2008

Defining Profiling. A New Type of Knowledge?

Mireille Hildebrandt

Chapter 2
Defining Profiling: A New Type of Knowledge?

Mireille Hildebrandt

In this first chapter a set of relevant distinctions will be made to explore old and new ways of profiling, making a first attempt to define the type of profiling that is the subject of this publication. The text explains how profiling or pattern recognition allows us to discriminate noise from information on the basis of the knowledge that is constructed, providing a sophisticated way of coping with the increasing abundance of data. The major distinctions discussed are between individual and group profiles (often combined in personalised profiling), between distributive and non-distributive group profiles and between construction and application of profiles. Having described automated profiling we will compare such machine profiling to organic and human profiling, which have been crucial competences for the survival of both human and non-human organisms. The most salient difference between organic and machine profiling may be the fact that as a citizen, consumer or employee we find ourselves in the position of being profiled, without access to the knowledge that is used to categorise and deal with us. This seems to impair our personal freedom, because we cannot adequately anticipate the actions of those that know about us what we may not know about ourselves.

2.1 Introduction

Profiling occurs in a diversity of contexts: from criminal investigation to marketing research, from mathematics to computer engineering, from healthcare applications for elderly people to genetic screening and preventive medicine, from forensic biometrics to immigration policy with regard to iris-scans, from supply chain management with the help of RFID-technologies to actuarial justice. Looking into these different domains it soon becomes clear that the term profiling is used here to refer to a set of technologies, which share at least one common characteristic: the use of algorithms or other techniques to create, discover or construct knowledge from huge sets of data. Automated profiling involves different *technologies* (hardware), such as RFID-tags, biometrics, sensors and computers as well as *techniques* (software), such as data cleansing, data aggregation and data mining. The technologies and techniques are integrated into profiling *practices* that allow both the construction

Vrije Universiteit Brussel, Erasmus University Rotterdam

M. Hildebrandt and S. Gutwirth (eds.), *Profiling the European Citizen: Cross-Disciplinary Perspectives.*

and the application of profiles. These profiles are used to make decisions, sometimes even without human intervention. The vision of Ambient Intelligence or ubiquitous networked environments depends entirely on autonomic profiling, the type of profiling that allows machines to communicate with other machines and to take decisions without human intervention.

In this chapter we will start with the identification of profiling as such, providing working definitions of profiling and some related terms. After that we will discuss the difference between group profiling and personalised profiling and the way they are mixed up in practice. On the basis of this initial exploration of automated profiling, such technological (machine) profiling will be compared with non-technological forms of profiling, in particular organic and human profiling. This should enhance our understanding of the difference between machine and human profiling, which is crucial for an adequate assessment of the opportunities and risks involved.

2.2 Identification of Profiling

In this volume the focus will be on automated profiling, which is the result of a process of data mining. Data mining – which will be discussed in detail in chapters 2 and 3 - is a procedure by which large databases are mined by means of algorithms for patterns of correlations between data. These correlations indicate a relation between data, without establishing causes or reasons.[2] What they provide is a kind of prediction, based on past behaviour (of humans or nonhumans). In that sense profiling is an inductive way to generate knowledge; the correlations stand for a probability that things will turn out the same in the future. What they do not reveal is why this should be the case. In fact, profilers are not very interested in causes or reasons, their interest lies in a reliable prediction, to allow adequate decision making. For this reason profiling can best be understood from a pragmatic perspective: it aims for knowledge that is defined by its effects, not for conceptual elaboration.[3] Another way to articulate the particular kind of knowledge produced by profiling is to see profiles as hypotheses. Interestingly, these hypotheses are not necessarily developed within the framework of a theory or on the basis of a common sense expectation. Instead, the hypothesis often *emerges* in the process of data mining, a change in perspective that is sometimes referred to as a discovery-driven approach,

[2] Correlations can of course be spurious (see http://www.burns.com/wcbspurcorl.htm), however, this does not mean that non-spurious correlations necessarily establish causal or motivational relationships between data.

[3] According to the founding father of American pragmatism, Charles Saunders Peirce, the Maxim of Pragmatism reads as follows: 'Consider what effects that might conceivably have practical bearings we conceive the object of our conception to have: then, our conception of those effects is the whole of our conception of the object' (Peirce, 1997:111). A pragmatic approach of knowledge should not be conflated with a populist or naïvely 'practical' attitudes to knowledge.

as opposed to the more traditional assumption-driven approach.[4] 'Data mining provides its users with answers to questions they did not know to ask' (Zarsky, 2002-2003:8). After correlations (hypotheses) have surfaced they are tested when the profiles are applied. This is why the construction and application of profiles are entangled in profiling practices (complementing the inductive process of generating profiles with the deductive process of testing them on new data).

Before supplying a working definition of profiling we need to define three terms, which are central in the context of profiling: data subject, subject and data controller. The central position in profiling is taken by what is called the *data subject*, which we define as the subject (human or non-human, individual or group) that a profile refers to. In the case of group profiling this means that the data subject may be the result of profiling, not necessarily pre-existing as a group that thinks of itself as a group. For instance, a category of blue-eyed women may emerge as a data subject, because as a category they correlate with a specific probability to suffer from breast cancer. This implies that we use the term data subject in a different way than is usual in data protection legislation, as in the case when the data subject is defined as 'an identified or identifiable natural person'.[5] We define a *subject* as the human or nonhuman individual of which data are recorded which are used to generate profiles and/or as the human or nonhuman individual to which a profile is applied.[6] The next – equally central – position is taken by the *data controller* (sometimes called *data user*), which we define as the subject (person or organisation) that determines the purposes of the processing of the data and the use that will be made of them (including the sale of data or of the profiles inferred from them).

A simple working definition of profiling could be:

> The process of 'discovering' correlations between data in databases that can be used to identify and represent a human or nonhuman subject (individual or group) and/or the application of profiles (sets of correlated data) to individuate and represent a subject or to identify a subject as a member of a group or category.

To understand the meaning of profiling, it may be helpful to add the purpose of profiling. Besides individuation, profiling mainly aims for risk-assessment and/or assessment of opportunities of individual subjects. This, however, cannot be taken for granted. If the interests of the data controller and subject differ it may well be that the interests of the data controller, who pays for the whole process, will take precedence. Thus – in the end – what counts are the risks and opportunities for the data controller. For this reason the purpose of profiling can best be formulated as:

---

[4] Custers (2004: 46), referring to B. Cogan, *Data Mining; Dig Deep for the Power of Knowledge*. Internet publication at www.BioinformaticsWorld.info/feature 3a.html.

[5] Par. 2 (a) Directive 95/46 European Community (D 95/46/EC).

[6] This means that in this context the term data subject is used in a different way compared to the way it is used in Data Protection legislation. What is called a subject in this text, is called a data subject in D 95/46/EC, meaning the subject whose data have been processed.

The assessment of risks and/or opportunities for the data controller (in relation to risks and opportunities concerning the individual subject).

This raises the question whether it is possible to empower a human subject to make her a data controller in her own right, with regard to profiles that can be inferred from her data and profiles that may be applied to her.

2.3 Group Profiling & Personalised Profiling

2.3.1 Groups: Communities and Categories

Profiling techniques generate correlations between data. For instance, a correlation may be found between people that are left-handed and have blue eyes and a specific disease or a correlation may be found between people that live in a certain neighbourhood and have a particular level of income or a correlation may be found between one's individual keystroke behaviour and regular visits to a specific type of pornographic website. To generate such correlations in a reliable way we need to collect, aggregate and store the relevant data over an extended period of time, perhaps by integrating different databases that contain such data. In the examples given, the correlations concern data of certain categories of subjects, for instance the category of people with blue eyes that are left-handed or the category of people that live in a certain neighbourhood. Once the process of data mining establishes the correlations, two interrelated things happen: (1) a certain category is constituted (2) as having certain attributes. The category is usually called a group and the set of attributes are called the group's profile.

Another possibility is that the data of an existing group of people, who form some kind of community, are collected, aggregated, stored and processed in order to find shared features. For instance, the members of a local church or the students living in a certain dormitory can be the target of profiling. In this case the process of data mining will not establish them as a group (which they already were) but it may generate correlations and certain attributes between them, such as a typical way of dressing, particular eating habits or specific travel habits.[7]

Group profiling can concern both communities (existing groups) and categories (e.g., all people with blue eyes). In the case of categories, the members of the group did not necessarily form a community when the process was initiated; in the case of communities the members of the group already formed a community (however unstructured). The fact that profiling may establish categories as sharing certain attributes may in fact lead to community building, if the members of such a category become aware of the profile they share. The fact that data controllers may target the members of a category in a certain way – without them being aware of this – may of course impact their behaviour as members of this category.

