely depend on real time monitoring and real time profiling, resulting in real time adaptation of the environment. In this contribution the author assesses the threats and opportunities of such autonomic profiling in terms of its impact on individual autonomy and refined discrimination and indicates the extent to which traditional data protection is ineffective as regards profiling.
1 Introduction: The Age of Identification

We live in the age of identification. Both government and business enterprise strive to develop effective tools for recurrent identification and authentication. Writers such as Scott (1998) in his ‘Seeing Like a State’, and Torpey (2000) in his ‘The Invention of the Passport. Surveillance, Citizenship and the State’, have described and analysed the insatiable need of the modern state for the registration of its citizens, originally seeking to attribute and implement tax obligations and subscription for the national army. The welfare state claims a need to maintain the line between those that are entitled to public benefits and those that have no such right, while it also claims to need identification for the prevention of fraud, crime and unlawful access in general, and for the attribution of liability, whether criminal or tort. E-government and e-health systems that aim to provide targeted services reiterate the quest for identification, although in this case the identification needed is more sophisticated and resembles what business undertakings seek when they develop targeted servicing and reinvent customer relationship management (CRM).

Business enterprise is less interested in registration of all inhabitants of a territory. Its focus is on acquiring relevant data about as many (potential) customers as possible as part of their marketing and sales strategies. As customer loyalty cannot be taken for granted, companies develop CRM in the hope of surviving the competitive arena of neo-liberal market economies. At the same time they try to establish which consumers may be persuaded to become their new customers under what conditions. It seems that they are less interested in unique identification of any particular customer then in a refined type of categorisation that allows them to provide targeted servicing at the right time and in the right place. Context is all is not just the key message of adherents to cultural theory. In fact, companies are not just after the attributes of predefined classes of (potential) customers, but would rather invest in finding out which classes they should discriminate. This is where profiling comes in.
In this contribution we will discuss how sophisticated machine profiling differs from the kind of profiling we do in our everyday life (sections 2 and 3). Next we will investigate profiling as the enabling technology for Ambient Intelligence and the Internet of Things (section 4). Profiling produces knowledge, rather than just data; section 5 will look into the threats posed by emerging knowledge-asymmetries due to the proliferation of profiling in smart environments. These threats touch on some of the fundamental tenets of democracy and rule of law, being the particular mélange of positive and negative freedom that allows citizens to develop their relative autonomy (section 6). To counter such threats the focus of legal scholars and practitioners should be extended from the protection of personal data to protection against the undesired application of profiles and the creation of transparency rights regarding group profiles (section 7). To exercise legal transparency rights, the technological infrastructure that is being constructed to facilitate smart environments must incorporate transparency-enhancing technologies (TETs). Section 8 will argue why the introduction of privacy enhancing technologies (PETs) should be complemented with TETs. Section 9 concludes with some closing remarks.

2 What else is new? Autonomic behaviour and autonomous action

Profiling is as old as life itself. Indeed one could say that the difference between living and lifeless material is the fact that living organisms are capable of self-constitution over and against an environment which is constituted as such by the act of self-constitution (Maturana and Varela 1991). In more simple terms: an organism and its environment co-create each other. Profiling is thus a crucial sign of life, because it consists of a reiterant identification of risks and opportunities by an organism in its environment. Profiling is the interplay between monitoring and adaptation: to survive and to celebrate life any organism must continuously adapt itself to changes in its surroundings, while it may also manage to adapt its surroundings to its own preferences (Maturana and Varela 1991). Monitoring one's context in this sense is a matter of pattern recognition, of discriminating noise from
information. Not all data are relevant or valid, and deciding which is the case will depend on the context and on the moment. Adequate profiling is always dynamic and caught up in the loop of recognising a pattern (constructing the profile) and testing its salience (applying the profile).

Interestingly enough, this organic profiling is not dependent on conscious reflection. One could call it a cognitive capacity of all living organisms, without claiming consciousness for an amoeba. One could also call it a form of intelligence based on the capacity to learn: monitoring and testing, subsequent adaptation and reiterant checking is what makes the living world function. This is what allows any organism to maintain its identity over the course of time, detecting opportunities to grow and spread as well as risks that need to be acted upon.