[7] Cp. Zarsky (2002-2003:9-15) on clustering and association rules.

2.3.2 Distributive and Non-distributive Profiles

To understand some of the implications of group profiling we have to discriminate between distributive and non-distributive profiles. A distributive profile identifies a group of which all members share all the attributes of the group's profile. This means that the group profile can be applied without any problem to a member of the group – in that sense it is also a personal profile. An example of a distributive profile is the category of bachelors that all share the attribute of not being married. A less tautological example is the category of oak trees that all develop a certain type of leaf. Being a member of a group with a distributive profile has potentially pervasive social and legal implications because the profile will apply without qualification to all members.

It should be obvious that apart from groups that are *defined* in terms of a shared attribute (e.g., the group of bachelors that share the attribute of not being married), most groups do not have distributive profiles. A non-distributive profile identifies a group of which not all members share all the attributes of the group's profile.[8] For instance, Hare's checklist for psychopaths is a non-distributive profile. It contains 20 items e.g., absence of guilt, superficial charm, pathological lying and poor aggression control, that have to be checked and scored on a 3-point scale (0: does not apply; 1: applies to some extent; 2: applies). A person whose profile counts 30 points or more is considered a psychopath, a profile that is said to be – statistically – predictive of violent criminal recidivism of released offenders. The category of persons who score 30 points or more on Hare's checklist has a non-distributive profile, because not every person in this group shares the same attributes. From a social and legal perspective this is very instructive, because it implies that one cannot apply the profile to members of the group without qualification (Edens, 2001).

It is important to realise that treating members of a group that has a non-distributive profiles as fitting the entire profile may have interesting effects. For instance, if people fit the profile of a high-income market segment, service providers may decide to offer them certain goods or provide access to certain services, which may reinforce their fit in the category. If the group profile is non-distributive and they in fact do not share the relevant attributes (e.g., they may live in a certain neighbourhood that is profiled as high-income, while in fact they have a very low income, being an au pair), they may actually be 'normalised' into the behaviour profiled as characteristic for this group.[9]

2.3.3 Actuarial Approach in Marketing, Insurance and Justice

The use of non-distributive profiles (the usual case) implies that profiles are always probabilistic. They basically describe the chance that a certain correlation will occur

[8] In terms of Wittgenstein, the members of the group have a family resemblance, they cannot be identified by means of a common denominator. Cp. Custers, 2004; Vedder, 1999.

[9] Cp. Vedder, 1999.

in the future, on the basis of its occurrence in the past. As indicated above, the correlation does not imply a causal or motivational relationship between the correlated data, they merely indicate the fact that the occurrence of one will probably coincide with the occurrence of the other. For instance, in genetic profiling, we may find that the presence of a certain gene correlates with a certain disease. Depending on the exact correlation (the percentage of cases in which it occurs) we may predict – in terms of probability – the chance that a person with the relevant gene will develop the relevant disease. In reality, of course, the correlations may be very complex, e. g., depending on a whole set of different factors in a non-linear way. The exponential increase in computer power, however, enables the storage of a nearly unlimited amount of data and allows computer scientists to develop very complex algorithms to mine these data.[10] This has led to relatively new developments in marketing, insurance and justice, based on the targeted assessment of consumer preferences (leading to spam), targeted risk-assessment (concerning financial credibility) and on criminal profiling (leading to actuarial justice). In all three fields it becomes possible to take decisions (in customer relationship management, on refined types of price-discriminations, on categorised or even personalised interest-rates and insurance premiums, on targets for criminal investigation and on sentencing modalities) that are based on highly informed predictions of future behaviour. This approach to customers and citizens can be termed an actuarial approach, because it builds on highly sophisticated assessments of the risks and opportunities involved. The caveat of this approach is that it extrapolates from the past to the future on the basis of blind correlations, tending to see the future as determined by established probabilities, possibly disabling potentially better solutions that lie in the realm of low probabilities.

2.3.4 Personalisation and Ambient Intelligence

Mining the data from a variety of people allows categorising them into different types of groups, generating high rates of predictability concerning the behaviour of categories of people. Apart from group profiling, however, a second type of profiling has evolved, that mines the data of one individuated subject.[11] Behavioural biometrics is a good example of such profiling (discussed in detail in chapter 5).

[10] Data mining by means of algorithms or heuristics (see chapter 4) works with a set of instructions that has to be followed chronologically, this is called conventional computing. The end result of such a process is entirely predictable, even if our brains do not have the computing powers to apply the algorithm in as little time as the computer does. According to Stergiou and Siganos (1996), data mining by means of neural networks works with 'highly interconnected processing elements (neurones) working in parallel to solve a specific problem'. They claim that one of the advantages of this emerging technology is that problems that we do not understand can still be solved, while for the same reason the resolution of the problem is not predictable.

[11] In terms of the reply of Jaquet-Chiffelle this would be 'direct individual profiling'.

For instance, profiling the keystroke behaviour of one particular person may enable a service provider to 'recognise' this person as she goes online because of her behavioural biometric 'signature' and allows the service provider to check her online behaviour (discussed in detail in chapter 9) and thus to build up a very personal profile that can be used to offer specific goods and provide access to certain services. The profile can also be stored and sold to other interested parties, or be requested by the criminal justice or immigration authorities.

Such personalised profiling is the *conditio sine qua non* of Ambient Intelligence (AmI), the vision of a networked environment that monitors its users and adapts its services in real time, permanently learning to anticipate the user's preferences in order to adapt to them.[12] Ambient Intelligence presumes an RFID-tagged environment, and/or an environment enhanced with sensors and/or biometric appliances, all connected with online databases and software that allows a continuous process of real-time profiling. The intelligence is not situated in one device but emerges in their interconnections. The online world with its seemingly limitless capability to collect, aggregate, store and mine behavioural data thus integrates the offline world, creating a new blend of virtual and physical reality (ITU, 2005).[13] AmI environments may know your preferences long before you become aware of them and adapt themselves in order to meet those preferences. The AmI vision promises a paradise of user-centric environments, providing a continuous flow of customised services by means of ubiquitous and pervasive computing.[14] However, one does not need an overdose of imagination to foresee that such highly personalised profiling engenders unprecedented risks for users to be manipulated into certain preferences, especially in the case that users have no feed-back on what happens to the data they 'leak' while moving around in their animated environments.

2.4 Automated and Non-automated Profiling

2.4.1 Categorisation, Stereotyping and Profiling

Long before computers made their way into everyday life, criminal investigators composed profiles of unknown suspects, psychologists compiled profiles of people with specific personality disorders,[15] marketing managers made profiles of different

[12] As elaborated in chapter 6, personalised profiling can be a combination of individual and group profiling, or in terms of the reply of Jaquet-Chiffelle 'direct and indirect individual profiling'.

[13] See Mark Weiser's pioneering work on ubiquitous computing, for example, Weiser (1991: 94 – 104).

[14] The AmI vision has been propagated mainly by Philips and the European Commission, see for example, Aarts and Marzano 2003; ISTAG (Information Society Technology Advisory Group), 2001.

[15] On psychometrics (psychological testing), see for example, Rasch, 1980; Thorndike, 1971.

types of potential customers and managers profiled the potentials of their employees for specific jobs.[16] Adequate profiling seems to have been a crucial competence of professional occupation and business enterprise since their inception, perhaps most visible today in marketing and criminal investigation.[17] However, profiling is not just a professional, business or government preoccupation. As Schauer (2003) convincingly demonstrates in his *Profiles, Probabilities and Stereotypes*, profiling is a form of generalisation or categorisation we all apply routinely to get us through life. Habermas would probably speak of *Kontingenzbewältigung*, being the reduction of complexity in an environment that demands continuous choices of action, which would swamp us if we were to reflect on each of them. Schauer professes that categorisation is mainly a good thing, especially if it is based on a 'sound statistical basis' and his position has a strong appeal to our common sense. How could we move on in life if we did not take certain generalisations for granted, if we did not live by certain rules that are based on such generalisation – even if they do not always apply? Schauer warns against attempts to look at each and every case in isolation, attending to the particular instead of the general, glorifying what lawyers in Germany once called 'Einzelfallgerechtigkeit'. In his opinion routine assessments on the basis of generalisation are not only necessary to cope with complexity and multiplicity but they also provide *just* instead of *arbitrary* decisions, because of the appeal to a general standard, which creates a type of predictability (essential for e.g., legal certainty). In psychology the need to reduce the weight of recurring decisions is thought to be the cause of 'stereotyping', a healthy way to deal with the growing complexities of life. It means that we – unconsciously - group different events, things or persons into categories in order to assess what can be expected and to be able to decide how to act. Stereotyping allows anticipation. Following this line of thinking, categorisation and stereotyping are a kind of everyday profiling, based on experience and practical wisdom and if we believe Schauer, it also produces a kind of justice. In the next section I will take this line of thought one step further in claiming that profiling is not only a part of professional and everyday life but also a constitutive competence of life itself in the biological sense of the word.