So, profiling is not typically human though we have developed our own brand of profiling, termed stereotyping by cognitive psychologists, and categorisation by Schauer (2003) in his *Profiles, Probabilities and Stereotypes*. What is special about humans is their capacity – according to brain scientists neatly embodied in the prefrontal cortex – to reflect upon the profiles they come up with. This is a rare capacity, closely related to consciousness and language, and we shall not explore this domain much further, leaving it at the nexus of neurosciences and philosophy of mind. What still needs to be established is the fact that our capacity for conscious reflection on the profiles that we have unconsciously generated gives us the freedom to deliberate on them, to reject or to reinforce them and deliberately to apply them. This is what creates our freedom to act. We can become aware of the patterns that regulate our actions and review them to change our habits. Though most of our interactions are automated, handled autonomically by the habits that are inscribed in our body and brains, we can bring them to mind and scrutinize their relevance, validity, fairness and justice. This is what makes us into autonomous agents, capable of making a conscious choice for a course of action, deciding by which law to live.
Autonomos derives from the Greek auton nomos: self and law. We can live by our own law, and are therefore held accountable for our actions (Hildebrandt 2008).

3 What is new? Profiling machines

Automated profiling is new in two ways. First, we are not talking about profiling by organisms but about profiling by machines (Elmer 2004). Essentially these machines are software programs 'trained' to recover unexpected correlations in masses of data aggregated in large databases. Second, we are not talking about making a query in a database, summing up the attributes of a predefined category, but about discovering knowledge we did not know to be in the data (Zarsky 2002-2003; Custers 2004).

Automated profiling can be described as the process of knowledge discovery in databases (KDD), of which data mining (DM, using mathematical techniques to detect relevant patterns) is a part (Fayyad, Piatetsky-Shapiro et al. 1996). KDD is generally thought to consist of a number of steps:

1. recording of data
2. aggregation & tracking of data
3. identification of patterns in data (DM)
4. interpretation of the outcome
5. monitoring data to check the outcome (testing)
6. applying the profiles

Only the third step is what is called data mining in the sense of using mathematical algorithms to locate correlations, clusters, association rules and other patterns. An example of such profiling, using genetic algorithms, is driver fatigue detection by Jin,
Park et al. (2007). This type of profiling is also called behavioural biometric profiling (BBP) and uses a combination of pupil shape, eye movement frequency and yawn frequency to check tiredness of a driver. The data are mined by means of a feed-forward neural network and a back-propagation learning algorithm. To be honest we must note that BBP is still in an early stage of development, even though some results are highly interesting.¹ Both Zarsky (2002-2003) and Custers (2004) emphasize that the knowledge generated by profiling machines is new. Zarsky speaks of data mining “answering questions users did not know to ask” (Zarsky 2002-2003: 4). He especially focuses on the difference between classification based on predefined classes and data mining techniques, which provoke unexpected clusters. Custers (2004:56-58) argues that this type of knowledge is new in comparison with traditional social science, which starts with a hypothesis concerning a population that is tested by applying it to a sample. He points out that in the case of KDD the hypothesis emerges in the process of data mining and is tested on the population rather than a sample. He also indicates that when trivial information turns out to correlate with sensitive information, an insurance company or an employer may use the trivial information to exclude a person without this being evident as unfair discrimination (called masking). His last point is that the recording of data by means of ICT makes it nearly impossible to delete records, especially as they are often shared across contexts. KDD can thus trace and track correlations in an ever-growing mass of retained data and confront us with inferences drawn from past behaviour that would otherwise be lost to oblivion (Warner 2005).

So, we have two differences with autonomic organic profiling: (1) the profiling is done by machines and (2) the type of knowledge they generate differs from 'classic' empirical statistics. This raises several questions in relation to privacy and security, especially with regard to the effectiveness of data protection legislation. Before moving into these anticipated threats I will first describe the Vision of Ambient Intelligence and the Internet of Things, to explain why autonomic machine profiling

¹ See e.g. BBP for aggression detection by means of monitoring of sound, at http://www.soundintel.com/index-en.html.
may have a major impact on our lives. This should reinforce the need to discuss potential new threats to privacy, security and other basic tenets of democracy and the rule of law.

4 A Vision of Ambient Intelligence and The Internet of Things

Both the European Commission (ISTAG 2001) and, for instance, Philips (Aarts and Marzano 2003), have invested heavily in what is called the vision of Ambient Intelligence (AmI), vaguely defined by its 'key elements' (Aarts and Marzano 2003:14), being:

- embedded (many networked devices integrated into the environment)
- context-aware (these devices can recognize you and your situational context)
- personalized (they can be tailored towards your needs)
- adaptive (they may change in response to you)
- anticipatory (they can anticipate your desires without conscious mediation)

Other key elements often repeated in the context of AmI are: hidden complexity, the absence of keyboards or monitors, the fact that the environment itself becomes the interface, real time monitoring and proactive computing.

The enabling technologies of this smart environment are sensor technologies, RFID systems, nanotechnology and miniaturization. Together they create The Internet of Things (ITU 2005), supposed to turn the offline world online. The Internet of Things consists of things that are tagged and observed permanently while communicating their data through the network that connects them. We must keep in mind, though, that most of these technologies only generate an enormous amount of data, which may not reveal any knowledge until profiling technologies are applied. We may conclude that profiling technologies are the crucial link between an overdose of trivial data about our movements, temperature, and interaction with other people or things and
applicable knowledge about our habits, preferences and the state of the environment. Only after using data mining techniques on the interconnected databases can the things in our environment become smart things and start acting like agents in a multi-agent network (MAS). Profiling thus creates the added value in the mass of data, amongst which we do not yet know what is noise and what is information.

The vision of AmI depends on a seamless adjustment of the environment to our inferred habits and preferences. The idea is that we need not provide deliberate input as in the case of interactive computing, but are 'read' by the environment that monitors our behaviour. This presumes what Tennenhouse (2000) describes as proactive computing, diminishing as far as possible any human intervention. To adapt seamlessly the environment we cannot wait for a human interpreter but need profiling machines that draw their own conclusions about what we prefer when and where, hoping to thus solve the problem of endless choice and deliberation.

5 Threats: Knowledge is Power

It is important to discuss profiling as a separate issue, not to be conflated with data collection. Informational privacy is all too often reduced to a private interest in the hiding of personal data, while (1) privacy is not just a private good and (2) the hiding of personal data seems a Pavlovian reaction that will, however, not protect us from the impact of group profiling, while (3) the hiding of personal data will reduce the quality of the profiles (and the intelligence of the environment). If we think of privacy as a public good that is constitutive of human agency in a constitutional democracy then trading between privacy and security becomes a tricky thing, especially if we realise that security is also more than a private interest as it is one of the raisons d'être of the state. As to the trade-off between privacy and security two points can be made:
a loss of privacy may imply a loss of security, making the idea of a trade-off between the two even more complicated, and trading implies commodification and with Schwartz (2000) we may expect a market failure owing to the unequal access to information about the consequences of trading one's personal data (especially in the case of profiling).²

Profiling machines may spy on citizens, but why should citizens care about a machine watching their daily business? In Aml, most of the monitoring and adaptation will be a matter of machine to machine communication (M2M), while these machines will not be interested in who citizens are but in what profit can be gained from which category that citizens fit. How does this relate to privacy and security as what Schwartz (2000) calls constitutive (public) values, aiming to provide citizens with a kind of agency presumed in constitutional democracy? As for profiling, privacy and security both seem to revolve around the question “who is in control: citizens or profilers?” But again control is often reduced to hiding or disclosing personal data and this does not cover privacy and security as public values.