However, before describing the process of profiling from the perspective of the life sciences, we need to make some comments on Schauer's defence of categorisation and its relationship to profiling. In the introduction to his book he discusses 'Generalization Good and Bad'. He starts by drawing a distinction between generalisations with and without a statistical or factual basis; those without he calls *spurious*. He qualifies this distinction by indicating that in everyday life we may pronounce many generalisations without intending them to be taken as absolute. For instance, when

[16] See for example, Rafter and Smyth, 2001.

[17] For a history and overview of criminal profiling, see for example Turvey, 1999. For a history and overview of data mining in academic marketing research see for example Wilkie, William and Moore (2003: 116-146). For more practical research see for example, Peppers and Rogers, 1993 (about the integration of data mining and CRM (Customer Relationship Management) to achieve mass customisation.

we say that 'Bulldogs have bad hips', this – according to Schauer – may be a good generalisation, even though a majority of bulldogs do not have bad hips. 'As long as the probability of a dog's having hip problems given that the dog is a bulldog is greater than the probability of a dog's having hip problems given no information about the breed of dog, we can say that the trait of being a bulldog is *relevant*, and we can say that generalizing from that trait meets the threshold of statistical (or actuarial) soundness' (Schauer 2003: 11).[18] Thus we have what he calls universal generalisations, which denote a group of which all members share the generalised characteristic and non-universal generalisations, which denote a group of which a majority or a relevant minority share the generalised characteristic. Schauer then moves on to discuss prejudice or stereotype as a kind of generalisation, recognising that these terms are often used in a pejorative way. He seems to conclude that the use of a non-universal generalisation must not be rejected, while admitting that - depending on the context (sic!) – sometimes such prejudice or stereotyping can indeed be morally flawed. One example he gives is the case of racial profiling, though he seems to suggest that in this case the generalisation is not based on sound statistical or empirical evidence. The reason for the fact that acting on a nonspurious non-universal generalisation may – under certain circumstances - be morally wrong is that 'equality becomes important precisely because it treats unlike cases alike' (Schauer, 2003: 296). So, even if most ex-convicts or a relevant minority of them, are prone to commit crimes again, we may decide we want to treat them equally when they apply for a job, insurance or try to rent a house - equally to non-ex-convicts. A principle such as the presumption of innocence has the same function: even if we are quite sure that a person has committed a certain crime, government officials cannot treat this person as an offender until guilt has been proven according to law. I am not sure these are the examples Schauer would endorse to demonstrate the importance of the moral evaluation that may interfere with justified generalisation but he has explained in a clear voice how generalisation, equality and even community relate to each other. In the remainder of this chapter we can use some of the salient distinctions he makes to clarify the complexities of automated profiling and the implications it may have for fairness and equality.

2.4.2 Organic Profiling: A Critical Sign of Life

After concluding that profiling is part of professional as well as everyday life of human beings, I would like to make a brief excursion into the life sciences to highlight the importance of profiling for living organisms. As Van Brakel (1999) writes, biology and information theory have developed into an integrated domain, part of the life sciences. This is an important development, which may help us to understand the way automated (machine) profiling can generate knowledge, although not human knowledge.

[18] Emphasis of Schauer.

Both 'organic profiling' and automated machine profiling concerns the production of implicit knowledge, or at least knowledge that has not reached a human conscious.

In 1987 Maturana and Varela published a little book, *The Tree of Knowledge*, explaining *The Biological Roots of Human Understanding*.[19] For our purposes the theory of knowledge argued in their book is interesting because it explains knowledge as something that an observer attributes to *an organism that effectively deals with its environment*. For Maturana and Varela knowledge is constituted by the interactions between – for instance - a fly and its immediate environment, if this interaction is successful in the sense that it sustains the life of the fly. Their understanding of knowledge is enactive (knowledge and action 'cause' each other): only by acting, an organism finds out about its environment and in that sense even perception is a form of – entirely implicit - action.[20] To be more precise one could say that all living organisms, in order to survive, must continuously profile their environment to be able to adapt themselves and/or to adapt the environment. Profiling in this case means the process of extracting relevant information from the environment. However, *what counts as information depends on the knowledge the organism has built* on the basis of continuous interaction with its environment, because this knowledge determines what type of information is relevant and valid. This means that what counts as information at one point in time may be noise at another point in time and what counts as noise for one individual (organism) may be information for another. It also means that knowledge depends on both the environment and the organism and must be understood as fundamentally dynamic and context-dependent. Knowledge in this sense is always local knowledge. This does not mean that generalisation is out of bounds, quite on the contrary. To be able to act in an environment adequate generalisation is *necessary* but the question of which generalisation is *adequate* will depend on the context (and on the organism).

What is crucial at this point is that (1) profiling the environment happens without involving a conscious mind (2) profiling provides feed-back necessary to survive (3) profiling extracts information, depending on knowledge that allows one to discriminate between noise and information (4) profiling transforms information into knowledge and (5) information and knowledge always depend on both the organism and its environment, there is no view from anywhere.

2.4.3 Human Profiling: The Meaning of Autonomous Action

The small excursion into profiling by nonhuman organisms allows us to develop a keener eye for what makes knowledge *human* knowledge. If perception, information

[19] Revised edition of 1998. Matuna and Varela coined the term autopeiosis in 1973 to describe the process that constitutes living organisms. In *The Tree of Knowledge* they expound on their theory of biology by investigation the relationship between living organisms and their environment.

[20] Their theory of knowledge thus combines pragmatism and embodied phenomenology, rejecting both mentalism or naïve empiricism.

gathering, feed-back and even knowledge are not specific for the human animal, what is? Could it be that consciousness is the discriminating attribute, and if so, what difference does this make for profiling? Compared to a plant, a dog seems to have a different kind of awareness of the world. We may be inclined to call this awareness a consciousness. This is not the case because the dog is aware of being aware but because it seems to embody a unified self that is absent in a plant. The philosopher Helmuth Plessner (1975) described the difference by pointing out that all mammals have a central nervous system that seems to allow for a centralisation of the awareness, giving rise to a conscious presence in the world. The difference between humans and other mammals, according to Plessner, is the fact that a human is also conscious of being conscious, conscious of herself. This reflective attribute, which is often thought to derive from the fact that we use language to communicate with each other, is absent in other mammals or present to a different degree.

To assess why this difference is relevant for our study of profiling we need to connect our capacity for reflection with our capacity for intentional action (which we suppose to be less evident in other mammals).[21] Reflection implies that we can look back upon ourselves, which also implies that we can consider our actions as our actions, as it were, from a distance. Such reflection can be incorporated into our actions – even before we act. We may thus consciously reflect upon different courses of action and intentionally prefer one alternative to another. This is what allows intentional action and this seems to be the pre-condition for autonomous action: an action we have freely decided upon, an action within our own control. Auto is Greek for self, nomos is Greek for law so human autonomy implies intentional action and conscious reflection, two conditions for positive freedom.

Before moving on to the relevance of intentional action and conscious reflection for profiling, we need to keep in mind an important fact. *Most of our actions are neither intentional nor conscious*. We can move around freely in this world because we have acquired habits that are inscribed in our bodies, allowing us to act in a number of ways without giving it thought. However, the very small amount of actions we actually consciously intend, are distinctive for our moral competence – taking into account that conscious reflection is the incentive to create new habits which will again move out from the zone of intentional action, but did originate from it.