To come to terms with potential threats we need to look deeper into the asymmetries between citizens on the one hand and large organisations who have access to their profiles on the other hand. We are not referring to the asymmetry of effective access to personal data but the asymmetry of access to knowledge. Especially insofar as this knowledge is protected as part of a trade secret or intellectual property, the citizens to whom this knowledge may be applied have no access at all. Zarsky (2002-2003) has demonstrated – by analysing a set of examples – how this lack of access can lead to what he calls the 'autonomy trap'. Precisely because a person is unaware of the profiles that are applied to her, she may be induced to act in ways she would not have chosen otherwise. Imagine that my online behaviour is profiled and matched with a

² Schwartz 2000, at 745, refers to the Calabresi-Melamed analysis that states that property rules work well in the case of few parties, difficult valuations, low transaction costs, while liability rules work well in the case of many parties, monopoly, strategic bargaining and high transaction costs.
group profile that predicts that the chance that I am a smoker on the verge of quitting is 67%. A second profile predicts that if I am offered free cigarettes together with my online groceries and receive news items about the reduction of dementia in the case of smoking I have an 80% chance of not quitting. This knowledge may have been generated by tobacco companies, who may use it to influence my behaviour. In a way, this kind of impact resembles Pavlov's stimulus-response training: it does not appeal to reason but aims to discipline or induce me into profitable behaviour. My autonomy is circumvented as long as I am unaware of the knowledge that is used. Zarsky (2002-2003) also warns about unfair discrimination, based on refined profiling technologies that allow sophisticated market segmentation. Price discrimination may be a good thing in a free market economy, but the fairness again depends on consumers’ awareness of the way they are categorised. In order to have a fair and free market-economy some rules of the game must be established to prevent unequal bargaining positions, or else we have another market failure. In short the threats can be summarised as follows:

- privacy (which must not be reduced to hiding one's personal data)
- security (which cannot be traded for privacy as a loss of the one may cause the loss of the other)
- unfair discrimination (power relations must be balanced to provide equal bargaining positions)
- autonomy (our negative and positive freedom to act must be established and maintained, manipulation on the basis of knowledge that one is not aware of violates one's autonomy)
6 Democracy and the Rule of Law

A sustainable democracy presumes and maintains the rule of law. The rule of law is often defined in reference to the protection of human rights and limited government. With regard to the implications of profiling technologies the most relevant achievement of the rule of law seems to be the mix of what (Berlin 1969/1958) has coined negative and positive freedom. Positive freedom – *freedom to* – regards the freedom to participate in public decision-making processes or the freedom to achieve one's personal objectives; negative freedom or liberty – *freedom from* – regards the absence of unreasonable constraints imposed on a person. Positive freedom has a long history, while negative freedom – as a value of liberal democracy – is a relatively recent invention. To nourish a sustainable democracy we need both types of freedom, for which we need the rule of law (Gutwirth and De Hert 2005). The rule of law establishes constitutional protection of citizens’ rights and liberties over and against their government, safeguarded by an independent judiciary that shares the authority of the state. This is called the paradox of the Rechtsstaat: the state gives its authority to those that judge citizens who contest the way the state uses its authority in a given case.

Profiling can endanger both negative and positive freedom. Negative freedom is often equated with opacity, retreat to a private space, the right to oblivion and invisibility to the public eye. It refers to a space and time to regain one's strength, to reflect upon one's objectives and opinions. This negative freedom is matched with a need to act, to anticipate and participate in the public space, for which some measure of transparency is needed. Without transparency one cannot anticipate or take adequate action. In fact I would claim that negative freedom is an illusion as long as transparency is absent, as we may think that we are facing up to reality in the privacy of our own thoughts, while in fact we have no access to the knowledge needed to assess this reality. Thus profiling may endanger the intricate combination or negative and positive freedom whenever we:
- think we are alone, but are in fact watched by machines (observing our online behaviour, our keystroke behaviour and in an AmI world any move we make);
- think we are making private decisions based on a fair idea of what is going on, while in fact we have no clue as to why service providers, insurance companies or government agencies are dealing with us the way they do.

Referring to the discussion in section 1.2 we should admit that most of our interactions take place without conscious reflection; they are a type of autonomic behaviour that is the result of individual learning processes that enable us to move smoothly through everyday life. This, in itself, is not a violation of our negative or positive freedom. As a result of learning processes it may even be the result of the way we exercised our freedom in the past (Varela 1992). However, if freedom is related to the possibility of deliberate reflection on our choices of action, then we need to have access to the knowledge that impacts these choices. Targeted servicing, customisation and filtering of information could otherwise provide us with a comfortable, golden cage (Sunstein 2001); allowing us a reflexive life without reflection (Lessig 1999).

7 Legal pitfalls and challenges

Data have legal status. They are protected - at least personal data are. Europe tends to seek protection in a personal right, which opens the possibility to declare certain data to be inalienable, but in practice it seems to boil down to the fact that the leaking of personal data is taken to imply consent for storing and using them. Whatever the written safeguards in the data protection directive, in practice most people most of the time have not the least notion of what is happening with which data resulting in the application of which profiles. The US tends towards commodification, which opens up the possibility to trade with your personal data, which may provide at least a sense of citizen's control. However, as discussed above, one may expect a market failure in
the sense that owing to grotesque knowledge asymmetries the implied consent will be as ignorant as in the European case. In both cases one of the problems is that we have no access to the group profiles that have been inferred from the mass of data that is being aggregated and not the faintest idea how these profiles impact our chances in life. It may be time to reconsider the legal focus on the protection of personal data, as well as the focus of the privacy advocates who invest in privacy enhancing technologies. What we need is a complementary focus on the dynamically inferred group profiles that need not be derived from one's personal data at all, but may nevertheless contain knowledge about one's probable (un)healthy habits, earning capacity, risk-taking, life style preferences, spending habits, political associations etc.

Profiles have no clear legal status. That is, they may be protected from access via intellectual property rights of the profiler or be considered part of a company's trade secrets. Protection against, or at least access to profiles is very limited. In data protection legislation one can locate two ways to claim access to a profile. First one can argue that once a profile has been applied to an individual person it becomes a personal data, e.g. in the case of credit scoring practices. This however, does not concern the relevant group profile or its relation to other group profiles, nor the way the profile was generated (by use of which algorithm etc.). Second one can argue that autonomic application of profiles falls within the scope of art. 15 of the Data Protection Directive (D 46/95 EC). Paragraph 1 of this article reads:

Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him

In short, this article seems to grant European citizens a right not to be subjected to an automated decision in the case that this decision makes a difference to their life. However, this safeguard has four pitfalls. First, as Bygrave (2001) suggests, it may be
that if the citizen does not exercise the right, the automated decision is not a violation of the directive. Second, the 2nd paragraph of art. 15 reads:

Subject to the other Articles of this Directive, Member States shall provide that a person may be subjected to a decision of the kind referred to in paragraph 1 if that decision:

(a) is taken in the course of the entering into or performance of a contract, provided the request for the entering into or the performance of the contract, lodged by the data subject, has been satisfied or that there are suitable measures to safeguard his legitimate interests, such as arrangements allowing him to put his point of view; or

(b) is authorized by a law which also lays down measures to safeguard the data subject's legitimate interests.

This seems to create many loopholes for automated application of profiles. The third pitfall concerns that fact that as soon as the decision is not automated because of a (routine) human intervention the article no longer applies. In the case of autonomic profiling in an AmI environment this would not be an option, because the seamless real time adjustment of the environment rules out such human intervention. This brings us to the fourth and last pitfall: as long as one is not aware of being subject to such decisions one cannot exercise this right. The fact that art. 12 grants the right to know “the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions referred to in art. 15” does not really help if one doesn't know about the automated decisions in the first place. This is the case even if after application the profile may in fact be a personal data and fall within the scope of art. 11 and 12 which obligates the relevant decision maker to notify the data subject and provide access to the data. In other words: today's technological and
organisational infrastructure makes it next to impossible seriously to check whether and when the directive is violated, creating an illusion of adequate data protection, while in fact the US approach may deliver better results due to the constitutional protection that can be claimed (Hosein 2005).

So, we have a major challenge here. To render the directive effective we need a technological and organisational infrastructure that provides real time information about knowledge that may be applied to us, including the potential consequences. Only if such an infrastructure is in place can the rule of law, especially the particular mélange of positive and negative freedom, be sustained.