2.4.4 Machine Profiling: The Meaning of Autonomic Machine Behaviour

In 2001, Paul Horn, IBM's Senior Vice President, introduced the idea of *autonomic computing*. Interestingly, he chose a term that refers to biology, to the autonomic

[21] We cannot be too presumptious here, see for example de Waal, 2001.

nervous system, because it 'governs our heart rate and body temperature, thus freeing our conscious brain from the burden of dealing with these and many other low-level, yet vital, functions' (Kephart and Chess, 2003: 41). One of the objectives of autonomic computing is to prevent or resolve the advancing software complexity crisis, by creating a network that is capable of self-management: self-configuring; self-healing; self-optimising; self-protecting (CHOP). Visions of Ambient Intelligence (AmI), pervasive computing or the RFID *Internet of Things* depend on extended interconnectivity and we are being warned that without self-management the design of the integrated network architectures will become entirely impossible. Another objective is to allow the user of the system to collect the fruits of ubiquitous computing without being bothered with the flow of minor and major adjustments that need to be made to keep the system operational. Kephart and Chess (2003: 42) distinguish different stages in the development of autonomic systems, starting with automated functions that collect and aggregate data and ending with automation technologies that can move beyond advice on decision-making, taking a large amount of low-level and even high-level decisions out of human hands.

To target the difference between organic, human and machine profiling it is interesting to discuss automated profiling in terms of *autonomic machine behaviour*. With autonomic machine behaviour I mean the behaviour of machines that are part of a network of machines that exchange data and make decisions after processing the data. This need not incorporate the entire concept of autonomic computing with its CHOP attributes, but is based on what is called M2M talk (machine to machine communication) (Lawton, 2004:12-15). 'Machine' can be anything, such as a RFID-tag (radio frequency identification tag), a PDA (personal digital assistant) or a PC (personal computer). I call the behaviour autonomic in as far as the network of machines processes data, constructs knowledge and makes decisions without the intervention of a human consciousness. This autonomic machine behaviour will be part and parcel of ambient intelligent environments, which monitor subjects and adapt the environment in real time, necessitating autonomic machine decision making.

The most simple form of automated profiling is when profiles are generated and applied in the process of data mining, after which human experts sit down to filter the results before making decisions. In this case we have no autonomic machine behaviour, because decisions are taken by human intervention. It may, however, be the case that these decisions routinely follow the machine's 'advice', bringing the whole process very close to autonomic machine profiling.[22]

[22] Art. 15 of the Directive on Data Protection 95/46/EC attributes a 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, such as his performance at work, creditworthiness, reliability, conduct, etc.'. The fact that usually some form of routine human intervention is involved means that art. 15 is not applicable, even if such routine decisions may have the same result as entirely automated decision making.

2.4.5 Organic, Human and Machine Profiling: Autonomic and Autonomous Profiling

After discussing organic, human and machine profiling we can now draw some prudent conclusions. It seems that most organic profiling does not involve conscious reflection or intentional action. It is important to note that an important part of human existence itself is sustained by the autonomic nervous system, which continuously profiles its environment inside and outside the body, by means of operations that we are not aware of. On top of that human profiling is done 'automatically' to a very large extent. This automation or habit-formation is the result of a learning process that often starts with conscious attention shifting to implicit behaviour as soon as the habit is inscribed in our way of doing things. The competence to act on this basis is referred to as implicit or tacit knowledge (Polanyi, 1966). Machine profiling seems similar to organic profiling, in the sense that it does not involve conscious reflection, nor intentional action. However, organic profiling presumes an organic system that constitutes and sustains itself. Maturana and Varela (1991) have coined the term autopoiesis for this self-constitution.[23] Even if autonomic computing – as defined by IBM – can be successfully compared to the autonomous nervous system, we may have a problem in defining it as self-constituting as long as it needs an initial software architecture provided by human intervention.

In other words, machine profiling is like organic profiling to the extent that it is part of autonomic behaviour and like human profiling to the extent that human profiling is done implicitly. At the same time, machine profiling differs from human profiling in two salient ways: (1) other than human and organic profiling machine profiling is not part of an autopeiotic system that constitutes itself, (2) other than human profiling machine profiling does not integrate conscious reflection or intentional action.

2.5 Conclusions: From Noise to Information, From Information to Knowledge

As they say, we live in an information society and in a knowledge society. One of the challenges of the present age is how to deal with the overload of information, or rather, how to discriminate noise from information. Another challenge is how to

[23] The term has been introduced into sociology by Luhmann and Teubner, who also build on Heinz von Foerster (cybernetics), implying that not only individual cells or metacellular organisms form autopoietic systems, but also social systems. However, system theory and other sociological schools that claim that individuals are determined by the social system or the underlying structure do not seem to build on Maturana and Varela, who explicitly claim that *social* systems amplify the individual creativity of its components, arguing that the social system actually exists for these components (and not the other way round, as is the case in metacellular organisms, cf. Maturana and Varela (1998:199).

(re)construct knowledge out of the flows of noise and information, how to deal with the growing complexities of our scientific knowledge constructs and with the emerging unpredictability of the complex technological infrastructures built to face the increasing mobility of human and nonhuman imbroglios.

One of the answers to both questions is the use and further development of profiling technologies. They may incorporate the only way to reduce the overload of information, to make it 'manageable', to make sense out of it and to regain control of the effects of one's actions. In other words, they may provide the only way to adequately anticipate the consequences of alternative courses of action. If freedom presumes anything, it is precisely this: a reliable anticipation of the results of the choices we have. This is why legal certainty and scientific experiment create the freedom to act, allowing citizens to adapt their position in the world to the realities it contains. At this point in time scientific experiment already makes widespread use of profiling technologies and it may be the case that legal certainty will need profiling technologies to interpret the overload of legal cases and decisions[24] and to regain some control over the way one's personal data are put to use.[25]

The biggest challenge however may be how to constrain profiling practices in order to prevent the coming-of-age of a technological infrastructure that is entirely geared for dataveillance, normalisation and customisation – practically destroying the effectiveness of our rights to privacy, fairness and due process (Leenes and Koops, 2005: 329-340). It would be unwise to wait for such an infrastructure to be in place, before establishing constraints, as this may render effective restraint an illusion. In chapter 15 these issues will be discussed, in order to assess the implications of profiling for the identity of the European citizen.

2.6 Reply: Further Implications?

Thierry Nabeth*

In her essay, Mireille Hildebrandt raises the important issue of considering profiles as knowledge itself, and not as mere information. This implies a new way of considering profiles and profiling, as knowledge is subject to interpretation and meaning and inseparable from the rich social contexts in which it is embedded.

This perspective has some profound implications in the way the information society is going to extract, manipulate and exploit data of human beings, or shall we say knowledge and apply it to the design of new categories of applications and services. In the new information society, applications "know" the people and not the other way around.

* Institut Européen D'Administration Des Affaires (INSEAD)

[24] The sheer volume of case law that is published (online) would in the end destroy legal certainty, because no human individual would be able to find her way in the proliferating decisions.

[25] For example, by the use of private Identity Management Devices that enable tracking of one's data and can be used to restrict the leaking of personal data. To regain control written law in itself will not suffice; the right to hide certain data must be inscribed into the technologies that would otherwise threaten one's personal autonomy.

In our reply to Mireille Hildebrandt, we will further explore the implications of this shifting of conceptualisation of profiles from data to knowledge and the ensuing consequences of almost intimate understanding of people and groups. We will in particular try to understand in which cases profile-informed applications will be used to better serve people and groups, or will - on the contrary - be used to alienate them.

2.6.1 Introduction

The essay of Mireille Hildebrandt on the subject of "defining profiling" comes very much as a surprise but a pleasant one. One would initially expect a formal definition, some descriptions of algorithms, some indication of security issues and a series of illustrative examples, providing finally more of a description than an explanation of the concept of profiling. What she provides though is a much more profound attempt to understand the concept of "profiling" that borrows ideas from many different fields and areas such as philosophy, complexity, anthropology and cognition (theory of action). This perspective is particularly useful in providing readers with the conceptual tools that will help them to articulate the different parts of this volume: the description of profiling, including an in-depth discussion of algorithms and a first indication of risks in part I; a series of illustrative examples (applications) that were presented in part II and the wider implications for democracy and rule of law in part III.

2.6.2 Profiling as Knowledge

In the first part of her chapter, Mireille Hildebrandt engages the discussion on the nature of the knowledge generated by the profiling process. This "profile" knowledge originates from the automatic extraction from an important amount of information aiming at discovering patterns that will have some predictive capabilities. Indeed, the underlying assumption is that the function of profiling is to help to reveal some hidden "order of things" and therefore to provide an oracle that will predict how people will behave in the future. Indeed, if they have behaved in a certain way in the past, they will most probably behave the very same way in the future. Mireille Hildebrandt also very rightly points to one of the main limitations of this form of knowledge: profiling "knowledge" does not explain things, as it is of a more inductive nature. It is then suggested that profiling should therefore be complemented by well thought out profiling practices.

Mireille Hildebrandt then indicates what the different stakeholders of profiling are: the subject (human or nonhuman, group or category) to which a profile refers (referred to as the data subject), the entity of which data are used to generate profiles and to which a profile is applied (referred to as the subject) and the actor that is the initiator of the profiling and the exploiter of the profiling data (referred to as the data controller). This distinction is very important, since it raises the question of different actors that may have conflicting objectives.

If Mireille Hildebrandt presents profiling at the individual level via the concept of personalised profiling, we have to admit that profiling at the group level receives a much higher level of attention in this chapter. Indeed, even if she acknowledges the importance of personalised profiling, she does it principally for an application in an Ambient Intelligent (AmI) context. We personally believe that it would have been useful to generalise the reflection to a much broader context, such as the domain of e-learning or e-commerce to cite a few.[26] At the group level, one should distinguish between characteristics that belong to all the members of a group (referred to as distributive attributes) and from characteristics that only statistically belong to this group (referred to as non distributive attributes). It is of particular importance to identify the non-distributive nature of the knowledge (also known as non-monotonic logic[27] in artificial intelligence), since it can be at the origin of errors in segregating people due to the merely probabilistic nature of some characteristics.

Moreover, we should point out the danger of the segregating function of profiling. Even in the case where the characteristic is distributive and no error is made, how should we deal with using profiling, which usage can be directly associated to segregation? The answer to this question follows an interesting angle: profiling did not have to wait for the advent of the computer to appear, since society, and for instance the social process, can be considered as a big *profiling machine*. Societies use categorisation and generalisation to function better, since it allows anticipation. To answer the question of whether generalisation is a good or bad thing, we will follow the reasoning of Mireille Hildebrandt by excluding generalisations that do not have a statistical or a factual basis. We also agree with the idea that some ethical issues may apply, depending on the context and for instance, taking the example of the "presumption of innocence", the role of society should help to erase the inequality that originates from circumstance. Even if it is proved that someone who only has one parent is more likely to become a criminal, such knowledge should not be used as a tool to segregate this category of persons; for instance by reducing the level of protection provided by the presumption of innocence or by removing some of their rights.

2.6.3 A Knowledge Ecology Perspective

The second part of this chapter is very interesting since it situates profiling according to a systemic and knowledge ecology perspective (the term organic profiling is used in this chapter). This part in particular relates to all the theories of complexity

[26] See on this chapter 10.

[27] See http://plato.stanford.edu/entries/logic-nonmonotonic/ for a description of the non-monitonic logic concept.

and collective intelligence that have emerged during the last decades, e.g., with the work of Varela and others (those involved in the Santa Fe movement). Applied to the context of ambient intelligence, it draws a vision that is not far from the 'Universe' of the great Sci-Fi author Philip K. Dick, for which the separation between the real word and the virtual word tends to blur. In particular, with the advent of RFID and other similar devices, this vision proposes to dissolve the frontier between human and machine and in our belief, introduces the concept of the trans-humanity that will merge the human and the machine. Indeed, RFID represents the typical device helping to create the bridge between the physical and the digital world (with RFID, the virtual world has access to "sensors" relating what happens in the physical world).

The consequences of this vision of seeing the world as a system closely integrated in society rather than as a well identified "machine" are many. First, and as previously indicated, the distinction between the physical world and the virtual world becomes artificial and should no longer be made, since we are talking about the same world. Second, profiling the environment does not necessary "involve a conscious mind" that is controlled by a central body (such as a government) but can also happen quasi spontaneously in society by a variety of actors. Let us also add that the existence of technology, even if it is not a mandatory condition for profiling to happen, can have tremendous consequences. For instance, the combination of autonomy and profiling can lead to the concept of autonomic profiling, for which the profiling processes do not need to have the "man in the loop" and as a consequence risk the loss of control by humanity. Even if we do not believe in the taking over of society by machines – a doom scenario popular in Sci-Fi literature - a real risk exists that people will lose the ability to control what is happening because of the complexity (in particular if machines gain the capabilities to learn and adapt). As a consequence, humanity may very well become dependant on profiling processes (typically in ambient intelligence) as is the case with an addiction: being aware of the dangers but having no capabilities to act.

However, we do not believe that the consequences need to be apocalyptic. The systemic vision may also mean a shift from the idea of very "controlled" profiling systems counceived by engineers, to systems that are more "self-regulated" and for which the designers also include people from the social sciences or law field who know how to deal with less mechanical and less deterministic approaches. In this later case, engineering systems involving profiling would mean working on the level of the different feedback loops helping to regulate the systems that continuously evolve (and deciding which ones are acceptable), rather than very supervised systems in which profiling represents one of the critical parts of an effective mechanism. To conclude, it would be the responsibility of this new category of designers, able to reason in a more holistic way, to ensure that the profiling mechanisms are put in place to service the good of society and individuals (for instance profiling can enable better personalisation, or can help to reduce inequalities by exposing them), rather than a tool of which the role is only to enforce social control (and typically used to maintain people in their initial condition).

2.6.4 What to Conclude About This Chapter From Mireille Hildebrandt?

We feel there exists a risk, every time we enter into an epistemological, philosophical or complex discussion, to detach too much from reality. Many discourses about the nature of knowledge and complexity easily become very abstract and tend to lose their readers in abstractions with little possibility to apply in reality.However, in this case Mireille Hildebrandt was able to avoid this trap by providing an illustration of what may be the concrete consequences for reality, for instance when applying it to ambient intelligence environments. Her chapter is therefore successful in proposing a global picture of how to conceptualise profiling at a higher level without losing the ground of reality.

If we could add something to this chapter, it would probably consist first in incorporating research linking cognition and profiling and second in investigating the consequences and impact of technology on the evolution of "profiled" environments. In the first case we will refer to the work that we will call instant cognition, consisting in the unconscious perception / classification / generalisation that people perform in their everyday life, which leads to very effective results but is also subject to bias and is at the origin of many dysfunctions in society, such as racism without real intention (the reader is invited to read the book from Malcolm Gladwell for information on this subject).[28] In particular, it could be interesting to explore how the new profiling approaches can be used to counterbalance the biases we have indicated (making them more visible). In the second case, it would be interesting to investigate very futuristic scenarios exploring the limits of extreme profiling. For instance, what would be the consequences of very "efficient" profiling done by the society? The movie industry has already given us some food for the thought with movies such as *Gattaca*[29] or the *Minority report*[30] that explore the less positive consequences of profiling but we do not doubt that similar work can be conducted exploring the more positive side, such as improving the effectiveness of education or work via a better personalisation that profiling would authorise.

2.7 Reply: Direct and Indirect Profiling in the Light of Virtual Persons

David-Olivier Jaquet-Chiffelle*

In our reply, we elaborate the difference between individual and group profiling in a slightly different manner, by distinguishing and defining direct and indirect profiling. We study these two types of profiling in the light of virtual persons.

* VIP, Berne University of Applied Sciences and ESC, University of Lausanne, Switzerland

[28] Gladwell, M. 2005.

[29] Gattaca is a 1997 science fiction drama film that describes the vision of a society driven by liberal eugenics.

[30] Minority report describes a society able to predict the crimes before they happen and imprison people for crimes that they have not yet committed but intended to.

In direct profiling, data is typically collected for one single subject or a small group of subjects. Knowledge built on this data then only applies to this specific subject or this small group of subjects.

Direct profiling can be used to uniquely characterise a person within a population or to infer, for example, future behaviour, needs or habits of a specific target.

In indirect profiling, data is collected from a large population. Groups and categories of subjects with similar properties emerge from the collected data. Each group has its own identity defined through a small amount of information. The typical member of one group can be modelled using the concept of virtual persons. It is then sufficient to identify a subject as a member of the group, i.e., with the corresponding virtual person to be able to infer, for this subject, knowledge inherited from the group itself: probable behaviour, attributes, risks, etc.

2.7.1 Introduction

In this chapter, Mireille Hildebrandt presents three key concepts related to profiling, namely the *data subject*, the *subject* and the *data user*. First, we want to enlighten these concepts using the concept of virtual persons. Then we will refine the concepts of individual and group profiling by distinguishing direct and indirect profiling. The different types of profiling will be illustrated using the generic model based on the concept of virtual persons.

2.7.2 Individual and Group Profiling

Individual profiling is used either to identify an individual within a community or just to infer its habits, behaviour, preferences, knowledge, risks, potential or other social and economic characteristics. Forensic individual profiling, for example, covers both aspects. Commercial individual profiling on the other hand is more interested in the latter, the inference of knowledge or rules about the individual.

Group profiling is used either to find shared features between members of a predefined community or to define categories of individuals sharing some properties. Forensic group profiling could, for example, find common characteristics in the community of convicted murderers or define risk categories of individuals. More generally, group profiling often raises ethical issues as it can lead very quickly, for example, to discrimination.

Several techniques can be used together or separately to define a direct profile:

- Information collected about an individual may directly give some important attributes of his profile (age, gender, etc).
- Data mining techniques applied to the data collected about an individual may help to induce his habits, his preferences, etc.
- Data mining techniques also help to find correlations between large sets of data collected about groups of people. These correlations might allow in turn the creation of categories: for example individuals sharing some attributes, living downtown, earning more than 100,000 a year, etc. Profiles are defined by associating knowledge with each category.

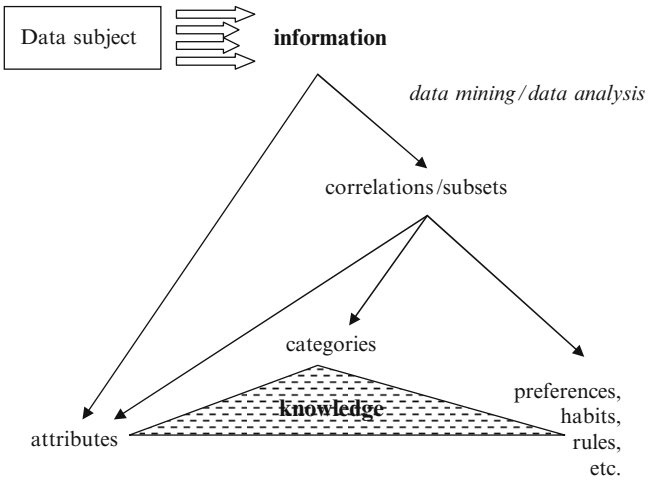

Fig. 2.1 From information to knowledge

Subsets are defined as elements sharing some properties. With each subset found in this process is associated its profile: attributes, rules, preferences, etc.

A *category* results from a process of generalisation. Each defined subset can be virtualised in a generalised subset or category, defined by the properties identifying the original subset; it inherits the profile of the original subset. The generalised subset may then exist independently of the original data subject.

2.7.3 Virtual Persons

We have elaborated the concept of *virtual persons* within the second work package of FIDIS *Identity of Identity*.[31] Virtual persons create an abstract layer allowing a more faithful description of many real-life scenarios appearing in our modern society. We want to apply this generic model to data mining and profiling, in particular to the data subject, the end user and the categories.

Virtual persons traditionally refer to characters in a MUD (Multi User Dungeon), MMORPG (Massively Multiplayer Online Role Playing Games) or other computer games.[32] These characters interact in a game; some of them rely on human beings (players) for their actions and/or behaviour, while others

[31] Jaquet–Chiffelle D.-O., Benoist, E., Anrig, B., Chapter 3 of Nabeth, T. et al. (eds), 2006a and Jaquet–Chiffelle D.-O., Benoist, E., Anrig B, in Jaquet – Chiffelle D.-O., Benoist, E., Anrig B. (eds.), 2006b: 6-7.

[32] http://dud.inf.tu-dresden.de/Anon_Terminology.html (Version 0.28; May 29, 2006).

might be directed by the game itself. For an external observer, it may be impossible to decide whether the subject behind a specific virtual person is a real player or just a computer programme. We see these virtual persons (characters) as masks used by subjects (human players, computer programmes) to act and/or interact within the game.

Laws also create a virtual world by associating rights, duties and/or responsibilities with virtual persons.[33] The one who is older than 18, the one who is married, the one who is president of a company, the person legally responsible… are typical examples of virtual persons living in the virtual legal world. These virtual persons are not linked to any physical or legal entity until the given conditions described in the law are fulfilled. Several physical or legal entities can be linked to these virtual persons as actions and/or transactions take place. Moreover, a single physical or legal entity may be linked to several virtual persons. These links are very often time dependant.

As an example, we consider *the person legally responsible* in a given transaction. The subject, i.e., the physical or legal entity behind this virtual person, could be the person executing the transaction himself; but it could be someone else, not necessarily visible: a tutor or the parents of a child – it could even be a company.

Such analogies between multiplayer games and real-life scenarios extend the field of application of virtual persons.

The concept of an abstract subject used by some authors is very close to our concept of virtual person. However, in using *virtual person*, we take advantage of the similarity between characters appearing in computer games and characters created in our daily life scenarios. Moreover, etymologically speaking, *person* comes from *personae* which means *mask*. Instead of adding a new theoretical term, we extend a well-known concept that is easy to imagine, even for non-specialists. Last but not least, in using two very distinct terms (subject and virtual persons), we avoid a possible confusion between both concepts and emphasise their differences: the virtual person is like a mask, the subject is the entity behind this mask.

In order to better understand the concept of virtual persons, we need a few core definitions. A *subject* is any physical or legal entity having – in a given context – some analogy with a physical person. Here subject is not opposed to object. Indeed, physical objects may satisfy our definition of a subject. In our definition, subjects typically act or play a role. Our subjects *are*, they *have*, they *do* or they *know* something just like physical persons.[34] Typical subjects are persons or groups of persons but can also be animals or computer programmes for example. A subject can be alive or not, can exist or not.

---

[33] In her article, Danièle Bourcier (2001: 847-871) introduces the concept of virtual persons in the context of artificial intelligence (intelligent programmes, software-agents, etc). She also refers to previous uses of this term in similar contexts. Our concept of virtual persons covers this approach while being more general and more adaptable to a wide variety of real-life scenarios.

[34] Our subjects look like the grammatical «subject» in a sentence as pointed out by Sarah Thatcher, London School of Economics, during the FIDIS WP2 workshop in Fontainebleau (December 2004).

A *virtual person* is a mask for a subject. In the context of individual authentication and/or identification, a *virtual person* is usually defined by what it *is* and/or what it *has* and/or what it *does* and/or what it *knows*.

The *one who knows your credit card PIN code* is a virtual person defined by what it knows. The subject behind this mask should be yourself and yourself only. However, the subject can be a group of persons (e.g., you gave the PIN code to other members of your family) or there might be no subject at all (e.g., you do not remember the PIN code).

More generally, a virtual person can also be defined by its *attribute(s)* and/or its *role(s)* and/or its *ability(-ies)* and/or its *acquisition(s)* and/or its *preference(s)* and/or its *habit(s)*, etc.

The *President of the United States* is a virtual person defined through its role; the subject linked to this virtual person might change after each Presidential election. However, rights, duties and responsibilities described in the law and associated with this role, i.e., with this virtual person, do not depend on who is elected.

Actually, a virtual person acquires its own existence, which is not necessarily correlated with the existence of any real subject behind it.

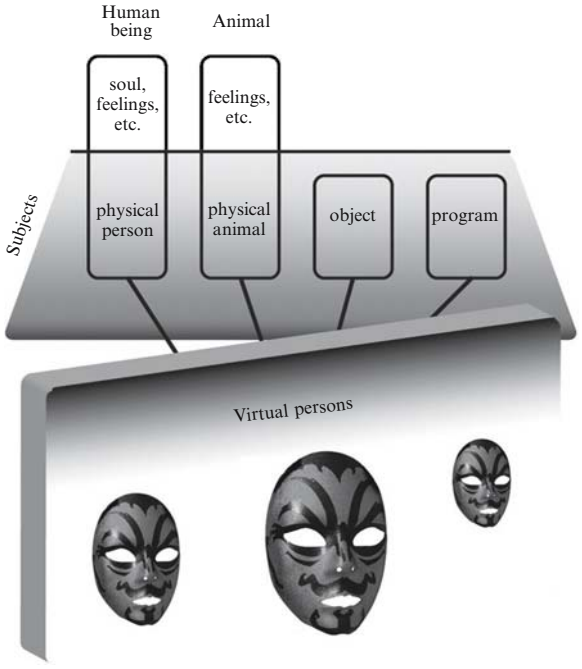

Fig. 2.2 Virtual persons

2.7.3.1 Virtual Persons Applied to Profiling

In her contribution, Mireille Hildebrandt gives the following definition for the *data subject*: "subject (human or non-human, individual or group) that data refer to." Her concept of *subject* is covered by our own definition of this term and is therefore compatible with our approach.

Using data mining techniques, subsets of elements sharing some properties can be defined. Virtual persons allow representation of the corresponding categories.

With each category, i.e., with each virtual person, is associated the inherited profile.

Indeed, with each virtual person is associated attributes, rules, preferences, etc. deduced from the correlations found via the data mining techniques: its *profile*. For example, people living downtown and earning more than 100,000 a year are likely to be more than 30 years old and not retired.

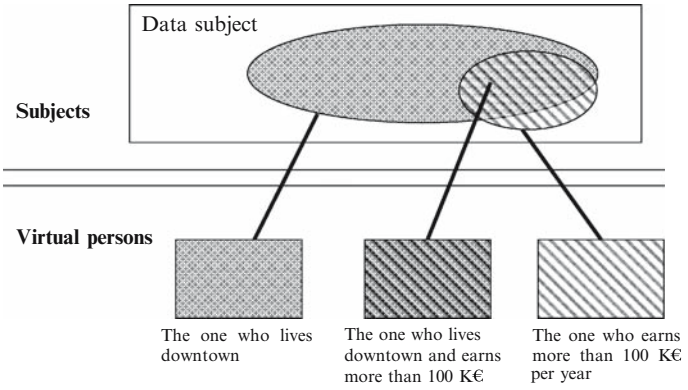

Fig. 2.3 Subsets of data subjects with their corresponding virtual persons (categories)

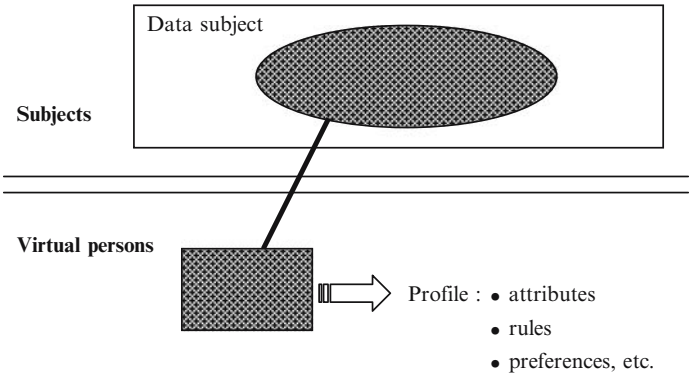

Fig. 2.4 Profile associated with a virtual person

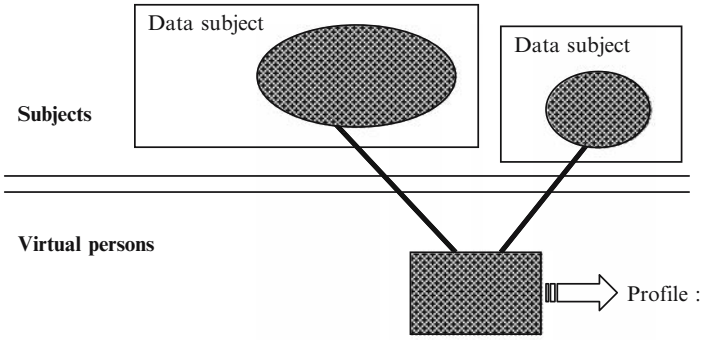

Fig. 2.5 Equivalent subsets defining the same category

At a later point, this knowledge may be used to infer probabilistic characteristics about what Mireille Hildebrandt calls the "end user" or the "profiled data subject."

Virtual persons acquire their own existence, which no longer depends on any specific, original subset of data subjects. Different data subjects can lead to equivalent subsets that define the same category, i.e., the same virtual person

2.7.4 Direct and Indirect Profiling

Profiling an end user consists in finding his profile by linking the end user to virtual persons.[35] Information gathered about the end user enables the data controller to find virtual persons linkable to this end user and to use the corresponding profiles for this end user. We consider that the classical distinction between individual and group profiling is not precise enough. We want to refine these concepts using direct and indirect profiling. *Direct profiling* occurs when the end user and the original data subject used to define the virtual person with its profile are the same. *Indirect profiling* aims at applying profiles deduced from other data subjects to an end user.

2.7.4.1 Direct Group Profiling

In the first chapter, Mireille Hildebrandt gives two examples of group profiling. In the case of a pre-existing community (members of a local church, students living in a certain dormitory) data are "collected, aggregated, stored and processed, in order

[35] In case of a direct profiling, it is essentially a direct construction of the profile. In case of an indirect profiling, it is the construction of the profile by applying a typical profile of a category.

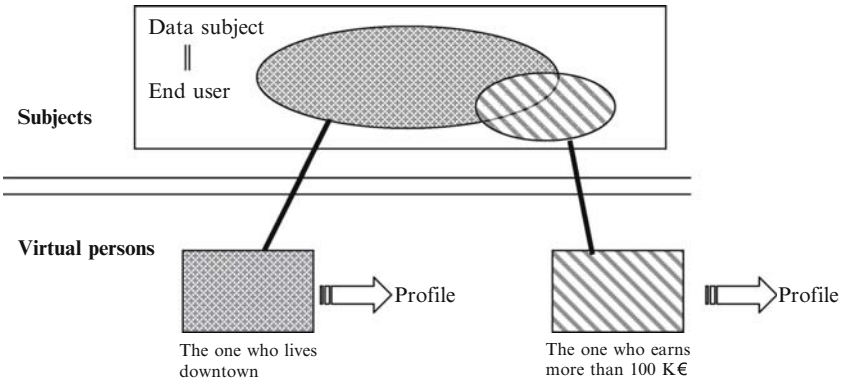

Fig. 2.6 Direct group profiling

to find shared features". Knowledge about this community (data subject) is established as the profile of the virtual person defined by this group.

When the end user is later the community itself, we have a typical example of a *direct group profiling*.

2.7.4.2 Indirect Group Profiling

Another example of group profiling given by Mireille Hildebrandt explains how data mining techniques find subsets of individuals in the group, who share certain attributes. This case illustrates the analogy between group profiling and the natural process of categorisation and generalisation of Schauer also described in the first chapter.

Each category defines a corresponding virtual person. The profile of this virtual person can then be applied (successfully or not) to any group (end user) linked to this virtual person. This gives a typical example for an *indirect group profiling*.

2.7.4.3 Direct Individual Profiling

In the case of individual profiling, the data subject contains one single element, the individual himself. Information is gathered about this individual and processed using data mining techniques, for example, in order to define his profile.

Knowledge in this profile derives directly and exclusively from the information about this individual. Such a profile will typically describe his habits and preferences, directly deduced from the observation of him.

This profile is then used for the individual himself in order to anticipate, for example, his actions, his behaviour or his preferences. This is what we call *direct individual profiling*.

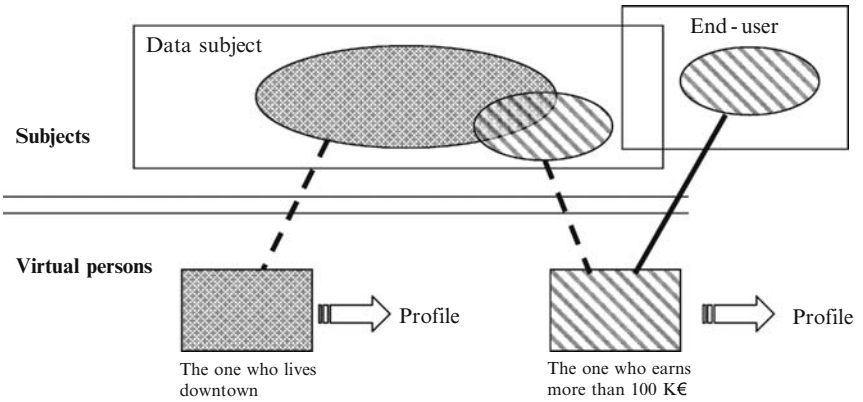

Fig. 2.7 Indirect group profiling

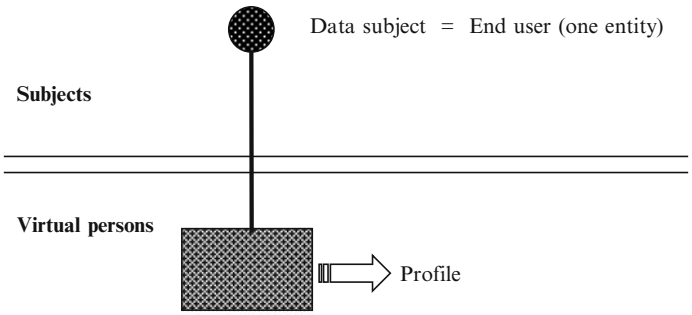

Fig. 2.8 Direct individual profiling

2.7.4.4 Indirect Individual Profiling

Direct individual profiling produces knowledge. This knowledge can in turn be mapped to pre-existing compatible virtual persons in order to infer probable profiles for this individual. These probable profiles come from pre-existing group profiles. This is what we call *indirect individual profiling*.

As an example, an insurance company might use group profiles in order to estimate risks associated with a potential client. If the person smokes, the group profile associated with the virtual person *the one who smokes* is used to infer probable characteristics of the potential client.

In a recent paper,[36] the authors explain how the knowledge of what a consumer watches on television (direct individual profiling) allows us to infer demographic characteristics about this consumer, such as his age or gender (indirect individual profiling).

[36] Spangler, W.E., Hartzel: K.S. and Gal-Or, M., 2006: 119-123.

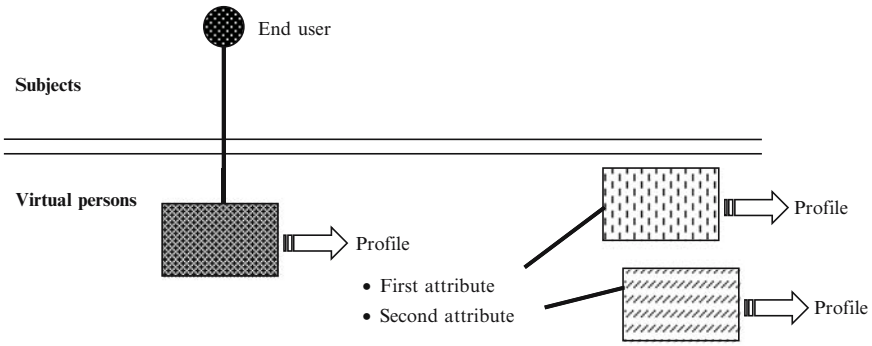

Fig. 2.9 Indirect individual profiling

The online bookstore *Amazon* gives personalised advice such as "people who have bought this book have also bought these others". Furthermore, it proposes personalised offers when the client is recognised through a cookie or when he enters his personal account. Those are typical examples of indirect individual profiling.

Real-time adaptive indirect individual profiling is part of the vision of the future AmI space, where the environment interacts with the individual in order, for example, to anticipate his needs.

2.7.5 Conclusion

Individual and group profiling have been refined using the new concept of direct and indirect profiling. While direct profiling is expected to be more reliable, indirect profiling uses the full potential of knowledge based on categorisation and generalisation.

We have shown how the generic model of virtual persons helps to describe profiling types. The four combined types of profiling have been illustrated using this model: direct group profiling, indirect group profiling, direct individual profiling and indirect group profiling.

2.8 Bibliography

Aarts, E. and Marzano, S., (eds.), *The New Everyday. Views on Ambient Intelligence*. 010 Publishers, Rotterdam, 2003.

Bourcier, D., 'De l'intelligence artificielle à la personne virtuelle: émergence d'une entité juridique?', *Droit et Société*, Vol. 49, l'Association française Droit et Société, Paris, 2001, pp. 847-871.

Custers, B., *The Power of Knowledge. Ethical, Legal, and Technological Aspects of Data Mining and Group Profiling in Epidemiology*, Wolf Legal Publishers, Nijmegen, 2004.

de Waal, F.B.M., *The Ape and the Sushi Master: Cultural Reflections of a Primatologist*, Basis Books, New York, 2001.

Edens, J.R., 'Misuses of the Hare Psychopathy Checklist-Revised in Court', *Journal of Interpersonal Violence*, Vol. 16, No. 10, Sage publications, London, 2001, pp. 1082-1094.

Gladwell, M., *Blink: The Power of Thinking Without Thinking*, Little, Brown and Company, Boston, 2005.

International Telecommunications Union (ITU), *The Internet of Things*, 7th Internet ITU report, 2005. Available (for purchase) at: http://www.itu.int/osg/ spu/publications/internetofthings/.

ISTAG (Information Society Technology Advisory Group), *Scenarios for Ambient Intelligence in 2010*, IST – IPTS report, EC, 2001. Available at: http://www.cordis.lu/ist/istag-reports.htm.

Jaquet – Chiffelle D.-O., Benoist, E., Anrig, B., 'Virtual? Identity', Chapter 3 of Nabeth, T. et al. (eds), *Set of use cases and scenarios, FIDIS Deliverable 2.2*, European Union IST FIDIS Project, 2006a. Available at: http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp2-del2.2_Cases__stories_and_Scenario.pdf

Jaquet – Chiffelle D.-O., Benoist, E., Anrig B,, 'Virtual persons applied to authorization, individual authenticatinon and identification', Jaquet – Chiffelle D.-O., Benoist, E., Anrig B. (eds.) *FIDIS brochure Deliverable 2.6*, '*Identity in a Networked World: Use Cases and Scenarios*', 2006b, pp. 6 and 7. Avalable at www.Vip.Ch

Kephart, J.O., Chess, D.M., 'The Vision of Autonomic Computing', *Computer*, Vol. 36, No. 1, IEEE Computer Society, Washington DC, 2003, pp. 41 - 50.

Lawton, G., 'Machine-to-Machine Technology Gears Up for Growth', *Computer,* Vol. 37, No. 9, IEEE Computer Society, Washington DC, Sept. 2004, pp. 12-15.

Leenes, R. and Koops, B.J., ' 'Code': Privacy's Death or Saviour?', *International Review of Law Computers & Technology*, Vol. 19, No. 3, Routledge, Oxford, 2005, pp. 329-340.

Maturana, H.R. and Varela, F.J., *Autopoiesis and Cognition: The Realization of the Living*, Reidel, Dordrecht, 1991.

Maturana, H. R., Varela, F.J., *The Tree of Knowledge. The Biological Roots of Human Understanding*, Shambala, Boston and London, 1998 (revised edition).

Peirce, Ch.S., *Pragmatism as a Principle and Method of Right Thinking. The 1903 Harvard Lectures on Pragmatism*, edited and introduced with a commentary by Patricia Ann Turrisi, State University of New York Press, Albany,1997.

Peppers, D., Rogers, M., *The One to One Future*, Currency, New York, 1993.

Plessner, H, *Die Stufen des Organischen under der Mensch. Einleitung in die philosophische Anthropologie*, Suhrkamp, Frankfurt, 1975.

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

Rafter, R. and Smyth, B., 'Passive Profiling from Server Logs in an Online Recruitment Environment', paper presented at *the IJCAI's Workshop on Intelligent Techniques for Web Personalization*, Seattle, Washington 4-6 August 2001. Available at http://maya.cs.depaul.edu/~mobasher/itwp01/papers/rafter.pdf.

Rasch, G., *Probabilistic Models for Some Intelligence and Attainment Tests*, University of Chicago Press, Chicago, Illinois, 1980.

Schauer, F., *Profiles Probabilities and Stereotypes*, Belknap Press of Harvard University Press, Cambridge, Mass. London, England, 2003.

Spangler, W. E., Hartzel K. S. and Gal-Or M., 'Exploring the Privacy Implications of Addressable Advertising and Viewer Profiling', *Communications of the ACM*, Vol. 49, No. 5, ACM Press, New York, 2006, pp. 119-123.

Stergiou, C. and Siganos, D., 'Neural Networks and their Users', *Surprise 96 Journal*, Vol. 4, Department of Science Technology and Medicin, Imperial College London, London, 1996. Available at: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html # Why%20use%20neural%20networks.

Thorndike, R.L. (ed.), *Educational Measurement*, American Council on Education, Washington, D. C., 2nd edition, 1971.

Turvey, B., *Criminal profiling: An introduction to behavioral evidence analysis*, Academic Press, New York, 1999.

Van Brakel, J., 'Telematic Life Forms', *Techné: Journal of the Society for Philosophy and Technology*, Vol. 4, No. 3, DLA, Blacksburg, 1999. Available at http://scholar.lib.vt.edu/ejournals/SPT/v4_n3html/VANBRAKE.html.

Vedder, A., "KDD: The challenge to individualism." *Ethics and Information Technology*, Volume 1, Number 4, Springer, 1999, pp. 275-281.

Weiser, M., 'The Computer for the Twenty-First Century', *Scientific American*, 265, 3, Scientific American Inc, New York, 1991, pp. 94-104.

Wilkie, William L. and Moore, E.S., 'Scholarly Research in Marketing: Exploring the "4 Eras" of Thought Development', *Journal of Public Policy & Marketing*, Vol. 22, No. 2, AMA Publications, Chicago, 2003, pp. 116-146.

Zarsky, Tal Z., '"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion', *Yale Journal of Law & Technology*, Vol. 5, 2002-2003. Available at: http://research.yale.edu/lawmeme/yjolt/files/20022003Issue/Zarsky.pdf