8 Transparency Enhancing Tools: From PETs to TETs

On 15th March 2007, after an extensive public consultation during 2006, the European Commission presented its Communication on RFID.3 The commission starts out with claiming a beneficial social contribution of RFID in a number fields: safety (e.g., food traceability, anti-counterfeiting of drugs), convenience (e.g., shorter queues in supermarkets, more accurate and reliable handling of luggage at airports, automated payment), accessibility (e.g., patients suffering from dementia and Alzheimer's disease), healthcare (increased quality of case and safety), retail and industry (supply chain management), protection of the environment (e.g. recycling). Next to the social contribution the Commission expects RFID to boost industrial innovation and growth potential. After this, the issues of data protection, privacy and security are dealt with in reference to the Data Protection Directive and the ePrivacy Directive.4 Regarding the Data Protection Directive the Commission notes that the Member States will have to ensure that the introduction of RFID applications complies with privacy and data protection legislation, necessitating the drawing up of specific codes of conduct. The

4 D 2002/58/EC.
Commission indicates that the national data protection authority and the European 'Article 29 Working Party'\(^5\) will have to review these codes of conduct and monitor their application.

What about the fact that the national data protection authority lacks the resources seriously to monitor what is happening? We only need refer to the transfer of transaction data of European banks and Swift to the US department of Treasury (UST) authorities to realise that organisations do not feel compelled to notify citizens of the way their data are used, unless forced by widespread publicity (ICPP 2006, Boon 2007).\(^6\) The Commission is of the opinion that the response to the challenges posed by RFID technologies should include the “adaption of design criteria that avoid risks to privacy and security, not only at the technological but also at the organisational and business process levels”.\(^7\) This is an interesting option. It refers to what has been called constructive technology assessment (CTA), initiated by e.g. Rip, Misa et al. (1995), building a case for 'upstream' involvement in technological design, i.e. not installing ethical commissions after the technology is a finished product but getting involved at the earliest possible stage of technological design.

RFID-systems are one of the enabling technologies of *The Internet of Things* and AmI (Hildebrandt and Meints 2006). They produce an immense amount of data about (change of) location and if linked to other data they provide a rich resource for profiling practices. Which technological and organisational infrastructure will provide the transparency of profiles that we argued above? The Commission mentions its support for privacy-enhancing technologies (PETs), “to mitigate privacy risks”.\(^8\) However, as stated above PETs focus on the hiding of data (anonymisation) and on the use of pseudonyms, which may provide a kind of what Nissenbaum (2004) has

\(^7\) COM(2007)96 final, at 6.
\(^8\) COM(2007)96 final, at 11.
coined contextual integrity. To counter the threats of autonomic profiling citizens will need more than the possibility of opting out, they will need effective transparency enhancing tools (TETs) that render accessible and assessable the profiles that may affect their life.\textsuperscript{9}

For this reason we end this contribution with an appeal to rethink the legal-technological infrastructure to provide profiles with an effective legal status, allowing citizens, whether as consumers, patients or target of government investigations, with the legal and technological tools to understand which profiles may impact their life in which practical ways.

9 Closing remarks

Advanced profiling technologies answer questions we did not raise. It generates knowledge we did not anticipate, but are eager to apply. As knowledge is power, profiling changes the power relationships between profilers and those being profiled. These asymmetries challenge the relative autonomy of individual citizens and allow an unprecedented dynamic segmentation of society, especially if Ambient Intelligence comes through: based on refined real time monitoring, followed by proactive adaptation of our smart environment. As long as we lack the legal and technological infrastructure to counter the emerging asymmetry we may find ourselves in a golden cage: an environment that anticipates our preferences before we become aware of them. This article argues that we urgently need to develop transparency-enhancing tools to match the proactive dimension of our smart environments. This will require substantial cooperation between social scientists, computer engineers, lawyers and policy makers with a clear understanding of what is at stake in terms of democracy and the rule of law.

\textsuperscript{9} Cp. Gutwirth and De Hert (2005) about the fact that data protection legislation is mainly a transparency tool, while privacy is considered to be an opacity tool. My point is that the transparency aimed for by the present generation of data protection regimes concerns personal data, without taking note of the result of the processing of such data, being highly sophisticated group profiles. It is those we urgently need to make transparent.
10 Bibliography:


